University of Groningen

From Voice to Speech

Nagels, Leanne

DOI:
[10.33612/diss.179062515](https://doi.org/10.33612/diss.179062515)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Nagels, L. (2021). *From Voice to Speech: the perception of voice characteristics and speech in children with cochlear implants*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.179062515>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

From Voice to Speech

The perception of voice characteristics and speech
in children with cochlear implants

Leanne Nagels

The work presented in this thesis has been carried out under the auspices of the Research School of Behavioral and Cognitive Neuroscience (BCN) and the Center for Language and Cognition Groningen (CLCG) of the University of Groningen.

This research was supported by a grant from the Netherlands Organization for Scientific Research (NWO) awarded to Prof. Dr. Deniz Başkent for the research programme Perception of Indexical Cues in Kids and Adults (PICKA) (grant no. 918.17.603).

Groningen Dissertations in Linguistics **209**

Cover page art by Samantha Janssen
Cover page frame from pngtree.com
Printed by Ipskamp Printing (Enschede)

© Leanne Nagels, 2021

rijksuniversiteit
 groningen

From Voice to Speech

The perception of voice characteristics and speech in
 children with cochlear implants

Proefschrift

ter verkrijging van de graad van doctor aan de
 Rijksuniversiteit Groningen
 op gezag van de
 rector magnificus prof. dr. C. Wijmenga
 en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 23 september 2021 om 11.00 uur

door

Leanne Leonarda Anna Maria Georgina Nagels

geboren op 3 mei 1993
 te Gennep

Promotores

Prof. dr. P. Hendriks

Prof. dr. D. Başkent

Beoordelingscommissie

Prof. dr. C. McGettigan

Prof. dr. H. Meister

Prof. dr. W.M. Lowie

Acknowledgements

Most of you who know me a bit know I am not a person who likes to be the center of attention, so I am very happy to share this spotlight with all the wonderful people who helped me along the way.

First of all, I would of course like to thank the two great women who made all of this possible, my promotors, Petra and Deniz. Lots of people were scared for me to have an interdisciplinary project, two promotors, and two more co-authors, but this never really seemed to be much of an issue, and I believe this is largely because both of you work so well together. It was a lot of fun to work on this PhD project and to get input from multiple perspectives. Petra, I really admire how even besides all your teaching duties, supervising your other PhDs, and your own research that you always manage to have such a good eye for detail. I also always enjoyed exchanging the latest GSH and CLCG gossip with you during my time as a PhD council member and the nice after-summer parties at your house. Deniz, thank you so much for involving me in your VICI project and giving me the opportunity to pursue my dream to study voice and speech perception in children with cochlear implants. Seeing how you always manage to collaborate so well with so many people in the academic world, which can sometimes be extremely competitive, has always been very inspiring to me and I really learned so much from you. I really enjoyed all the nice brunches at your place and our nice trips to CIAP in Lake Tahoe and VIHAR in London.

I am also very grateful to my amazing supervisors and co-authors, Etienne and Debi. Etienne, thank you so much for sharing all your incredible knowledge about the technical workings of cochlear implants with me and being a great sparring partner for all the statistical analyses. Debi, thank you for all your great ideas like using the CRM experiment, and for your valuable feedback and native speaker input. I am really glad that both of you were part of the team.

I would also like to thank the Cochleair Implantatieteam Noord Nederland (CINN), especially Rolien Free, Bert Maat, and Gera Nijhof, for helping me to recruit all the CI children for the studies. A big thank you also to BSO Huis de B and Basisschool de Brink for helping me to recruit normal-hearing children for the studies. Especially a big thank you to my sister and primary school teacher, Ilse, whose popularity as a teacher helped a lot. Finally, I want to thank all the parents and children who took part in the studies. I would also like to thank the members of the reading committee, Dr. Carolyn McGettigan, Dr. Hartmut Meister, and Dr. Wander Lowie, for taking the time to read and review my dissertation, and for giving great feedback.

To my amazing linguistics colleagues, thank you for all the good times we had together and hopefully will still have in the future. I always enjoyed our high tea parties and fun activities so much. Anna and Amélie, you were my first two Harmony office mates, and I always looked up to you, even though you started only one and two months before me. It's a very tough act to follow such talented researchers, but I will try my very best. I would also like to thank my second Harmony office mates, Dorothee, Vera, and Irene, for always giving very good PhD and life advice and also for telling me to get back to work whenever I was talking way too much again. Finally, I would like to thank my third Harmony office mates, lots of offices throughout the years, Sophia, Iris, and our SemCog-adopted PhD, Abby. Doing a PhD is definitely not a simple job, but working together with you girls really made it much easier and fun.

The greatest thing about having an interdisciplinary PhD project is that you get not one but two sets of amazing colleagues. I would like to thank all the members of the KNO research group for the interesting talks and very useful feedback, especially on the more technical CI-related parts. It was so nice to be part of this very interdisciplinary research group, talking about emotion perception one day and tinnitus in mice the next day. A big thank you to all my (former) UMCG office mates, Daniël, Nawal, Enja, Floor, and Joëlle. Thank you for always listening to my complaints and giving me good advice on all sorts of PhD and life issues. For the other KNO PhDs, Sina, Minke, Nick, Julie, Elif, Jose, Punit, Dora, Shahin, Elouise, Marita, Nelleke, Luke, and Ada, thank you for all the nice coffee breaks and after-work drinks. Finally, I would also especially like to thank Paolo and Anita. Anita, thank you for introducing me to the world of speech perception in CI users via my master thesis project and for always giving me helpful career advice. I feel very lucky to have had such a supportive and inspiring supervisor like you. Paolo, thank you for always helping me with all kinds of programming issues and errors, and also for the fun conversations with lots of dark humor.

I would also like to thank all my friends from Limburg and Brabant, although most of us are kind of scattered around the country by now, all my lovely salsa friends, and my former EMCL master's crew. It was so nice to always have your support to fall back on. I especially want to thank all of my personal psychologists Lisa, Simone, and Lotte for always helping me to keep things in perspective and giving me good advice, and my amazing personal IT specialist and advisor, Jackie. A special thanks also to the best Portuguese person in the world, sorry Cristiano Ronaldo but it is not you, Catarina, thank you for being such a good friend and working through all our PhD and life issues together.

I am forever grateful to my family, especially to my parents and my sister, who always loved and supported me no matter what path I chose in life and always tried to help me in every way they could. I would also like to thank my other family members for always asking me what in the world I was doing now again, which helped me so much with communicating my research to others and writing up this dissertation. A special thanks also goes to my grandparents, who in many ways laid the foundation for my career in linguistics by making me able to understand the local dialect perfectly but not being able to speak it, an issue that still intrigues me today.

Last but certainly not least, I would like to thank my partner in crime and also my best friend, Sebastiaan. I simply do not know how I could have managed all of this without you. You really deserve a medal for putting up with my emotional outbursts and high constant stress levels during the past year. I love you, and let's go on a nice long holiday soon!

Contents

Acknowledgements	v
1 Introduction	15
1.1 Limitations in voice perception via CIs.....	19
1.1.1 Spectrotemporal resolution of the CI signal.....	19
1.1.2 Voice characteristics perception via CIs	23
1.2 Development of voice perception in NH children	27
1.2.1 The perceptual weighting of voice-related cues	29
1.2.2 Relationships between different voice perception abilities	31
1.3 Aim of this dissertation.....	32
1.4 Scope of this dissertation.....	33
2 Development of voice perception is dissociated across gender cues in school-age children	47
2.1 Introduction.....	49
2.2 Discrimination of differences in F0 and VTL cues.....	53
2.2.1 Methods.....	53
2.2.2 Results.....	56
2.2.3 Discussion	58
2.3 Weighting of F0 and VTL cues for voice gender categorization	59
2.3.1 Methods.....	59
2.3.2 Results.....	61
2.3.3 Discussion	65
2.4 General discussion	67

3 School-age children benefit from voice gender cue differences for the perception of speech in competing speech.....	75
3.1 Introduction.....	77
3.2 Method	83
3.2.1 Participants.....	83
3.2.2 Stimuli and apparatus	84
3.2.3 Procedure	86
3.2.4 Data analysis	88
3.3 Results.....	90
3.3.1 F0 and VTL benefit for perceiving speech in competing speech	90
3.3.2 Adult-like benefit from F0 and VTL differences and adult-like accuracy scores.....	93
3.3.3 Adult-like benefit from F0 and VTL differences for interpolated accuracy scores.....	97
3.3.4 Correlations of F0 and VTL benefit, and overall accuracy scores with JNDs	99
3.4 Discussion	101
 4 The perception of voice gender cues and speech in competing speech in children with cochlear implants	 115
4.1 Introduction.....	118
4.1.1 Effects of CI signal limitations on the perception of voice characteristics	119
4.1.2 Developmental effects on the perception of voice characteristics	123
4.1.3 Present study.....	126
4.2 Discrimination of differences in F0 and VTL cues.....	127
4.2.1 Methods.....	127
4.2.2 Results.....	134
4.2.3 Discussion	138
4.3 Weighting of F0 and VTL differences for voice gender categorization	139
4.3.1 Methods.....	139
4.3.2 Results.....	141

4.3.3 Discussion	146
4.4 Benefit from target-masker voice differences	149
4.4.1 Methods.....	149
4.4.2 Results.....	152
4.4.3 Discussion	156
4.5 General discussion	159
5 Development of vocal emotion categorization in school-age children	175
5.1 Introduction.....	177
5.2 Materials & Methods	180
5.2.1 Participants.....	180
5.2.2 Stimuli and Apparatus.....	181
5.2.3 Procedure	182
5.2.4 Data analysis	183
5.3 Results.....	184
5.3.1 NH Dutch and English data	184
5.3.2 Data of CI children.....	186
5.4 Discussion.....	187
5.4.1 Age effects.....	187
5.4.2 Language effects.....	187
5.4.3 CI children	188
5.5 Conclusions.....	189

6 Discussion	195
6.1 The development of voice characteristics perception in NH children	198
6.1.1 Development of discrimination and categorization of voice gender cues in NH children.....	198
6.1.2 Development of a benefit from voice gender cue differences for perceiving speech in competing speech in NH children	201
6.1.3 Development of categorization of vocal emotions in NH children.....	203
6.1.4 Summary of findings for NH children	206
6.2 The development of voice characteristics perception in CI children.....	207
6.2.1 Development of discrimination and categorization of voice gender cues and a benefit from voice gender cue differences for perceiving speech in competing speech in CI children.....	207
6.2.2 Development of categorization of vocal emotions in CI children	211
6.2.3 Summary of findings for CI children.....	212
6.3 Future perspectives.....	213
6.4 Conclusions.....	214
 Nederlandse samenvatting.....	 223
List of publications.....	235
Groningen dissertations in linguistics (GRODIL)	237

1 | Introduction

Speech is a complex auditory signal which conveys not only linguistic messages but also information about the speaker. For instance, listeners can perceive a speaker's age, affective state, regional accent, and gender merely based on their voice (see a review by Scott & McGettigan, 2016). These indexical voice-related cues are often presumed to play only a minor role in the perception of speech due to *speaker normalization*, i.e., listeners can generally recognize different phonemes and words despite the acoustic variations in production related to speakers' voices (Johnson, 2005). However, even normal-hearing (NH) adults show better recognition of words that are produced by a familiar speaker compared to words that are produced by an unfamiliar speaker (Nygaard & Pisoni, 1998). In noisy environments, such as a crowded café, the perception of speakers' voice characteristics also helps listeners to segregate different speech streams and effectively process speech (Bregman, 1994; Cherry, 1953). As speech perception rarely takes place in absolute silence, voice characteristics are essential for the perception of speech in daily life.

When the spectral and temporal information that are provided by the speech signal are degraded, the perception of speech and indexical voice-related cues becomes even more challenging, which is the case for listeners with a cochlear implant (CI). A CI is a prosthetic hearing device that allows listeners with severe to profound hearing loss to regain at least partial auditory perception. The device is primarily used when the cochlea in the inner ear is malfunctioning and regular hearing aids can no longer provide auditory sensation through amplification of the acoustic signal. Instead, a CI bypasses the outer and middle ear, and transforms acoustic signals into electric pulses that directly stimulate the auditory nerve via the inserted electrode array in the inner ear (see a review by Başkent et al., 2016b). A prerequisite for successful cochlear implantation therefore is that a sufficient number of auditory nerve fibers are still functioning and relatively healthy. Although CIs have improved greatly over the last few decades, with improvements in the technology, patient rehabilitation, and our knowledge of the nerve-electrode information transmission, the CI signal still lacks a substantial amount of spectrotemporal information that is available via normal hearing. Despite the spectrotemporally degraded CI signal, many CI users can perceive speech relatively well in quiet (Blamey et al., 2013; Lazard et al., 2012; Wilson & Dorman, 2008), but they often continue to face difficulties with perceiving speech in noisy environments.

A striking case of interference from background noise is that of competing speech that is produced by other speakers talking in the background or two people talking at the same time while the listener's attention is directed to one of them. These types of multitalker situations are commonly encountered in daily life, i.e., *the*

cocktail party problem (Cherry, 1953). Although NH adult listeners are often able to deal with these situations with relative ease, CI users seem to have great difficulty with understanding speech in the presence of competing background speech (El Boghdady et al., 2019; Meister et al., 2020; Stickney et al., 2004). In addition, CI users' ability to perceive indexical voice-related cues, such as differences in speakers' voice characteristics (Cleary et al., 2005; Cleary & Pisoni, 2002; Gaudrain & Başkent, 2018), vocal emotions (Chatterjee et al., 2015; Nakata et al., 2012), or prosodic differences (Chatterjee & Peng, 2008; Everhardt et al., 2020; Meister et al., 2009; Peng et al., 2012), is generally poorer compared to NH listeners.

Interpreting the underlying cause of CI users' difficulties with perceiving voice and speech cues has been challenging because of an extensive amount of individual variability in their speech perception abilities. While factors that contribute to the success of CI users' speech perception outcomes have been quantified on a population level, such as the duration of deafness, these factors cannot predict or fully account for the variability among individual CI users (Blamey et al., 2013; Holden et al., 2013; Lazard et al., 2012). The variability in CI users' speech perception abilities seems to partially originate from differences in the peripheral perceptual mechanisms that are primarily related to the device (e.g., the active number of electrodes, the insertion depth of the electrode array, the proximity of the electrodes to the nerves to be stimulated, and the specific signal processing algorithms that determine how acoustic speech signals are converted into electric stimulation patterns), and the etiology of hearing loss (e.g., duration of auditory deprivation or neural damage). Furthermore, because the electrically transmitted speech cues are degraded compared to those provided via normal acoustic hearing in many ways, cognitive mechanisms are required to compensate for the degraded speech cues for the perception of speech. Hence, in addition to the aforementioned factors, cognitive and developmental factors seem to play an important role in perceiving speech via a CI. However, research on these aspects is still relatively new (Başkent et al., 2016a; Nagels et al., 2020; Pisoni et al., 2018).

In clinical practice and research with CI users, a distinction is usually made between prelingually deaf CI children and postlingually deaf CI adults. Prelingually deaf CI children are often congenitally deaf or became deaf at a young age during the period of language development, usually described as implanted before the age of six, while postlingually deaf CI adults became deaf later in life. Especially in developed countries, deafness is often identified soon after birth as a result of newborn hearing screenings, which has generally led to earlier CI implantation (Bruijnzeel et al., 2017) and better speech and language outcomes for CI children (Kirk et al., 2002; Nicholas & Geers, 2007; Svirsky et al., 2007), likely due to effects

of neural auditory plasticity (Manrique et al., 1999). An important difference between prelingually deaf CI children and postlingually deaf CI adults concerning their voice perception abilities is that the latter group of CI users usually acquired language via a normal acoustic speech signal. Besides the acquisition of language, CI children learned auditory categories, such as those related to speakers' voice gender or their vocal emotions, via the spectrotemporally degraded CI signal. Cognitive-developmental factors therefore likely affect voice and speech perception in CI children, particularly because many voice perception abilities also develop to adult-like levels in NH children during the school-age years, such as voice discrimination (Buss et al., 2017; Cleary et al., 2005; Levi & Schwartz, 2013) or vocal emotion recognition (Sauter et al., 2013; Tonks et al., 2007). Hence, studying the perception of voice and speech cues in children with CIs requires not only the systematic investigation of the peripheral perceptual limitations imposed by the CI (section 1.1) but also the developmental effects related to central cognitive development on different voice perception abilities in NH children (section 1.2).

1.1 Limitations in voice perception via CIs

1.1.1 Spectrotemporal resolution of the CI signal

The first CI implantation was performed by the American otologist William House and neurosurgeon John Doyle in 1961 (Mudry & Mills, 2013). At this point, CIs only had a single active electrode and provided highly limited frequency information, where the implant user could only make use of the temporal speech envelope i.e., slow fluctuations in the amplitude of the signal over time, for understanding speech. Despite these limitations, the implant still provided better means of auditory communication to the user than without the implant, especially when the user supplemented the severely degraded speech signal that was perceived via the CI with visual lipreading cues. Thirteen years later, the first multi-electrode device was developed and implanted. Using multiple electrodes made it possible to activate the auditory nerve in different parts of the cochlea, and thus different frequency regions, by partitioning the original acoustic signal into different frequency bands. This mimicked normal acoustic hearing more closely and enabled CI users to perceive differences in timbre and formant frequencies better, which are, for instance, required for the recognition of vowels (Clark, 2015).

Nowadays, CIs consist of an external part that is worn behind the ear and an internal part, the implant, that is surgically placed beneath the skin (the receiver-stimulator) and in the cochlea (electrode array). The external part contains a

directional microphone, a speech processor, and a transmitter, while the internal part consists of a receiver-stimulator and an electrode array (Figure 1.1). Acoustic signals are perceived via the microphone located above the outer ear (pinna), and subsequently coded and bandpass filtered into different frequency bands by the speech processor. The temporal envelope of each frequency band is then extracted and used to modulate an electric carrier pulse train. The electric pulse trains of different frequency bands are sent to the receiver-stimulator, which is connected to the transmitter via a magnetic coil and located beneath the skin. The receiver-stimulator decodes the processed signal and sends the pattern of electric pulses to the inserted electrode array in the scala tympani of the cochlea to stimulate the auditory nerve.

Figure 1.1: Ear with a cochlear implant reproduced and adapted from the image published by Kral (2013) under a CC BY 4.0 license.

Frequency components are processed in different regions of the basilar membrane in the cochlea based on their spectral content because of the changing width and stiffness of the basilar membrane (Figure 1.2). High frequencies cause a maximum vibration of the basilar membrane (peak response) near the base of the cochlea, which is narrow and stiff, while low frequencies cause peaks near the apex, which is wider and more flexible. This spatial organization for the analysis of different frequencies in the cochlea, and also in the brainstem and the cortex, is called *tonotopy*. In NH listeners, the vibrations of the basilar membrane activate the

auditory receptor cells, i.e., *hair cells*, which are attached to the basilar membrane. The activated hair cells subsequently depolarize, which results in an electrical potential that stimulates the auditory nerve. The electrical signal is then sent to the auditory brainstem and the auditory cortex.

In CI users, the electrode array that is inserted in the scala tympani aims to use the natural tonotopic organization of the cochlea. Electric pulses are sent to electrodes in different frequency regions on the basilar membrane based on the frequency bands of the acoustic signal. The CI electrode array attempts to mimic the tonotopic mapping in normal acoustic hearing in this manner. However, the direct stimulation of the nerve by the electrical pulses from the electrode array causes nerve responses that are inherently different from those that are produced by the hair cells in normal acoustic hearing. The activated electrodes create temporospatial patterns of activation, which only partially resemble that of acoustic hearing, that have to be interpreted as speech sounds by the brain.

Figure 1.2: Overview of the tonotopic organization of the basilar membrane in an uncoiled cochlea to show the structure of the basilar membrane for illustrative purposes, reproduced from Kern et al. (2008) and published under a CC BY 2.5 license. The basilar membrane is narrow and stiff around the base and becomes broader and more flexible towards the apex.

Although the electric pulses that are provided by the CI stimulate the auditory nerve, like the electric signals produced by the hair cells in normal acoustic hearing, the spectrotemporal information is degraded. The spectral resolution of the CI speech signal is not necessarily limited by the number of CI electrodes, which

usually varies between 12 and 24 electrodes. The main limiting factor is the interaction between the activation of different electrodes, which comes from the spread of excitation from individual contact points and causes spectral smearing, i.e., a wider and smoother distribution of energy across frequencies (Başkent et al., 2016b; Fu & Nogaki, 2005). Figure 1.3 shows the spectrograms of a normal acoustic speech signal and a noise-vocoded CI-simulated speech signal, which illustrates the wider spread of energy across frequencies and the degraded spectrotemporal details in the CI speech signal. While the spectrogram of the normal acoustic speech signal shows the individual harmonics (frequency components, horizontal lines in Figure 1.3a) and formant frequencies (resonance peaks in the vocal tract, bright orange lines in Figure 1.3a), the noise-vocoded CI-simulated speech signal only provides very broad and imprecise spectrotemporal information (nearly indistinguishable horizontal lines in Figure 1.3b). In addition, as the electric pulses are presented at a fixed rate by most CI speech processors, no temporal fine structure information, i.e., information about the periodicity (repetition rate) of the acoustic waveform, is provided via the CI. Temporal information is only provided by the temporal envelope modulations of each frequency channel, which are extracted by the CI processor. As a result, voice and speech cues are represented to a lesser extent in the CI signal compared to the information provided via normal hearing.

Figure 1.3: The narrow-band spectrograms of a normal acoustic speech signal (a) and a noise-vocoded 16-channel Cochlear CI24 clinical frequency map speech signal (b). The used speech signal was a recording of the Dutch sentence: *Laat de kat zien waar de zwarte acht is.* [Show the cat where the black eight is.], which was a stimulus from our speech in competing speech perception experiment described in Chapters 3 and 4 of this dissertation. The horizontal lines that are visible in the normal acoustic speech signal show individual harmonics (frequency components) and the bright orange lines show the formant frequencies (resonance peaks in the vocal tract). In the noise-vocoded CI-simulated signal, the bright orange areas show the proximity of frequencies to the formant frequencies.

1.1.2 Voice characteristics perception via CIs

Many of the acoustic cues that define a speaker's voice are learned and acquired by speakers over time via experience, such as their accent, speaking rate, or aspects of their voice quality. However, the variation in speakers' voice is also restricted by relatively unchangeable, physiologically based characteristics, the two most influential acoustic cues being speakers' mean fundamental frequency (F0) and their vocal-tract length (VTL) (Kreiman & Sidtis, 2011). Together, these acoustic cues primarily define speakers' so-called *perceived voice gender* (Skuk & Schweinberger, 2014; Smith & Patterson, 2005; Titze, 1989), i.e., the gender of the speaker as perceived by others based on the speaker's voice characteristics. A speaker's mean F0 is determined by the vibration rate of the vocal folds (Figure 1.4) and affects the perceived pitch of a voice. Voiced parts of speech are characterized by a harmonic structure, in which frequency components are multiples. Thus, if the F0 increases, the spacing between successive harmonics becomes broader and vice versa, but the spectral envelope remains unchanged (Figure 1.5, Panel E vs. Panel F or Panel G vs. Panel H). The vibration rate of a speaker's vocal folds largely depends on the length and mass of the vocal folds. Before the age of 12, the mean F0 values of male and female children's voices largely overlap (Perry et al., 2001; Vorperian & Kent, 2007). An increase in testosterone at puberty stimulates the growth of the laryngeal cartilages and the length and the mass of the vocal folds in male and female children (Harries et al., 1998). The vocal folds of males become longer and thicker than those of females, leading to a typical difference in the mean F0 of male and female adult speakers' voices of around 70%, approximately 12 semitones (st) (Klatt & Klatt, 1990). NH listeners can perceive the mean pitch of a voice via 'place pitch', which is based on the tonotopic location of the excitation of populations of neurons along the basilar membrane that corresponds to different frequencies. In addition, 'temporal pitch' can be used for voice pitch perception. This process relies on the fact that auditory nerve fibers tend to generate action potentials, i.e., spikes in activation, at a certain rate, which is called *phase locking*. The periodicity of a signal can be determined based on the time interval between the successive action potentials which evokes a pitch percept (Oxenham, 2013).

For CI users, the perception of place pitch is highly restricted because of the limited number of frequency channels and the earlier mentioned channel interactions that cause spectral smearing (Fu & Nogaki, 2005). Intermediate pitch percepts or *virtual channels*, i.e., different frequency percepts by the simultaneous activation of two adjacent electrodes, have been reported for CI users, but there is considerable variability among CI users' ability to perceive these differences

(Donaldson et al., 2005; Laneau & Wouters, 2004; Townshend et al., 1987). In addition, although the electrode array mimics the function of the hair cells and uses the natural tonotopic organization of the cochlea, there are often discrepancies in terms of spacing between the frequency bands allocated to the electrodes and their tonotopic location on the basilar membrane for CI users (Başkent & Shannon, 2004). As electric pulses are presented at a fixed rate by most CI processors, CI users mainly perceive the temporal pitch of a speech signal via the temporal envelope modulations (Geurts & Wouters, 2001; Laneau et al., 2004), which results in a relatively weak pitch percept, also in NH listeners, and only applies to signals below 300 Hz for most CI users (Kong et al., 2009; Zeng, 2002). Previous attempts to enhance the temporal envelope modulations improved CI users' voice pitch processing but negatively affected their recognition of vowels due to interference with the spectral information that is provided by the CI (Green et al., 2004, 2005).

Figure 1.4: A sagittal view of the human speech production system adapted from Gray (1878). The location of the vocal folds (related to speakers' mean F0) is marked in green and the vocal-tract (related to speakers' VTL), ranging from the vocal folds to the aperture of the oral and nasal cavities, is marked in blue.

Figure 1.5: Schematic representations of the waveform (A–D and I–L) and spectrum (E–H and M–N) of the vowel /a/ for different combinations of F0 and VTL in a normal acoustic speech signal (left panel) or a CI-simulated speech signal (right panel) reproduced from Başkent et al. (2016b).

A speaker's VTL consists of the distance between their vocal folds and the aperture of their oral and nasal cavities, which strongly correlates with the speaker's overall height (Fitch & Giedd, 1999). The vocal tract functions as an acoustic filter for the air pressure waves that are generated by the vibrations of the vocal folds, as described in the source-filter model of speech production (Chiba & Kajiyama, 1958; Fant, 1960), and influences the values of the formant frequencies of the speech signal (Titze, 1989). Besides estimating speakers' height, listeners use VTL cues for formant frequency normalization to interpret vowels correctly (Irino & Patterson, 2002; Ladefoged & Broadbent, 1957). This normalization process allows listeners to perceive vowels produced by men, women, and children as the same vowel, even though there are sizeable differences in formant frequencies. If speakers' VTL becomes shorter and their formant frequencies increase, the spectral envelope is shifted towards the higher frequencies and vice versa, but the harmonic structure, i.e., the spacing between individual harmonics, under the spectral envelope remains the same (Figure 1.5, Panel E vs. Panel G or Panel F vs. Panel H). For children, the F0 and VTL voice parameters largely overlap across genders until puberty, although there seem to be small differences in formant frequencies that could be related to differences in the resonator width and size (Bennett & Weinberg, 1979; Huber et al., 1999; Perry et al., 2001; Vorperian & Kent, 2007). During puberty, children's VTL grows and their formant frequencies decrease around 30% for males and 20% for females compared to their voice parameters at four years of age (Fitch & Giedd,

1999; Huber et al., 1999). Male speakers' VTL is generally around 20% larger than that of female speakers, which is equal to a difference of approximately 3.6 st. NH listeners can estimate a speaker's VTL by the distribution of the formant frequencies in an audio signal based on the spectral envelope.

According to the work by Gaudrain and Başkent (2015) using noise-vocoded CI-simulated speech, VTL perception in CI users heavily relies on the spectral resolution of the CI speech signal, while the contribution of temporal cues is minimal. A difference of 3.6 st in VTL would approximately produce a shift in the location of stimulation of one electrode (Fuller et al., 2014). Hence, differences in speakers' VTL that are smaller than 3.6 st, which is equal to the typical difference between male and female speakers' VTL, may not be accessible to most CI users. This suggestion is in agreement with the results of Gaudrain and Başkent (2018) which showed that postlingually deaf CI adults had a mean just-noticeable difference (JND) of 7.19 st in VTL. Thus, the CI adults who took part in this study would not be able to distinguish speakers' voice gender based on VTL differences according to their discrimination thresholds. However, it is unknown if the same would apply to prelingually deaf children with CIs.

As aforementioned, prelingually deaf CI children and postlingually deaf CI adults acquired language in different manners, which seems to affect their processing of spectrotemporal information (DiNino & Arenberg, 2018; Jung et al., 2012; Landsberger et al., 2018), and may also lead to differences in their processing of voice and speech cues, such as those related to speakers' voice gender. While postlingually deaf CI adults have to adapt their already established acoustic representations of speakers' voice gender that are based on normal acoustic hearing, prelingually deaf CI children formed acoustic representations of speakers' voice gender that are primarily based on the spectrotemporally degraded CI signal. Yet, it is uncertain if this would enable CI children to use the acoustic cues in the CI signal more effectively than CI adults, possibly also due to their higher brain plasticity (Manrique et al., 1999), or less effectively, as their acoustic representations may be less robust than those of CI adults.

1.2 Development of voice perception in NH children

Although the peripheral auditory system of NH children is already largely mature around 6 months of age (Abdala & Keefe, 2012; Moore & Linthicum Jr., 2007), their voice perception abilities continue to develop to adult-like levels during the school-age years in many aspects. The conclusions drawn from previous studies on the development of voice perception in NH infants and children are seemingly

contradictory. Infants perform surprisingly well on voice perception tasks, while school-age children do not show adult-like voice perception abilities yet (Fecher et al., 2019). For instance, infants already show a preference for hearing their mother's voice over other female voices (DeCasper & Fifer, 1980; Fernald & Kuhl, 1987) and for hearing their native language over other languages (Johnson et al., 2011). They also seem able to reliably distinguish male from female speakers' voices very early on (Floccia et al., 2000; Miller, 1983; Miller et al., 1982), and are sensitive to large differences in speakers' mean F0 (Spence & Freeman, 1996), timbre differences associated with VTL (Vestergaard et al., 2009), and prosodic differences (Fernald & Kuhl, 1987). Even though it is quite remarkable that infants already show these preferences and voice discrimination abilities after such a short period of auditory exposure, their perception is still far from adult-like concerning more complex voice perception tasks. For instance, unfamiliar voice recognition (Fecher et al., 2019; Mann et al., 1979) and the discrimination of more subtle differences in speakers' voices, such as small differences in their mean F0 (Buss et al., 2017) or their vocal emotions (Sauter et al., 2013; Tonks et al., 2007), only become adult-like during the school-age years.

The relatively long development towards adult-like voice perception abilities in NH children seems to depend heavily on their developing cognitive abilities, rather than changes in the auditory periphery (Litovsky, 2015; Werner, 2007). Werner and Leibold (2011) propose that this relatively long development is most likely caused by children's immature processing efficiency. Processing efficiency is described by Werner and Leibold as a combination of several central cognitive processes, such as general attention span, memory, motivation, and selective attention, which seem to partially overlap and interact with each other (Cowan, 1998). Another explanation for children's immature voice perception abilities is that children may use different processing strategies and a different weighting of acoustic cues compared to adults. This hypothesis has particularly been explored in research on children's perceptual weighting of acoustic cues for the categorization of phonemes, which seems to differ from adults even beyond the school-age years (Hazan & Barrett, 2000; Nitttrouer & Miller, 1997). According to this hypothesis, children need more auditory experience to learn which acoustic cues they should attend to so they can adapt their perceptual weighting strategies. Considering that CI children only receive spectrotemporally degraded auditory input, CI children may need more auditory experience than NH children to reach the same performance level. This may also partially explain why CI children's voice perception abilities are usually poorer than those of their NH age-matched peers (Cleary et al., 2005; Cleary & Pisoni, 2002).

1.2.1 The perceptual weighting of voice-related cues

A model that provides a concrete explanation for children's different perceptual weighting of acoustic cues is the Developmental Weighting Shift (DWS) model for the categorization of phonemes by Nittrouer and Miller (1997). The DWS model specifically suggests that children may primarily rely on dynamic acoustic cues, i.e., acoustic cues that change spectrally over time, such as formant transitions, as these properties can help them in recognizing syllabic structures in the speech signal. After gaining more auditory and linguistic experience, children start to attend more to static acoustic cues, i.e., acoustic cues that do not involve a spectral change over time, such as fricative noise, that are particularly informative for phonetic structures.

To illustrate their hypothesis, Nittrouer and Miller (1997) conducted an experiment with children and adults examining the phonemic categorization of fricatives using fricative-vowel (FV) syllables, e.g., /sɑ/, /su/, /ʃɑ/, /ʃu/. Figure 1.6 shows the narrow-band spectrograms of these four FV syllables. Nittrouer and Miller (1997) created a stimulus continuum with different amounts of fricative noise, which distinguishes a /s/ from a /ʃ/ sound (Figure 1.6), and transitions in the third formant frequency (F3), which distinguishes /ɑ/ from /u/ (Figure 1.6). The F3 also varies more depending on the preceding fricative (/s/ vs. /ʃ/) for FV syllables containing an /u/ vowel compared to FV syllables containing an /ɑ/ vowel. This difference makes it more informative to rely on F3 formant transitions for the categorization of FV syllables containing an /u/ vowel than FV syllables containing an /ɑ/ vowel. The results showed that children generally weighted F3 formant transitions more than adults, but they weighted fricative noise cues to a lesser extent than adults for categorizing the four different FV syllables. Children showed a larger effect of F3 transitions on their categorizations than adults, i.e., larger distinctions between /sɑ/ vs. /su/ and /ʃɑ/ vs. /ʃu/, but they showed a smaller effect of fricative noise cues, i.e., smaller distinctions between /sɑ/ vs. /ʃɑ/ and /su/ vs. /ʃu/. Thus, children do not yet seem to be aware that the F3 formant transitions are less informative for determining the preceding fricative in FV syllables containing an /ɑ/ sound than those containing an /u/ sound. It seems children need more time to learn that they should rely more on the fricative noise cues when discriminating FV syllables containing an /ɑ/ sound than F3 transitions, like adults. Nittrouer and Miller (1997) concluded that children seem to rely primarily on differences in dynamic acoustic cues at first and have to learn to increase their weighting of static cues in environments in which these are more informative to demonstrate adult-like perceptual weighting.

Figure 1.6: The narrow-band spectrograms of the fricative-vowel syllables /sa/, /su/, /ʃa/, and /ʃu/. When we compare the panels horizontally from left to right (/sa/ vs. /su/ and /ʃa/ vs. /ʃu/), a change in the formant frequencies (horizontal lines) is visible due to the different vowels. When we compare the panels vertically from top to bottom (/sa/ vs. /ʃa/ and /su/ vs. /ʃu/), an increase in fricative noise in the spectrum (darkness) can be seen. The change in the F3 varies more between FV syllables with an /u/ vowel (/su/ vs. /ʃu/) than for FV syllables with an /a/ vowel (/sa/ vs. /ʃa/).

The principles of the DWS model and the distinction between dynamic and static acoustic cues may also apply to children’s perceptual weighting of F0 and VTL cues for voice gender categorization. Speakers can easily modulate their mean F0 by using the laryngeal muscles to adjust the length and tension of the vocal folds, and therefore speakers’ mean F0 can vary widely. These characteristics make F0 a dynamic acoustic cue. The patterns in speakers’ mean F0, i.e., F0 contours or intonation, are also linguistically important by accenting syllables in words and sentences (Rosen, 1992; Soderstrom et al., 2003), and thus help listeners to recognize syllabic structures, similar to the dynamic formant transitions in the DWS model. In contrast to F0, speakers’ VTL can only be modified to a very limited extent, which narrows the variation to which listeners are exposed compared to F0 and makes it a relatively more static acoustic cue. In addition, listeners use speakers’ VTL to normalize formant frequencies and correctly interpret vowels (Irino &

Patterson, 2002; Ladefoged & Broadbent, 1957), which also provides information about the phonetic structure of speech segments, similar to the static fricative noise cues in the DWS model. Finally, VTL may also be a more informative cue for voice gender categorization, as speakers' mean F0, unlike their VTL, can be modulated easily within the same speaker, and therefore often overlaps across gender.

1.2.2 Relationships between different voice perception abilities

The exact relationship between children's ability to discriminate differences in speakers' voice characteristics and their abilities to categorize or use these differences for perceiving speech in competing speech remains a topic of discussion. While voice discrimination is a lower-order perception task, which only requires the listener to compare two or more stimuli in real-time, categorization relies on the stored representations of auditory categories and concepts that listeners form over time (Gibson, 1969; Pisoni, 1973). Besides categorization, the use of differences in speakers' voice characteristics for perceiving speech in competing speech seems to rely heavily on cognitive mechanisms in addition to the peripheral perceptual processing of speech (Arbogast et al., 2002; Brungart et al., 2001, 2006; Cooke et al., 2008; Mattys et al., 2009; Scott et al., 2004). For instance, Flaherty et al. (2019) found that children benefitted less than adults from voice differences in the mean F0 of the target speaker and a two-talker speech masker. However, children's benefit from F0 differences was unrelated to their F0 discrimination thresholds. As children demonstrated a benefit that was equal or even larger than that observed in adults in previous studies when there were differences in the gender of the target and the masker (Leibold et al., 2018; Wightman & Kistler, 2005), Flaherty et al. (2019) proposed that children may require larger acoustic differences, such as those in VTL, to show a benefit. This idea is also in agreement with earlier findings that children seem to require a higher amount of spectrotemporal information for processing noise-vocoded speech than adults to show the same speech recognition performance (Eisenberg et al., 2000; Mlot et al., 2010). Hence, children's reduced sensitivity to differences in speakers' voice characteristics may play a role in explaining the development of their voice perception abilities. The development in the voice and speech perception abilities of NH children also emphasizes the need for systematic investigation of these developmental effects to better understand the voice and speech perception abilities of CI children.

1.3 Aim of this dissertation

The aim of this dissertation is firstly to assess the development in NH and CI children's perception of speakers' voice characteristics as a function of age, taken as a proxy for cognitive development. More specifically, we focus on their perception of F0 and VTL cues, which are the primary acoustic cues related to speakers' voice gender, and vocal emotions. *Perception* is used as an overarching term throughout this dissertation for different voice perception abilities, such as voice discrimination, voice categorization, and a benefit from target-masker voice differences for perceiving speech in competing speech. Studying the development of the perception of voice characteristics in NH children is a prerequisite for understanding the complex interplay between the perceptual limitations imposed by the spectrotemporally degraded CI signal and developmental effects on CI children's perception of voice characteristics. In addition, the second aim of this dissertation is to examine how NH and CI children's ability to discriminate differences in voice gender cues affects other voice perception abilities. To be precise, we examine the relationship with their perceptual weighting of voice gender cues for categorization and their ability to perceive speech in competing speech based on differences in speakers' voice gender cues. Hence, the four main research questions that will be addressed in this dissertation are:

- How do the perception of voice gender cues (voice discrimination, voice categorization, and the benefit from target-masker voice differences for perceiving speech in competing speech) and vocal emotions (voice categorization) develop in NH school-age children as a function of age?
- How does NH children's ability to discriminate differences in voice gender cues relate to their perceptual weighting of voice gender cues for categorization and their ability to benefit from target-masker differences in voice gender cues for perceiving speech in competing speech?
- How does the development of an auditory system based on the spectrotemporally degraded CI signal affect the development of the perception of voice gender cues and vocal emotions in CI children?
- How does CI children's ability to discriminate differences in voice gender cues relate to their perceptual weighting of voice gender cues for categorization and their ability to benefit from target-masker differences in voice gender cues for perceiving speech in competing speech?

1.4 Scope of this dissertation

This dissertation presents four experimental studies that assess the development of voice discrimination, voice categorization, and the benefit from target-masker voice differences for perceiving speech in competing speech. We examine how these voice perception abilities, which varied from lower-order perception tasks to higher-order cognitive tasks, develop in NH and CI children as a function of age. Furthermore, we investigate the potential relationships between these different voice perception abilities by testing the same group of NH and CI children on all four experiments. The four experimental studies examine the development in the following specific voice perception abilities: (1) Discrimination of differences in voice gender cues, (2) Perceptual weighting of voice gender cues for categorization, (3) Benefitting from target-masker differences in voice gender cues for perceiving speech in competing speech, or (4) Categorization of vocal emotions. The research questions are addressed in the individual chapters as follows:

In **Chapter 2**, we investigate the development of voice gender cue discrimination and categorization in NH children during the school-age years. More specifically, we conduct an experimental study with NH children to assess how their ability to discriminate differences in F0 and VTL cues develops during the school-age years as a function of age, and how it is related to the perceptual weight attributed to F0 and VTL cues for voice gender categorization. We address these questions by examining the just-noticeable differences (JNDs), i.e., discrimination thresholds that NH children are able to perceive using a 3-interval 3-alternative forced-choice (3I-3AFC) adaptive procedure. Besides examining NH children's discrimination abilities, a lower-order perception task, we studied their perceptual weight attributed to F0 and VTL for voice gender categorization, a higher-order cognitive task, using a visual-auditory match-to-sample task. Adult-like discrimination and weighting may be acquired earlier for F0 compared to VTL due to its dynamic nature and potentially higher perceptual salience, partially based on the DWS model for phonemic categorization (Nittrouer & Miller, 1997). On the other hand, VTL may be a more informative cue for voice gender categorization, as speakers' F0 values can be modified easily within the same speaker, unlike their VTL, which causes a large overlap between male and female speakers' voices across gender. Children's perceptual weight attributed to F0 and VTL cues may differ from adults due to differences in their representations of voice gender categories, which are likely less robust, similar as hypothesized for phonemic categories (Hazan & Barrett, 2000; Nittrouer & Miller, 1997). Finally, we study the relation between the two measures to investigate if NH children's perceptual weight attributed to F0 and

VTL cues for voice gender categorization is related to their corresponding JNDs or if these two perceptual abilities rely on different mechanisms.

In **Chapter 3**, we extend the study in Chapter 2 by examining the development of the benefit from target-masker differences in voice gender cues for the perception of speech in competing speech in NH children during the school-age years. We study the development in the benefit that NH children have from voice differences in F0 and VTL cues between two competing speakers while tasked with comprehending one of them, and how this benefit is related to their F0 and VTL JNDs presented in Chapter 2. Previous research has shown that NH adult listeners benefit from target-masker voice differences in F0 and VTL for perceiving speech in competing speech (Başkent & Gaudrain, 2016; Darwin et al., 2003). However, as NH children seem less sensitive to differences in F0 and VTL cues than adults (Chapter 2), they may not benefit from the target-masker differences in F0 and VTL for the perception of speech in competing speech or to a lesser extent than adults.

In **Chapter 4**, we conduct the same experiments that are described in Chapters 2 and 3 with CI children to study how the development of their perception of voice gender cues differs from that observed in NH children. Developmental effects on the perception of voice characteristics may also be present in CI children, but their perceptual abilities may also differ from NH children in some aspects due to the development of an auditory system based on the spectrotemporally degraded CI speech signal and perceptual limitations. CI children may have reduced access to F0 and VTL cues, like postlingually deaf CI adults (e.g., Gaudrain & Başkent, 2018), which can lead to a different perceptual weighting of F0 and VTL cues for voice gender categorization and a reduced benefit from target-masker voice differences in F0 and VTL for perceiving speech in competing speech. However, CI children may differ from CI adults in these aspects because of the different manners in which they acquired language, which also seems to affect their processing of spectrotemporal information (DiNino & Arenberg, 2018; Jung et al., 2012; Landsberger et al., 2018).

In **Chapter 5**, we conduct a final experimental study on how the ability to categorize vocal emotions develops in NH children and CI children as a function of age. We address this question by studying NH and CI children's abilities to correctly categorize three core emotions (happy, angry, and sad) based on pseudospeech sentence productions by six different Dutch speakers. In the literature, the ability to recognize speakers' vocal emotions is commonly referred to by the term *vocal emotion recognition*. In this dissertation, we instead use the term *vocal emotion categorization* to describe what participants have to do in our task, as the task specifically tests the listener's ability to categorize speakers' vocal emotions into

distinct categories and thus resembles the earlier mentioned voice gender categorization task. Similar to voice gender categorization, the categorization of vocal emotions is a higher-order cognitive task that relies on long-term auditory exposure and stored representations of auditory categories (Gibson, 1969; Pisoni, 1973). Based on earlier reports (Sauter et al., 2013; Tonks et al., 2007), we expect to see a strong development in the ability of NH children to categorize vocal emotions as a function of age during the school-age years. This development may be related to differences in children's perceptual weighting of acoustic cues and less robust representations of auditory categories compared to adults. The development as a function of age may be less strong in CI children due to potential perceptual limitations related to the spectrotemporally degraded CI signal (Nakata et al., 2012).

Finally, in **Chapter 6**, a discussion is presented of the overall findings of the experimental studies that are included in this dissertation, which test different aspects of voice perception using multiple perceptual tasks. The results generally show strong developmental effects on the different voice perception abilities of NH children, which continue to improve to adult-like levels during the school-age years, even though their peripheral hearing system is already largely mature at birth. For CI children, the development of their voice perception abilities is even more difficult to interpret because of the combination of developmental effects and perceptual limitations imposed by the spectrotemporally degraded CI signal. Together, these results illustrate the complex interactions between different voice perception abilities and the intricate relationship between auditory peripheral and developmental factors in NH and CI children.

Bibliography

- Abdala, C., & Keefe, D. H. (2012). Morphological and functional ear development. In L. Werner, R. R. Fay, & A. N. Popper (Red.), *Human Auditory Development* (pp. 19–59). Springer New York.
- Arbogast, T. L., Mason, C. R., & Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America*, 112(5), 2086–2098.
- Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., Sarampalis, A., Wagner, A., & Gaudrain, E. (2016a). Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging: How and to what degree can it be achieved? *Trends in Hearing*, 20, 1–16.
- Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, 139(3), EL51–EL56.
- Başkent, D., Gaudrain, E., Tamati, T. N., & Wagner, A. (2016b). Perception and psychoacoustics of speech in cochlear implant users. In A. T. Cacace, E. de Kleine, A. G. Holt, & P. van Dijk (Red.), *Scientific Foundations of Audiology: Perspectives from Physics, Biology, Modeling, and Medicine* (pp. 285–319). Plural Publishing, Inc.
- Başkent, D., & Shannon, R. V. (2004). Frequency-place compression and expansion in cochlear implant listeners. *The Journal of the Acoustical Society of America*, 116(5), 3130–3140.
- Bennett, S., & Weinberg, B. (1979). Sexual characteristics of preadolescent childrens' voices. *The Journal of the Acoustical Society of America*, 65(1), 179–189.
- Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. F., ... Lazard, D. S. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiology & Neurotology*, 18(1), 36–47.
- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press.
- Bruijnzeel, H., Bezdjian, A., Lesinski-Schiedat, A., Illg, A., Tzifa, K., Monteiro, L., Volpe, A. della, Grolman, W., & Topsakal, V. (2017). Evaluation of pediatric cochlear implant care throughout Europe: Is European pediatric cochlear implant care performed according to guidelines? *Cochlear Implants International*, 18(6), 287–296.
- Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America*, 120(6), 4007–4018.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, 110(5), 2527–2538.

- Buss, E., Flaherty, M. M., & Leibold, L. J. (2017). Development of frequency discrimination at 250 Hz is similar for tone and /ba/ stimuli. *The Journal of the Acoustical Society of America*, 142(1), EL150–EL154.
- Chatterjee, M., & Peng, S.-C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research*, 235(1–2), 143–156.
- Chatterjee, M., Zion, D. J., Deroche, M. L. D., Burianek, B. A., Limb, C. J., Goren, A. P., Kulkarni, A. M., & Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research*, 322, 151–162.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5), 975–979.
- Chiba, T., & Kajiyama, M. (1958). *The vowel: Its nature and structure* (Vol. 652). Phonetic society of Japan Tokyo.
- Clark, G. M. (2015). The multi-channel cochlear implant: Multi-disciplinary development of electrical stimulation of the cochlea and the resulting clinical benefit. *Hearing Research*, 322, 4–13.
- Cleary, M., & Pisoni, D. B. (2002). Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results. *Annals of Otology, Rhinology & Laryngology*, 111(5_suppl), 113–118.
- Cleary, M., Pisoni, D. B., & Kirk, K. I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 48(1), 204–223.
- Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, 123(1), 414–427.
- Cowan, N. (1998). *Attention and memory: An integrated framework*. Oxford University Press.
- Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America*, 114(5), 2913–2922.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208(4448), 1174–1176.
- DiNino, M., & Arenberg, J. G. (2018). Age-related performance on vowel identification and the spectral-temporally modulated ripple test in children with normal hearing and with cochlear implants. *Trends in Hearing*, 22, 2331216518770959.
- Donaldson, G. S., Kreft, H. A., & Litvak, L. (2005). Place-pitch discrimination of single-versus dual-electrode stimuli by cochlear implant users. *The Journal of the Acoustical Society of America*, 118(2), 623–626.
- Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America*, 107(5), 2704–2710.

- El Boghdady, N., Gaudrain, E., & Başkent, D. (2019). Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users? *The Journal of the Acoustical Society of America*, 145(1), 417–439.
- Everhardt, M. K., Sarampalis, A., Coler, M., Baskent, D., & Lowie, W. (2020). *Meta-Analysis on the Identification of Linguistic and Emotional Prosody in Cochlear Implant Users and Vocoder Simulations*. LWW.
- Fant, G. (1960). *Acoustic theory of speech production*. Mouton.
- Fecher, N., Paquette-Smith, M., & Johnson, E. K. (2019). Resolving the (apparent) talker recognition paradox in developmental speech perception. *Infancy*, 24(4), 570–588.
- Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10(3), 279–293.
- Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106(3), 1511–1522.
- Flaherty, M. M., Buss, E., & Leibold, L. J. (2019). Developmental effects in children’s ability to benefit from F0 differences between target and masker speech. *Ear and Hearing*, 40(4), 927–937.
- Floccia, C., Nazzi, T., & Bertoni, J. (2000). Unfamiliar voice discrimination for short stimuli in newborns. *Developmental Science*, 3(3), 333–343.
- Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology*, 6(1), 19–27.
- Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., & Başkent, D. (2014). Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology*, 15(6), 1037–1048.
- Gaudrain, E., & Başkent, D. (2015). Factors limiting vocal-tract length discrimination in cochlear implant simulations. *The Journal of the Acoustical Society of America*, 137(3), 1298–1308.
- Gaudrain, E., & Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear and Hearing*, 39(2), 226–237.
- Geurts, L., & Wouters, J. (2001). Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *The Journal of the Acoustical Society of America*, 109(2), 713–726.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton-Century-Crofts.
- Gray, H. (1878). *Anatomy of the human body* (Vol. 8). Lea & Febiger.
- Green, T., Faulkner, A., & Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *The Journal of the Acoustical Society of America*, 116(4), 2298–2310.
- Green, T., Faulkner, A., Rosen, S., & Macherey, O. (2005). Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification. *The Journal of the Acoustical Society of America*, 118(1), 375–385.

- Harries, M., Hawkins, S., Hacking, J., & Hughes, I. (1998). Changes in the male voice at puberty: Vocal fold length and its relationship to the fundamental frequency of the voice. *The Journal of Laryngology & Otology*, 112(5), 451–454.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4), 377–396.
- Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., Gotter, B. D., Vanderhoof, S. S., Mispagel, K., Heydebrand, G., & Skinner, M. W. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear and Hearing*, 34(3), 342–360.
- Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., & Johnson, K. (1999). Formants of children, women, and men: The effects of vocal intensity variation. *The Journal of the Acoustical Society of America*, 106(3), 1532–1542.
- Irino, T., & Patterson, R. D. (2002). Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform. *Speech Communication*, 36(3), 181–203.
- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, 14(5), 1002–1011.
- Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 363–389). Blackwell Publishing Ltd.
- Jung, K. H., Won, J. H., Drennan, W. R., Jameyson, E., Miyasaki, G., Norton, S. J., & Rubinstein, J. T. (2012). Psychoacoustic Performance and Music and Speech Perception in Prelingually Deafened Children with Cochlear Implants. *Audiology and Neurotology*, 17(3), 189–197.
- Kern, A., Heid, C., Steeb, W.-H., Stoop, N., & Stoop, R. (2008). Biophysical parameters modification could overcome essential hearing gaps. *PLoS Computational Biology*, 4(8), e1000161.
- Kirk, K. I., Miyamoto, R. T., Lento, C. L., Ying, E., O'Neill, T., & Fears, B. (2002). Effects of age at implantation in young children. *The Annals of Otology, Rhinology & Laryngology*, 111(5), 69–73.
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, 87(2), 820–857.
- Kong, Y.-Y., Deeks, J. M., Axon, P. R., & Carlyon, R. P. (2009). Limits of temporal pitch in cochlear implants. *The Journal of the Acoustical Society of America*, 125(3), 1649–1657.
- Kral, A. (2013). Auditory critical periods: A review from system's perspective. *Neuroscience*, 247, 117–133.
- Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons.
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29(1), 98–104.

- Landsberger, D. M., Padilla, M., Martinez, A. S., & Eisenberg, L. S. (2018). Spectral-temporal modulated ripple discrimination by children with cochlear implants. *Ear and Hearing*, 39(1), 60–68.
- Laneau, J., & Wouters, J. (2004). Multichannel place pitch sensitivity in cochlear implant recipients. *Journal of the Association for Research in Otolaryngology*, 5(3), 285–294.
- Laneau, J., Wouters, J., & Moonen, M. (2004). Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *The Journal of the Acoustical Society of America*, 116(6), 3606–3619.
- Lazard, D. S., Vincent, C., Venail, F., Heyning, P. V. de, Truy, E., Sterkers, O., Skarzynski, P. H., Skarzynski, H., Schauwers, K., O’Leary, S., Mawman, D., Maat, B., Kleine-Punte, A., Huber, A. M., Green, K., Govaerts, P. J., Fraysse, B., Dowell, R., Dillier, N., ... Blamey, P. J. (2012). Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: A new conceptual model over time. *PLOS ONE*, 7(11), 1–11.
- Leibold, L. J., Buss, E., & Calandruccio, L. (2018). Developmental effects in masking release for speech-in-speech perception due to a target/masker sex mismatch. *Ear and Hearing*, 39(5), 935–945.
- Levi, S. V., & Schwartz, R. G. (2013). The development of language-specific and language-independent talker processing. *Journal of Speech, Language, and Hearing Research*, 56(3), 913–925.
- Litovsky, R. (2015). Development of the auditory system. In *Handbook of Clinical Neurology* (Vol. 129, pp. 55–72). Elsevier.
- Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, 27(1), 153–165.
- Manrique, M., Cervera-Paz, F. J., Huarte, A., Perez, N., Molina, M., & García-Tapia, R. (1999). Cerebral auditory plasticity and cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 49, S193–S197.
- Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59(3), 203–243.
- McCullough, E. A., Clopper, C. G., & Wagner, L. (2019). Regional dialect perception across the lifespan: Identification and discrimination. *Language and Speech*, 62(1), 115–136.
- Meister, H., Landwehr, M., Pyschny, V., Walger, M., & Wedel, H. von. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *International Journal of Audiology*, 48(1), 38–48.
- Meister, H., Walger, M., Lang-Roth, R., & Müller, V. (2020). Voice fundamental frequency differences and speech recognition with noise and speech maskers in cochlear implant recipients. *The Journal of the Acoustical Society of America*, 147(1), EL19–EL24.
- Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development*, 6(2), 313–330.

- Miller, C. L., Younger, B. A., & Morse, P. A. (1982). The categorization of male and female voices in infancy. *Infant Behavior and Development*, 5(2), 143–159.
- Mlot, S., Buss, E., & Hall III, J. W. (2010). Spectral integration and bandwidth effects on speech recognition in school-aged children and adults. *Ear and Hearing*, 31(1), 56–62.
- Moore, J. K., & Linthicum Jr., F. (2007). The human auditory system: A timeline of development. *International Journal of Audiology*, 46(9), 460–478.
- Mudry, A., & Mills, M. (2013). The early history of the cochlear implant: A retrospective. *JAMA Otolaryngology–Head & Neck Surgery*, 139(5), 446–453.
- Nagels, L., Bastiaanse, R., Başkent, D., & Wagner, A. (2020). Individual differences in lexical access among cochlear implant users. *Journal of Speech, Language, and Hearing Research*, 63(1), 286–304.
- Nakata, T., Trehub, S. E., & Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *The Journal of the Acoustical Society of America*, 131(2), 1307–1314.
- Nicholas, J. G., & Geers, A. E. (2007). Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to profound hearing loss. *Journal of Speech, Language, and Hearing Research*, 50(4), 1048–1062.
- Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *The Journal of the Acoustical Society of America*, 101(4), 2253–2266.
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3), 355–376.
- Oxenham, A. J. (2013). Revisiting place and temporal theories of pitch. *Acoustical science and technology / edited by the Acoustical Society of Japan*, 34(6), 388–396.
- Peng, S.-C., Chatterjee, M., & Lu, N. (2012). Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends in Amplification*, 16(2), 67–82.
- Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, 109(6), 2988–2998.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253–260.
- Pisoni, D. B., Kronenberger, W. G., Harris, M. S., & Moberly, A. C. (2018). Three challenges for future research on cochlear implants. *World Journal of Otorhinolaryngology-Head and Neck Surgery*, 3(4), 240–254.
- Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 336(1278), 367–373.
- Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, 31(1), 97–113.
- Scott, S. K., & McGettigan, C. (2016). The voice: From identity to interactions. In D. Matsumoto, H. C. Hwang, & M. G. Frank (Eds.), *APA handbook of nonverbal communication*. (2015-02851-011; pp. 289–305). American Psychological Association.

- Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *The Journal of the Acoustical Society of America*, 115(2), 813–821.
- Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language & Hearing Research*, 57(1), 285–296.
- Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, 118(5), 3177–3186.
- Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49(2), 249–267.
- Spence, M. J., & Freeman, M. S. (1996). Newborn infants prefer the maternal low-pass filtered voice, but not the maternal whispered voice. *Infant Behavior and Development*, 19(2), 199–212.
- Stickney, G. S., Zeng, F.-G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America*, 116(2), 1081–1091.
- Svirsky, M. A., Chin, S. B., & Jester, A. (2007). The effects of age at implantation on speech intelligibility in pediatric cochlear implant users: Clinical outcomes and sensitive periods. *Audiological Medicine*, 5(4), 293–306.
- Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, 85(4), 1699–1707.
- Tonks, J., Williams, W. H., Frampton, I., Yates, P., & Slater, A. (2007). Assessing emotion recognition in 9–15-years olds: Preliminary analysis of abilities in reading emotion from faces, voices and eyes. *Brain Injury*, 21(6), 623–629.
- Townshend, B., Cotter, N., Van Compernelle, D., & White, R. L. (1987). Pitch perception by cochlear implant subjects. *The Journal of the Acoustical Society of America*, 82(1), 106–115.
- Vestergaard, M. D., Háden, G. P., Shtyrov, Y., Patterson, R. D., Pulvermüller, F., Denham, S. L., Sziller, I., & Winkler, I. (2009). Auditory size-deviant detection in adults and newborn infants. *Biological Psychology*, 82(2), 169–175.
- Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, 50(6), 1510–1545.
- Werner, L. A. (2007). Issues in human auditory development. *Journal of Communication Disorders*, 40(4), 275–283.
- Werner, L. A., & Leibold, L. J. (2011). Auditory development in normal-hearing children. In J. S. Gravel, R. Sewald, & A. M. Tharpe (Red.), *Handbook of Pediatric Audiology*. Sage Publications.

- Wightman, F. L., & Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*, 118(5), 3164–3176.
- Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: A remarkable past and a brilliant future. *Hearing Research*, 242(1), 3–21.
- Zeng, F.-G. (2002). Temporal pitch in electric hearing. *Hearing Research*, 174(1), 101–106

2 | Development of voice perception is dissociated across gender cues in school-age children

Published as:

L. Nagels, E. Gaudrain, D. Vickers, P. Hendriks, and D. Başkent (2020). Development of voice perception is dissociated across gender cues in school-age children. *Scientific Reports*, 10(1).

Abstract

Children's ability to distinguish speakers' voices continues to develop throughout childhood, yet it remains unclear how children's sensitivity to voice cues, such as differences in speakers' gender, develops over time. This so-called voice gender is primarily characterized by speakers' mean fundamental frequency (F0), related to glottal pulse rate, and vocal-tract length (VTL), related to speakers' size. Here we show that children's acquisition of adult-like performance for voice discrimination, a lower-order perception task, and voice categorization, a higher-order cognitive task, differs across voice gender cues. Children's discrimination was adult-like around the age of 8 for VTL but still differed from adults at the age of 12 for F0. Children's perceptual weight attributed to F0 for voice gender categorization was adult-like around the age of 6 but around the age of 10 for VTL. Children's discrimination and weighting of F0 and VTL were only correlated for 4- to 6-year-olds. Hence, children's development of discrimination and weighting of voice gender cues are dissociated, i.e., adult-like performance for F0 and VTL is acquired at different rates and does not seem to be closely related. The different developmental patterns for voice discrimination and voice categorization highlight the complexity of the relationship between perceptual and cognitive mechanisms of voice perception.

2.1 Introduction

Voice cues enable listeners to recognize and distinguish speakers, which is imperative for speech-related tasks, such as the perception of speech in noisy environments. At an early age, children are already sensitive to differences in voice cues, especially for familiar speakers (DeCasper & Fifer, 1980) and in their native language (Johnson et al., 2011). Furthermore, infants are already sensitive to differences in voice cues, such as fundamental frequency (F0) (Olsho et al., 1988) or voice pitch (Spence & Freeman, 1996), timbre differences associated with vocal-tract length (Vestergaard et al., 2009), or prosody (Fernald & Kuhl, 1987). On the other hand, there is a clear prolonged development in children's ability to encode and recognize speakers' voice characteristics (Creel & Jimenez, 2012; Fecher et al., 2019; Mann et al., 1979). Hence, it is unclear how children's sensitivity to differences in voice cues develops over time.

Children's ability to discriminate differences in voice-related acoustic cues, such as F0 (Buss et al., 2017; Flaherty et al., 2019; Jensen & Neff, 1993; Maxon & Hochberg, 1982) or temporal cues (Irwin et al., 1985; Wightman et al., 1989),

develops throughout childhood. Yet, the specific age at which children's voice discrimination thresholds are adult-like has been debated, partially due to differences in experiment stimuli. Earlier research has primarily used non-voice stimuli, such as pure tones (Jensen & Neff, 1993; Maxon & Hochberg, 1982), or octave-band (Wightman et al., 1989) and narrow-band noises (Irwin et al., 1985), where the results were interpreted and extrapolated for voice perception. In contrast, research directly using voice stimuli has been scarce (Buss et al., 2017; Flaherty et al., 2019). Further challenges come from task demands, as these can influence outcomes, and the high variability among children's performance, even after controlling for effects of age (Deroche et al., 2012). Children's ability to categorize voice and speech cues, a higher-order cognitive task, also continues to develop even after 12 years of age, e.g., as observed for phonemic categorization (Hazan & Barrett, 2000; Nitttrouer & Miller, 1997). These higher-order cognitive identification tasks depend on long-term exposure to voice cues and stored representations of voice categories in addition to perceptual processing (Gibson, 1969; Pisoni, 1973). Despite these shared features, the source of the relatively slow development in sensitivity to differences in voice cues and the relation between discrimination and categorization of voice cues are not well understood. Some parts of the auditory system only become mature during or after childhood, such as the auditory cortex, as evidenced by developmental effects on the latency of evoked cortical potentials (Moore & Linthicum Jr., 2007), but children's sensitivity to subtle acoustic cues seems to be also largely determined by cognitive development and auditory experience (Litovsky, 2015; Werner, 2007).

Mann, Diamond, and Carey (1979) have proposed that children's difficulties with encoding speakers' voice characteristics are caused by changes in central auditory processing related to right hemisphere maturation. This explanation is partially supported by findings of recent neuroimaging studies which show that listeners rely on areas in both hemispheres to process differences in speakers' voices, but predominantly on the right hemisphere (Belin et al., 2004; Kreitewolf et al., 2014). Furthermore, Mann et al. (1979) proposed that the development in children's ability to encode unfamiliar speakers' voice characteristics may be caused by differences in their processing strategies and their reliance on different acoustic cues compared to adults. This hypothesis is in agreement with the Developmental Weighting Shift (DWS) model for the categorical perception of phonemes by Nitttrouer and Miller (1997), which states that children weigh dynamic acoustic cues of speech (such as formant transitions) more than static acoustic speech cues (such as noise segments of consonants) due to their higher perceptual salience and more informative properties. The usage of static acoustic cues seems to require more

auditory experience and language exposure for children to attend to these cues. A similar explanation may apply to the development of children's sensitivity to different acoustic voice cues, such as voice gender cues.

The perceived gender of a voice results from a combination of many acoustic features (Skuk & Schweinberger, 2014; Titze, 1989). However, it is primarily defined by speakers' mean F0, related to the glottal pulse rate, and their vocal-tract length (VTL), related to the size of the speaker (Fitch & Giedd, 1999). The mean F0 of female voices is around 70%, approximately 12 semitones (st), equal to one octave, higher than male voices (Klatt & Klatt, 1990), but the range of F0 values of male and female speakers' voices largely overlaps, as F0 can be modulated easily within the same speaker. Therefore, listeners are exposed to a wide range of F0 variations. F0 contours indicate a word's meaning in tonal languages and play a large role in prosodic differences that convey, among others, the emotions and attitudes of speakers, and they are an important cue for infants to recognize their mother's voice (Spence & Freeman, 1996). Therefore, F0, with its information-bearing dynamic fluctuations, seems to parallel dynamic formant transitions in the DWS model and may consequently be subject to similar principles.

The female vocal-tract is around 20% shorter than the male vocal-tract, causing a difference of around 3.6 st in formant frequencies (Titze, 1989). Estimation of speakers' VTL is essential for listeners to recognize speech, as the formant frequency distributions of vowels depend on speakers' VTL size (Irino & Patterson, 2002). Differences in speakers' mean F0 or their VTL facilitate the perception of speech in competing speech maskers for adult listeners (Başkent & Gaudrain, 2016; Darwin et al., 2003). But differently than F0, VTL cues can be modified only to a very limited extent within the same speaker, which narrows the range of differences and fluctuations that listeners are exposed to in VTL relative to F0, resembling the static cues of the DWS model. Previous research on voice gender categorization in adult listeners has also indicated that differences in dynamic F0 cues may be more perceptually salient (Lass et al., 1976; Pernet & Belin, 2012) than static VTL cues. Finally, there are differences in the mean F0 and VTL values of male and female voices across speakers of different languages and modalities. For an overview of speakers' mean F0 values across different languages and modalities see Traunmüller and Eriksson (1995), and for an overview of differences in speakers' average height per country, the main determinant of speakers' VTL (Fitch & Giedd, 1999), see Roser et al. (2020).

For children's voices, voice gender is differentiated almost exclusively by formant frequencies, as there is a large overlap in the mean F0 values of male and female children's voices until the age of 12 (Perry et al., 2001; Vorperian & Kent,

2007). For instance, Perry et al. (2001) found that the formant frequencies of 8-year-old boys were approximately 9% lower than those of 8-year-old girls, while their F0 values did not significantly differ from each other. As there are also no clear differences in the VTL of male and female children before 12 years of age, Vorperian and Kent (2007) have proposed that the gender differences in children's formant frequencies may arise from differences in the resonator width and size instead of the length of the vocal-tract. For face recognition, Anastasi and Rhodes (2005) found that children and adults are better at recognizing faces from people of their own age. This 'own-age bias' may also apply to children's ability to recognize voices. Consequently, children may attend to different acoustic cues than adults, and hence have different representations of voice gender categories than adults.

In the present study, we investigated if children's ability to discriminate differences in voice gender cues (Experiment 1) and their perceptual weighting of voice gender cues for categorization (Experiment 2) develop similarly for F0 as for VTL cues, the primary cues of speakers' voice gender, and at what age these abilities are adult-like. Previous studies have shown that children's discrimination of F0 continues to develop throughout childhood, but some of these studies measured pure-tone frequency discrimination using non-voice stimuli (Allen et al., 1989; Maxon & Hochberg, 1982) instead of discrimination of more realistic voice pitch cues using voice stimuli (Buss et al., 2017; Flaherty et al., 2019). Regarding VTL, discrimination has only been studied before in infants and only using EEG, which, while providing some evidence for the processing of the acoustic traits associated with VTL, does not necessarily reflect perception (Vestergaard et al., 2009). Thus, the development of children's ability to discriminate differences in voice gender cues has not been previously studied in a systematic way. Furthermore, the relation between voice discrimination, a lower-order perception task, and voice categorization, a higher-order cognitive process, is not well understood. If children's ability to discriminate F0 and VTL cues is indeed reduced, their weighting may differ from adults' due to perceptual limitations. However, as discrimination is a lower-order perception task and categorization is a higher-order cognitive task, the development of both abilities may also be dissociated. We conducted two experiments to measure children's just-noticeable differences (JNDs), i.e., discrimination thresholds, for F0 and VTL (Experiment 1) and their perceptual weighting of F0 and VTL cues for the categorization of speakers' voice gender (Experiment 2). We hypothesized that children are more sensitive to differences in F0 than VTL, as F0 is a dynamic cue that varies widely among and within speakers, and therefore listeners are exposed to wide variations in F0. These characteristics may make F0 more perceptually salient than VTL, which can be modified only to a

very limited extent within the same speaker, and hence lead to a higher weighting of F0, similar to the principles of the DWS model. Alternatively, due to the high overlap in speakers' F0 values across gender, VTL may be a more informative cue for determining speakers' voice gender, and hence receive a higher weighting than F0. In addition, we expected children to show different weighting of F0 and VTL cues for voice gender categorization than adults due to differences in their representations of voice gender categories, based on differences in exposure to voice cues and an 'own-age bias' (Anastasi & Rhodes, 2005). Finally, we investigated if children's discrimination and weighting for voice gender categorization develop in a similar manner for F0 and VTL, as these abilities may rely on different auditory processes.

2.2 Discrimination of differences in F0 and VTL cues

2.2.1 Methods

Participants

Fifty-eight children between the ages of 4 and 12, divided into four different participant age groups (Table 2.1), and fifteen adults between the ages of 20 and 30 took part in the study. All participants were native speakers of Dutch, and reported no history of hearing or language disorders. We screened participants' hearing thresholds at 20 dB HL using pure-tone audiometry at octave frequencies between 500 and 4000 Hz to ensure they were normal-hearing. Table 2.1 summarizes participants' demographic characteristics. Children's vocabulary size was measured using the Dutch version of the Renfrew Word Finding Vocabulary Test (Renfrew, 1995) to ensure age-normal linguistic development in our test population.

Table 2.1: Demographic characteristics of participants per age group. Age is given in years, rounded to two decimal places. Vocabulary corresponds to scores on the Renfrew Word Finding Vocabulary Test (maximum of 50 points).

Participant age groups	Number of participants	Age (median)	Vocabulary size (median)
Children 4-6 years	13	5.08	36
Children 6-8 years	13	7.17	40
Children 8-10 years	16	9.00	45
Children 10-12 years	16	11.10	46
Adults	15	24.40	N/A

A written informed consent form for study participation was signed by the parents and/or legal guardians of children and by adult participants prior to data collection. Ethical approval of the study was given by the Medical Ethical Review Committee of the University Medical Center Groningen (METc 2016.689). All experiments and methods were performed in accordance with the relevant guidelines and regulations.

Stimuli and apparatus

The stimuli consisted of three-syllable CVCVCV pseudowords, for instance, ‘ba-ki-mo’, that were spliced from Dutch CVC words, taken from the NVA corpus (Bosman & Smoorenburg, 1995). The stimuli were produced by a female native speaker of Dutch with a mean F0 of 242 Hz and an estimated VTL size of approximately 13.5 cm, based on the average height of Dutch women of 168.72 cm (Roser et al., 2020). For the experiment, we focused on the differences in F0 and VTL relative to the original speaker’s voice. As both F0 and VTL alterations lead to differences in frequencies, we expressed the differences as ratios measured on a logarithmic scale in semitones (st) instead of differences in Hertz or centimeters. The word recordings were spliced into 61 CV syllables, equalized for root-mean-square (RMS) level, and analyzed using STRAIGHT (Kawahara & Irino, 2005), to obtain the F0 contour and the spectral envelope. The method that we used is the same as employed by Gaudrain and Başkent (2018).

For each trial, three CV syllable pairs were randomly selected, normalized to a duration of 200 ms and a mean F0 of 242 Hz, and resynthesized with STRAIGHT using the new F0 and VTL parameters. This resynthesis procedure was performed for all stimuli to prevent potential artifacts arising from the resynthesis procedure itself. The three syllables were then concatenated with 50 ms of silence separating the individual syllables. The overall F0 contour of a CVCVCV pseudoword was adjusted by shifting the natural F0 contour of each CV syllable by random steps of $\frac{1}{3}$ st, which could be $-\frac{1}{3}$ st, 0 st or $+\frac{1}{3}$ st, to make them sound more natural. After adjusting the F0 contour of each CV syllable, the overall F0 contour of the pseudoword was centered to zero again so that it would not change the mean F0. For the F0 manipulation, the F0 contour (in Hz) was multiplied by a specified factor, which depended on the applied change in mean F0, to preserve the original F0 fluctuations and only adjust the mean F0. For the VTL manipulation, the spectral envelope was compressed toward the lower frequencies to generate an overall spectral shift. The modified F0 contour and spectral envelope were then recombined using a pitch synchronous overlap-add (PSOLA) resynthesis method.

In each trial, the same pseudoword with the same syllable structure was used for the target pseudoword and the two reference pseudowords. The target pseudoword differed in mean F0 or VTL from the two reference pseudowords. Besides this manipulation, there were only small differences in the mean F0 contour of the pseudowords to make them sound more natural. Depending on the condition, the target pseudoword was either produced with a lower F0 or a larger VTL value relative to the original female speaker's voice, which made the target pseudoword sound more male-like. All signal processing was performed in Matlab using a sampling frequency of 44.1 kHz.

Procedure

We used a 3-interval, 3-alternative forced-choice (3I-3AFC) adaptive procedure to measure participants' JNDs for F0 and VTL. The experiment started with two training sessions, each consisting of three trials, and was followed by two experiment sessions of approximately 25 to 45 trials, depending on participants' responses. The experiment sessions were presented to participants in a randomized order, either measuring their F0 JND or their VTL JND first.

Each experiment session started with a difference of 12 st in F0 or VTL relative to the original female speaker's voice. The voice difference decreased by one step size after two consecutive correct responses and increased by one step size after an incorrect response (2-down, 1-up). The initial step size was 2 st but after 15 trials with the same step size or when the difference became smaller than two times the step size, the step size was divided by $\sqrt{2}$. The experiment session ended after eight reversals. Finally, we calculated the geometric mean of the voice difference values at the last six reversals to determine participants' JND, which corresponds to the 70.7% correct discrimination point on the psychometric function (Levitt, 1971).

Stimuli were presented to participants on a touchscreen laptop via Sennheiser HD 380 Pro headphones at a sound level of 65 dBA. A child-friendly interface was constructed using Matlab to maintain children's attention (Figure 2.1b). In each trial, three sea animals appeared on the screen and subsequently produced the same CVCVCV pseudoword. Participants were instructed to click on the sea animal whose voice differed from the other two sea animals. Visual feedback on response accuracy was provided to participants. Children were tested in a quiet room at their homes, and adults were tested in a quiet testing room at the University of Groningen.

Data analysis

Participants' JNDs for F0 and VTL were analyzed using the lme4 package (version 1.1.12, Bates et al., 2015) in R (version 3.4.1, R Core Team, 2020). A linear mixed-effects model with a two-way interaction between *voice cue* (F0 or VTL) and *age*, and a random intercept per participant for children's data was computed to examine if children's discrimination of F0 and VTL cues developed similarly. We evaluated the models using backward stepwise selection with ANOVA Chi-Square tests, starting with the full factorial model, in lme4 syntax: `log(JND) ~ voice cue * age + (1 | participant)`. We used log-transformed JNDs, which are more normally distributed than the raw values, as the JND procedure does not allow crossing the zero line. Furthermore, the log-transformation reduced the differences in variability across voice cues and decoupled it from the mean. *Voice cue* indicated whether the JND was for F0 or VTL, while *age* represented children's age in years. In addition, to investigate at what age children's discrimination of F0 and VTL no longer differs from adults, we used the DescTools package (version 0.99.25, Signorell et al., 2018) to perform two Dunnett's Tests on subsets of F0 and VTL JNDs with *log(JND)* as an outcome variable and *age group* as a fixed effect.

2.2.2 Results

Development of the ability to discriminate differences in F0 and VTL cues

Figure 2.1a shows participants' JNDs for F0 (left) and VTL (right) as a function of their age (dots) and the median JND per participant age group (boxplots), Figure 2.1b shows the experiment interface, and Figure 2.1c shows the F0 and VTL JND adaptive tracks of one adult and one child participant. We first performed a Shapiro-Wilk test to confirm that the assumption for LMMs of normally distributed residuals for our data was met ($W = 0.98$, $P = 0.13$). Model selection revealed that the model with *voice cue* and *age* as fixed effects had a better and more parsimonious fit than the model with the two-way interaction between *voice cue* and *age* ($\chi^2(1) = 0.02$, $P = 0.89$), or the models without *voice cue* ($\chi^2(1) = 24.80$, $P < 0.001$) or without *age* ($\chi^2(1) = 28.01$, $P < 0.001$) as fixed effects. The significant effects of *voice cue* and *age* indicate that children's JNDs were significantly lower for VTL than for F0, ($F(1,56) = 30.95$, $t = -5.56$, $P < 0.001$), and their F0 and VTL JNDs decreased significantly with age ($F(1,56) = 36.01$, $t = -6.00$, $P < 0.001$). However, the rate of improvement, i.e., the slope size, did not significantly differ across voice cues, which is evidenced by the lack of a significant interaction between *voice cue* and *age*.

Figure 2.1: Differences in JNDs for F0 and VTL, and the experiment interface of Experiment 1. (a) Log-transformed JNDs for F0 (left panel) and VTL (right panel). The dots show individual data points located at participants' age in years, rounded to two decimal places ($N_{\text{children}} = 58$, $N_{\text{adults}} = 15$). The boxplots depict the median JND value per age group, and the box shows the lower and upper quartiles. The whiskers show the lowest and highest data points within plus or minus 1.5 times the interquartile range. (b) The experiment interface. As participants give correct answers, they collect sea animals around the central yellow fish. The illustrations were made by Jop Luberti for the purpose of this study. This image is published under the CC-BY-NC 4.0 license. (c) The F0 and VTL JND adaptive tracks of one adult (NHA012) and one child participant (NHK034).

Adult-like discrimination of differences in F0 and VTL cues

We performed two Dunnett's Tests on subsets of F0 and VTL JNDs per participant age group. For F0, the JNDs of adults significantly differed from the JNDs of children of all age groups (4-6 years difference = 1.85, $P < 0.001$; 6-8 years difference = 1.38, $P < 0.001$; 8-10 years difference = 1.01, $P < 0.001$; 10-12 years difference = 0.75, $P < 0.01$). For VTL, adults' JNDs significantly differed from the JNDs of 4- to 8-year-olds (4-6 years difference = 1.07, $P < 0.001$; 6-8 years difference = 0.62, $P < 0.01$) but not from the JNDs of 8- to 12-year-olds (8-10 years difference = 0.30, $P = 0.30$; 10-12 years difference = 0.09, $P = 0.96$).

2.2.3 Discussion

Our results show that children's discrimination of F0 and VTL improves gradually with age, but discrimination of VTL becomes adult-like earlier than their discrimination of F0. Adults' VTL JNDs did not significantly differ from those of 8- to 12-year-old children, whereas adults' F0 JNDs differed from children's F0 JNDs at all ages. The improvement in children's ability to discriminate differences in F0 with age is in line with previously reported research (Buss et al., 2017; Flaherty et al., 2019; Jensen & Neff, 1993; Maxon & Hochberg, 1982). The differences in the reported ages at which children's discrimination of F0 differences is adult-like are likely due to differences in stimuli and task demands. The perception of VTL differences has only been studied before in infants and adults, using an objective EEG mismatched negativity paradigm (Vestergaard et al., 2009). Vestergaard et al. (2009) found that newborn infants are already sensitive to large differences in VTL, but it was not possible to quantify infants' ability to discriminate VTL and to assess whether it was adult-like yet. To our knowledge, there is no prior information regarding the development of school-age children's discrimination of VTL differences.

Our findings do not support our expectation that children would show better discrimination of differences in F0 than VTL cues due to exposure to wide variations in F0, even within the same speaker, compared to small variations in VTL, which only differs markedly between different speakers. According to the DWS model (Nitttrouer & Miller, 1997), children are more sensitive to the most salient and informative acoustic cue, which can be F0 because of its high perceptual salience (Lass et al., 1976; Pernet & Belin, 2012) or VTL because of the smaller overlap across genders and potentially higher informativeness for determining speakers' voice gender. On a linguistic level, VTL also seems to be more informative, as it is required for the correct interpretation of vowels (Irino & Patterson, 2002), while F0 primarily provides paralinguistic information in non-tonal languages. However, the DWS model is also based on categorization rather than discrimination of acoustic cues. Acoustic dynamics may have different effects on discrimination than on categorization. Evidence for a potential trade-off effect between the discrimination and categorization of acoustic cues based on acoustic dynamics is provided by Mirman et al. (2004). They found that adult listeners could more accurately discriminate sounds that differed in steady-state acoustic cues than rapidly-changing, dynamic acoustic cues, but they showed an opposing pattern for categorization. Therefore, dynamic cues may be more difficult to discriminate than static cues but perhaps still more perceptually salient for categorization.

Another explanation is that children may have different representations of voice gender than adults, as discrimination abilities are also influenced by the categories that listeners have of the particular acoustic dimensions (Goldstone, 1994; Kuhl, 1991). Voice gender differences between male and female children's voices are primarily caused by differences in formant frequencies, possibly due to differences in the resonator width and size (Vorperian & Kent, 2007), as their F0 and VTL values do not systematically vary until puberty (Perry et al., 2001; Vorperian & Kent, 2007). As a result, children may have voice gender representations that are primarily based on voice gender differences among children's voices instead of adults' voices.

2.3 Weighting of F0 and VTL cues for voice gender categorization

2.3.1 Methods

Participants

The same group of participants that took part in the first experiment also participated in the second experiment.

Stimuli and apparatus

The stimuli consisted of four CVC words, *bus* [bus], *vaak* [often], *leeg* [empty], and *pen* [pen], produced by the same female speaker with a mean F0 of 201 Hz and taken from the same NVA corpus as used for Experiment 1 (Bosman & Smoorenburg, 1995). The recordings were first equalized in RMS and modified using STRAIGHT (Kawahara & Irino, 2005). The F0 and VTL values were manipulated independently from each other in the same manner as the stimuli of Experiment 1. For the condition in which stimuli had differences of 0.0 st in F0 and 0.0 st in VTL, stimuli were also resynthesized using the same procedure to prevent any artifacts from the resynthesis procedure itself.

The mean F0 was decreased by 0.0 st, 6.0 st, or 12.0 st, resulting in mean F0 values of 201 Hz, 142 Hz, and 100 Hz respectively, while the VTL was increased by 0.0 st, 1.8 st, or 3.6 st, corresponding to the estimated VTL sizes of 13.5 cm, 15.1 cm, and 16.6 cm. Thus, there were nine different voice combinations for the four selected words, resulting in 36 stimuli. The values for the differences in F0 and VTL were chosen based on earlier studies on voice gender categorization by Fuller et al.

(2014) and consistent with earlier findings by Smith et al. (2007), Smith and Patterson (2005), and Peterson and Barney (1952). Seven of the nine voice combinations that were used in our experiment were within 95% of the adult population based on these values. Furthermore, Fuller et al. (2014) showed that a combined difference of a decrease of 12 st in F0 and an increase of 3.6 st in VTL using this resynthesis procedure reliably changes the perceived voice gender of speakers for adult listeners. The number of “male” categorizations was less than 10% when only the F0 of the speaker was decreased by 12 st and around 30% when only the VTL of the speaker was increased by 3.6 st.

Procedure

A visual-auditory match-to-sample task was used to investigate children’s weighting of F0 and VTL cues for voice gender categorization. The experiment consisted of a training session of 5 trials and an experiment session of 36 trials with items that were presented in a randomized order. In each trial, a stimulus word was produced with voice characteristics in which the F0 and VTL parameters were more female, i.e., conforming more to the original speaker’s voice parameters, more male-like, because of a decrease in F0 and an increase in VTL, or somewhat ambiguous.

The stimuli were presented using the same touchscreen laptop and Sennheiser HD 380 pro headphones as were used in Experiment 1. The experiment was conducted using a child-friendly interface that was developed in Matlab (Figure 2.2d). After hearing the auditory stimulus, participants were presented with either a male or female face on the computer screen. Participants could then indicate if the voice and the face were of the same gender by pressing a green check mark or a red cross when the gender of the voice and face differed. No feedback on the accuracy of responses was provided to participants. The second experiment was conducted in the same testing environment after participants completed the first experiment.

Data analysis

We analyzed participants’ perceptual weighting of F0 and VTL cues for voice gender categorization across age. To calculate participants’ *cue weights*, we first normalized the F0 and VTL differences relative to the original speaker’s voice in st and defined them as $\delta F0 = -\Delta F0 \div 12 - 0.5$ and $\delta VTL = \Delta VTL \div 3.6 - 0.5$. This normalization was done to make F0 and VTL functionally equivalent for model fitting, despite the differences in range. With these normalized cues, the original female speaker’s voice had a $\delta F0$ value of -0.5 and a δVTL value of -0.5 , while the

most male-sounding voice (difference of -12 st in F0 and +3.6 st in VTL) had a $\delta F0$ of +0.5 and a δVTL of +0.5. We then extracted the coefficients of participants by fitting a mixed-effects logistic regression model with random intercepts and slopes for $\delta F0$ and δVTL per participant, in lme4 syntax: `response ~ ( $\delta F0$  +  $\delta VTL$  | participant)`. This model gives coefficient predicting values on a logit scale, i.e., log odds ratios based on the natural log, relative to the normalized $\delta F0$ and δVTL ranges. We converted participants' coefficients for $\delta F0$ and δVTL into Berkson units per semitone (Bk/st) by scaling the factors so they correspond to $\log_2$ odds ratios per st (Hilkuysen et al., 2012). This conversion makes the differences in cue weights easier to interpret, as an increase of 1 Bk per st is equal to doubling the percentage of man categorizations relative to that of woman categorizations. Individual Bk units of children were then analyzed by fitting generalized linear mixed-effects models with random intercepts per participant. Models were compared using backward stepwise model selection with ANOVA Chi-Square tests, starting with the full factorial model with *cue weight* (Bk/st) as an outcome variable and a two-way interaction between *voice cue* (F0 or VTL) and *age*, in lme4 syntax: `cue weight ~ voice cue * age + (1 | participant)`.

In addition, we used two Dunnett's Tests on participant age group subsets of F0 and VTL cue weights to investigate at what age children's cue weights were adult-like. Finally, we determined two Pearson's correlation coefficients relating children's F0 and VTL JNDs to their F0 and VTL cue weights. We used the residuals of participants' JNDs and cue weights, extracted from four linear models, to ensure that the correlations between children's JNDs and cue weights for F0 and VTL were not caused by a general effect of age. This method partials out the effects of age on the differences in JNDs and cue weights. As outcome variables, we used either $\log(JNDs)$ or *cue weights* and only *age* as a fixed effect. These analyses were repeated for F0 and VTL. The residuals of each model were then used to calculate the correlations between JNDs and cue weights for each of the two voice cues.

2.3.2 Results

Development of the weighting of F0 and VTL cues for voice gender categorization

Figure 2.2a shows participants' cue weights for F0 (left panel) and VTL (right panel) as a function of their age (dots) and participant age group (boxplots). Figure 2.2b shows the correlation between participants' cue weight residuals for F0 (left panel) and VTL (right panel) as a function of their JND residuals (Experiment 1) and

participant age group. Figure 2.2c shows the average categorization judgments as a function of differences in F0 and VTL for each participant age group. Model selection demonstrated that the full model with a two-way interaction between *voice cue* and *age* had a significantly better fit than the model with *voice cue* and *age* as fixed effects ($\chi^2(1) = 16.4$, $P < 0.001$). The significant effect of *voice cue* shows that children's weighting of VTL was overall higher than their weighting of F0 (*Estimate* = 0.20, $t = 2.74$, $P < 0.01$). On average, the ratio between the F0 and VTL cue weights differed by 2% between children and adults. Yet, children gave 71% of the weight given to F0 by adults and 63% of the weight given to VTL. Furthermore, the significant effect of *age* shows that children's weighting of F0 and VTL cues became more adult-like with age for both cues (*Estimate* = 0.02, $t = 2.64$, $P < 0.01$). Finally, the significant interaction between *voice cue* and *age* (*Estimate* = 0.02, $t = 4.29$, $P < 0.001$) indicates that there was an overall larger increase in children's weighting of VTL with age relative to their weighting of F0. However, since there was also a significant fixed effect of *voice cue*, which indicates that participants' overall weighting of F0 differs from that of VTL (when expressed in Bk/st), the interaction could simply stem from the same multiplicative effect of age being applied at different intercepts. In a signal-detection-theory model (Green & Swets, 1966), the ratio between cue weights will be constant if the responses are affected by a single source of internal noise which is common to both cues. We thus further explored this interaction by looking at the *relative* contribution of the two cues. For this, we calculated, for each participant, the ratio between the weighting of F0 and VTL. We found that, after removing data of four 4- to 6-year-old children who showed extremely singular weighting values, the ratios between participants' weighting of F0 and VTL did not differ across age ($F(1,52) = 0.65$, $P = 0.42$), which suggests that the interaction between *voice cue* and *age* was likely caused by the differences in the overall weighting of F0 and VTL cues.

Figure 2.2: Cue weights for F0 and VTL on the voice gender categorization task, the correlations between participants' JNDs and cue weights, and the experiment interface of Experiment 2. (a) Participants' cue weights for F0 (left panel) and VTL (right panel). The dots show individual data points at participants' age in years, rounded to two decimal places ($N_{\text{children}} = 58$, $N_{\text{adults}} = 15$) and the boxplots show the median cue weights per age group. The box shows the lower and upper quartiles, and the whiskers show the lowest and highest data points within plus or minus 1.5 times the interquartile range. (b) The relationship between participants' F0 JND and cue weight residuals (left panel) and VTL JND and cue weight residuals. (c) Average gender categorization judgments as a function of differences in F0 (x-axis) and VTL (y-axis) in st for each participant age group. Blue corresponds to 100% "man" categorizations and yellow corresponds to 100% "woman" responses. (d) The experiment interface. Participants were instructed to click on the green check mark if the perceived gender of the voice and the portrayed face matched and on the red cross if these differed. The illustrations were made by Jop Luberti for the purpose of this study. This image is published under the CC-BY-NC 4.0 license.

Adult-like weighting of F0 and VTL cues

The Dunnett's Test that was performed on the subset of F0 cue weights indicated that adults' F0 cue weights only significantly differed from 4- to 6-year-old children's cue weights (4-6 years: difference = -0.51, $P < 0.001$). For VTL cue weights, the Dunnett's Test indicated that adults' cue weights significantly differed from 4- to 10-year-old children's cue weights but not from 10- to 12-year-old children's cue weights (4-6 years: difference = -1.08, $P < 0.001$; 6-8 years: difference = -0.59, $P < 0.01$; 8-10 years: difference = -0.39, $P < 0.05$; 10-12 years: difference = -0.37, $P = 0.061$).

Relationship between discrimination and categorization of voice gender cues

The two Pearson's correlation coefficients of children's JNDs and cue weights showed moderate correlations between their F0 JNDs and cue weights (*Pearson's* $r = -0.49$, $P < 0.001$) and their VTL JNDs and cue weights (*Pearson's* $r = -0.48$, $P < 0.001$). However, the correlations were only significant for the group of 4- to 6-year-old children for F0 (*Pearson's* $r = -0.63$, $P < 0.05$) and for VTL (*Pearson's* $r = -0.59$, $P < 0.05$). Hence, the reported correlations mainly seem to be driven by the generally low performance of some of the youngest children on both tasks. For this reason, we performed two additional Pearson's correlations using the residuals of children's F0 and VTL JNDs (Experiment 1) and their F0 and VTL cue weights (Experiment 2) to prevent a correlation resulting from a general effect of age. Figure 2.2b shows the correlation between participants' F0 JND and cue weight residuals and their VTL JNDs and cue weight residuals (Experiment 2). We found that there was still a significant moderate correlation between the residuals of participants' JND and cue weight residuals for F0 (*Pearson's* $r = -0.37$, $P < 0.01$) and a small correlation for VTL (*Pearson's* $r = -0.28$, $P < 0.05$) using children's residuals. In addition, we looked again at the correlations between JND residuals and cue weight residuals for F0 and VTL for each participant age group, and we found that, as for the non-residuals, there was only a significant correlation for the 4- to 6-year-olds for F0 (*Pearson's* $r = -0.64$, $P < 0.05$) and for VTL (*Pearson's* $r = -0.61$, $P < 0.05$) JNDs and cue weights even after taking a general effect of age into account.

2.3.3 Discussion

Our results show that children's cue weights for F0 only differed from adults' cue weights for the youngest age group between 4 and 6 years of age, while children's cue weights for VTL differed from adults' cue weights between 4 and 10 years of age. We also investigated if participants' JNDs and cue weights for F0 and VTL were related to each other. After correcting for a general effect of age, we only found a significant moderate and small overall correlation between children's JNDs and cue weights for F0 and VTL, respectively, but our additional correlation analyses for each participant age group showed that only 4- to 6-year-old children's F0 and VTL JNDs and cue weights were significantly correlated.

Contrary to our first experiment, the results of our second experiment are in line with our expectation that children's weighting of F0 cues for voice gender categorization is adult-like earlier than their weighting of VTL cues. F0 is a more dynamic acoustic cue than VTL and may be a more perceptually salient cue for children to use for voice gender categorization (Lass et al., 1976; Nittrouer & Miller, 1997; Pernet & Belin, 2012). Hence, children may learn earlier which F0 values correspond to male and female voices compared to VTL values, as they may attend more to F0 differences. In keeping with the findings of Mirman et al. (2004), the different developmental patterns for F0 and VTL provide evidence for a trade-off effect between the discrimination and categorization of acoustic cues based on their acoustic dynamics (steady-state, static acoustic cues versus rapidly-changing, dynamic acoustic cues). A potential explanation for this observation is related to the fact that, in NH adults, the VTL JNDs represent a larger proportion of the typical male-female difference (about 34%) than F0 JNDs (about 9%). As a result, VTL appears less robust to perceptual degradations than F0, as was observed in adult cochlear implant users (Fuller et al., 2014; Gaudrain & Başkent, 2018). In the present study, the larger VTL JNDs observed in the younger age groups, which were sometimes above the smallest step size of 1.8 st in VTL, may have hindered children's access to the VTL cue in the voice gender categorization task. However, as participants' JNDs corresponded to the 70.7% correct discrimination point on the psychometric function, participants may have still been able to perceive F0 and VTL differences that were below their discrimination thresholds that were measured in Experiment 1.

Children's weighting was adult-like earlier for F0 than for VTL, but they used both acoustic cues for categorizing speakers' voice gender at all ages. This finding implies that children may have different representations of voice gender and their corresponding F0 and VTL values than adults. Most children were consistent in

their categorizations of the clearly male- and female-sounding voices, but they were less consistent in categorizing the other, more ambiguous-sounding, voices. A similar finding was reported by Hazan and Barrett (2000) who found that 6- to 12-year-old children's categorization of phonemes, based on a range of phonemic contrasts, was not as consistent and clear as it was for adults. One could also consider that children's representations of voice gender may differ from adults because their representations could be predominantly based on voice gender differences between other children rather than adults (Anastasi & Rhodes, 2005). As the voice gender of child voices cannot be recognized reliably based on differences in their F0 or VTL values until the age of 12 (Perry et al., 2001; Vorperian & Kent, 2007), this may also explain why they show different weighting of these voice cues than adults.

Furthermore, as mentioned by Hazan and Barrett (2000), children's limited attention span seems to play merely a minor role in explaining their inconsistent responses, as most children categorized the clearly male-sounding and female-sounding voices consistently, apart from some children in the youngest age group. Children may not have clear representations of what a male or female voice sounds like yet for various reasons. For instance, they may interact primarily with other young children for whom the differences in their perceived voice gender are still very limited (Perry et al., 2001; Vorperian & Kent, 2007). However, research by Cartei et al. (2019) shows that children adapt their F0 and VTL values when asked to imitate a speaker from the opposite sex, although these adaptations only led to relatively small acoustic changes. These findings seem to indicate that children are already aware of the contributions of F0 and VTL to the perceived gender of adult voices. Participants' JNDs and cue weights were also only moderately and weakly correlated for F0 and VTL after correcting for a general effect of age, and only for 4- to 6-year-old children when looking at individual participant age groups. It is also difficult to interpret the causality of these results. Categorization of voice cues may be driven by the ability to discriminate differences in these voice cues, but both abilities may also develop independently and improve merely as a function of age. As we only found a significant correlation among 4- to 6-year-old children, of which some showed general low performance on both tasks, children's different weighting of F0 and VTL seems to be likely caused by their less robust representations of voice gender categories or task performance rather than their lower discrimination abilities.

2.4 General discussion

The goal of the present study was to investigate the development of children's ability to discriminate differences in voice gender cues and their weighting of F0 and VTL cues for voice gender categorization. Our results show that there are dissociations in the development of children's discrimination of F0 and VTL and their weighting of F0 and VTL for voice gender categorization. In our first experiment, we found that children's discrimination of VTL becomes adult-like earlier than their discrimination of F0. In contrast, the results of our second experiment showed that children's weighting for voice gender categorization becomes adult-like earlier for F0 than VTL. Hence, for both discrimination and perceptual weighting for voice gender categorization, we found differences in the development of F0 and VTL. However, the dissociation between the development of children's discrimination of F0 and VTL cues did not correspond to the developmental pattern for F0 and VTL that we found for voice gender categorization. Finally, we only found a moderate and a small overall significant correlation between children's F0 and VTL JNDs, i.e., discrimination thresholds, and cue weights after correcting for a general effect of age and merely a significant correlation for 4- to 6-year-old children when looking at individual participant age groups.

The discrepancy between the developmental patterns for discrimination and categorization of voice gender cues suggests that these abilities rely on different auditory processes. Evidence for the latter theory is provided by Mirman et al. (2004), who reported a similar trade-off effect between the discrimination versus categorization based on the acoustic dynamics of voice cues. Dynamic acoustic cues may be more perceptually salient than static cues, which makes listeners rely on them more for categorization. However, due to their rapidly-changing and dynamic nature, differences in dynamic cues may be more difficult to perceive for listeners than static cues. In addition, children's discrimination of VTL may become adult-like earlier because it has a linguistically more prominent role, as it is essential for the correct interpretation of vowels (Irino & Patterson, 2002) while F0 gives more subtle paralinguistic information in non-tonal languages.

Another factor that may have played a role is that children's representations of voice gender are based on voice gender distinctions among children rather than adults. Fleming et al. (2014) found that adult listeners are better at distinguishing speakers' voices in their native language than in a different language because of a higher familiarity with the phonology of their native language. This finding demonstrates that even adults show a bias in voice perception and discrimination

based on auditory experience. Voice gender in children is primarily determined by differences in formant frequencies, possibly caused by differences in the resonator width and size (Vorperian & Kent, 2007), as there are no clear differences in male and female children's F0 and VTL voice parameters before puberty (Perry et al., 2001; Vorperian & Kent, 2007). Hence, children may have different categories of voice gender than adults. Future research should investigate whether adults show similar weighting of F0 and VTL as children for categorizing the voice gender of children's voices instead of adults' voices.

Investigating the relationship between different tasks and their underlying cognitive processes is also eminent for research on other perceptual learning mechanisms, as similar differences in developmental patterns and parameters may apply. For instance, Leibold et al. (2018) found that children showed a similar or larger benefit from target-masker differences in voice gender for the perception of speech in competing speech than adults, while Flaherty et al. showed children did not benefit from differences in speakers' mean F0 only. A potential explanation for this discrepancy given by the authors was that children may rely more on combined differences in F0 and VTL cues to segregate speakers based on voice gender. However, our results from Experiment 2 suggest that children use and rely on differences in both F0 and VTL for categorizing speakers' voice gender when the differences are sufficiently large. Understanding these developmental patterns is also particularly relevant for research with clinical populations, such as children with hearing impairments (Cleary et al., 2005; Deroche et al., 2019), as their difficulties in perceiving and recognizing voices may originate from different sources, such as perceptual limitations or difficulties with allocating attention.

In conclusion, we found that children's discrimination and categorization of voice gender cues develop differently for F0 and VTL. In addition, we found opposing developmental patterns for F0 and VTL for the two abilities, implying that these abilities inherently rely on different auditory processes and differ based on the nature of the acoustic cue. Generally, we also found that the discrimination and categorization of voice gender cues take a long time to mature, which corresponds with the earlier reported findings on the discrimination and categorization of phonemes (Hazan & Barrett, 2000; Nitttrouer & Miller, 1997). In addition, our findings emphasize that voice perception research should take into account how the development of various perceptual abilities, for instance, voice discrimination, voice categorization, and voice recognition, are related and on which parameters and cognitive mechanisms these abilities rely.

Bibliography

- Allen, P., Wightman, F., Kistler, D., & Dolan, T. (1989). Frequency resolution in children. *Journal of Speech, Language, and Hearing Research*, 32(2), 317–322.
- Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review*, 12(6), 1043–1047.
- Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, 139(3), EL51–EL56.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–51.
- Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3), 129–135.
- Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology*, 34(5), 260–284.
- Buss, E., Flaherty, M. M., & Leibold, L. J. (2017). Development of frequency discrimination at 250 Hz is similar for tone and /ba/ stimuli. *The Journal of the Acoustical Society of America*, 142(1), EL150–EL154.
- Cartei, V., Garnham, A., Oakhill, J., Banerjee, R., Roberts, L., & Reby, D. (2019). Children can control the expression of masculinity and femininity through the voice. *Royal Society Open Science*, 6(190656), 1–9.
- Cleary, M., Pisoni, D. B., & Kirk, K. I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 48(1), 204–223.
- Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, 113(4), 487–509.
- Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America*, 114(5), 2913–2922.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208(4448), 1174–1176.
- Deroche, M. L. D., Lu, H.-P., Kulkarni, A. M., Caldwell, M., Barrett, K. C., Peng, S.-C., Limb, C. J., Lin, Y.-S., & Chatterjee, M. (2019). A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. *Scientific Reports*, 9(1), 1–12.
- Deroche, M. L. D., Zion, D. J., Schurman, J. R., & Chatterjee, M. (2012). Sensitivity of school-aged children to pitch-related cues. *The Journal of the Acoustical Society of America*, 131(4), 2938–2947.
- Fecher, N., Paquette-Smith, M., & Johnson, E. K. (2019). Resolving the (apparent) talker recognition paradox in developmental speech perception. *Infancy*, 24(4), 570–588.
- Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10(3), 279–293.

- Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106(3), 1511–1522.
- Flaherty, M. M., Buss, E., & Leibold, L. J. (2019). Developmental effects in children's ability to benefit from F0 differences between target and masker speech. *Ear and Hearing*, 40(4), 927–937.
- Fleming, D., Giordano, B. L., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences*, 111(38), 13795–13798.
- Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., & Başkent, D. (2014). Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology*, 15(6), 1037–1048.
- Gaudrain, E., & Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear and Hearing*, 39(2), 226–237.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton-Century-Crofts.
- Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123(2), 178–200.
- Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley & Sons, Inc.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4), 377–396.
- Hilkuysen, G., Gaubitch, N., Brookes, M., & Huckvale, M. (2012). Effects of noise suppression on intelligibility: Dependency on signal-to-noise ratios. *The Journal of the Acoustical Society of America*, 131(1), 531–539.
- Irino, T., & Patterson, R. D. (2002). Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform. *Speech Communication*, 36(3), 181–203.
- Irwin, R. J., Ball, A. K. R., Kay, N., Stillman, J. A., & Rosser, J. (1985). The development of auditory temporal acuity in children. *Child Development*, 56(3), 614–620.
- Jensen, J. K., & Neff, D. L. (1993). Development of basic auditory discrimination in preschool children. *Psychological Science*, 4(2), 104–107.
- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, 14(5), 1002–1011.
- Kawahara, H., & Irino, T. (2005). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In *Speech Separation by Humans and Machines* (pp. 167–180). Springer. https://doi.org/10.1007/0-387-22794-6_11
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, 87(2), 820–857.

- Kreitewolf, J., Gaudrain, E., & von Kriegstein, K. (2014). A neural mechanism for recognizing speech spoken by different speakers. *Neuroimage*, 91, 375–385.
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93–107.
- Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *The Journal of the Acoustical Society of America*, 59(3), 675–678.
- Leibold, L. J., Buss, E., & Calandruccio, L. (2018). Developmental effects in masking release for speech-in-speech perception due to a target/masker sex mismatch. *Ear and Hearing*, 39(5), 935–945.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467–477.
- Litovsky, R. (2015). Development of the auditory system. In *Handbook of Clinical Neurology* (Vol. 129, pp. 55–72). Elsevier.
- Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, 27(1), 153–165.
- Maxon, A. B., & Hochberg, I. (1982). Development of psychoacoustic behavior: Sensitivity and discrimination. *Ear and Hearing*, 3(6), 301–308.
- Mirman, D., Holt, L. L., & McClelland, J. L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *The Journal of the Acoustical Society of America*, 116(2), 1198–1207.
- Moore, J. K., & Linthicum Jr., F. (2007). The human auditory system: A timeline of development. *International Journal of Audiology*, 46(9), 460–478.
- Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *The Journal of the Acoustical Society of America*, 101(4), 2253–2266.
- Olsho, L. W., Koch, E. G., Carter, E. A., Halpin, C. F., & Spetner, N. B. (1988). Pure-tone sensitivity of human infants. *The Journal of the Acoustical Society of America*, 84(4), 1316–1324.
- Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Frontiers in Psychology*, 3, 1–11.
- Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children’s voices. *The Journal of the Acoustical Society of America*, 109(6), 2988–2998.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253–260.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>

- Renfrew, C. E. (1995). *Word Finding Vocabulary Test*. Speechmark Publishing.
- Roser, M., Appel, C., & Ritchie, H. (2020). *Human Height*. Our World in Data.
<https://ourworldindata.org/human-height>
- Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arppe, A., Baddeley, A., Barton, K., Bolker, B., & Borchers, H. W. (2018). *DescTools: Tools for descriptive statistics* (R package version 0.99.25) [Computer software].
- Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language & Hearing Research*, 57(1), 285–296.
- Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, 118(5), 3177–3186.
- Smith, D. R. R., Walters, T. C., & Patterson, R. D. (2007). Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. *The Journal of the Acoustical Society of America*, 122(6), 3628–3639.
- Spence, M. J., & Freeman, M. S. (1996). Newborn infants prefer the maternal low-pass filtered voice, but not the maternal whispered voice. *Infant Behavior and Development*, 19(2), 199–212.
- Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, 85(4), 1699–1707.
- Traunmüller, H., & Eriksson, A. (1995). The frequency range of the voice fundamental in the speech of male and female adults. *Unpublished manuscript*.
- Vestergaard, M. D., Håden, G. P., Shtyrov, Y., Patterson, R. D., Pulvermüller, F., Denham, S. L., Sziller, I., & Winkler, I. (2009). Auditory size-deviant detection in adults and newborn infants. *Biological Psychology*, 82(2), 169–175.
- Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, 50(6), 1510–1545.
- Werner, L. A. (2007). Issues in human auditory development. *Journal of Communication Disorders*, 40(4), 275–283.
- Wightman, F., Allen, P., Dolan, T., Kistler, D., & Jamieson, D. (1989). Temporal resolution in children. *Child Development*, 60(3), 611–624.

3 | School-age children benefit from voice gender cue differences for the perception of speech in competing speech

Published as:

L. Nagels, E. Gaudrain, D. Vickers, P. Hendriks, and D. Başkent (2021). School-age children benefit from voice gender cue differences for the perception of speech in competing speech. *The Journal of the Acoustical Society of America*, 149(5).

Abstract

Differences in speakers' voice characteristics, such as mean fundamental frequency (F0) and vocal-tract length (VTL) that primarily define speakers' so-called perceived voice gender, facilitate the perception of speech in competing speech. Perceiving speech in competing speech is particularly challenging for children, which may relate to their lower sensitivity to differences in voice characteristics than adults. This study investigated the development of the benefit from F0 and VTL differences in school-age children (4-12 years) for separating two competing speakers while tasked with comprehending one of them, and also the relationship between this benefit and their corresponding voice discrimination thresholds. Children benefited from differences in F0, VTL, or both cues at all ages tested. This benefit proportionally remained the same across age, although overall accuracy continued to differ from adults. Additionally, children's benefit from F0 and VTL differences, and their overall accuracy were not related to their discrimination thresholds. Hence, although children's voice discrimination thresholds and speech in competing speech perception abilities develop throughout the school-age years, children already show a benefit from voice gender cue differences early on. Factors other than children's discrimination thresholds seem to relate more closely to their developing speech in competing speech perception abilities.

3.1 Introduction

In daily life, we are often presented with sounds originating from different sources and locations but overlapping in temporal and spectral characteristics. Although listening to one particular speech signal from this mixture may be demanding, adult listeners are generally able to effectively process the speech signal components that belong to the same source by segregating and grouping them, by principles of *auditory stream segregation* (Bregman, 1994). The perception of speech in competing background speech as opposed to competing noise differs in that the masking signal can interfere beyond pure perceptual obliteration of the target signal, also known as “energetic masking”. For competing speech maskers, the masking signal can largely overlap with the target speech signal in their spectrotemporal properties and the masking signal can cause lexical-semantic interference, also sometimes called “informational or perceptual masking” (Carhart et al., 1969; Mattys et al., 2009; Pollack, 1975). When listeners process the target and masking signals, they have to

inhibit the information that is provided by the masker on a cognitive level to interpret the target signal correctly (Kidd et al., 2008; Schneider et al., 2007). Therefore, speech stream segregation in the presence of competing noise seems to be primarily a matter of peripheral perceptual processing, namely the correct grouping and interpretation of target speech components (Bronkhorst, 2015; Carhart et al., 1969). In contrast, speech stream segregation in the presence of competing speech seems to additionally and heavily rely on central cognitive mechanisms, such as the allocation of attention to the target signal and the inhibition of masker interference. The similarities and differences between the effects of competing noise and competing speech maskers on listeners' ability to perceive speech and required cognitive resources involved have been studied extensively in adult listeners (Arbogast et al., 2002; Brungart et al., 2001, 2006; Cooke et al., 2008; Evans et al., 2016; Mattys et al., 2009; Ruggles et al., 2011; Scott et al., 2004; Swaminathan et al., 2015).

For children, there seems to be a discrepancy in the acquisition of adult-like speech perception for both masker types. It takes children considerably longer developmentally to reach an adult-like level for perceiving speech in competing speech compared to perceiving speech in competing noise, which is associated with their developing cognitive abilities, such as the inhibition of masker interference (Bonino et al., 2013; Buss et al., 2017a, 2019; Corbin et al., 2016; Hall et al., 2002; Leibold and Buss, 2013). The ability to perceive speech in competing speech also often develops less linearly in children than their ability to perceive speech in competing noise (Corbin et al., 2016). Listeners' susceptibility to the informational masking provided by a competing speech masker also largely varies across individuals, even among adult listeners (Swaminathan et al., 2015). In addition, psychometric functions for perceiving speech in competing speech, i.e., accuracy as a function of the target-to-masker ratio (TMR), seem to be less steep for children compared to adults (MacPherson and Akeroyd, 2014; Sobon et al., 2019). Bonino et al. (2013) found that 8- to 10-year-old children performed similarly to adults in recognizing disyllabic words in a speech-shaped noise masker, but worse than adults in multi-talker babble or two-talker speech maskers. It seems that the underlying cause of this discrepancy specifically concerns informational masking and might be of a central cognitive nature (McCreery et al., 2019, 2020; Sobon et al., 2019). For instance, the ability to segregate different speech streams based on differences in speakers' voice characteristics may help with optimally using attentional mechanisms for understanding the target speech signal better. As the ability to discriminate subtle differences in voice cues seems to develop in children during

the school-age years (Buss et al., 2017b; Cleary et al., 2005; Flaherty et al., 2019; Nagels et al., 2020a), their ability to benefit from voice differences between the target and the masker for perceiving speech in competing speech may be limited, potentially affecting further processing stages.

Spatial differences and speech onset asynchrony are strong cues for speech stream segregation in adult listeners (Freyman et al., 2001; Kidd et al., 2005; Lee and Humes, 2012; Zobel et al., 2019). Another cue that particularly improves speech stream segregation in competing speech maskers is differences in speakers' voice characteristics (Bird and Darwin, 1998; Broadbent, 1952; Brungart, 2001; Helfer and Freyman, 2009; Zekveld et al., 2014). For instance, adults are better at perceiving speech in competing speech when the masker speakers are of a different sex than the target speaker (Brungart, 2001). Several studies have followed up on these findings by investigating the relative contribution of speakers' mean fundamental frequency (F0) and vocal-tract length (VTL) to the results (Başkent & Gaudrain, 2016; Darwin et al., 2003). Speakers' mean F0 is defined by the vibration rate of speakers' vocal folds and affects the perceived voice pitch, while speakers' VTL strongly correlates with speakers' height (Fitch and Giedd, 1999), and affects their formant frequencies (Kreiman and Sidtis, 2011). Together, these two voice cues are the primary acoustic features that define speakers' perceived sex or so-called voice gender (Fitch and Giedd, 1999; Skuk and Schweinberger, 2014; Titze, 1989). In the studies by Başkent and Gaudrain (2016) and Darwin et al. (2003), a single-talker competing speech masker was created by taking the target speaker's voice and manipulating only the mean F0 or VTL voice parameters to keep all other speaker-specific acoustic features consistent. The results demonstrated that adult listeners also benefit when the masker differs from the target in either mean F0 only or VTL only.

Children also benefit similarly to — or more than — adults when the masker speech is produced by speakers of a different sex than the target speaker (Leibold et al., 2018; Wightman and Kistler, 2005). Even two-and-a-half-year-old toddlers already show this benefit from talker-sex differences for speech stream segregation (Newman and Morini, 2017). However, Flaherty et al. (2019) found that, when a two-talker speech masker differed from the target speech in only mean F0, children did not benefit from target-masker F0 differences of -3, -6, or -9 semitones until 7 years of age. In addition, 8- to 12-year-old children showed a reduced benefit relative to 13- to 15-year-old children and adults. Therefore, Flaherty et al. (2019) have suggested that children may rely more on the combination of F0 and VTL differences to determine speakers' voice gender, and hence do not benefit from

differences in speakers' mean F0 only. More evidence for this hypothesis is provided by a follow-up study in which speakers' VTL was also manipulated (Flaherty et al., 2021). The results showed that similarly to F0 differences, young children did not benefit from differences in speakers' VTL only, but they did benefit from a change in the two voice cues together, although this benefit was still lower than that observed in older children. This argument is also in line with the results of our previous study that indicate children weigh both F0 and VTL cues to categorize speakers' voice gender (Nagels et al., 2020a). Children may also not be sensitive enough to the acoustic variations induced by only mean F0 or only VTL differences to benefit from these during speech stream segregation and could have been relying on additional acoustic differences for speech segregation in the aforementioned studies by Leibold et al. (2018), and Wightman and Kistler (2005). Hence, children's ability to perceive speech in competing speech, as opposed to competing noise, may relate to the development in how well they can discriminate differences in voice cues, which improves during the school-age years (Buss et al., 2017b; Flaherty et al., 2019; Nagels et al., 2020a).

Children's ability to discriminate pitch differences based on pure-tones (Jensen and Neff, 1993; Maxon and Hochberg, 1982) or mean F0 based on voice stimuli (Buss et al., 2017b; Flaherty et al., 2019; Nagels et al., 2020a), or VTL cues based on voice stimuli (Nagels et al., 2020a) develops during the school-age years. Higher-order cognitive mechanisms of voice perception also develop during this period, for instance, children's ability to recognize unfamiliar voices (Creel and Jimenez, 2012; Fecher et al., 2019; Mann et al., 1979) or their weighting of voice and speech cues for categorization tasks (Floccia et al., 2009; Hazan & Barrett, 2000; Nagels et al., 2020a; Nitttrouer & Miller, 1997). However, the exact relationship between children's perception and discrimination of voice cue differences, and their ability to use these differences for speech stream segregation is still not well understood. Flaherty et al. (2019) did not find a significant correlation between individual children's ability to discriminate F0 cues and their benefit from F0 differences for perceiving speech in competing speech. This finding indicates that voice discrimination and the functional usage of voice cue differences may, in fact, develop independently from each other.

An additional factor that may play a role in the development of children's ability to perceive speech in competing speech is their general language development (Klein et al., 2017; McCreery et al., 2017, 2020). When a target speech stimulus is presented simultaneously with a masker, some parts of the target are obliterated or inaccessible to the listener. This missing information can be restored by the listener

by relying on the acoustic and linguistic redundancy inherent to speech and language. Speech information can sometimes be superfluous to listeners, as speech cues are coded in multiple ways and words can often be predicted based on their probability, i.e., word frequency or neighborhood density, or sentential context (Başkent et al., 2016). The ability to restore missing segments depends on the listeners' language abilities, which are not fully developed yet for school-age children. Support for development in children's speech restoration abilities is provided by studies using a gating paradigm in which parts of words or sentences have been gated off, and the proportion of word segments that listeners need for correct word recognition is examined. Young children require a greater amount of word segments for correct word recognition than older children and adults, who seem to use word probability and sentential context information more effectively (Craig et al., 1993; Elliott et al., 1987; Metsala, 1997). On the other hand, using a perceptual restoration paradigm (Samuel, 1996; Warren, 1970), Newman (2004) found that the perceptual restoration abilities of five-year-old children were equal to those of adults. Nitttrouer and Boothroyd (1990) also showed that four- to six-year-old children used lexical and syntactic constraints to the same extent as adults for perceiving speech in competing noise. Nevertheless, Buss et al. (2019) observed a discrepancy in young children's ability to benefit from sentential context for perceiving speech in competing noise compared to competing speech. Young children benefitted equally from sentential context as older children and adults for perceiving speech in competing noise but less so for perceiving speech in competing speech. Buss et al. (2019) suggested that the high cognitive demands associated with perceiving speech in competing speech may prevent a benefit from sentential context in young children. In addition, a correlation between the ability to perceive speech in competing noise or competing speech seems to depend on the complexity of the linguistic stimuli that are used (Klein et al., 2017; McCreery et al., 2017, 2020). Deducing from such observations, it is not entirely evident yet how language development is related to concurrent speech perception and if developmental effects are always present or depend on the specific task and materials that are used.

In the present study, we investigated how the benefit from differences in speakers' mean F0 and VTL for perceiving speech in competing speech develops in children during the school-age years (4 to 12 years of age). Furthermore, we examined if children's benefit from F0 and VTL differences relates to their ability to discriminate these voice cues by using their F0 and VTL discrimination thresholds (taken from Nagels et al., 2020a). We used a child-friendly version of the coordinate response measure (CRM, used earlier by, for instance, (Bolia et al., 2000;

Brungart, 2001; Hazan et al., 2009; Moore, 1981; Saleh et al., 2013; Welch et al., 2015) with a single-talker speech masker, which was created by manipulating the F0 and VTL parameters of the target speaker's voice, and presented at three fixed TMRs. Based on previous research, we expected that children's ability to perceive speech in competing speech would improve during the school-age years (Bonino et al., 2013; Buss et al., 2017a; Corbin et al., 2016; Hall et al., 2002; Leibold and Buss, 2013). Furthermore, if children only show substantial benefit from a combined change in F0 and VTL cues, as suggested by Flaherty et al. (2019, 2021), children's performance will improve when F0 and VTL cues are manipulated together, but not when only one voice cue is manipulated. Finally, if children's benefit from voice gender cue differences directly depends on their discrimination abilities, a significant correlation between these measures is expected.

We also collected vocabulary size scores as a marker of language development (Marchman and Fernald, 2008). Currently, the effects of language development on children's ability to perceive speech in competing speech remain unclear. While some studies have reported that children's perceptual restoration abilities and use of lexical and syntactic constraints are adult-like (Newman, 2004; Nitttrouer and Boothroyd, 1990), other studies have reported that young children use word probability and sentential context information less effectively than older children and adults (Buss et al., 2019; Craig et al., 1993; Elliott et al., 1987; Metsala, 1997). Vocabulary size could have some effects on children's ability to perceive speech in competing speech (Klein et al., 2017; McCreery et al., 2017, 2020), as vocabulary size is age-specific and increases during the school-age years. However, as we tested young children from 4 years of age and older, our stimuli for the current study consisted of simple closed-set sentence materials, where the child only had to identify color terms and number words, words that children would be familiar with already and posing a closed-set of response options. As a result, the effects of vocabulary size on school-age children's performance may be minimal in this study.

3.2 Method

The experiment was part of a larger project on the perception of indexical cues in kids and adults (PICKA) for which we collected data from the same population of children and adults for a number of studies on voice and speech perception (Nagels et al., 2020a, 2020b).

3.2.1 Participants

Fifty-eight Dutch children between 4 and 12 years of age and fifteen Dutch adults between 20 and 29 years of age took part in the experiment. The selected age range for children was based on the ages at which children attend primary school in the Netherlands (4 to 12 years of age) and would therefore be expected to be able to perform the experiments. However, three of the five 4-year-old children who completed the other PICKA experiments did not fully complete the current experiment due to attentional and motivational issues. All PICKA measures were done on the same day during one testing session of approximately 60 to 90 minutes. The speech in competing speech perception task was the last and longest task, and lasted approximately 15 to 20 minutes. The partial data of these children were not included in the data analysis, reducing the number of child participants to fifty-five. Also, the vocabulary size of one 5-year-old participant was not measured, but their data were included in the analysis. The demographic characteristics of participants, categorized into five specified age groups used in parts of the data analysis, are summarized in Table 3.1. The age groups each spanned two years. For instance, the 4- to 6-year-old age groups consisted of children who were 4 years of age or older but younger than 6 years of age (≥ 4 years and < 6 years). We primarily used age as a continuous variable for data analysis, but age groups were used to approximate the age at which children showed adult-like performance. We recruited child participants via local primary schools and after-school care centers, and adult participants via online advertisements. All participants were monolingually raised native speakers of Dutch and reported no hearing or language disorders.

To ensure that all participants had normal hearing, we used a portable Interacoustics AS608B screening audiometer to conduct a short 20 dB HL pure-tone audiometric screening at octave frequencies between 500 Hz and 4000 Hz. The raw scores of children's vocabulary size were measured using the Dutch version of the Renfrew Word Finding Vocabulary Test (Renfrew, 1995) which had a maximum achievable score of 50 points. Before participants took part in the experiment, they were provided with detailed information about the study, and a

written informed consent form was signed by the adult participants and by the parents or legal guardians of the child participants. Ethical approval of the study was given by the Medical Ethical Review Committee of the University Medical Center Groningen (METc 2016.689).

Table 3.1: Demographic characteristics of participants per age group. Age is given in years, rounded to two decimal places. Vocabulary corresponds to scores on the Renfrew Word Finding Vocabulary Test (maximum of 50 points).

Participant age groups	Number of participants	Age (median; range)	Vocabulary size (median; range)
Children 4-6 years	10	5.42; 4.08 - 5.83	36; 29 - 41
Children 6-8 years	13	7.17; 6.25 - 7.83	40; 37 - 46
Children 8-10 years	16	9.00; 8.08 - 9.83	45; 40 - 49
Children 10-12 years	16	11.10; 10.00 - 12.00	46; 41 - 48
Adults	15	24.40; 20.8 - 29.17	N/A

3.2.2 Stimuli and apparatus

We used a CRM task (first used by Moore, 1981; (Bolia et al., 2000; Brungart et al., 2001; Hazan et al., 2009; Saleh et al., 2013; Welch et al., 2015) with sentence stimuli adapted from the English stimuli used by Hazan et al. (2009) and Welch et al. (2015), translated into Dutch. The 48 target sentences consisted of a carrier phrase in which one of six colors and one of eight numbers were mentioned, e.g., *Laat de hond zien naar de rode* (color) *twee* (number) *is*. [Show the dog where the red (color) two (number) is.]. The six basic colors were all disyllabic words in Dutch (*rode*, *zwarte*, *groene*, *blauwe*, *witte*, and *gele*) [red, black, green, blue, white, and yellow] and the eight numbers were all monosyllabic words in Dutch (1-10; but excluding *zeven* [seven] and *negen* [nine] which are disyllabic words in Dutch). Using these closed-set sentence stimuli with words that are highly familiar and acquired early in life (Brysbaert et al., 2014) makes the test-retest reliability of the CRM in general moderate to high (Saleh, 2013; Semeraro et al., 2017). For the masker speech, we used a second set of 48 sentences with the same structure in which the call sign *hond* [dog] was replaced by *kat* [cat]. Sentence chunks ranging from 150 to 300 ms were randomly selected from these sentences and concatenated after applying 50-ms raised cosine ramps, and avoiding sentences with the same color or number. We used sentence chunks instead of complete sentences for the masker speech, similar to El Boghdady et al. (2019), to make the masker differ in structure from the target.

We expected this would make the task easier to comprehend for children and help avoid potential confusion about which speaker they should attend to. For that purpose, the masker also started 750 ms before the target, end ended 250 ms after the target. The target and masker stimuli were produced by a female native speaker of Dutch with a standard Dutch accent, a mean F0 value of 242 Hz, and an approximated VTL size of about 13.6 cm based on the speaker's height of 166 cm (Fitch & Giedd, 1999). The recordings were made in an anechoic room at a sampling rate of 44.1 kHz. The duration of the target stimuli ranged from 2.14 seconds to 2.49 seconds with a mean duration of 2.27 seconds.

As differences in F0 are commonly measured in Hertz and differences in VTL in centimeters, we expressed the F0 and VTL differences in semitones (st). This manipulation and comparison between cues were also done in prior studies from our research group (El Boghdady et al., 2019; Fuller et al., 2014; Gaudrain & Başkent, 2018; Nagels et al., 2020a). There were four different voice conditions produced for the masker speech: (1) the same voice parameters as the target speaker, (2) a difference of -12 st in F0, (3) a difference of +3.8 st in VTL, or (4) a combined difference of -12 st in F0 and +3.8 st in VTL, relative to the female target speaker's voice. These voice differences correspond to mean F0 values of approximately 242 Hz and 121 Hz, and VTL sizes of approximately 13.6 cm and 16.7 cm. The specified differences of -12 st in F0 and +3.8 st in VTL are consistent with earlier findings by Smith et al. (2007), Smith and Patterson (2005), and Peterson and Barney (1952). Earlier studies from our research group confirmed that these differences in F0 and VTL, when processed in the same way as in this study, reliably change the perceived gender of a voice for normal-hearing adult listeners (Fuller et al., 2014; Nagels et al., 2020a).

All sentence stimuli were analyzed and resynthesized via STRAIGHT software (Kawahara & Irino, 2005) using Matlab. The mean F0 of the sentences was first normalized to a value of 242 Hz. Subsequently, the F0 contour and the spectral envelope of the sentences were extracted. We adjusted the masker speech stimuli to match the specified differences in F0 and VTL of one of the four masker speech voice conditions. We also resynthesized the masker speech stimuli with no differences in F0 and VTL to prevent giving any unfair advantage to this condition in case artifacts may have been arising from the resynthesis procedure itself. For the F0 manipulation, the original F0 fluctuations were preserved by multiplying the overall F0 contour by the specified change in mean F0 to only alter the mean F0 of the sentence. For the VTL manipulation, an overall spectral shift of the formant frequencies was produced by compressing the spectral envelope toward the lower

frequencies to approximate a perceived change in speakers' VTL. The actual VTL size of the speaker could not be measured and hence had to be estimated. We then used STRAIGHT's pitch synchronous overlap-add (PSOLA) resynthesis method to recombine the modified F0 contour and spectral envelope.

3.2.3 Procedure

As mentioned before, the current speech in competing speech perception experiment was part of a larger project on children's perception of indexical cues. We had also examined children's ability to discriminate differences in F0 and VTL cues and their perceptual weighting of F0 and VTL cues for voice gender categorization prior to this experiment during the same test session (Nagels et al., 2020a). For the discrimination experiment, we measured the just-noticeable differences (JNDs) in F0 and VTL that children could perceive via a 3-interval, 3-alternative forced-choice (3I-3AFC) adaptive procedure using CVCVCV pseudowords, for instance, 'ba-ki-mo'. The F0 and VTL parameters of the stimuli were manipulated using the same STRAIGHT procedure as described in the previous section for the current experiment. The initial voice difference was 12 st in F0 or VTL values relative to the original female speaker's voice. After two consecutive correct responses, the voice difference decreased two step sizes, and after an incorrect response, the voice difference increased one step size (2-down, 1-up). The step size initially had a value of 2 st but after 15 trials with the same step size or when the difference became smaller than two times the step size, the step size was divided by $\sqrt{2}$. The geometric mean of the voice difference values at the last six out of eight reversals was calculated to determine participants' JND, which corresponds to the 70.7% correct discrimination point on the psychometric function (Levitt, 1971). Our results showed that the discrimination of differences in F0 was adult-like around the age of 8 for VTL, while the ability to discriminate differences in F0 was still not adult-like at the age of 12 for most children. More information about the methods and the procedure of these experiments is presented in Nagels et al. (2020a).

The speech in competing speech perception experiment started with a practice session consisting of eight trials to familiarize participants with the task. Participants heard the target sentences without the masker during the first three practice trials and a combination of the target and masker speech with different F0 and VTL parameters with a TMR of +6 dB for the remaining practice trials. We did not set a criterion level for performance before moving on to the experiment trials. The experiment session consisted of seven items per TMR (-6 dB, 0 dB, or +6 dB) for

each of the four masker speech voice conditions, resulting in a total number of 84 trials ($7 \text{ items} \times 3 \text{ TMRs} \times 4 \text{ voice conditions}$). The TMR values were chosen based on the adaptive TMR values of Flaherty et al. (2019), which were mostly between -6 dB and +6 dB, and pilot testing with four children between 6 and 12 years of age who were not included in the study. All items were presented in a randomized order in a single block. The total duration of the experiment was approximately 15 minutes, including two optional breaks.

The experiment was conducted on a Dell XPS 13 inch touchscreen laptop using a child-friendly experiment interface (Figure 3.1) created in Matlab (MathWorks Inc., 2012). The stimuli were presented to participants via Sennheiser HD 380 Pro headphones. Child participants were tested in a quiet room at their homes, and adult participants were tested in a quiet testing room at the University of Groningen. Before the experiment, young children were asked to name all six basic colors and eight numbers used to ensure they knew the correct words. We instructed participants to attend only to the target speech which started 750 ms later than the masker speech and contained the carrier phrase, e.g., *Laat de hond zien waar de (color) (number) is.* [Show the dog where the (color) (number) is.]. The participants were told to press the color-number combination button that was mentioned in the target speech as fast as possible after hearing the stimulus. They could choose between 48 buttons ($6 \text{ colors} \times 8 \text{ numbers}$) and would receive 1 point if they had both the color and number correct, and 0 points when they had the color, number, or neither of them correct, similar to, for instance, Brungart et al. (2001). Thus, the probability of giving a correct response due to chance was 2.08%. After a response, the experiment would continue to the next trial. Participants did not receive any feedback on the accuracy of their responses during the practice or experiment sessions.

Figure 3.1: The experiment response interface.

3.2.4 Data analysis

Children’s accuracy scores were analyzed as a function of age, TMR, and target-masker differences in F0 and VTL. We normalized the target-masker differences in F0 and VTL in semitones by defining them as $\delta F0 = -\Delta F0 \div 12 - 0.5$ and $\delta VTL = \Delta VTL \div 3.6 - 0.5$, to make the differences in F0 and VTL commensurate for model fitting. As a result, $\delta F0$ and δVTL values of -0.5 corresponded to the original female speaker’s voice parameters, while $\delta F0$ and δVTL values of $+0.5$ corresponded to the manipulated male-sounding voice. Subsequently, we fitted a mixed-effects logistic regression model using the lme4 package (Bates et al., 2015) in R (R Core Team, 2020) with random intercepts per participant. We only fitted the model on children’s data without those of adults due to the non-continuous distribution of age. We used ANOVA Chi-Square tests to assess the improvement in models due to the deletion of individual factors, starting from the full factorial model, in lme4 syntax: `correct ~  $\delta F0$  *  $\delta VTL$  * TMR *  $\log(\text{age})$  + (1 | participant)`. The outcome variable *correct* indicated whether the participant had both the color and number correct (1 point) or not (0 points). The predictor variables $\delta F0$ and δVTL consisted of the normalized differences in F0 and VTL, *TMR* represented the different TMRs of -6 dB, 0 dB, and +6 dB, and $\log(\text{age})$ indicated children’s log-transformed age in years, rounded to two decimal places.

Furthermore, we used a Dunnett’s Test using the DescTools package (Signorell et al., 2018) to compare the mean benefit from target-masker differences in F0 and VTL of different child age groups to that of adults and approximate at what age children’s benefit is adult-like. For this analysis, we converted participants’ statistical

benefit from target-masker differences in F0 and VTL into Berkson units per semitone (Bk/st) (Hilkhuyzen et al., 2012). The conversion into Bk/st makes the differences in participants' benefit from F0 and VTL changes easier to interpret, as an increase of 1 Bk/st corresponds to doubling the odds of getting a correct response for each semitone of voice difference. To obtain the voice-benefits, we first computed two mixed-effects logistic regression models with random intercepts and slopes for $\delta F0$ per participant, in lme4 syntax: `correct ~ ( $\delta F0$  | participant)`, and for δVTL per participant, in lme4 syntax: `correct ~ ( $\delta VTL$  | participant)`. Subsequently, we extracted participants' model coefficients and scaled them to correspond to $\log_2$ odds per semitone, i.e., $\delta F0 \text{ coefficient} \div (12 \times \log(2))$ and $\delta VTL \text{ coefficient} \div (3.8 \times \log(2))$, because the logit is based on the natural log. In addition, we examined at what age children's accuracy scores were approximately adult-like by performing a Dunnett's Test comparing the mean logit-transformed accuracy scores of the specified child age groups to that observed in adults. We applied a logit-transformation on the accuracy scores to take into account the near ceiling-level scores of adults and older children, primarily in the +6 dB TMR condition.

We performed an additional analysis in which we interpolated participants' accuracy scores across TMRs to the same performance level for all participants, including adults, as the proportion of the benefit from target-masker differences seems to depend on the overall accuracy scores. Particularly because adult participants often demonstrated ceiling-level performance, except in the -6 dB TMR condition, there was less room for improvement in their performance, which makes the proportion of their benefit from target-masker differences relatively small compared to that observed in children, despite the fact that the logistic model would take the saturation into account. We calculated the mean logit-transformed accuracy scores of children and adults, and interpolated these across TMRs to a logit score of 1.79 (equal to an accuracy score of 85.7% correct) in the same-voice condition with no target-masker differences in F0 or VTL. This specific score was chosen for interpolation, as it resulted in the fewest outliers. The scores of six children were all below 85.7% and the scores of three children and three adults were all above 85.7% in the same-voice condition. We computed a linear mixed-effects model to examine participants' benefit of F0 and VTL differences once their accuracy scores were interpolated, in lme4 syntax: `interpolated logit score ~  $\delta F0$  *  $\delta VTL$  *  $\log(\text{age})$  + (1 | participant)`. Furthermore, we used a Dunnett's Test to compare the mean benefit from target-masker differences in F0 and VTL of the specified child age groups to that of adults based on their interpolated logit accuracy scores.

Finally, we performed several correlation analyses to investigate if children's benefit from target-masker differences in F0 and VTL and their overall ability to perceive speech in competing speech were related to their F0 and VTL discrimination thresholds (Nagels et al., 2020a). To prevent a correlation coming forth from merely a general effect of age, we used the residuals of children's benefit from target-masker differences in F0 and VTL, their overall accuracy scores, and their F0 and VTL discrimination thresholds. To calculate the residuals, we used linear regression models with children's F0 or VTL voice-benefits (in Bk/st), their overall accuracy scores, and children's F0 or VTL JNDs as outcome variables and only age as a fixed effect, e.g., `overall accuracy scores ~ age`. Subsequently, we calculated the Pearson's correlation coefficient for the correlations between the residuals of children's F0 and VTL voice-benefits and their corresponding JNDs, and the correlations between the residuals of their overall accuracy scores, and F0 and VTL JNDs. In addition, we examined if children's benefits from F0 and VTL differences were related, and if children's vocabulary size potentially affected their ability to perceive speech in competing speech by calculating the Pearson's correlation coefficient between the residuals of children's vocabulary scores and their overall accuracy scores.

3.3 Results

3.3.1 F0 and VTL benefit for perceiving speech in competing speech

Figure 3.2 shows the accuracy scores of participants in percentage points as a function of TMR, masker speech voice condition, and age. While adults demonstrated near ceiling-level performance in most conditions, most 4- to 6-year-olds could not perform the task in the condition with no target-masker differences in F0 or VTL at a TMR of -6 dB. Figure 3.3 shows the increase in participants' accuracy scores in percentage points as a function of the masker speech voice condition with respect to the same-voice condition, averaged across TMRs. We used backward stepwise model comparison using ANOVA Chi-Square tests to select the best fitting, most parsimonious model for children's total number of correct responses by deleting one factor at a time from the four-way interaction in the full model. The model comparison analysis indicated that the model with three-way interactions between $\delta F0$, δVTL , and *TMR*, and between $\delta F0$, *TMR*, and *age*, and a two-way interaction between δVTL and *age* was the best fitting model, in lme4 syntax: `correct ~  $\delta F0$  *  $\delta VTL$  * TMR +  $\delta F0$  * TMR * log(age) +  $\delta VTL$  * log(age) + (1 | participant)`.

The best model shows children's accuracy scores significantly increased as the TMR became more advantageous ($\zeta = 2.87$, *Estimate* = 0.20, *SE* = 0.071, $P < 0.01$) and as a function of age ($\zeta = 11.80$, *Estimate* = 2.83 *SE* = 0.239, $P < 0.001$). Children's accuracy scores also improved as a result of target-masker differences in δVTL ($\zeta = 2.50$, *Estimate* = 1.59, *SE* = 0.637, $P < 0.05$). There were significant two-way interactions between *TMR* and $\delta F0$ ($\zeta = -3.03$, *Estimate* = -0.43, *SE* = 0.141, $P < 0.01$) and between *TMR* and δVTL ($\zeta = -3.97$, *Estimate* = -0.08, *SE* = 0.021, $P < 0.001$) demonstrating that the beneficial effect of target-masker differences in F0 or in VTL on accuracy scores became smaller as the TMR became more advantageous. Furthermore, there was a significant two-way interaction between $\delta F0$ and δVTL ($\zeta = -5.65$, *Estimate* = -1.16, *SE* = 0.205, $P < 0.001$), which shows that a combined target-masker difference in F0 and VTL increased children's accuracy scores less than the additive effect of individual cues. There was also a significant three-way interaction between $\delta F0$, δVTL , and *TMR* ($\zeta = -1.99$, *Estimate* = -0.08, *SE* = 0.041, $P < 0.05$), indicating the beneficial effect of a combined target-masker difference in F0 and VTL on children's accuracy scores decreased as the TMR became more advantageous. Finally, the three-way interaction between $\delta F0$, *TMR*, and *age* was significant ($\zeta = 2.33$, *Estimate* = 0.17, *SE* = 0.071, $P < 0.05$), showing that the beneficial effect of target-masker differences in F0 on accuracy scores became smaller as the TMR became higher, but that this differed across ages. However, based on Figure 3.2, this effect mainly seems to be caused by some 4- to 6-year-old children who showed a detrimental effect of F0 differences in the +6 dB TMR condition. To summarize, children's accuracy scores generally improved with age, higher TMRs, and target-masker differences in either or both F0 and VTL cues. The benefit from target-masker differences in F0 and VTL decreased for children as the TMR became more advantageous. We did not find any differences in the benefit that children derived from target-masker differences across age, meaning the size of the improvement in accuracy as a result of differences in F0 and VTL was the same for children of all ages.

Figure 3.2: Accuracy scores of participants for perceiving speech in competing speech in percentage points as a function of age, TMR, and voice condition ($N_{\text{children}} = 55$, $N_{\text{adults}} = 15$). The panels from top to bottom show the accuracy scores in percentage points for the -6 dB TMR (upper panels), 0 dB TMR (middle panels), and +6 dB TMR (lower panels) conditions. Each row consists of four plots that show the accuracy scores per masker speech voice condition, arranged from the condition with target-masker differences of 0 st in F0 and VTL (left panels) to the condition with target-masker differences of -12 st in F0 and +3.8 st in VTL (right panels). The boxplots show the median accuracy scores of participants per age group and the lower and upper quartiles. The dots represent individual data points at participants' age, and the whiskers indicate the lowest and highest data points within plus or minus 1.5 times the interquartile range.

Figure 3.3: The improvement in the accuracy scores of participants, averaged across TMRs, as a function of the masker speech voice condition ($N_{\text{children}} = 55$, $N_{\text{adults}} = 15$) shown across the panels and as a function of age within each panel. Each panel shows the improvement in individual participants' accuracy scores with respect to their accuracy scores in the voice condition with no target-masker differences in F0 or VTL. The dots represent individual data points at participants' age.

3.3.2 Adult-like benefit from F0 and VTL differences and adult-like accuracy scores

Figures 3.4a and 3.4b show participants' benefit from target-masker differences in F0 and VTL cues in Bk/st for each TMR condition, and Figure 3.4c shows participants' accuracy scores for each TMR condition presented in percentage points. Positive Bk/s/s reflect a benefit from F0 or VTL differences, while negative Bk/s/s reflect a detrimental effect of F0 or VTL differences. Based on Figures 3.4a and 3.4b, we observe that almost all children demonstrated a benefit of F0 and VTL differences in the -6 dB and 0 dB TMR conditions. The Dunnett's Test results are presented in Table 3.2. Since we compared the results of adults to those of different child age groups instead of using age as a continuous factor, this analysis only gives a rough approximation of when children's performance is adult-like rather than a

precise estimate. For F0, children between 4 and 10 years of age overall benefited more from F0 differences than adults, but the benefit from F0 differences derived from children between 10 to 12 years of age did not differ from that of adults. When we examined children's benefit from F0 differences in the different TMR conditions, we found that all children benefited more from F0 differences than adults in the -6 dB and the 0 dB conditions. In the +6 dB TMR condition, 4- to 6-year-old and 8- to 10-year-old children differed from adults by, in fact, showing a small but significant detrimental effect of F0 differences on their accuracy scores. For VTL, children benefited more from VTL differences than adults at all tested ages. Children also showed a greater benefit from VTL differences than adults in the -6 dB TMR condition at all tested ages and only 10- to 12-year-old children did not show a larger benefit than adults in the 0 dB TMR condition. Finally, similar to F0 differences, 4- to 6-year-old children and 8- to 10-year-old children showed a small but significant detrimental effect of VTL differences on their accuracy scores. In summary, children's overall benefit from F0 was significantly larger than adults' between 4 and 10 years of age but no longer differed from adults' after 10 years of age. The overall benefit that children had of VTL differences was significantly larger than that observed in adults for children at all tested ages. All children showed a larger benefit from F0 and VTL differences than adults in the -6 dB condition, 10- to 12-year-old children no longer showed a larger benefit from VTL differences in the 0 dB condition, and only 4- to 6-year-old and 8- to 10-year-old children differed from adults by showing a small but significant detrimental effect of F0 and VTL differences in the +6 dB condition.

For participants' overall logit-transformed accuracy scores, the Dunnett's Test indicated that children's total number of correct responses was lower than adults' at all tested ages, even for the oldest children tested. In the -6 dB and 0 dB TMR conditions, all children's accuracy scores were lower than adults'. However, in the +6 dB TMR condition, 6- to 8-year-old and 10- to 12-year-old children's accuracy scores did not differ from those of adults. The accuracy scores of 4- to 6-year-old and 8- to 10-year-old children differed from adults at all TMRs. To summarize, children's overall accuracy scores and their accuracy scores in the -6 dB and 0 dB conditions were generally lower compared to adults. In the +6 dB condition, only 4- to 6-year-old and 8- to 10-year-old children had lower accuracy scores than adults.

Table 3.2: The Dunnett's Test analysis results for the differences between adults' and children's mean benefit from target-masker differences in F0 and VTL, and their accuracy scores per age group, overall and per TMR condition.

	Age group	Overall	TMR -6 dB	TMR 0 dB	TMR +6 dB
<i>F0 benefit</i>	4-6 years	0.07, $P < 0.01^{**}$	0.19, $P < 0.001^{***}$	0.13, $P < 0.001^{***}$	-0.05, $P < 0.001^{***}$
	6-8 years	0.05, $P < 0.001^{***}$	0.14, $P < 0.001^{***}$	0.06, $P < 0.001^{***}$	-0.01, $P = 0.07$
	8-10 years	0.05, $P < 0.001^{***}$	0.12, $P < 0.001^{***}$	0.03, $P < 0.05^{*}$	-0.02, $P < 0.05^{*}$
	10-12 years	0.02, $P = 0.35$	0.06, $P < 0.05^{*}$	0.05, $P < 0.001^{***}$	-0.002, $p = 0.97$
<i>VTL benefit</i>	4-6 years	0.33, $P < 0.001^{***}$	0.46, $P < 0.001^{***}$	0.42, $P < 0.001^{***}$	-0.10, $P < 0.001^{***}$
	6-8 years	0.22, $P < 0.001^{***}$	0.35, $P < 0.001^{***}$	0.20, $P < 0.001^{***}$	-0.03, $P = 0.06$
	8-10 years	0.17, $P < 0.001^{***}$	0.24, $P < 0.001^{***}$	0.16, $P < 0.001^{***}$	-0.03, $P < 0.05^{*}$
	10-12 years	0.09, $P < 0.001^{***}$	0.16, $P < 0.001^{***}$	0.09, $P = 0.06$	-0.004, $P = 0.98$
<i>Overall logit accuracy scores</i>	4-6 years	-2.93, $P < 0.001^{***}$	-3.18, $P < 0.001^{***}$	-2.70, $p < 0.001^{***}$	-2.16, $P < 0.001^{***}$
	6-8 years	-2.03, $P < 0.001^{***}$	-2.45, $P < 0.001^{***}$	-1.39, $p < 0.001^{***}$	-0.58, $P = 0.07$
	8-10 years	-1.70, $P < 0.001^{***}$	-1.85, $P < 0.001^{***}$	-1.12, $p < 0.001^{***}$	-0.69, $P < 0.05^{*}$
	10-12 years	-0.99, $P < 0.001^{***}$	-1.02, $P < 0.001^{***}$	-0.72, $p < 0.05^{*}$	-0.09, $P = 0.98$

Figure 3.4: Participants' benefit from target-masker differences in F0 and VTL in Bk/st, and their accuracy scores per TMR condition ($N_{\text{children}} = 55$, $N_{\text{adults}} = 15$). (a) Participants' benefit from target-masker differences in F0 for perceiving speech in competing speech in Bk/st overall and per TMR condition. (b) benefit from target-masker differences in VTL for perceiving speech in competing speech in Bk/st overall and per TMR condition. The boxplots show participants' median benefit from F0 or VTL in Bk/st per age group, and the lower and upper quartiles of the values that were obtained. The dashed line at the value of 0 was included to indicate if participants benefitted from voice difference, i.e., positive value, or not, i.e., negative value. (c) Participants' accuracy scores in percentage points per TMR condition. The boxplots show participants' median accuracy scores per age group.

In all plots, the dots represent individual data points at participants' age, and the whiskers indicate the lowest and highest data points within plus or minus 1.5 times the interquartile range.

3.3.3 Adult-like benefit from F0 and VTL differences for interpolated accuracy scores

Figure 3.5a shows participants' mean accuracy scores interpolated across TMRs to 85.7% correct in the same-voice condition, equal to the mean accuracy score of adults in the -6 dB TMR condition, and Figure 3.5b shows the TMRs of individual participants which were used for the interpolation. Note that twelve participants had scores that were either all below or above 85.7% correct which explains why not all data points are centered to 85.7% correct in the upper left panel of Figure 3.5a. We fitted the model on the data of all participants, children and adults combined, and used log-transformed age values to reduce the effects of the non-continuous distribution in age. The backward step-wise model comparison analysis showed the model with *age* as a fixed factor and a two-way interaction between $\delta F0$ and δVTL was the best fitting model, in lme4 syntax: `interpolated logit score ~ log(age) +  $\delta F0$  *  $\delta VTL$  + (1 | participant)`.

There was a significant effect of $\delta F0$ ($t = 3.45$, $Estimate = 0.19$, $SE = 0.057$, $P < 0.001$) and a significant two-way interaction between $\delta F0$ and δVTL ($t = -4.57$, $Estimate = -0.53$, $SE = 0.116$, $P < 0.001$) that demonstrate participants' accuracy scores improved due to target-masker differences in F0 and a combined difference in F0 and VTL. We did not find a significant main effect of δVTL , although it was near-significant ($t = 1.80$, $Estimate = 0.10$, $SE = 0.058$, $P = 0.07$). We also only found a significant main effect of *age* on participants' accuracy scores ($t = 4.92$, $Estimate = 0.37$, $SE = 0.075$, $P < 0.001$). This lack of an interaction between *age* with $\delta F0$ and δVTL indicates that age did not affect the benefit participants derived from target-masker differences in F0 and VTL once their overall accuracy scores were interpolated to the same performance level in the same-voice condition.

Furthermore, to evaluate at what age children benefited similarly from F0 and VTL differences as adults, we performed a Dunnett's Test on participants' benefit, i.e., the difference between their interpolated logit-transformed accuracy scores in the masker speech voice condition with -12 st in F0 and 3.8 st in VTL minus the accuracy scores in the same-voice condition. These results indicated that the benefit only differed from adults for 4- to 6-year-old children (4-6 years difference = -0.40, $P < 0.01$; 6-8 years difference = 0.07, $P = 0.94$; 8-10 years difference = 0.08, $P = 0.87$; 10-12 years difference = -0.05, $P = 0.98$). Intriguingly, as can be seen in Figure 3.5a, some 4-

to 6-year-old children showed a detrimental effect of target-masker differences in F0. However, based on Figure 3.2, this only seemed to be the case in the +6 dB TMR condition.

Figure 3.5: Mean accuracy scores of participants per age group and voice condition interpolated across TMRs ($N_{\text{children}} = 55$, $N_{\text{adults}} = 15$). (a) Participants' mean accuracy scores per age group and voice condition interpolated to 85.7% correct in the same-voice condition. The four panels show the accuracy score for each masker speech voice condition. (b) Participants' TMRs at 85.7% correct in the same-voice condition which were used for the interpolation. The boxplots show the median accuracy scores of participants per age group and the lower and upper quartiles. The dots represent individual data points at participants' age, and the whiskers indicate the lowest and highest data points within plus or minus 1.5 times the interquartile range.

3.3.4 Correlations of F0 and VTL benefit, and overall accuracy scores with JNDs

Figure 3.6a shows the correlations between the residuals of children's benefit from target-masker differences in F0 and VTL, and the residuals for their corresponding JNDs. The correlation analysis between the residuals of children's F0-difference benefit (Bk/st) and their F0 JNDs, i.e., after partialling out the effect of age, indicated there was no significant correlation between the two measures (*Pearson's* $r = 0.14$, $P = 0.32$). For the residuals of participants' VTL-difference benefit and their VTL JNDs, we also did not find a significant correlation between the two measures (*Pearson's* $r = 0.19$, $P = 0.15$). Thus, children's benefit from target-masker differences in F0 and VTL was not directly related to their respective discrimination thresholds.

Figure 3.6b shows the correlations between children's overall accuracy score residuals, and their F0 and VTL JND residuals. The correlation analysis between the residuals of children's overall accuracy scores and their F0 JNDs indicated there was no significant correlation between the two measures (*Pearson's* $r = -0.26$, $P = 0.06$). There was also no significant correlation between the residuals of children's overall accuracy scores and their VTL JNDs (*Pearson's* $r = -0.14$, $P = 0.32$). Thus, children's overall accuracy scores were also not directly related to their F0 and VTL discrimination thresholds. In addition, Figure 3.6c shows the correlation between children's benefit from target-masker differences in F0 and VTL. These two measures were also not significantly correlated (*Pearson's* $r = 0.05$, $P = 0.73$). Finally, Figure 3.6d shows the correlation between the residuals of children's overall accuracy scores and the raw scores of their vocabulary size as measured by the Renfrew Word Finding Vocabulary Test. The correlation analysis indicated there was no significant correlation between the two measures (*Pearson's* $r = 0.24$, $P = 0.08$).

Figure 3.6: The relationship between the residuals of children's benefit from target-masker differences in F0 and VTL and their corresponding JNDs, and the relationships between the residuals of children's overall accuracy scores with their F0 JNDs, VTL JNDs, and vocabulary scores ($N_{\text{children}} = 55$). (a) The relationships

between the residuals of children's benefit from target-masker differences in F0 (left panel) and VTL (right panel) in Bk/st and their corresponding JNDs. (b) The relationships between the residuals of children's overall accuracy scores and their F0 (left panel) and VTL JNDs (right panel). (c) The relation between the residuals of children's benefit from target-masker differences in F0 (left panel) and VTL (right panel) in Bk/st. (d) The relation between the residuals of children's overall accuracy scores and the raw scores of their vocabulary size based on the Renfrew Word Finding Vocabulary Test. In all plots, the central line represents a linear regression line and the surrounding area shows the 95% confidence intervals. The shape of the data points indicates the age group of participants.

3.4 Discussion

In the current study, we investigated how children's benefit from target-masker differences in speakers' mean F0 and VTL for the perception of speech in competing speech develops during the school-age years (4 to 12 years of age). Our results show that the accuracy scores of children improved as a function of age and TMR as expected based on previous research (Bonino et al., 2013; Buss et al., 2017a, 2019; Corbin et al., 2016; Flaherty et al., 2019, 2021; Hall et al., 2002; Leibold et al., 2018; Leibold & Buss, 2013; Wightman & Kistler, 2005). Children of all age groups showed a benefit from target-masker differences in either or both F0 and VTL. The benefit from target-masker differences in F0 and VTL also decreased as the TMR became more advantageous, likely due to the increase in children's overall accuracy scores. The size of the benefit from target-masker differences in F0 and VTL mainly seemed to be larger in children than in adults because of their overall lower accuracy scores, which were not adult-like at all tested ages and hence left more room for improvement. This explanation was confirmed by the lack of an effect of age on participants' benefit when their mean accuracy scores were interpolated across TMRs to the same performance level. Finally, we examined if the benefit from target-masker differences in F0 and VTL for children was related to their respective discrimination thresholds. Our results indicate that there were no significant correlations between children's benefit from F0 and VTL differences and their respective JNDs, or between their benefit from F0 differences and their benefit from VTL differences. Moreover, we did not find any correlations for children's overall accuracy scores with their F0 JNDs, VTL JNDs, or vocabulary scores.

Our findings indicate that children of all age groups showed a benefit from target-masker differences in either or both F0 and VTL for perceiving speech in competing speech. The benefit from voice differences was particularly large for

young children at the -6 dB TMR, the highest masker speech level tested. While the performance of 4- to 6-year-old children was 7.14% on average in the same-voice condition, their performance increased substantially as a result of introducing target-masker differences in F0 (average performance of 47.1%), VTL (average performance of 38.6%), or both F0 and VTL (average performance of 64.3%). This considerable improvement in performance as a result of voice difference benefit is in agreement with previous research indicating that most children are able to distinguish different speakers when there are large differences between their voices (Cleary et al., 2005; Creel & Jimenez, 2012). Children benefitted from F0 and VTL differences at all tested ages, although their benefit only became as small as that observed in adults around 10 years of age. However, the observed differences in the benefit from F0 and VTL differences between children and adults mainly seemed to be caused by the differences in their overall accuracy scores. This explanation is also in line with our findings that children's overall accuracy scores did not reach an adult-like level for all tested ages. When we interpolated participants' accuracy scores across TMRs to a performance of 85.7% correct in the same-voice condition, we did not observe any significant differences across age. However, the results from the Dunnett's Test indicated that 4- to 6-year-old children benefitted less than adults, but this seems to be caused by a slightly detrimental effect that was observed in a small number of young children.

Counterintuitively, for three out of ten 4- to 6-year-old children, there seemed to be a detrimental effect of target-masker differences in F0, although this was only observed in the +6 TMR condition. The underlying cause of this detrimental effect could be the specific combination of the lower presentation level of the masker compared to the target (6 dB lower) and the manipulation of the masker speech voice. While the favorable TMR should have given an advantage in selecting the target speech stream and inhibiting the masker speech stream, which is the case for older children and adults, it seems not to have worked in this manner for a small number of young children within our participant group. Some young children may not have mastered the effective use of selective attention mechanisms yet, which would allow them to focus on the target and ignore the masker like other children and adults. Given that the voice of the target speaker remained the same throughout the experiment, the exposure to these voice parameters was higher than the manipulated voice parameters of maskers, which may have caused a so-called novelty effect due to the change in voice gender (Darwin et al., 2003). This effect could make the masker speech stream more distracting for children rather than facilitating the segregation and selection of the target speech stream. Supporting this

idea, there are some studies on the mechanisms of selective attention in the processing of visual cues in children that suggest that the interference of distractors becomes larger when the perceptual load of the task becomes lower (Huang-Pollock et al., 2002; Lavie, 2005). Tasks with a high perceptual load engage the full capacity of attentional resources for the processing of task-relevant stimuli, while tasks with a low perceptual load give more opportunity for the processing of also task-irrelevant stimuli. On the other hand, this detrimental effect of voice differences was mainly observed in a very small number of participants, three out of ten children within the youngest group of 4- to 6-year-olds, and hence it could also have been a simple consequence of brief periods of inattention (Lutfi et al., 2003; Sussman & Steinschneider, 2009). Wightman and Kistler (2005) also addressed some of the difficulties involved in estimating psychometric function data from young children, such as higher individual variability in their performance related to the rapid development that takes place at this age, and the upper asymptote not reaching 100% correct due to general attention span.

Our findings partially contradict the earlier results from Flaherty et al. (2019, 2021), which showed that young children do not benefit from target-masker differences in F0 or VTL and benefit less than older children and adults from combined differences in F0 and VTL. The F0 differences tested in their studies, namely -3 st, -6 st, and -9 st, and VTL differences using scaling factors of 0.84 and 1.16 were smaller than the -12 st difference in F0 and 3.8 st difference in VTL that we have used. However, differences in speakers' mean F0 of -6 st and -9 st seem to be well within the range of most school-age children's discrimination thresholds (Buss et al., 2017b; Flaherty et al., 2019; Nagels et al., 2020a). The mean F0 discrimination threshold was 6.18 st for the youngest age group of 4- to 6-year-old children in the current study, as was previously reported by Nagels et al. (2020a). Participants' JNDs also correspond to the 70.7% correct discrimination point on the psychometric function, so voice differences below their JND could still be expected to be somewhat audible. In addition, similarly to Flaherty et al. (2019), we did not find any significant correlations between children's discrimination thresholds and their benefit of target-masker differences in F0 or VTL, or with their overall accuracy scores. The lack of a significant correlation also suggests that children's reduced sensitivity to differences in voice cues does not seem to be the primary factor explaining their poorer perception of speech in competing speech compared to adults. Instead, children's ability to segregate different speech streams seems to rely on different voice perception mechanisms than those involved in voice discrimination, such as selective auditory attention and inhibition. In

agreement with this explanation, Sussman et al. (2007), and Sussman and Steinschneider (2009) reported a similar discrepancy between children's frequency discrimination abilities and their ability to use frequency differences to segregate two streams of pure tones. Nevertheless, it should be kept in mind that the target-masker voice differences of -12 st in F0 and 3.8 st in VTL in the current study were well above most children's discrimination thresholds. A closer relationship between discrimination thresholds and a benefit from target-masker voice differences may be present if more subtle voice differences that are nearer to children's discrimination thresholds were used.

Another explanation for the discrepancy between our results and those of Flaherty et al. (2019, 2021) could be the use of a carrier phrase to mark the target sentences in the current study. Using the same consistent carrier phrase most likely helped children to segregate the different speech streams and use the voice differences more effectively for keeping track of the target speech stream (Bonino et al., 2013; Freyman et al., 2004; Sussman-Fort & Sussman, 2014). Bonino et al. (2013) reported a large improvement of 16.8% in the accuracy scores for children's perception of speech in competing speech when a carrier phrase was used. An idea for future research would also be to make the task more difficult for adults, for example by modulating the onset asynchrony between the target and masker sentences (Lee & Humes, 2012), and hence minimizing the chance for ceiling performance. Finally, our masker speech consisted of sentence chunks instead of full sentences and a single-talker speech masker instead of a two-talker speech masker. These design differences could have led to less effective overall masking in the current study (Buss et al., 2017a; Litovsky, 2005; Rosen et al., 2013). For instance, the single-talker masker may have provided more opportunity for "glimpsing" acoustic information due to temporary reductions in masking (Rosen et al., 2013), although findings by Buss et al. (2017a) suggest that young children cannot use these low-level glimpses as efficiently as adults. Further research should examine the individual contributions of these design parameters to children's benefit from voice differences for perceiving speech in competing speech.

Furthermore, we did not find any significant correlations between children's overall ability to perceive speech in competing speech and their vocabulary scores in the current study. However, we chose to use simple sentences with the same carrier phrase and sentence structure as target sentences, and words for basic colors and numbers up to 10 as target words in these sentences to ensure that children could perform the task at all tested ages. Such words are acquired very early in life, around three or four years of age (Brysbaert et al., 2014). If there are effects of

vocabulary size on the ability to perceive speech in competing speech, the use of simple closed-set sentence materials with very familiar words likely limited these effects in the current study, as such effects have been reported when more complex open-set sentence stimuli were used (Klein et al., 2017; McCreery et al., 2017, 2020). Another reason for choosing these simple sentence materials was that we are planning to follow up on these findings by testing children with cochlear implants, who often have smaller vocabulary sizes than their age-matched peers (Fagan & Pisoni, 2010; Lund, 2016), using the same testing materials and procedure.

Finally, children may have more difficulties with perceiving speech in competing speech due to their limited acoustic and language experience compared to adults, as was suggested earlier by Corbin et al. (2016). Children may need more time and experience to develop representations of speech that are robust enough to withstand the perceptual obliteration caused by masking. This idea is also supported by the fact that children seem to require a greater spectrotemporal resolution for speech perception in degraded speech conditions (Eisenberg et al., 2000; Mlot et al., 2010) and use word probability and sentential context less effectively than older children and adults (Buss et al., 2019; Craig et al., 1993; Elliott et al., 1987; Metsala, 1997). On the other hand, a reduced ability to perceive speech in competing speech may also negatively affect children's language development, as it limits the amount of language input they receive. This matter of causality has primarily been addressed with respect to the language development of children with hearing impairments, such as for children with cochlear implants (Fagan & Pisoni, 2010; Geers et al., 2017) but has received little attention for children with normal hearing. Having a lower ability to perceive speech in competing speech may also negatively affect the language development of children with normal hearing.

In conclusion, children seem to benefit from target-masker differences in speakers' F0 or VTL at all tested ages in the current study. Their benefit from differences in F0 and VTL was larger than that observed in adults until the age of 10, but when we interpolated participants' accuracy scores to the same performance level, the benefit was proportionally the same across all tested ages. Children's overall accuracy scores also did not become adult-like for all tested ages, which indicates that the ability to perceive speech in competing speech continues to develop even for the oldest tested age group of 10- to 12-year-old children. Moreover, we did not find any correlations between children's benefit from target-masker differences in F0 and VTL and their corresponding voice cue JNDs or between children's overall accuracy scores and F0 JNDs, VTL JNDs, or vocabulary scores. These findings suggest that children's ability to perceive speech in

competing speech is not directly related to their voice discrimination abilities. Instead, other factors such as those related to selective auditory attention and inhibition seem to be more closely associated with the developing speech in competing speech perception abilities of school-age children.

Bibliography

- Arbogast, T. L., Mason, C. R., & Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America*, 112(5), 2086–2098.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–51.
- Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, 107(2), 1065–1066.
- Bonino, A. Y., Leibold, L. J., & Buss, E. (2013). Release from perceptual masking for children and adults: Benefit of a carrier phrase. *Ear and Hearing*, 34(1), 3–14.
- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press.
- Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception, & Psychophysics*, 77(5), 1465–1487.
- Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America*, 120(6), 4007–4018.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, 110(5), 2527–2538.
- Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, 150, 80–84.
- Buss, E., Leibold, L. J., Porter, H. L., & Grose, J. H. (2017a). Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing. *The Journal of the Acoustical Society of America*, 141(4), 2650–2660.
- Buss, E., Flaherty, M. M., & Leibold, L. J. (2017b). Development of frequency discrimination at 250 Hz is similar for tone and /ba/ stimuli. *The Journal of the Acoustical Society of America*, 142(1), EL150–EL154.
- Buss, E., Hodge, S. E., Calandruccio, L., Leibold, L. J., & Grose, J. H. (2019). Masked sentence recognition in children, young adults, and older adults: Age-dependent effects of semantic context and masker type. *Ear and Hearing*, 40(5), 1117–1126.
- Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual masking in multiple sound backgrounds. *The Journal of the Acoustical Society of America*, 45(3), 694–703.
- Cleary, M., Pisoni, D. B., & Kirk, K. I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 48(1), 204–223.

- Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, 123(1), 414–427.
- Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear and Hearing*, 37(1), 55–63.
- Craig, C. H., Kim, B. W., Pecyna Rhyner, P. M., & Bowen Chirillo, T. K. (1993). Effects of word predictability, child development, and aging on time-gated speech recognition performance. *Journal of Speech, Language, and Hearing Research*, 36(4), 832–841.
- Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, 113(4), 487–509.
- Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America*, 114(5), 2913–2922.
- Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America*, 107(5), 2704–2710.
- El Boghdady, N., Gaudrain, E., & Başkent, D. (2019). Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users? *The Journal of the Acoustical Society of America*, 145(1), 417–439.
- Elliott, L. L., Hammer, M. A., & Evan, K. E. (1987). Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults. *Perception & Psychophysics*, 42(2), 150–157.
- Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., & Scott, S. K. (2016). Getting the Cocktail Party Started: Masking Effects in Speech Perception. *Journal of Cognitive Neuroscience*, 28(3), 483–500.
- Fagan, M. K., & Pisoni, D. B. (2010). Hearing experience and receptive vocabulary development in deaf children with cochlear implants. *The Journal of Deaf Studies and Deaf Education*, 15(2), 149–161.
- Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106(3), 1511–1522.
- Flaherty, M. M., Buss, E., & Leibold, L. J. (2019). Developmental effects in children’s ability to benefit from F0 differences between target and masker speech. *Ear and Hearing*, 40(4), 927–937.
- Flaherty, M. M., Buss, E., & Leibold, L. J. (2021). Independent and combined effects of fundamental frequency and vocal tract length differences for school-age children’s sentence recognition in a two-talker masker. *Journal of Speech, Language, and Hearing Research*, 64(1), 206–217.

- Floccia, C., Butler, J., Girard, F., & Goslin, J. (2009). Categorization of regional and foreign accent in 5- to 7-year-old British children. *International Journal of Behavioral Development*, 33(4), 366–375.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 115(5), 2246–2256.
- Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., & Başkent, D. (2014). Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology*, 15(6), 1037–1048.
- Gaudrain, E., & Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear and Hearing*, 39(2), 226–237.
- Geers, A. E., Mitchell, C. M., Warner-Czyz, A., Wang, N.-Y., Eisenberg, L. S., & Team, the Cd. I. (2017). Early sign language exposure and cochlear implantation benefits. *Pediatrics*, 140(1), 1–9.
- Hall, J. W., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear and Hearing*, 23(2), 159–165.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4), 377–396.
- Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & Shakespeare, B. (2009). Speech perception abilities of adults with dyslexia: Is there any evidence for a true deficit? *Journal of Speech, Language, and Hearing Research*, 52(6), 1510–1529.
- Hilkuysen, G., Gaubitch, N., Brookes, M., & Huckvale, M. (2012). Effects of noise suppression on intelligibility: Dependency on signal-to-noise ratios. *The Journal of the Acoustical Society of America*, 131(1), 531–539.
- Huang-Pollock, C. L., Carr, T. H., & Nigg, J. T. (2002). Development of selective attention: Perceptual load influences early versus late attentional selection in children and adults. *Developmental Psychology*, 38(3), 363–375.
- Kawahara, H., & Irino, T. (2005). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In *Speech Separation by Humans and Machines* (pp. 167–180). Springer. https://doi.org/10.1007/0-387-22794-6_11
- Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational Masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Red.), *Auditory Perception of Sound Sources* (pp. 143–189). Springer US.
- Klein, K. E., Walker, E. A., Kirby, B., & McCreery, R. W. (2017). Vocabulary facilitates speech perception in children with hearing aids. *Journal of Speech, Language, and Hearing Research*, 60(8), 2281–2296.
- Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, 9(2), 75–82.

- Lee, J. H., & Humes, L. E. (2012). Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background. *The Journal of the Acoustical Society of America*, 132(3), 1700–1717.
- Leibold, L. J., & Buss, E. (2013). Children’s identification of consonants in a speech-shaped noise or a two-talker masker. *Journal of Speech, Language, and Hearing Research*, 56(4), 1144–1155.
- Leibold, L. J., Buss, E., & Calandruccio, L. (2018). Developmental effects in masking release for speech-in-speech perception due to a target/masker sex mismatch. *Ear and Hearing*, 39(5), 935–945.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467–477.
- Litovsky, R. Y. (2005). Speech intelligibility and spatial release from masking in young children. *The Journal of the Acoustical Society of America*, 117(5), 3091–3099.
- Lund, E. (2016). Vocabulary knowledge of children with cochlear implants: A meta-analysis. *Journal of Deaf Studies and Deaf Education*, 21(2), 107–121.
- Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., & Callahan, M. R. (2003). One factor underlies individual differences in auditory informational masking within and across age groups. *Perception & Psychophysics*, 65(3), 396–406.
- MathWorks Inc. (2012). *MATLAB: The language of technical computing. Desktop tools and development environment, version 2012a*. MathWorks.
- Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59(3), 203–243.
- McCreery, R. W., Miller, M. K., Buss, E., & Leibold, L. J. (2020). Cognitive and linguistic contributions to masked speech recognition in children. *Journal of Speech, Language, and Hearing Research*, 63(10), 3525–3538.
- McCreery, R. W., Spratford, M., Kirby, B., & Brennan, M. (2017). Individual differences in language and working memory affect children’s speech recognition in noise. *International Journal of Audiology*, 56(5), 306–315.
- Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & Cognition*, 25(1), 47–56.
- Mlot, S., Buss, E., & Hall III, J. W. (2010). Spectral integration and bandwidth effects on speech recognition in school-aged children and adults. *Ear and Hearing*, 31(1), 56–62.
- Moore, T. J. (1981). Voice communication jamming research. *AGARD Conference Proceedings 331: Aural Communication in Aviation*, 2:1-2:6.
- Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2020a). Development of voice perception is dissociated across gender cues in school-age children. *Scientific Reports*, 10(1), 1–11.
- Nagels, L., Gaudrain, E., Vickers, D., Lopes, M. M., Hendriks, P., & Başkent, D. (2020b). Development of vocal emotion recognition in school-age children: The EmoHI test for hearing-impaired populations. *PeerJ*, 8, 1–14.

- Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *The Journal of the Acoustical Society of America*, 101(4), 2253–2266.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184.
- Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, 57(S1), S5.
- R Core Team. (2020). R: *A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Renfrew, C. E. (1995). *Word Finding Vocabulary Test*. Speechmark Publishing.
- Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, 133(4), 2431–2443.
- Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proceedings of the National Academy of Sciences*, 108(37), 15516–15521.
- Saleh, S. M. I. (2013). *The efficacy of fitting cochlear implants based on pitch perception* [Doctoral thesis, University College London].
<https://discovery.ucl.ac.uk/id/eprint/1400212/6/Thesis%20S.M.I.%20Saleh%202013.pdf>
- Saleh, S. M., Saeed, S. R., Meerton, L., Moore, D. R., & Vickers, D. A. (2013). Clinical use of electrode differentiation to enhance programming of cochlear implants. *Cochlear Implants International*, 14(S4), 16–18.
- Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, 18(7), 559–572.
- Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *The Journal of the Acoustical Society of America*, 115(2), 813–821.
- Semeraro, H. D., Rowan, D., Besouw, R. M. van, & Allsopp, A. A. (2017). Development and evaluation of the British English coordinate response measure speech-in-noise test as an occupational hearing assessment tool. *International Journal of Audiology*, 56(10), 749–758.
- Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arppe, A., Baddeley, A., Barton, K., Bolker, B., & Borchers, H. W. (2018). *DescTools: Tools for descriptive statistics* (R package version 0.99.25) [Computer software].
- Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, 118(5), 3177–3186.
- Smith, D. R. R., Walters, T. C., & Patterson, R. D. (2007). Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. *The Journal of the Acoustical Society of America*, 122(6), 3628–3639.

- Sussman, E., & Steinschneider, M. (2009). Attention effects on auditory scene analysis in children. *Neuropsychologia*, 47(3), 771–785.
- Sussman, E., Wong, R., Horváth, J., Winkler, I., & Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. *Hearing Research*, 225(1), 117–127.
- Sussman-Fort, J., & Sussman, E. (2014). The effect of stimulus context on the buildup to stream segregation. *Frontiers in Neuroscience*, 8, 1–8.
- Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, J. G., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports*, 5(1), 11628.
- Welch, G. F., Saunders, J., Edwards, S., Palmer, Z., Himonides, E., Knight, J., Mahon, M., Griffin, S., & Vickers, D. A. (2015). Using singing to nurture children’s hearing? A pilot study. *Cochlear Implants International*, 16(S3), 63–70.
- Wightman, F. L., & Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*, 118(5), 3164–3176.

4 | The perception of voice gender cues and speech in competing speech in children with cochlear implants

Abstract

Although cochlear implants (CIs) have improved greatly over the last few decades, the speech signal transmitted via CIs still lacks spectrotemporal fine structure, limiting the robust perception of speech cues and indexical voice-related cues. Voice characteristics contribute to a better perception of speech in competing speech by helping listeners to distinguish target from masker speech. Postlingually deaf CI adults have difficulties with voice discrimination and benefit less from voice differences for the perception of speech in competing speech. In prelingually deaf children with CIs, not much is known yet about their perception of voice characteristics and how it develops. Unlike postlingually deaf CI adults, CI children became deaf during the period of language development, and hence their auditory system developed primarily based on spectrotemporally degraded speech input. The voice perception abilities of CI children may develop as a function of age, like those of normal-hearing (NH) children during the school-age years, or their abilities may plateau at a certain age due to perceptual limitations. High brain plasticity during childhood could also make CI children better at using the available acoustic cues than CI adults or the lack of exposure to a normal acoustic speech signal could make it more difficult for them to learn which acoustic cues they should attend to. In this study, we examined the development of voice perception as a function of age by the systematic investigation of NH and CI children's perception of speakers' mean fundamental frequency (F0) and vocal-tract length (VTL), the primary acoustic cues related to speakers' so-called perceived voice gender. We examined children's ability to discriminate differences in F0 and VTL (Experiment 1), the perceptual weight they attributed to F0 and VTL cues for voice gender categorization (Experiment 2), and their benefit from target-masker voice differences in F0 and VTL for perceiving speech in competing speech (Experiment 3). Additionally, we examined how these different voice perception abilities, which range from lower-order perception to higher-order cognition, developed with age and how these were related to each other. Our results show that the F0 and VTL discrimination thresholds of CI children were generally higher than those of their NH age peers, but these improved with age and were better on average than those of the CI adults that were tested by Gaudrain and Başkent (2018). The perceptual weight attributed to F0 and VTL cues for voice gender categorization did not differ across groups, although CI children with high F0 discrimination thresholds did show a lower weighting of F0 cues. Finally, perceiving speech in competing speech was overall more difficult for CI children compared to their NH age peers, but they benefitted to a similar degree from target-masker voice

differences in F0 and VTL. Thus, although CI children's F0 and VTL discrimination abilities were generally poorer than those of their NH age peers, their weighting of F0 and VTL cues for voice gender categorization and their benefit from target-masker differences in F0 and VTL were similar to NH children. Our results also imply that high brain plasticity in CI children results in using the acoustic cues that are available in the spectrotemporally degraded CI signal better than CI adults. Together, these results show that F0 and VTL cues are represented in the CI signal to a certain degree, and prelingually deaf CI children seem to acquire the ability to effectively use the reduced F0 and VTL cues for voice gender categorization and to better perceive speech in competing speech, unlike most postlingually deaf CI adults.

4.1 Introduction

Even though CIs have improved considerably over the last few decades, speech perception often remains challenging for CI users due to the limited spectrotemporal fine structure cues in signals transmitted by the CI (see a review by Başkent et al., 2016). Besides the perception of speech cues, the spectrotemporally degraded speech signal also limits CI users' perception of indexical voice-related cues, such as prosody (Chatterjee & Peng, 2008; Everhardt et al., 2020; Meister et al., 2009; Peng et al., 2012) or vocal emotions (Chatterjee et al., 2015; Luo et al., 2007; Most & Aviner, 2009). CI users' ability to perceive differences in speakers' voice characteristics, such as their voice gender, was also observed to be poorer compared to NH listeners (El Boghdady et al., 2019; Fuller et al., 2014; Gaudrain & Başkent, 2018; Meister et al., 2009, 2016, 2020; Skuk et al., 2020). The perception of voice characteristics can help listeners in segregating different speech streams and perceiving speech in competing background speech (Bregman, 1994; Cherry, 1953).

Previous studies have mainly been focused on the perception of voice characteristics in postlingually deaf CI adults, who became deaf later in life after the completion of language development. Not much is known yet about the perception of voice characteristics in prelingually deaf CI children, who are usually congenitally deaf or became deaf early in life during the period of language development, considered as before 6 years of age in most studies. For postlingually deaf CI adults, the perception of voice characteristics is determined by perceptual limitations imposed by the CI and their adaptation to the spectrotemporally degraded CI signal. These CI users had developed an auditory perception system based on normal

acoustic hearing, which requires them to adapt their already established acoustic representations to use the reduced acoustic cues that are available via the spectrotemporally degraded CI signal. For prelingually deaf CI children, cognitive-developmental factors, which also seem to affect the voice perception abilities of NH school-age children, may additionally affect their perception of voice characteristics. However, it is unclear to what extent the perception of voice characteristics in CI children can improve as a function of age, and if their perceptual abilities plateau due to perceptual limitations. Unlike postlingually deaf CI users, the auditory perception system of prelingually deaf CI children is primarily based on the spectrotemporally degraded CI signal. Combined with the high brain plasticity during childhood, this may make CI children better at using the acoustic cues that are available in the degraded CI signal than CI adults. Alternatively, the lack of exposure to normal acoustic hearing may make their acoustic representations less robust than those of CI adults. Hence, in this study, we have investigated to what degree prelingually deaf CI children are able to perceive voice cues and benefit from voice differences for perceiving speech in competing speech, in light of factors that could play a role, such as perceptual limitations and cognitive development. Furthermore, we examined three voice perception abilities that involve different cognitive skills, ranging from voice discrimination, a lower-order perception task, to voice categorization and the perception of speech in competing speech, two higher-order cognitive tasks.

4.1.1 Effects of CI signal limitations on the perception of voice characteristics

As the CI speech input signal is spectrotemporally degraded, CI users' perception of voice characteristics, such as speakers' mean fundamental frequency (F0) or its perceptual correlate 'voice pitch', is reduced (see a review by Moore & Carlyon, 2005). Speakers' mean F0 is one of the primary acoustic cues that define a speaker's voice and also their so-called perceived voice gender (Skuk & Schweinberger, 2014; Smith & Patterson, 2005; Titze, 1989). A typical difference between male and female speakers' mean F0 is approximately 70%, which is around 12 semitones (st) or one octave. The mean F0 of a voice is determined by the rate at which the vocal folds open and close, also known as the glottal-pulse rate. Besides contributing to speakers' voice gender, F0 contours or modulations also convey prosodic information and indicate a word's meaning in tonal languages. Voice cues that rely heavily on F0 differences are perceived to a lesser extent by CI users compared to NH listeners, for instance, lexical tones (Peng et al., 2004; Wei et al., 2004), prosody

(Chatterjee & Peng, 2008; Everhardt et al., 2020; Meister et al., 2009; Peng et al., 2012), or pitch contours (Deroche et al., 2016, 2019). Most of the studies on pitch discrimination in CI users have used non-voice stimuli, such as pure-tones (Gfeller et al., 2007; Zeng et al., 2014), harmonic complexes (Deroche et al., 2014; Laneau et al., 2004), or pulse trains (Chatterjee & Peng, 2008; Zeng, 2002) or artificial speech cues like synthetic vowels (Geurts & Wouters, 2001; Green et al., 2004, 2005), and made inferences about natural voice and speech perception. Although the acoustic information transferred by these stimuli may be easier to control compared to natural voice stimuli, which vary across many acoustic dimensions, e.g., breathiness or coarseness (Kreiman & Sidtis, 2011), natural voice stimuli may better reflect voice perception in daily life. Using voice stimuli, Gaudrain and Başkent (2018) found that postlingually deaf CI adults showed an average just-noticeable difference (JND) in mean F0 of 9.19 st, which was 4.7 times larger than that observed in NH listeners.

Besides mean F0 discrimination, Gaudrain and Başkent (2018) investigated postlingually deaf CI adults' discrimination of vocal-tract length (VTL) cues. Speakers' VTL is determined by the distance between their vocal folds and the aperture of their oral and nasal cavities which closely relates to speakers' height (Fitch & Giedd, 1999). The VTL of speakers affects their formant frequencies. Together with the mean F0, speakers' VTL predominantly defines their perceived voice gender (Skuk & Schweinberger, 2014; Smith & Patterson, 2005; Titze, 1989). The common difference between male and female speakers' VTL is around 20%, which is equal to 3.6 st. Gaudrain and Başkent (2018) found that CI adults had an average VTL JND of 7.19 st, which was 4.2 times larger than the average VTL JND of NH listeners. Although the discrimination of F0 and VTL cues seems to be nearly equally impaired in CI adults, the impact of the elevated JNDs on the perception of speakers' voice gender is considerably larger for VTL than F0. While the typical voice gender difference of 12 st in speakers' mean F0 can be perceived by most CI adults based on their discrimination thresholds, the common difference between male and female speakers' VTL of 3.6 st is inaudible for almost all CI adults.

Unlike voice discrimination, a lower-order perception task that involves the direct comparison of voices based on their acoustic features, voice categorization is a higher-order cognitive task that requires the comparison of stimuli to stored acoustic representations of voice categories (Gibson, 1969; Pisoni, 1973). Studies on voice gender categorization in CI users have shown that most CI users can categorize the voice gender of different speakers fairly reliably when there are

sizeable differences in F0, but they have difficulties when speakers' mean F0 has a somewhat ambiguous in-between gender value (Fu et al., 2005; Kovačić & Balaban, 2009, 2010; Massida et al., 2013). The specific contribution of mean F0 and VTL cues to voice gender categorization was further investigated in studies in which the F0 and VTL parameters of the same speaker were artificially manipulated using a resynthesis procedure to keep all other voice-related cues, such as speaking rate, constant (Fuller et al., 2014; Meister et al., 2016; Skuk et al., 2020). The results of these studies confirmed that postlingually deaf CI adults seem able to reliably categorize speakers' voice gender when there are substantially large differences in mean F0, i.e., at the outer ends of the stimulus continuum. When speakers' mean F0 is somewhat ambiguous, NH listeners primarily rely on speakers' VTL for voice gender categorization. However, CI adults show an overall reduced weighting of VTL cues across the whole stimulus continuum, which results in a deviant voice gender categorization pattern relative to that observed in NH listeners (Fuller et al., 2014; Meister et al., 2016; Skuk et al., 2020). Nevertheless, CI adults' weighting of VTL cues seems to increase when sentences instead of words are used as stimuli (Meister et al., 2016). Fuller et al. (2014) reported that a difference of 3.6 st in VTL, the typical difference between male and female speakers' VTL size, would approximately produce a stimulation shift of one electrode. Hence, differences that are smaller than 3.6 st in VTL may not be accessible to most CI users. Combined with the results of Gaudrain and Başkent (2018), these findings suggest that VTL cues are not adequately transmitted, and as a result, CI users cannot reliably determine speakers' voice gender based on VTL cues.

In contrast to postlingually deaf CI adults, few studies have examined prelingually deaf CI children's categorization of speakers' voice gender. As aforementioned, prelingually deaf CI children and postlingually deaf CI adults developed an auditory perception system in different manners. These differences also seem to affect their processing of spectrotemporal information (DiNino & Arenberg, 2018; Gifford et al., 2018; Jung et al., 2012; Landsberger et al., 2018) and may also lead to differences in their processing of voice and speech cues, such as those related to speakers' voice gender. Kovačić and Balaban (2010) reported that prelingually deaf CI children's ability to categorize the voice gender of different speakers based on differences in their mean F0 was highly variable and also inversely related to the duration of the period of auditory deprivation preceding CI implantation. A longer period of auditory deprivation may impede the establishment of robust voice gender representations in long-term memory by reducing the amount of auditory input and exposure that listeners receive. Thus,

early implantation seems to have a positive impact on CI children's ability to categorize speakers' voice gender. Furthermore, postlingually deaf CI adults have to adapt their already established acoustic representations of speakers' voice gender based on normal acoustic hearing and correctly interpret VTL differences in the spectrotemporally degraded CI signal. Prelingually deaf CI children may be better at learning how to use the cues that are available in the spectrotemporally degraded CI signal because of their higher brain plasticity than postlingually deaf CI adults (Manrique et al., 1999). This explanation is also in line with the results of Zaltz et al. (2018), who found that CI adults who were prelingually deafened were better at discriminating differences in VTL cues when they were implanted earlier. Hence, VTL cues or other speaker size information may be present in the spectrotemporally degraded CI signal, although reduced and likely distorted, but postlingually deaf CI adults may not correctly interpret these cues as speaker size differences. This alternative explanation of why postlingually deaf CI adults show a reduced weighting of VTL cues for voice gender categorization cannot be discarded entirely yet.

Differences in speakers' voice gender also contribute to listeners' ability to perceive speech in competing speech (Bregman, 1994; Cherry, 1953), which is a more realistic listening condition that is encountered in daily life than speech perception in quiet. Perceiving speech in competing speech heavily relies on higher-order cognitive abilities, such as selective auditory attention and the inhibition of masker interference (Kidd et al., 2008; Schneider et al., 2007). CI users seem to have great difficulties with perceiving speech in competing speech, which has been hypothesized to be partly caused by their poorer voice discrimination abilities. Voice differences are one of the most evident cues that can help NH adult listeners to distinguish target from masker speech, such as differences in the gender of the target and the masker (Brungart, 2001; Brungart et al., 2001). For CI users, some studies have reported a benefit from differences in the gender of the target and the masker (Cullington and Zeng, 2008; Misurelli and Litovsky, 2015 (for CI children); Visram et al., 2012), while other studies did not find such a benefit (Bernstein et al., 2016; Stickney et al., 2004). Furthermore, NH listeners also derive benefit from target-masker voice differences in F0 or VTL when the voice parameters of the same speaker are artificially manipulated using a resynthesis procedure and all other voice cues remain consistent (Başkent & Gaudrain, 2016; Darwin et al., 2003; El Boghdady et al., 2019). Using such as resynthesis procedure makes it possible to specifically examine the contribution of F0 and VTL cues to listeners' benefit from voice differences for perceiving speech in competing speech. For CI users, the

results of such a benefit from target-masker voice differences in speakers' F0 or VTL are inconclusive. While some studies found that target-masker differences in F0 improved CI users' perception of speech in competing speech (Meister et al., 2020; Pyschny et al., 2011), other studies did not find such an effect (Auinger et al., 2017; El Boghdady et al., 2019; Stickney et al., 2007). Target-masker differences in VTL also do not seem to improve CI users' perception of speech in competing speech (El Boghdady et al., 2019). Meister et al. (2020) have suggested that the discrepancies in the results of different studies may partly be caused by the different masker stimuli that were used. Most studies that reported a benefit from voice differences in CI users have used different sentences as maskers (Cullington & Zeng, 2008; Misurelli & Litovsky, 2015; Pyschny et al., 2011; Visram et al., 2012), while the studies that did not find a benefit used the same single sentence as a masker (Auinger et al., 2017; Stickney et al., 2004, 2007), or concatenated speech chunks (El Boghdady et al., 2019) which may provide more energetic instead of informational masking. However, other stimuli characteristics (e.g., open-set vs. closed-set speech perception), the testing parameters (e.g., male and female speakers, or target-masker voice differences in F0 or VTL), and the auditory task that is used (e.g., repetition of the complete target sentence or repetition of specific target words) may also have a considerable impact on the outcomes (Litovsky et al., 2017). Finally, especially for CI children, who usually have smaller vocabulary sizes than their NH age peers (Fagan & Pisoni, 2010; Lund, 2016), the linguistic complexity of the stimuli that are used is important to consider when assessing listeners' ability to perceive speech in competing speech.

4.1.2 Developmental effects on the perception of voice characteristics

Most of the studies that have been conducted with CI children have been focused on assessing their general speech and language development using standardized tests, such as the Peabody Picture Vocabulary Test (PPVT) (Davidson et al., 2011; Fagan & Pisoni, 2010; Geers et al., 2003; Svirsky et al., 2000). Few studies have examined the perception of voice-related cues in prelingually deaf CI children, such as their mean F0 or voice pitch discrimination, and, to our knowledge, none have looked at their discrimination of VTL cues. Assessing the voice and speech perception abilities of prelingually deaf CI children is complicated by the effects of both perceptual limitations related to the spectrotemporally degraded CI signal and cognitive-developmental factors, such as selective auditory attention, inhibition, and auditory working memory (Kronenberger et al., 2013). These cognitive-developmental factors even seem to affect lower-order voice perception in NH

children, as evidenced by the development of their ability to discriminate subtle differences in speakers' voices, such as differences in their mean F0 (Buss et al., 2017a; Nagels et al., 2020a) or VTL (Nagels et al., 2020a), during the school-age years. In our previous study with NH school-age (4-12 years) children, we found that children's ability to discriminate differences in speakers' VTL was adult-like around 8 years of age, but their ability to discriminate differences in speakers' mean F0 was not adult-like yet at 12 years of age. For voice gender categorization, children's perceptual weight attributed to mean F0 cues, i.e., the impact of changes in mean F0 cues on voice gender categorization judgments, was adult-like around 6 years of age and their perceptual weight attributed to VTL cues was adult-like around 10 years of age. Thus, the F0 and VTL discrimination thresholds and the perceptual weight attributed to F0 and VTL cues for voice gender categorization develop in NH children during the school-age years, and the observed developmental patterns differ for the two abilities across voice cues. In a follow-up study, we also tested the ability of the same NH children to benefit from target-masker voice differences in F0 and VTL for perceiving speech in competing speech (Nagels et al., 2021). We found that NH children benefitted from target-masker voice differences in F0 and VTL at all tested ages, although their overall accuracy scores continued to differ from adults. Similar to voice gender categorization, the benefit from voice gender cue differences did not relate closely to NH children's ability to discriminate differences in voice gender cues, illustrating the different developmental trajectories in these three voice perception abilities (voice discrimination, voice categorization, and the benefit from target-masker voice differences for perceiving speech in competing speech).

While most CI children seem to be able to distinguish the voices of male speakers from those of female speakers significantly above chance (Osberger et al., 1991; Staller et al., 1991), NH children typically show ceiling-level performance on similar tasks (Bennett & Montero-Diaz, 1982). Nevertheless, newborn hearing screening has drastically improved the early identification of hearing loss and early CI intervention. These clinical practices are associated with better speech and language outcomes for CI children (Kirk et al., 2002; May-Mederake, 2012; Nicholas & Geers, 2007; Svirsky et al., 2007), and likely improved the ability to distinguish male from female speakers' voices in CI children compared to when these earlier studies were conducted. However, also more recent studies on CI children's voice perception abilities indicate that speaker discrimination remains difficult for CI children (Cleary et al., 2005; Cleary & Pisoni, 2002). Cleary et al. (2005) examined speaker discrimination in NH and CI children by simultaneously changing a female

speaker's mean F0 and formant frequencies (equal to the concept of "VTL cues" in the current study). They resynthesized sentences via STRAIGHT software (Kawahara & Irino, 2005) to create a stimulus continuum that results in a change of the perceived gender of the speaker. Children were asked to indicate whether two sentences were produced by the same speaker or by two different speakers. Although a subgroup of 8 out of 16 CI children who were tested showed a response pattern that was relatively similar to that of NH children, the other CI children did not show reliable speaker-category boundaries based on differences in F0 and VTL across the whole stimulus continuum. These results imply that F0 and VTL cues are transmitted via CIs to some degree, although some CI children do not seem to be able to use these cues effectively for speaker discrimination. Cleary et al. (2005) suggest that the characteristics of children's compromised auditory pathways, such as the degree of auditory nerve survival, or differences in the placement of the electrode array may contribute to the observed differences.

Several other studies have examined prelingually deaf CI children's perception of acoustic and voice-related cues, such as (affective) prosody (Hopyan-Misakyan et al., 2009; Nakata et al., 2012), musical pitch and timbre (Jung et al., 2012; Sjöberg et al., 2017; Stordahl, 2002; Torppa et al., 2014), mean pitch (Deroche et al., 2014; Kopelovich et al., 2010), pitch contours (Deroche et al., 2016, 2019), or lexical tones (Barry et al., 2002; Ciocca et al., 2002; Peng et al., 2004). The general findings of these studies are that most CI children show poorer perception compared to their NH age peers, but CI children's perception often improves as a function of age. However, age seems to only partially explain the variability in the perceptual abilities of CI children. The aforementioned studies on the discrimination of mean pitch, pitch contours, and lexical tones also did not find any effects of age at CI implantation and duration of CI use on CI children's pitch processing abilities, even though these factors seem to affect their general speech perception outcomes (Kirk et al., 2002; Nicholas & Geers, 2007; Svirsky et al., 2007). Interestingly, previous studies testing spectral ripple discrimination, a non-linguistic task that examines spectral resolution, did not find any effects of chronological age on the perceptual abilities of CI children (DiNino & Arenberg, 2018; Jung et al., 2012; Landsberger et al., 2018), although DiNino and Arenberg (2018) did find an effect of hearing age, i.e., the number of years after children received their CI. There may still be some effects of training and auditory experience on CI children's perceptual abilities. For instance, Deroche et al. (2019) found that Taiwanese CI children showed better discrimination of pitch contours than American CI children, demonstrating that adequate training and experience, such as proficiency in a tonal

language in this case, can still improve CI children's perception of pitch cues. However, no tonal language advantage was found for mean pitch discrimination as opposed to pitch contours (Deroche et al., 2014), likely because pitch contours bear a stronger resemblance to lexical tones than mean pitch cues. Taken together, these results imply that the perception of voice characteristics in CI children is affected by perceptual limitations imposed by the spectrotemporally degraded CI signal as well as cognitive-developmental factors related to age and auditory experience.

4.1.3 Present study

The difficulties that prelingually deaf CI children have with perceiving speakers' voice characteristics are twofold. Their perception of voice characteristics may be poorer because of the perceptual limitations that are imposed by the spectrotemporally degraded CI speech signal, as indicated by postlingually deaf CI adults' limited perception of voice-related cues, such as prosodic cues (Chatterjee & Peng, 2008; Everhardt et al., 2020; Meister et al., 2009) or voice gender cues (El Boghdady et al., 2019; Fuller et al., 2014; Gaudrain & Başkent, 2018; Meister et al., 2016, 2020; Skuk et al., 2020). In addition, cognitive-developmental factors seem to play a role in the perception of voice characteristics, as evidenced by the development of NH children's discrimination of differences in voice gender cues during the school-age years (Flaherty et al., 2019; Nagels et al., 2020a) and the observed effects of chronological age on CI children's discrimination of pitch cues (Deroche et al., 2014; Kopelovich et al., 2010).

In the present study, we investigated CI children's perception of the primary acoustic cues that define speakers' voice gender, namely speakers' mean F0 and their VTL. CI children performed three voice perception experiments, which varied from lower-order perception tasks to higher-order cognitive tasks. Specifically, we investigated the development in CI children's F0 and VTL discrimination thresholds (Experiment 1), their perceptual weight attributed to F0 and VTL cues for voice gender categorization (Experiment 2), and their benefit from target-masker differences in F0 and VTL cues for perceiving speech in competing speech (Experiment 3) as a function of age. The voice perception abilities of CI children were compared to data collected from NH school-age children using the same experiments, which were presented in previous studies that were part of the same project on the perception of indexical cues in kids and adults (PICKA) (Nagels et al., 2020a, 2020b, 2021). In this study, we examined how the tested abilities developed as a function of age in CI children, and if the three tested abilities develop similarly as in NH children. Furthermore, we investigated if these different

perceptual abilities were related to each other. The high amount of individual variability in CI children's discrimination abilities (Cleary et al., 2005; Deroche et al., 2014; Kopelovich et al., 2010) may make the development in CI children's performance on the three experiments as a function of age less consistent than in NH school-age children (Nagels et al., 2020a, 2021). The developmental patterns for CI children's ability to discriminate, weigh, and benefit from F0 and VTL cues may also differ from those observed in NH school-age children, as they likely have less access to F0 and VTL cues because of the spectrotemporally degraded CI signal, like postlingually deaf CI adults (El Boghdady et al., 2019; Fuller et al., 2014; Gaudrain & Başkent, 2018; Meister et al., 2016, 2020; Skuk et al., 2020). However, CI children's perception of F0 and VTL cues may differ from that of CI adults due to the different manners in which they developed their auditory perception system and formed representations of voice gender categories. Prelingually deaf CI children show higher brain plasticity than postlingually deaf CI adults (Manrique et al., 1999) and were exposed to the spectrotemporally degraded CI signal from an early age. These aspects may make CI children better at capitalizing on the reduced F0 and VTL cues that are available via the CI signal than postlingually deaf CI adults. Alternatively, the lack of exposure to a normal acoustic speech signal could make their acoustic representations of speakers' voice gender less robust than those of CI adults.

4.2 Discrimination of differences in F0 and VTL cues

4.2.1 Methods

Participants

Fourteen prelingually deaf CI children between 4 and 16 years of age took part in the current study. We extended the age range for testing CI children to 17 years of age, as our preceding studies with NH school-age children demonstrated that even 12-year-old NH children did not always show adult-like performance yet on some tasks (Nagels et al., 2020a, 2021). All except one CI child were congenitally deaf. This CI child had deafness in one ear and profound hearing loss in the other ear until the age of three and became bilaterally deaf one year later. All except one CI child, a different CI child than previously mentioned, were bilaterally implanted. Besides prelingual deafness and chronological age, the only other inclusion criterion for participation was that CI children should be generally healthy and able to comprehend and use spoken language. An overview of more detailed information about CI children's demographics and hearing history can be found in Table 4.1.

The age at implantation of CI children for their first implant ranged from 9 months to 4 years and 3 months of age. The median age at which they received their implant was 12 months. This age is slightly higher than the age at which the largest proportion of CI children are implanted in the Netherlands, which is between 6 and 11 months of age (Bruijnzeel et al., 2017). We found significant correlations between CI children's chronological age and their age at implantation (*Pearson's* $r = -0.51$, $P < 0.01$), and with their hearing age (*Pearson's* $r = -0.95$, $P < 0.001$). These correlations indicate that younger CI children were implanted earlier in life than older CI children, likely because of improved newborn hearing screenings and guidelines for cochlear implantation.

Table 4.1: Demographic characteristics of CI children ($N_{CI\ children} = 14$). Age is given in years, rounded to two decimal places. The hearing age indicates the difference between CI children's chronological age and their age at first implantation, i.e., the number of years after children received their CI. The abbreviations AS (Auris Sinistra; left ear), AD (Auris Dextra; right ear), and AU (Auris Utraque; both ears) describe the characteristics of the implants for each ear. For the etiologies, PS is an abbreviation for Perrault Syndrome, CMV is an abbreviation of Cytomegalovirus, and Genetic Cx26 indicates that the cause of deafness was a mutation of the protein connexin 26 (Cx26) that is found on the Gap junction beta-2 (GJB2) gene. Vocabulary corresponds to scores on the Renfrew Word Finding Vocabulary Test (maximum of 50 points).

Subject	Age (yrs)	Age at first implantation (yrs)	Hearing age (yrs)	Implant(s)	Processor	Strategies	Etiologies	Vocabulary score
CIK001	11.08	0.75	10.25	CI512 (AS)/CI24RE (AD)	CP1000	MP3000 (AS) / ACE (AD)	Genetic	47
CIK002	11.34	4.17	7.17	CI24RE (AD)	CP910/920	MP3000 (AD)	Auditory	31
CIK003	11.92	0.92	11.00	CI512 (AS)/CI24RE (AD)	CP910/920	ACE (AU)	Unknown	49
CIK004	16.92	2.00	14.92	Unknown	CP1000	Unknown	Genetic PS	46
CIK005	6.42	0.92	5.42	CI24RE (AU)	CP1000	MP3000 (AU)	Genetic	28
CIK006	9.17	2.34	6.67	CI24RE (AU)	CP910/920	MP3000 (AU)	Unknown	37
CIK007	6.08	0.92	5.17	CI24RE (AU)	CP1000	MP3000 (AU)	Genetic	39
CIK008	5.92	0.83	5.08	CI24RE (AU)	CP1000	MP3000 (AU)	Genetic Cx26	35
CIK009	4.00	1.00	3.00	CI24RE (AU)	CP910/920	MP3000 (AU)	Genetic Cx26	26
CIK010	14.08	2.00	12.08	Unknown	CP1000	Unknown	Genetic Cx26	46
CIK011	8.58	1.00	7.67	Unknown	Unknown	Unknown	Genetic	37
CIK012	10.67	1.00	9.67	Unknown	Unknown	Unknown	Genetic	43
CIK013	14.92	4.25	10.67	Unknown	CP1000	Unknown	Likely CMV	45
CIK014	12.17	1.25	10.92	Unknown	Unknown	Unknown	CMV	37

The control group consisted of fifty-eight NH children between 4 and 12 years of age, and fifteen NH adults between 20 and 29 years of age. The data of NH participants were presented and published in earlier studies that were part of the PICKA project (Nagels et al., 2020a, 2021). We categorized NH and CI children into six different age groups that each spanned two or three years (4-6 years, 6-8 years, 8-10 years, 10-12 years, 12-14 years, and 14-17 years), in line with our previous studies with NH children. We primarily used age as a continuous variable for data analysis, but age groups were used in the figures for visualization of the data and the calculation of the z-scores of CI children. Before participating in the study, we screened NH participants' audiometric thresholds at 20 dB HL using pure-tone audiometry at octave frequencies between 500 and 4000 Hz. All NH and CI participants were monolingual native speakers of Dutch. The vocabulary size of NH and CI children was determined using the Dutch version of the Renfrew Word Finding Vocabulary Test (Renfrew, 1995). This vocabulary test is a short measure of productive vocabulary using 50 line-drawn pictures (maximum of 50 points).

The parents and legal guardians of NH and CI children, NH and CI children who were older than 12 years of age, and adult participants signed a written informed consent form before they took part in the study, in agreement with the regulations of the Dutch Medical Research Involving Human Subjects Act (WMO). Ethical approval of the study was received from the Medical Ethical Review Committee of the University Medical Center Groningen (METc 2016.689).

Stimuli and Apparatus

For this experiment, the same stimuli and resynthesis procedure were used as in the study by Gaudrain and Başkent (2018) with postlingually deaf CI adults. To create the stimuli, we used 61 CV syllables that were spliced from Dutch CVC words, taken from the NVA speech corpus (Bosman & Smoorenburg, 1995), which are commonly used in Dutch ENT clinics to measure CI users' speech perception outcomes. In each trial, three CV syllables were randomly selected and concatenated to create trisyllabic CVCVCV pseudowords, e.g., *ba-ki-mo*. The recordings were produced by a female native speaker of Dutch with no discernable regional accent. The mean F0 of the speaker was 201 Hz, and the estimated VTL of the speaker was 13.5 cm based on the average height of Dutch women of 168.72 cm (Roser et al., 2020). Even though differences in speakers' mean F0 are commonly expressed in Hertz and differences in VTL in centimeters, changes in both voice cues result in frequency differences. Hence, we expressed the applied differences in F0 and VTL as ratios measured on a logarithmic scale in st. The comparison between the

perception of these two voice cues has also been made in previous studies from our research group (El Boghdady et al., 2019; Fuller et al., 2014; Gaudrain & Başkent, 2018; Nagels et al., 2020a, 2021).

We used STRAIGHT software (Kawahara & Irino, 2005) in Matlab to implement the voice differences in F0 and VTL. The CV syllables were first analyzed to extract the F0 contour and spectral envelope, equalized for root-mean-square (RMS) level, and normalized to a duration of 200 ms and a mean F0 of 242 Hz. For each trial, the F0 contour and spectral envelope of three randomly selected CV syllables were resynthesized with the new F0 and VTL parameters using STRAIGHT. We performed this resynthesis procedure even for the reference pseudowords with F0 and VTL voice parameters equal to those of the original speaker to ensure no perceptual differences would result from the resynthesis procedure itself. We then concatenated the three CV syllables with 50 ms of silence in between the individual syllables. The overall F0 contour of each CVCVCV pseudoword was modified by changing the F0 contour of each CV syllable with random steps of $\frac{1}{3}$ st, e.g., $-\frac{1}{3}$ st, 0 st, or $+\frac{1}{3}$ st, to make the pseudowords sound less artificial. The overall F0 contour of the pseudoword was then centered back to zero again to keep the same mean F0. Subsequently, the mean F0 of the pseudowords was changed by multiplying the F0 contour (in Hz) by the intended change in F0, which preserved the existing F0 fluctuations. The VTL of the pseudowords was adjusted by compressing the spectral envelope towards the lower frequencies to generate an overall spectral shift while retaining the same ratio between formant frequencies. After these manipulations, we recombined the modified F0 contour and spectral envelope using STRAIGHT's pitch synchronous overlap-add (PSOLA) resynthesis method.

The same pseudoword with the same CVCVCV syllable combination was used to create the target and the two reference pseudowords for each trial. The target pseudoword differed from the two reference pseudowords in either mean F0 or VTL. Besides this difference, there were only small differences of $\frac{1}{3}$ st in the mean F0 contour of each pseudoword, as mentioned earlier. The target pseudoword either had a lower F0 or a larger VTL value compared to the reference pseudowords, depending on the JND that was being measured, which made the target pseudoword sound more male-like.

Procedure

Participants' just-noticeable differences (JNDs) in F0 and VTL cues were measured via a 3-interval 3-alternative forced-choice (3I-3AFC) adaptive procedure. Participants first performed two practice sessions, which each consisted of three trials, to familiarize themselves with the task. The practice sessions were followed by two experiment sessions which consisted of approximately 25 to 45 trials, depending on participants' responses, and measured participants' F0 and VTL JNDs. The total duration of the experiment was around 15 minutes. The order in which the F0 or VTL JNDs were measured for each participant was randomized. Each experiment session started with a voice difference of -12 st in F0 or +12 st in VTL, which made the target pseudoword voice sound more male-like than the two reference pseudowords. Participants' JND was determined using a 2-down 1-up adaptive staircase procedure. The voice difference would decrease with one step size after two successively correct responses and would increase with one step size after an incorrect response. The initial step size value was 2 st, but after 15 consecutive trials with the same step size or when the difference became smaller than twice the step size, the step size value was divided by $\sqrt{2}$. After eight reversals, i.e., an incorrect response followed by a correct response, the experiment session ended. The geometric mean of the voice difference values of the last six reversals was calculated, which corresponds to the 70.7% discrimination point of the psychometric function (Levitt, 1971).

NH and CI children were tested in a quiet room at their homes, and NH adults were tested in a quiet testing room at the University of Groningen. The experiment was performed on a Dell XPS 13 inch touchscreen laptop via a child-friendly game interface that was created in Matlab. In each trial, three sea animals appeared on the screen and sequentially produced the same CVCVCV-structured pseudoword. Two of the sea animals produced the reference pseudoword and one sea animal would produce the target pseudoword with a different F0 or VTL value. Participants were instructed to click on the sea animal whose voice differed from the other two sea animals. Visual feedback on the accuracy of their response was provided to participants. A red box appeared around the selected sea animal when the response was incorrect, or the sea animal would go into one of the seven bubbles when the response was correct. The stimuli of all three experiments were presented via Sennheiser HD 380 Pro headphones for NH children and adults, and via Logitech Z200 loudspeakers for CI children. We instructed CI children to use the settings they most frequently used and to keep the settings consistent during all experiments. The presentation level of the stimuli was calibrated to a sound level of 65 dBA.

Data analysis

We examined if CI children's ability to discriminate differences in F0 and VTL cues develops in the same manner as in NH children by analyzing their log-transformed F0 and VTL JNDs using the lme4 package (version 1.1.21, Bates et al., 2015) in R (version 3.6.1, R Core Team, 2020). We computed a linear mixed-effects regression model using the JND data of children, not those of adults due to the discontinuous distribution of age, with a three-way interaction between *age*, *voice cue* (F0 vs. VTL), and *group* (NH vs. CI), and a random intercept per participant, in lme4 syntax: `log(JND) ~ voice cue * age * group + (1 | participant)`. We compared different models by deleting one factor at a time from the full model and evaluating the fit of the models using ANOVA Chi-Square tests. Children's JNDs were log-transformed for this analysis, as zero-crossings are not possible using the JND procedure. The variable *voice cue* indicated whether the JND value represented children's F0 or VTL discrimination threshold. The variable *age* contained children's age in years, rounded to two decimal places, while the variable *group* indicated if children were NH or used a CI.

Furthermore, to estimate how much CI children's F0 and VTL JNDs deviated from those of their NH age peers, we calculated the 50% prediction intervals of NH children's log-transformed JNDs as a function of their chronological age. We used the `predict` function of the `car` package (Fox et al., 2012) in R on a linear model with fixed effects of *age* and *voice cue* to calculate the 50% prediction intervals of NH children's F0 and VTL JNDs. We also calculated the z-scores of CI children based on the mean F0 and VTL JND values and standard deviations of NH children per age group. The z-scores of CI children were calculated by subtracting the mean F0 and VTL JND values of NH children and NH adults of the same chronological age group (4-6 years, 6-8 years, 8-10 years, 10-12 years, and adults) and dividing the outcome by the corresponding standard deviation, i.e., $\text{z-score} = (\text{log-transformed F0 JND CI child} - \text{mean log-transformed F0 JND of NH age peers}) \div \text{standard deviation F0 JND of NH age peers}$. We used the mean F0 and VTL JND and standard deviation values of 10- to 12-year-old NH children to calculate the z-score of one 12-year-old CI child. The mean F0 and VTL JND and standard deviation values of NH adults were used to calculate the z-scores of three 14- to 17-year-old CI children. In addition, we calculated the z-scores of CI children based on their hearing age. Some CI children belonged to a different age group based on their hearing age than based on their chronological age, and a different NH age group was used to calculate their z-scores in this case.

4.2.2 Results

Figure 4.1 shows the F0 and VTL JNDs of CI children (shown as individual data points) and how these fit into the distribution of NH children's JNDs (shown as boxplots) based on their chronological age (upper panels) and hearing age (lower panels), i.e., the number of years after CI children received their CI. We used a linear mixed-effects model to analyze the data of NH children and CI children to compare the potential differences in the developmental patterns as a function of chronological age across the two groups. The model comparison analysis demonstrated that the model with only fixed effects of *age*, *voice cue*, and *group* and no interactions had the best fit, in lme4 syntax: `log(JND) ~ voice cue + age + group + (1 | participant)`. This model was better than the models without fixed effects of *age* ($\chi^2(1) = 24.9$, $P < 0.001$), *voice cue* ($\chi^2(1) = 23.9$, $P < 0.001$), or *group* ($\chi^2(1) = 15.3$, $P < 0.001$) or those with two-way or three-way interactions.

There was a significant effect of *voice cue* ($F(1,72) = 28.3$, $t = -5.32$, $P < 0.001$), which indicates that children's JNDs were lower for VTL than for F0 in both NH and CI children. We also found a significant effect of *age* ($F(1,72) = 29.8$, $t = -5.46$, $P < 0.001$), which shows that children's JNDs overall decreased as a function of chronological age, and there were no differences in the steepness of this decrease across groups or voice cues. Finally, the significant effect of *group* ($F(1,72) = 17.1$, $t = 4.13$, $P < 0.001$) demonstrates that CI children's JNDs were overall higher, i.e., their discrimination thresholds were poorer, than those observed in NH children, but this difference was equal across age and voice cues. To summarize, although the F0 and VTL JNDs of CI children were higher compared to those of their NH children, the development of their JNDs as a function of chronological age and across voice cues did not significantly differ between groups.

Figure 4.1: The log-transformed JNDs of CI children (dots) and NH children and adults (boxplots) for F0 (left panels) and VTL (right panels) as a function of their chronological age (a) and their hearing age (b) ($N_{\text{CI children}} = 14$; $N_{\text{NH children}} = 58$; $N_{\text{NH adults}} = 15$). The dots show individual data points of CI children located at their chronological and hearing age in years, rounded to two decimal places. The purple dot shows the mean F0 and VTL JNDs of the postlingually deaf CI adults that were tested by Gaudrain and Bařkent (2018). The boxplots indicate the median JND values of NH children and adults per age group, and the boxes show the lower and upper quartiles. The whiskers show the lowest and highest data points within plus or minus 1.5 times the interquartile range. The shaded areas surrounding the boxplots show the 50% prediction interval of NH children’s JNDs based on a linear mixed model of their JNDs as a function of chronological age and voice cue.

These analyses indicate that CI children's JNDs were generally higher than those of NH children. Visual inspection of the data suggests that the development as a function of chronological age and the differences across voice cues are less clear and consistent within the sample of CI children compared to our population of NH children (Figure 4.1). Hence, we examined the differences between NH and CI children's JNDs more closely. First of all, Figure 4.1 shows the 50% prediction intervals which were based on NH children's JNDs. These prediction intervals demonstrate that approximately seven CI children have F0 JNDs within the prediction intervals and five CI children have VTL JNDs within the prediction intervals. We also looked at CI children's hearing age instead of their chronological age, which corrects for the period of auditory deprivation preceding CI implantation. As we have argued that the development in NH children's JNDs seems to be mostly due to cognitive-developmental factors related to, among others, auditory and linguistic experience (Nagels et al., 2020a), using CI children's hearing age may give a fairer comparison between NH and CI children. Based on their hearing age, eight CI children have F0 JNDs within the prediction intervals and seven CI children have VTL JNDs within the prediction intervals. Nevertheless, replacing chronological age with hearing age in the main analysis resulted in an F statistic that was similar in magnitude ($F(1,72) = 30.8$, $t = -5.55$, $P < 0.001$), and hence CI children still had significantly higher JNDs than NH children based on their hearing age.

To further explore these observations, we calculated the z-scores of CI children based on the mean JND values and standard deviations of NH children and adults of the same chronological or hearing age group (Figure 4.2). Since we currently do not have any data of NH children between 12 and 16 years of age, the data of 10- to 12-year-old NH children were used to calculate the z-score of one 12-year-old CI child. The data of NH adults were used to calculate the z-scores of three 14- to 16-year-old CI children. Based on their hearing age, some CI children belonged to a different age group than based on their chronological age, and a different NH age group was used to calculate their z-scores in this case. The dashed lines at the values of -1.96 and +1.96 indicate (as opposed to prediction intervals) the confidence intervals in which 95% of NH children's JNDs fall. Four CI children have F0 JNDs that are above the 95% confidence intervals based on their chronological age, and three CI children have values above this range based on their hearing age. Furthermore, six CI children have JNDs that are above the 95% confidence intervals based on their chronological age, and four CI children have values above this range based on their hearing age. The number of CI children with JNDs outside

of the confidence intervals is only slightly larger for VTL than for F0 based on their chronological age and equal across voice cues based on their hearing age.

Figure 4.2: The z-scores of the log-transformed JNDs of CI children ($N_{\text{CI children}} = 14$) for F0 (left panels) and VTL (right panels) as a function of their chronological age (a) and their hearing age (b), calculated based on the mean JND values of NH children per age group. The dots show individual data points of CI children located at their chronological age in years, rounded to two decimal places. One 12-year-old CI child was categorized under the 10-12 age group, and the three CI children above 14 years of age were categorized as adults for this purpose. The dashed lines at -1.96 and +1.96 mark the confidence intervals in which 95% of NH children's JNDs fall.

4.2.3 Discussion

Our results show that CI children's F0 and VTL JNDs were generally higher than those of their NH age peers. However, the improvement in the JNDs of CI children as a function of chronological age and the developmental pattern across voice cues were not significantly different from NH children. Although the JNDs of CI children were generally higher than those of their NH age peers, a small number of CI children still showed JNDs that were approximately within the NH age-appropriate range, particularly based on their hearing age instead of their chronological age. The number of outlying z-scores for CI children's JNDs was slightly larger for VTL than for F0 based on their chronological age but equal across voice cues based on their hearing age.

The generally higher JNDs of CI children and the improvement of their JNDs as a function of chronological age correspond with the findings from previous studies on pitch discrimination in prelingually deaf CI children using non-voice stimuli (Deroche et al., 2014; Kopelovich et al., 2010), and speaker discrimination based on simultaneous changes in F0 and VTL cues (Cleary et al., 2005). Similar to these studies, chronological age could not explain the observed individual variability in the F0 and VTL discrimination thresholds of CI children as well as for NH children. Most CI children had poorer discrimination thresholds than their NH age peers, but some CI children had discrimination thresholds that were within the 50% prediction intervals of their NH age peers. As described in the participant section, the CI participant group was relatively homogeneous considering device-related factors. Therefore, it seems unlikely that the CI signal by itself is the only cause of the overall poorer discrimination thresholds of CI children. Instead, F0 and VTL cues seem to be delivered to a certain degree by the CI, but some CI children may not be able to use these cues as effectively as others. These results are also partly in line with those of Cleary et al. (2005), who mentioned that differences in CI children's compromised auditory pathways, such as the degree of auditory nerve survival or individual differences in the placement of the electrode array, could explain differences in their voice perception abilities.

The reduced ability of CI children relative to their NH age peers to discriminate F0 or pitch differences is also in agreement with earlier findings from studies with postlingually deaf CI adults (Chatterjee & Peng, 2008; Gaudrain & Başkent, 2018; Geurts & Wouters, 2001; Gfeller et al., 2007; Green et al., 2004, 2005; Laneau et al., 2004; Zeng, 2002; Zeng et al., 2014). However, the F0 and VTL JND values of prelingually deaf CI children are noticeably lower, i.e. their discrimination thresholds were better, than those obtained from postlingually deaf CI adults by

Gaudrain and Başkent (2018) using a similar testing procedure. While postlingually deaf CI adults had mean JNDs of 9.19 st for F0 and 7.19 st for VTL, the prelingually deaf CI children had mean JNDs of 5.92 st for F0 and 4.10 st for VTL. Zaltz et al. (2018) also reported that CI adults who were prelingually deafened showed better VTL discrimination thresholds when they were implanted earlier, which implies that early exposure to the spectrotemporally degraded CI signal and high brain plasticity are associated with better VTL perception. Combined with the improvement of CI children's JNDs as a function of age, our results suggest that CI children as a group may make better use of the reduced F0 and VTL cues that are available via the spectrotemporally degraded CI signal than postlingually deaf CI adults. Besides high brain plasticity (Manrique et al., 1999), these differences may also be related to the fact that CI children do not have to adapt already existing auditory representations that are based on normal acoustic hearing like CI adults. Peripheral factors related to the underlying etiology of deafness (mostly genetic in CI children) or related to the device such as differences in CI stimulation strategies (Wouters et al., 2015) or bilateral instead of unilateral CI implantation may also play a role. These factors are discussed more elaborately in the general discussion section.

4.3 Weighting of F0 and VTL differences for voice gender categorization

4.3.1 Methods

Participants

The same participants who performed the first experiment also took part in the second experiment, except for the youngest CI child (4 years of age) who did not fully complete the second and the third experiment.

Stimuli and Apparatus

We used the same stimuli and resynthesis procedure for the voice gender categorization experiment as in the earlier study by Fuller et al. (2014) with postlingually deaf CI adults. The stimuli consisted of four CVC words, *bus* [bus], *vaak* [often], *leeg* [empty], and *pen* [pen], which were taken from the NVA corpus recordings (Bosman & Smoorenburg, 1995). The stimuli were produced by the same female speaker as in the voice JND experiment, with a mean F0 of 201 Hz. We manipulated the F0 and VTL parameters and resynthesized all stimuli following

the same STRAIGHT resynthesis procedure as used for the voice JND experiment.

The experiment consisted of nine different voice conditions. We used three F0 parameters of 0 st, -6 st, and -12 st, corresponding to mean F0 values of 201 Hz, 142 Hz, and 100 Hz, and three VTL parameters of 0 st, 1.8 st, and 3.6 st, corresponding to estimated VTL sizes of 13.5 cm, 15.1 cm, and 16.6 cm. The F0 and VTL parameter values were based on the study by Fuller et al. (2014) and in agreement with earlier findings by Smith et al. (2007), Smith and Patterson (2005), and Peterson and Barney (1952). Fuller et al. (2014) confirmed that a combined change of -12 st in F0 and 3.6 st in VTL reliably changed the perceived voice gender for NH adult listeners using this speaker's voice.

Procedure

Participants performed the voice gender categorization experiment after completing the voice JND experiment. The experiment consisted of a visual-auditory match-to-sample task in which participants' perceptual weight attributed to F0 and VTL differences for voice gender categorization was examined. The experiment was preceded by a practice session of five trials in which five randomly selected items were presented to participants to familiarize them with the task. The experiment consisted of 36 trials, i.e., four items (one item per CVC word) per voice condition, with items that were presented in a randomized order. The duration of the experiment was approximately 6 minutes.

A child-friendly interface was created to administer the voice gender categorization experiment in Matlab. The participants would first hear an auditory stimulus that sounded like a female, male, or somewhat ambiguous in-between gender voice, depending on the F0 and VTL parameters of the voice condition. After the auditory stimulus was played, a male or female face would appear on the computer screen. We instructed the participants to click on the green checkmark when the gender of the voice and face matched, and on the red cross when the gender of the voice and face differed. The participants did not receive any feedback on the accuracy of their responses.

Data analysis

For the analysis of participants' weighting of F0 and VTL cues for voice gender categorization, we calculated their cue weights. We first normalized the differences in F0 and VTL to make them functionally equivalent for model fitting by defining them as $\delta F0 = -\Delta F0 \div 12 - 0.5$ and $\delta VTL = \Delta VTL \div 3.6 - 0.5$. After this normalization, the female-sounding voice had $\delta F0$ and δVTL values of -0.5, while the male-sounding voice (difference of -12 st in F0 and +3.6 st in VTL) had $\delta F0$ and δVTL values of +0.5. Secondly, we computed a mixed-effects logistic regression model with *response* ("woman" corresponding to 1, "man" corresponding to 0) and random intercepts and slopes for $\delta F0$ and δVTL per participant. We extracted the coefficients from this model, which are predictive values on a logit scale, i.e., log odds ratios based on the natural log, relative to the normalized $\delta F0$ and δVTL ranges. Finally, we converted the coefficients into Berkson units per semitone (Bk/st) by scaling the factors to correspond to log2 odds ratios per semitone (Hilkuysen et al., 2012). An increase of 1 Bk/st corresponds to doubling the percentage of woman categorizations relative to man categorizations in the current study. We then analyzed the Bk/st units of children by computing a linear mixed-effects model with *cue weight* (Bk/st) as an outcome variable, a three-way interaction between *age*, *voice cue*, and *group* (NH vs. CI), and random intercepts per participant, in lme4 syntax: `cue weight ~ voice cue * age * group + (1 | participant)`. We performed a model comparison analysis to evaluate which model had the best fit.

Furthermore, we examined how much the cue weights of CI children deviated from those of their NH age peers by calculating the 50% prediction intervals. The 50% prediction intervals of NH children's cue weights were calculated based on a linear model containing a two-way interaction between *age* and *voice cue*. We also computed the z-scores of CI children's cue weights based on the mean cue weights of NH children of the same chronological and hearing age via the same method as used in the previous experiment. Finally, we calculated the Pearson correlation coefficients between CI children's log-transformed JNDs and their cue weights for F0 and VTL to investigate the relationship between these two abilities.

4.3.2 Results

Figure 4.3 shows the F0 and VTL cue weights of CI children for voice gender categorization (Bk/st units shown as individual data points) and how they fit within the distribution of NH children's cue weights (shown as boxplots) based on their

chronological age and hearing age. The shaded area shows the 50% prediction intervals of NH children's cue weights based on a linear model with a two-way interaction between *age* and *voice cue*. In addition, Figure 4.4 shows the gender categorization judgments of individual CI children as a function of F0 and VTL differences and the mean categorization judgments of NH children and adults per age group. The model comparison analysis showed that the model with only a two-way interaction between *voice cue* and *age* was the best fitting model, in lme4 syntax: `cue weight ~ voice cue * age + (1 | participant)`. This model had a better fit than the model without the two-way interaction between *age* and *voice cue* ($\chi^2(1) = 13.54, P < 0.001$), while the addition of *group* as a fixed effect did not improve the model ($\chi^2(1) = 0.15, P = 0.69$).

There was a significant effect of *age* on NH and CI children's cue weights ($F(1,71) = 16.3, t = 2.10, P < 0.05$), which shows that the F0 and VTL cue weights of NH and CI children increased, and thus became more adult-like, as a function of chronological age. No significant effect of *voice cue* alone was found ($F(1,71) = 0.005, t = -0.07, P = 0.95$). The two-way interaction between *voice cue* and *age* was significant ($F(1,71) = 14.9, t = 3.86, P < 0.001$), which, combined with the lack of a significant main effect of *voice cue* indicates that children's cue weights mainly increased as a function of age for VTL but not so much for F0. To summarize, the developmental pattern in children's cue weights did not differ between NH and CI children. Furthermore, children's cue weights only appear to improve significantly with chronological age for VTL and not so much for F0. This observation is also in line with the findings from our earlier analysis of only NH children's cue weights (Nagels et al., 2020a), which showed that their cue weights were adult-like already around 6 years of age for F0, and thus changed little with age, but were only adult-like around 10 years of age for VTL.

Figure 4.3: The cue weights in Bk/st units, i.e., the weight attributed to F0 and VTL for voice gender categorization, of CI children (dots) and NH children and adults (boxplots) for F0 (left panels) and VTL (right panels) as a function of their chronological age (a) and their hearing age (b) ($N_{CI \text{ children}} = 13$, $N_{NH \text{ children}} = 58$; $N_{NH \text{ adults}} = 15$). The dots show individual cue weights of CI children located at their chronological and hearing age in years, rounded to two decimal places. The boxplots indicate the median cue weights of NH children and adults per age group, and the boxes show the lower and upper quartiles. The whiskers show the lowest and highest data points within plus or minus 1.5 times the interquartile range. The shaded areas show the 50% prediction intervals of NH children's cue weights based on a linear model of their cue weights as a function of chronological age and voice cue.

A

B

Figure 4.4: The voice gender categorization judgments of individual CI children (a) and the average categorization judgments of NH children and adults per participant age group, and those of the postlingually deaf CI adults that were tested by Fuller et al. (2014) (b) as a function of differences in F0 (x-axis) and VTL (y-axis) ($N_{CI\text{ children}} = 13$, $N_{NH\text{ children}} = 58$; $N_{NH\text{ adults}} = 15$). Blue corresponds to 100% “man” categorizations and yellow corresponds to 100% “woman” categorizations.

The 50% prediction intervals of NH children’s cue weights are shown in Figure 4.3. Based on these prediction intervals, we can see that only two CI children attributed less weight to F0 cues than predicted for 50% of the NH children and three CI children attributed less weight to VTL cues. Figure 4.4 also seems to suggest that CI children, like NH children, weigh both F0 and VTL cues for their voice gender categorization judgments.

The z-scores of CI children based on the mean cue weights and standard deviations of NH children per age group are shown in Figure 4.5. One 12-year-old CI child was again categorized under the 10- to 12-year-old age group and three 14- to 16-year-old CI children were categorized under the adult age group for

calculating their z-scores based on their chronological age. In agreement with our previous observations, only one 10- to 12-year-old CI child attributed more perceptual weight to VTL differences according to the z-scores calculated based on CI children's chronological age. In addition, two children attributed more perceptual weight to F0. When we look at CI children's hearing age, three CI children attributed more perceptual weight to F0, but none attributed more or less perceptual weight to VTL. In sum, unlike postlingually deaf CI adults, who mainly rely on F0 differences for voice gender categorization and not so much on VTL differences (Fuller et al., 2014; Meister et al., 2016; Skuk et al., 2020), none of the CI children attributed significantly less perceptual weight to F0 or VTL cues than their NH age peers.

Figure 4.5: The z-scores of the cue weights of CI children for F0 (left panels) and VTL (right panels) as a function of their chronological age (a) and their hearing age (b), calculated based on the mean cue weights of NH children and adults per age group ($N_{\text{CI children}} = 13$). The dots show individual data points of CI children located at their chronological and hearing age in years, rounded to two decimal places. One 12-year-old CI child was categorized under the 10-12 age group, and the three CI children above 14 years of age were categorized as adults for this purpose. The dashed lines at -1.96 and +1.96 mark the confidence intervals in which 95% of NH children's cue weights fall.

To investigate whether the perceptual weight that CI children attributed to differences in F0 and VTL cues for voice gender categorization could be limited by their F0 and VTL discrimination thresholds, we calculated the Pearson correlation coefficients for the two measures for F0 and VTL. For F0, there was a relatively strong significant relationship between CI children's JNDs and their corresponding cue weights (*Pearson's* $r = -0.58$, $P < 0.05$). This correlation was negative, so CI children's perceptual weight attributed to F0 differences was lower as their JNDs were higher. For VTL, there was no significant relationship between CI children's JNDs and their cue weights (*Pearson's* $r = -0.36$, $P = 0.23$). However, as observed in our earlier analyses, age seems to affect CI children's JNDs and cue weights. To ensure that the correlations were not caused by general age effects, we computed linear regression models with CI children's F0 or VTL cue weights, or their F0 or VTL JNDs as outcome variables, and only age as a fixed effect, e.g., **F0 cue weight ~ age**, and extracted the residuals from these models. We found that there was still a relatively large significant correlation between the residuals of children's JND and cue weight residuals for F0 (*Pearson's* $r = -0.59$, $P < 0.05$) but not for VTL (*Pearson's* $r = -0.25$, $P = 0.41$). To summarize, CI children with high F0 JNDs tended to show lower F0 cue weights.

4.3.3 Discussion

The results from the second experiment show that the perceptual weight that CI children attributed to F0 and VTL cues for voice gender categorization did not differ from that of NH children. For both NH and CI children, the cue weights increased to adult-like levels during the school-age years. Particularly children's VTL cue weights seemed to increase as a function of chronological age, which is in line with our previous finding that the cue weights of NH children are already adult-like around 6 years of age for F0 and only around 10 years of age for VTL (Nagels et al., 2020a). The z-scores also confirm that few CI children showed F0 and VTL cue weights that differed from their NH age peers and that those who did, in fact, attributed more perceptual weight to F0 and VTL differences. Finally, we found that there was a significant negative correlation between CI children's JNDs and their cue weights for F0 but not for VTL. This correlation indicates that CI children with high F0 discrimination thresholds attributed less perceptual weight to F0 differences for voice gender categorization.

Almost all CI children categorized the voices with the original female speakers' voice parameters exclusively as "woman" and those with the most male-sounding voice parameters exclusively as "man". This finding seems to contradict the observations from previous studies on voice gender categorization in CI children (Cleary et al., 2005; Osberger et al., 1991; Staller et al., 1991). The discrepancy between our results with those of Osberger et al. (1991) and Staller et al. (1991) may be a simple consequence of improved CI technology and better speech and language outcomes due to the earlier identification of hearing loss and CI intervention (Kirk et al., 2002; May-Mederake, 2012; Nicholas & Geers, 2007; Svirsky et al., 2007). However, also in slightly more recent work by Cleary et al. (2005), half of the tested CI children did not show reliable discrimination between the voices at the outer ends of their voice gender continuum. A potential explanation of this discrepancy may be that the mean age at implantation of CI children was 12 months of age in the current study and 36 months of age in the study by Cleary et al. (2005). As was demonstrated earlier by Zaltz et al. (2018), age at implantation seems to have a considerable impact on VTL perception. In addition, the tasks seem to measure different aspects of voice perception. Cleary et al. (2005) asked children if two voices belonged to the same speaker or to two different speakers, which examines children's knowledge of possible within-speaker variability in addition to between-speaker variability, while we asked children to indicate if the speaker was a man or a woman, which examines their acoustic representations of speakers' voice gender. These different task demands likely contributed to the different outcomes, as such dissociations have also been observed in NH adult listeners (Johnson et al., 2020; Lavan et al., 2020). For instance, it is possible that children were able to hear a voice difference in some voice conditions in the study by Cleary et al. (2005) but assumed this was possible within-speaker variability. However, this still does not explain why some CI children were unable to categorize the voices at the outer ends of the voice gender continuum in which there was a clear difference in the perceived voice gender. Another possibility is that the step sizes of the changes in mean F0 and VTL were larger in the current study than those applied by Cleary et al. (2005), which may have made the voice gender differences clearer to CI children.

Our results show that prelingually deaf CI children weigh both F0 and VTL cues for voice gender categorization, unlike postlingually deaf CI adults, who mainly rely on differences in F0 and not so much on those in VTL for voice gender categorization (Fuller et al., 2014; Meister et al., 2016; Skuk et al., 2020). This finding is also in agreement with our results from the first experiment, which showed that

CI children had better F0 and VTL discrimination thresholds and therefore seem able to better pick up on differences in these voice cues than the postlingually deaf CI adults that were tested by Gaudrain and Başkent (2018). Hence, the results from both experiments suggest that CI children seem to use the reduced F0 and VTL cues that are transmitted via the spectrotemporally degraded CI signal better than CI adults. Previous research on vowel recognition in CI adults and NH adults that were exposed to noise-vocoded CI-simulated speech made a somewhat similar observation concerning the suboptimal use of acoustic cues (Iverson et al., 2006; McGettigan et al., 2014). Both participant groups did not seem to optimally use acoustic cues that are relatively well preserved in the noise-vocoded and the spectrotemporally degraded CI signals for vowel recognition, such as vowel duration and voicing cues. Specific training may be required to learn postlingually deaf CI users to alter their acoustic representations and attribute more weight to acoustic cues that are more reliable.

In addition, CI children may have formed more robust representations for speakers' voice gender based on the spectrotemporally degraded CI signal. Postlingually deaf CI adults had to adapt their already established acoustic representations of speakers' voice gender that are based on normal acoustic hearing. As neural plasticity decreases with age, adapting these already existing acoustic representations of voice gender categorization could have led to less efficient use of the reduced VTL cues or speaker size information that are available in the degraded CI signal. Somewhat in line with this explanation is the finding by Meister et al. (2016) that postlingually deaf CI adults attribute more weight to VTL cues for voice gender categorization when sentence instead of word stimuli are used, although still to a lesser degree than NH adults. According to Meister et al. (2016, 2020), sentence stimuli may provide a larger variety of phonetic and suprasegmental cues than word stimuli to CI users, which can facilitate the processing of F0 and VTL differences. Together these findings suggest that postlingually deaf CI adults may require more processing time or more acoustic information to use VTL cues, while prelingually deaf CI children can already efficiently extract this information based on short word stimuli. Nevertheless, peripheral limitations may still affect the perceptual weight that CI children attributed to F0 differences, as is shown by the significant correlation with their F0 discrimination thresholds. The perceptual weight that was attributed to F0 differences was lower for CI children with poorer F0 discrimination thresholds. However, this still did not lead to significantly deviant voice gender categorization patterns compared to those observed in NH children.

4.4 Benefit from target-masker voice differences

4.4.1 Methods

Participants

The participants who took part in the first two experiments also performed the third experiment, except for one 4-year-old CI child and three 4-year-old NH children who did not fully complete this experiment.

Stimuli and Apparatus

To measure participants' ability to perceive speech in competing speech, we used the same stimuli and procedure as in our previous study with NH children and adults (Nagels et al., 2021). A coordinate response measure (CRM, used earlier by, for instance, Bolia et al., 2000; Brungart, 2001; Hazan et al., 2009; Moore, 1981; Saleh et al., 2013; Welch et al., 2015) was used with a single-talker speech masker and sentence stimuli that were adapted from the English stimuli of Hazan et al. (2009) and Welch et al. (2015), and translated to Dutch. The 48 target sentences consisted of a carrier phrase in which a color and a number were mentioned, e.g., *Laat de hond zien waar de rode (color) twee (number) is.* [Show the dog where the red (color) two (number) is.]. We used six disyllabic words for basic colors (*rode, zwarte, groene, blauwe, witte, and gele*) [red, black, green, blue, white, and yellow] and eight monosyllabic number words (1-10; but excluding *zeven* [seven] and *negen* [nine], which are disyllabic words in Dutch). We recorded another set of 48 sentences for the masker speech stimuli in which the call sign *hond* [dog] was replaced by *kat* [cat]. Sentence chunks ranging from 150 to 300 ms were randomly selected from these sentences and concatenated after applying 50-ms raised cosine ramps to create the masker speech. We used sentence chunks instead of complete sentences to simplify the task for children, as it was easier to explain which sentence they should pay attention to. The masker also started 750 ms before the target and ended 250 ms after the target. This procedure was also used before by El Boghdady et al. (2019). The stimuli were recorded in an anechoic room at the University Medical Center Groningen at a sampling rate of 44.1 kHz. The stimuli were produced by a different speaker than the NVA corpus speaker who produced the stimuli of the first and second experiments. This speaker was also a female Dutch native speaker with no discernable regional accent. The mean F0 value of the stimuli was 242 Hz, and the estimated VTL was 13.4 cm based on the speaker's height of 166 cm. The mean duration of the stimuli was 2.27 seconds and ranged from 2.14 to 2.49 seconds.

We measured the benefit that participants derived from voice differences in the F0 and VTL of the target and the masker by manipulating the F0 and VTL parameters of the masker. The same STRAIGHT resynthesis procedure was applied as in the first and second experiments, to manipulate the F0 and VTL parameters of the stimuli. We used four different voice conditions for the masker relative to the target: (1) no differences in F0 or VTL, (2) a difference of -12 st in F0, (3) a difference of +3.8 st in VTL, or (4) a combined difference of -12 st in F0 and +3.8 st in VTL (equal to a male-sounding voice). These differences resulted in mean F0 values of 242 Hz and 121 Hz and estimated VTL values of 13.6 cm and 16.7 cm.

Procedure

The experiment was performed after the completion of the first two experiments. Participants first performed a practice session, which consisted of eight trials. In the first three trials, the target speech was presented without the competing speech masker to familiarize participants with the target sentences and experiment interface. For the remaining five trials of the practice session, the target and masker were presented at a target-to-masker ratio (TMR) of +6 dB with target-masker voice differences in F0 and VTL. After the practice session, the participants performed the experiment. The design of the experiment differed from our previous study with NH children in the TMRs that were used (Nagels et al., 2021). We initially replaced the -6 dB TMR condition with a +12 dB TMR condition to ensure that there would be a relatively easy condition that would keep CI children motivated. However, the first three CI children that were tested showed relatively good accuracy scores, even in the 0 dB TMR condition, which made us decide to add a -6 dB TMR condition to the experiment. Therefore, the experiment consisted of 84 trials with seven items per voice condition and per TMR (0 dB, +6 dB, and +12 dB) for the first three tested CI children, and 112 trials for the other CI children that participated with also seven items per TMR and per voice condition ($7 \text{ items} \times 4 \text{ TMRs} \times 4 \text{ voice conditions}$). The target stimuli were calibrated and presented at a fixed level of 65 dBA. All items were presented in a randomized order and within a single block with two optional breaks. The duration of the experiment was around 15 to 20 minutes.

Participants were instructed to pay attention to the color and the number that were mentioned in the target sentence with the carrier phrase (e.g., *Laat de hond zien waar de rode twee is.* [Show the dog where the red two is.]) and to press the correct color-number combination button on the touchscreen as fast as possible. After participants gave a response, the experiment would proceed with the next trial. The participants could choose between 48 different color-number combination buttons.

They would receive one point when the color and number were both correct and 0 points if the color or number were incorrect, similar as applied by, for instance, Brungart et al. (2001). The probability of giving a correct response due to chance was 2.08%. The participants did not receive any feedback on the accuracy of their responses.

Data analysis

We analyzed CI children's benefit from target-masker voice differences in F0 and VTL cues for the perception of speech in competing speech to examine if this benefit developed in the same manner as in NH children. To be able to make a comparison between the two groups, we did not include the data of CI children in the +12 TMR condition. We first normalized the differences in F0 and VTL to make them equivalent for model fitting by defining them as $\delta F0 = -\Delta F0 \div 12 - 0.5$ and $\delta VTL = \Delta VTL \div 3.6 - 0.5$. We computed a mixed-effects logistic regression model for the accuracy of children's responses with random intercepts per participant, and fixed effects of *group*, $\delta F0$, δVTL , *TMR*, and *age*, in lme4 syntax: `correct ~ group +  $\delta F0$  +  $\delta VTL$  + age + TMR + (1 | participant)`. We used forward stepwise selection tests to evaluate the fit of different models with ANOVA Chi-Square tests.

In our previous study with NH children and adults (Nagels et al., 2021), we observed that the benefit from target-masker differences partially depends on participants' overall accuracy scores, as there is less room for improvement when accuracy scores are high overall. Hence, we interpolated the accuracy scores of NH and CI children to the same performance level in the same-voice condition with no target-masker differences in F0 or VTL. We used a linear mixed-effects model on the interpolated accuracy scores of children to examine their benefit from target-masker differences in F0 and VTL when their overall performance was interpolated across TMRs to the same performance level. Children's logit-transformed accuracy scores were interpolated across TMRs to a logit score of 0.290 (equivalent to an accuracy score of 57.2% correct) in the same-voice condition. This specific value caused the fewest outliers, whose scores could not be interpolated to 57.2% in the same-voice condition. Finally, we calculated the Pearson correlation coefficients between CI children's F0 and VTL JNDs (experiment 1) and their benefit from target-masker differences in F0 and VTL (in Bk/st, i.e., an increase of 1 Bk/st corresponds to doubling the odds of getting a correct response for each semitone of voice difference), and their overall accuracy scores.

4.4.2 Results

Figure 4.6 shows the accuracy scores of CI children (shown as individual data points) and how these fit within the distribution of accuracy scores of NH children and adults (shown as boxplots) in percentage points as a function of TMR, masker voice condition, and chronological age. The model comparison analysis showed that the model with two-way interactions between *group* and *age*, *group* and *TMR*, $\delta F0$ and *TMR*, δVTL and *TMR*, and $\delta F0$ and δVTL was the best fitting model, in lme4 syntax: `correct ~ group * age + group * TMR +  $\delta F0$  * TMR +  $\delta VTL$  * TMR +  $\delta F0$  *  $\delta VTL$  + (1 | participant)`. This model had a better fit than the models without any of the two-way interactions or the models with any three-way interactions.

We found significant main effects of *age* ($\chi^2 = 119.9$, $\zeta = 10.95$, $P < 0.001$), *TMR* ($\chi^2 = 371.1$, $\zeta = 19.26$, $P < 0.001$), $\delta F0$ ($\chi^2 = 78.2$, $\zeta = 8.84$, $P < 0.001$), and δVTL ($\chi^2 = 42.93$, $\zeta = 6.55$, $P < 0.001$). These effects showed that the accuracy scores of NH and CI children improved as a function of chronological age, as the TMR became more advantageous, and as a function of target-masker differences in F0 or VTL. Furthermore, there were significant two-way interactions between *group* and *age* ($\chi^2 = 15.2$, $\zeta = -3.90$, $P < 0.001$), and between *group* and *TMR* ($\chi^2 = 15.4$, $\zeta = 3.92$, $P < 0.001$). These interactions indicate that the accuracy scores of CI children improved less with age compared to NH children, but improved more as the TMR became higher compared to NH children. Finally, there were significant two-way interactions between $\delta F0$ and *TMR* ($\chi^2 = 22.48$, $\zeta = -4.74$, $P < 0.001$), δVTL and *TMR* ($\chi^2 = 12.39.4$, $\zeta = -3.52$, $P < 0.001$), and $\delta F0$ and δVTL ($\chi^2 = 28.54$, $\zeta = -5.34$, $P < 0.001$). These interactions demonstrate that the benefit that NH and CI children derived from target-masker differences in F0 and VTL decreased as the TMR became more advantageous. Their benefit also became smaller when there were target-masker differences in both F0 and VTL cues as opposed to differences in only one voice cue. In sum, the accuracy scores of NH and CI children improved as a function of chronological age, as the TMR became more advantageous, and as a result of target-masker differences in F0 or VTL. The accuracy scores of CI children improved less with age than those of NH children but improved more as the TMR became higher. Finally, children's benefit from target-masker differences in F0 and VTL became smaller as the TMR became more advantageous, and was smaller when there was a combined difference in F0 and VTL instead of a difference in only one voice cue.

Figure 4.6: The accuracy scores of participants for the speech perception in competing speech experiment in percentage points as a function of chronological age, TMR, and voice condition ($N_{CI\ children} = 13$, $N_{NH\ children} = 55$, $N_{NH\ adults} = 15$). The panels from top to bottom show the accuracy scores in percentage points for the -6 dB TMR (first-row panels), 0 dB TMR (second-row panels), and +6 dB TMR (third-row panels) conditions for CI children and NH children and adults, and the +12 dB TMR (fourth-row panels) condition for CI children. Each panel consists of four plots that show the accuracy scores per voice condition for each TMR condition, arranged from the condition with no target-masker differences in F0 and VTL (left-sided plots) to the condition with target-masker differences of -12 st in F0 and +3.8 st in VTL (right-sided plots).

Since the benefit from target-masker differences partially depends on participants' overall accuracy scores and CI children generally had lower accuracy scores than their NH age peers, we interpolated children's accuracy scores to the same performance level to analyze the differences in their benefit. Children's accuracy scores were interpolated across TMRs to a logit score of 0.290 (equivalent to an accuracy score of 57.2%) in the same-voice condition. Figure 4.7a shows the interpolated logit scores of children as a function of their chronological age and age group for each voice condition, and Figure 4.7b shows the TMR values of individual children that correspond to these scores. The logit scores of ten NH children and two CI children (who were not tested with the -6 dB TMR condition) were all above 57.2% in the same-voice condition, which explains why not all data points are centered to 57.2% in the upper left panel of Figure 4.7a. The model comparison analysis showed that the model with a two-way interaction between $\delta F0$ and δVTL , and fixed effects of *group* and *age*, was the best fitting model, in lme4 syntax: `correct ~  $\delta F0$  *  $\delta VTL$  + group + age + (1 | participant)`.

The interpolated logit scores of NH and CI children improved as a result of target-masker voice differences in F0 ($t = 9.60$, $Estimate = 0.73$, $SE = 0.076$, $P < 0.001$) or VTL ($t = 7.30$, $Estimate = 0.56$, $SE = 0.076$, $P < 0.001$). The significant two-way interaction between $\delta F0$ and δVTL ($t = -4.61$, $Estimate = -0.70$, $SE = 0.153$, $P < 0.001$) indicates that the benefit from target-masker differences in F0 and VTL decreased when the two voice cues were manipulated together instead of a change in one voice cue. Furthermore, the interpolated logit scores of NH and CI children improved as a function of chronological age ($t = 5.21$, $Estimate = 0.11$, $SE = 0.020$, $P < 0.001$). However, the lack of an interaction between *age* with $\delta F0$ or δVTL indicates that the benefit from target-masker differences in F0 and VTL remained the same across age. Finally, the interpolated logit scores of CI children were overall lower than those of NH children, even though the scores were interpolated to the same logit score in the same-voice condition ($t = -2.38$, $Estimate = -0.32$, $SE = 0.134$, $P < 0.05$). However, after removing one outlying data point, i.e., the data point of a 6-year-old CI child (second panel of Figure 4.7a), this effect was no longer significant ($t = -1.98$, $Estimate = -0.26$, $SE = 0.134$, $P = 0.052$). To summarize, NH and CI children showed a benefit from target-masker differences in F0 and VTL that was similar in magnitude and remained the same across age. The interpolated logit scores of NH and CI children improved as a function of chronological age, and these did not differ across groups once one outlying data point of one 6-year-old CI child was removed.

Figure 4.7: The accuracy scores of CI and NH children per age group and voice condition interpolated across TMRs to 57.2% correct in the same-voice condition ($N_{\text{CI children}} = 13$, $N_{\text{NH children}} = 55$). (a) The four panels show the accuracy scores of NH and CI children per voice condition. The dots show individual data points of CI children located at their chronological age in years, rounded to two decimal places. The boxplots show the median accuracy scores of NH children per age group and the lower and upper quartiles. The whiskers indicate the lowest and highest data points within plus or minus 1.5 times the interquartile range. (b) The TMRs of NH and CI children at 57.2% correct in the same-voice condition that were used for the interpolation.

To investigate if the F0 and VTL discrimination thresholds of CI children could have limited their benefit from target-masker differences in F0 and VTL, or affected their overall ability to perceive speech in competing speech, we computed the Pearson correlation coefficients between these measures. We calculated the Pearson correlation coefficient of CI children's log-transformed JNDs and their benefit

from target-masker differences in F0 and VTL, and their overall accuracy scores. Our results show no significant correlation between CI children's F0 JNDs and their benefit from F0 differences (*Pearson's* $r = 0.01$, $P = 0.93$), but there was a significant correlation between their VTL JNDs and their benefit from VTL differences (*Pearson's* $r = 0.27$, $P < 0.05$). However, we also computed the Pearson correlation coefficients using the residuals of these measures, correcting for general age effects like we did in the second experiment. We computed linear regression models with CI children's F0 or VTL JNDs, their F0 or VTL benefit (Bk/st), or their overall accuracy scores as outcome variables and only *age* as a fixed effect, e.g., **F0 benefit** ~ **age**, and extracted the residuals from these models. When we used the residuals of CI children's JNDs and their benefit, we neither found significant correlations between the JNDs and the benefit for F0 (*Pearson's* $r = 0.04$, $P = 0.89$) nor for VTL (*Pearson's* $r = 0.43$, $P = 0.14$). Thus, the initial significant correlation between the VTL JNDs and benefit from target-masker differences in VTL mainly seems to be related to general effects of age. Furthermore, there was no significant correlation between CI children's overall accuracy scores and their F0 JNDs (*Pearson's* $r = -0.52$, $P = 0.07$) or their VTL JNDs (*Pearson's* $r = -0.47$, $P = 0.10$). There was also no significant correlation between the residuals of CI children's overall accuracy scores and those of their F0 JNDs (*Pearson's* $r = -0.35$, $P = 0.25$) or VTL JNDs (*Pearson's* $r = -0.04$, $P = 0.89$). To conclude, we only found a significant correlation between CI children's VTL JNDs and their benefit from target-masker differences in VTL. However, this correlation was no longer significant when we used the residuals of their JNDs and their benefit, and hence mainly seemed to be caused by a general effect of age.

4.4.3 Discussion

The results of our third experiment demonstrate that, although the accuracy scores of CI children were generally lower compared to those of their NH age peers, CI children showed a clear benefit from target-masker differences in F0 and VTL for perceiving speech in competing speech. The accuracy scores of NH and CI children improved as a function of chronological age, increasing the TMR, and larger target-masker differences in F0 and VTL. The improvement in accuracy scores with age was smaller for CI children compared to NH children, but the improvement as a function of TMR was larger. The analysis using the interpolated accuracy scores of NH and CI children indicated that their overall accuracy scores improved with age, but the benefit that they derived from target-masker differences in F0 and VTL remained the same across age and was the same across groups. Finally, we only

found a significant correlation between the VTL JNDs of CI children and their benefit from target-masker differences in VTL, but this correlation was no longer significant when we used the residuals of these measures which corrected for general age effects.

We found that CI children benefitted from target-masker differences in F0 and VTL to a similar degree as NH children when their accuracy scores were interpolated to the same performance level. These results are also in line with the findings of Misurelli and Litovsky (2015) that show CI children benefitted when the gender of the target and the masker differed. Our results complement these previously reported findings by showing that CI children can also benefit from target-masker voice differences in speakers' mean F0 only or their VTL only while all other voice cues remain unchanged. The results indicate that the target-masker differences in F0 and VTL cues in the current study seem to be represented in the spectrotemporally degraded CI signal to a certain extent that allows CI children to better perceive speech in competing speech. Furthermore, although the F0 and VTL discrimination thresholds of CI children were overall poorer than those of their NH age peers, their benefit from target-masker differences in F0 and VTL and their voice discrimination thresholds were not closely related. Hence, the poorer voice discrimination abilities of CI children do not seem to be the primary cause of their difficulties with perceiving speech in competing speech. For NH children, a similar discrepancy between their discrimination abilities and their benefit from target-masker differences has been observed (Flaherty et al., 2019; Nagels et al., 2021; Sussman et al., 2007; Sussman & Steinschneider, 2009). Nevertheless, the lack of a significant correlation could also be caused by the fact that the target-masker differences in F0 and VTL that were used in this experiment were relatively large, i.e., differences of -12 st in F0 and +3.8 st in VTL. Based on the results of the first and second experiments, these differences are above the F0 and VTL discrimination thresholds of most CI children, and they seem to be able to use these differences effectively for voice gender categorization. The benefit that CI children derived from target-masker voice differences for perceiving speech in competing speech could still differ from their NH age peers when the voice differences are more subtle and nearer to their discrimination thresholds.

Although the benefit from target-masker differences in F0 and VTL did not differ across groups when NH and CI children's accuracy scores were interpolated, the overall accuracy scores of CI children were still significantly lower than those of their NH age peers. Besides perceptual limitations imposed by the spectrotemporally degraded CI signal, having difficulties with perceiving speech in

competing noise or speech may also give CI children less opportunity for incidental language learning compared to their NH age peers. Auditory exposure and linguistic factors, such as vocabulary size, seem to play an important role in the perception of speech in competing speech in NH children by affecting their ability to restore masked speech segments (Corbin et al., 2016; Klein et al., 2017; McCreery et al., 2017, 2019). Lower auditory exposure may result in less well-developed speech restoration abilities in CI children compared to those of their NH age peers and hence contribute to their difficulties with perceiving speech in competing noise or speech. Furthermore, other cognitive mechanisms, such as selective auditory attention and the inhibition of masker interference, that are involved in auditory stream segregation may play a role (Buss et al., 2017b; Sussman et al., 2007). For instance, the temporary reductions in masking for competing speech maskers improve the perception of speech in competing speech for NH adults, i.e., ‘glimpsing’ (Rosen et al., 2013), but not so much for NH children (Buss et al., 2017b; Litovsky, 2005). As the temporal resolution is likely poorer for CI children due to the spectrotemporally degraded CI signal, they may be even less able to use glimpsing for speech stream segregation than NH children of the same age. However, in general, perceiving speech in competing speech seems to still improve in CI children as a function of chronological age which implies that cognitive-developmental factors can still help to overcome their reduced perceptual abilities to a certain degree. Based on the results of the current study, it is difficult to conclude to what extent the ability to perceive speech in competing speech can improve in CI children, and if their abilities will plateau at a certain age. More research is required to further examine the observed developmental trend in CI children’s ability to perceive speech in competing speech by testing a larger number of CI children or, ideally, by collecting longitudinal data.

Consistent with the results of the first and second experiments, the results of the third experiment also suggest that there are discrepancies between the perceptual abilities of prelingually deaf CI children and those of postlingually deaf CI adults. El Boghdady et al. (2019), using a similar resynthesis procedure, found that postlingually deaf CI adults rarely showed a benefit from target-masker differences in F0 and VTL for perceiving speech in competing speech. In contrast, our results demonstrate that CI children show a benefit from target-masker differences in F0 and VTL similar to their NH age peers. As mentioned earlier, the results of a potential benefit from target-masker voice differences in postlingually deaf CI adults have been inconclusive. While some studies reported that most CI adults did not show a benefit from voice differences related to speakers’ voice

gender (speakers of different genders: Bernstein et al., 2016; Stickney et al., 2004; target-masker F0 differences: Auinger et al., 2017; El Boghdady et al., 2019; Stickney et al., 2007; or target-masker VTL differences: El Boghdady et al., 2019), other studies did find such a benefit (speakers of different genders: Cullington & Zeng, 2008; target-masker F0 differences: Meister et al., 2020; Pyschny et al., 2011). Differences in the experiment design seem to play an important role in explaining the discrepancies in the results of different studies (Litovsky et al., 2017). In our previous study with only NH children (Nagels et al., 2021), we also found that NH school-age children benefitted from target-masker differences in F0 and VTL at all tested ages, while they did not show a benefit from F0 and VTL differences in the studies by Flaherty et al. (2019, 2021). Particularly using simple closed-set sentence materials with a carrier phrase may have made it easier for children to use the voice differences to keep track of the target speech stream in the current study (Bonino et al., 2013; Freyman et al., 2004). As El Boghdady et al. (2019) used open-set sentence stimuli without a carrier phrase in their study with postlingually deaf CI adults, the differences in stimuli should be taken into consideration when comparing the results with those of the current study with prelingually deaf CI children. Yet, combined with the results from the previous two experiments, our results imply that CI children may also outperform CI adults in this aspect of voice perception.

4.5 General discussion

In the current study, we investigated the development in the perception of the primary acoustic cues that define speakers' voice gender, namely speakers' mean F0 and their VTL, in prelingually deaf CI children. We conducted three voice perception experiments that varied from lower-order perception tasks to higher-order cognitive tasks. We examined CI children's F0 and VTL discrimination thresholds (Experiment 1), the perceptual weight they attributed to F0 and VTL cues for voice gender categorization (Experiment 2), and their benefit from target-masker differences in F0 and VTL cues for perceiving speech in competing speech (Experiment 3). The results from CI children were compared to the data collected from NH children in earlier studies using the same experiments (Nagels et al., 2020a, 2021). The results from our first experiment show that the F0 and VTL discrimination thresholds of CI children were generally higher than those of their NH age peers, but their thresholds improved as a function of age and the same developmental pattern across voice cues was observed. Although the discrimination

thresholds of CI children were poorer, the results of our second experiment indicate that this did not lead to a deviant weighting of F0 and VTL cues for voice gender categorization. The F0 and VTL cue weights also developed as a function of age similar to NH children. However, CI children with high F0 discrimination thresholds showed lower F0 cue weights, indicating that poorer discrimination of F0 differences resulted in a lower perceptual weighting of F0 cues for voice gender categorization. Finally, the results of our third experiment show that the ability to perceive speech in competing speech was generally poorer in CI children relative to their NH age peers. Nevertheless, when NH and CI children's accuracy scores were interpolated to the same performance level, their benefit from target-masker differences in F0 and VTL did not differ across age or groups. The ability to perceive speech in competing speech improved as a function of age in both NH and CI children, but their accuracy scores and their benefit from target-masker differences in F0 and VTL were not related to their F0 and VTL discrimination thresholds.

Although we found significant effects of age on the perceptual abilities of CI children for all three experiments, there still seems to be a large amount of variability in their perceptual abilities that remains unexplained. Some device-related factors may play a role in explaining this variability, such as the CI stimulation strategy that was used by CI children. All CI children that took part in the current study had devices from Cochlear and were using either the ACE or MP3000 stimulation strategies, making the CI participant group relatively homogeneous concerning these aspects, particularly compared to postlingually deaf CI adults. The CI stimulation strategies affect the spectrotemporal resolution of the perceived speech signal (Wouters et al., 2015), which could also have an impact on the abilities to perceive voice characteristics and speech. The MP3000 strategy is based on the ACE strategy but additionally attempts to select the most perceptually relevant channels to reduce the spread in excitation between electrodes and increase the spectral resolution. Nevertheless, previous studies did not find any differences in speech intelligibility or CI users' preference between the two CI stimulation strategies (Buechner et al., 2011). Yet, differences in CI stimulation strategies may be particularly important to consider when we compare the results of CI children to those of postlingually deaf CI adults, who often use a wider range of device types and stimulation strategies. Furthermore, all except one CI child was bilaterally implanted, while many CI adults are unilaterally implanted in the Netherlands. This difference may especially be important for comparing the abilities to perceive speech in competing speech of CI children and CI adults (Brown & Balkany, 2007).

Other factors that may play a role are differences in the underlying etiology (Pyman et al., 2000), or the degree of auditory nerve survival and adaptation to mismatches in the nerve-electrode array (Gordon et al., 2011). Furthermore, many environmental factors have been demonstrated to play a role, such as the amount of time CI children use their CI on a daily basis, socioeconomic status, and parental involvement in the rehabilitation process (Boons et al., 2012; Busch et al., 2020; Geers et al., 2003; Niparko et al., 2010; Spencer, 2004). Other cognitive abilities, such as selective auditory attention, inhibition, and auditory working memory, also seem to contribute to CI children's perceptual abilities (Kronenberger et al., 2013). Our analysis of the potential influence of these factors is slightly limited because of the small sample size, and at least for CI users, relatively homogeneous participant group in terms of device-related aspects and age at implantation. Nevertheless, these are important factors that should be kept in mind in future research in which a larger number and more heterogeneous group of CI children can be tested.

One of the most interesting findings of this study is the discrepancy between the results of prelingually deaf CI children in the current study and those of previous work from our research group with postlingually deaf CI adults using the same STRAIGHT resynthesis procedure (El Boghdady et al., 2019; Fuller et al., 2014; Gaudrain & Başkent, 2018). The CI children in the current study had lower mean F0 and VTL discrimination thresholds than the postlingually deaf CI adults that were tested by Gaudrain and Başkent (2018). In addition, unlike postlingually deaf CI adults, who mainly relied on F0 differences for voice gender categorization and not so much on VTL differences (Fuller et al., 2014; Meister et al., 2016; Skuk et al., 2020), CI children relied on differences in both F0 and VTL and their F0 and VTL cue weights did not differ from those of their NH age peers. Finally, CI children's benefit from target-masker differences in F0 and VTL for perceiving speech in competing speech did not differ from their NH age peers when their overall accuracy scores were interpolated to the same performance level, unlike postlingually deaf CI adults, who generally did not show a benefit (El Boghdady et al., 2019). Based on the high VTL JNDs of CI adults (Gaudrain & Başkent, 2018), the explanation that VTL cues are not transmitted via the CI seemed more plausible at the time than the alternative that the distorted VTL cues are not interpreted correctly as speaker size differences. However, the results of CI children in the current study support the latter explanation that VTL cues are available in the spectrotemporally degraded CI signal to a certain degree. Together, these findings suggest that CI children are able to use the reduced F0 and VTL cues that are available in the spectrotemporally degraded CI signal better than postlingually deaf

CI adults. The early exposure to the spectrotemporally degraded CI signal and high brain plasticity (Manrique et al., 1999) may have resulted in this better use of reduced F0 and VTL cues. Also, adapting existing auditory representations of speakers' voice gender that are based on normal acoustic hearing may, in fact, impede learning which acoustic cues related to voice gender are available and how these are represented in the spectrotemporally degraded CI signal for CI adults. Specific auditory training could perhaps support CI adults in learning to adjust their existing auditory representations and optimally use acoustic cues that are relatively well preserved in the spectrotemporally degraded CI signal (Iverson et al., 2006; McGettigan et al., 2014).

To conclude, although CI children's F0 and VTL discrimination thresholds were poorer than those of their NH age peers, their weighting of F0 and VTL cues for voice gender categorization and their benefit from target-masker differences in F0 and VTL for perceiving speech in competing speech were similar. The results of the current study show the importance of cognitive-developmental factors on voice perception in CI children, which are also present in NH children, besides only perceptual limitations imposed by the spectrotemporally degraded CI signal. Furthermore, our results imply that prelingually deaf CI children are better at capitalizing on the reduced F0 and VTL cues that are available in the spectrotemporally degraded CI signal than postlingually deaf CI adults. These results also provide evidence for greater perceptual learning in CI children than in CI adults, possibly related to early exposure to spectrotemporally degraded speech and high brain plasticity (Manrique et al., 1999).

Bibliography

- Auinger, A. B., Riss, D., Liepins, R., Rader, T., Keck, T., Keintzel, T., Kaider, A., Baumgartner, W.-D., Gstöettner, W., & Arnoldner, C. (2017). Masking release with changing fundamental frequency: Electric acoustic stimulation resembles normal hearing subjects. *Hearing Research*, 350, 226–234.
- Barry, J. G., Blamey, P. J., Martin, L. F. A., Lee, K. Y.-S., Tang, T., Ming, Y. Y., & Hasselt, C. A. van. (2002). Tone discrimination in Cantonese-speaking children using a cochlear implant. *Clinical Linguistics & Phonetics*, 16(2), 79–99.
- Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, 139(3), EL51–EL56.
- Başkent, D., Gaudrain, E., Tamati, T. N., & Wagner, A. (2016). Perception and psychoacoustics of speech in cochlear implant users. In A. T. Cacace, E. de Kleine, A. G. Holt, & P. van Dijk (Red.), *Scientific Foundations of Audiology: Perspectives from Physics, Biology, Modeling, and Medicine* (pp. 285–319). Plural Publishing, Inc.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–51.
- Bennett, S., & Montero-Díaz, L. (1982). Children’s perception of speaker sex. *Journal of Phonetics*, 10(1), 113–121.
- Bernstein, J. G. W., Goupell, M. J., Schuchman, G. I., Rivera, A. L., & Brungart, D. S. (2016). Having two ears facilitates the perceptual separation of concurrent talkers for bilateral and single-sided deaf cochlear implantees. *Ear and Hearing*, 37(3), 289–302.
- Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, 107(2), 1065–1066.
- Bonino, A. Y., Leibold, L. J., & Buss, E. (2013). Release from perceptual masking for children and adults: Benefit of a carrier phrase. *Ear and Hearing*, 34(1), 3–14.
- Boons, T., Brokx, J. P. L., Dhooge, I., Frijns, J. H. M., Peeraer, L., Vermeulen, A., Wouters, J., & van Wieringen, A. (2012). Predictors of spoken language development following pediatric cochlear implantation. *Ear and Hearing*, 33(5), 617–639.
- Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology*, 34(5), 260–284.
- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press.
- Brown, K. D., & Balkany, T. J. (2007). Benefits of bilateral cochlear implantation: A review. *Current Opinion in Otolaryngology & Head & Neck Surgery*, 15(5), 315–318.
- Brujnzeel, H., Bezdjian, A., Lesinski-Schiedat, A., Illg, A., Tzifa, K., Monteiro, L., Volpe, A. della, Grolman, W., & Topsakal, V. (2017). Evaluation of pediatric cochlear implant care throughout Europe: Is European pediatric cochlear implant care performed according to guidelines? *Cochlear Implants International*, 18(6), 287–296.

- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, 109(3), 1101–1109.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, 110(5), 2527–2538.
- Buechner, A., Beynon, A., Szyfter, W., Niemczyk, K., Hoppe, U., Hey, M., Brokx, J., Eyles, J., Van de Heyning, P., Paludetti, G., Zarowski, A., Quaranta, N., Wesarg, T., Festen, J., Olze, H., Dhooge, I., Müller-Deile, J., Ramos, A., Roman, S., ... Smoorenburg, G. (2011). Clinical evaluation of cochlear implant sound coding taking into account conjectural masking functions, MP3000™. *Cochlear Implants International*, 12(4), 194–204.
- Busch, T., Vermeulen, A., Langereis, M., Vanpoucke, F., & van Wieringen, A. (2020). Cochlear implant data logs predict children's receptive vocabulary. *Ear and Hearing*, 41(4), 733–746.
- Buss, E., Flaherty, M. M., & Leibold, L. J. (2017a). Development of frequency discrimination at 250 Hz is similar for tone and /ba/ stimuli. *The Journal of the Acoustical Society of America*, 142(1), EL150–EL154.
- Buss, E., Leibold, L. J., Porter, H. L., & Grose, J. H. (2017b). Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing. *The Journal of the Acoustical Society of America*, 141(4), 2650–2660.
- Chatterjee, M., & Peng, S.-C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research*, 235(1–2), 143–156.
- Chatterjee, M., Zion, D. J., Deroche, M. L. D., Burianek, B. A., Limb, C. J., Goren, A. P., Kulkarni, A. M., & Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research*, 322, 151–162.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5), 975–979.
- Giocca, V., Francis, A. L., Aisha, R., & Wong, L. (2002). The perception of Cantonese lexical tones by early-deafened cochlear implantees. *The Journal of the Acoustical Society of America*, 111(5), 2250–2256.
- Cleary, M., & Pisoni, D. B. (2002). Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results. *Annals of Otology, Rhinology & Laryngology*, 111(5_suppl), 113–118.
- Cleary, M., Pisoni, D. B., & Kirk, K. I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 48(1), 204–223.
- Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear and Hearing*, 37(1), 55–63.

- Cullington, H. E., & Zeng, F.-G. (2008). Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. *The Journal of the Acoustical Society of America*, 123(1), 450–461.
- Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America*, 114(5), 2913–2922.
- Davidson, L. S., Geers, A. E., Blamey, P. J., Tobey, E., & Brenner, C. (2011). Factors contributing to speech perception scores in long-term pediatric CI users. *Ear and Hearing*, 32(1 Suppl), 19S–26S.
- Deroche, M. L. D., Kulkarni, A. M., Christensen, J. A., Limb, C. J., & Chatterjee, M. (2016). Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants. *Frontiers in Neuroscience*, 10, 1–15.
- Deroche, M. L. D., Lu, H.-P., Kulkarni, A. M., Caldwell, M., Barrett, K. C., Peng, S.-C., Limb, C. J., Lin, Y.-S., & Chatterjee, M. (2019). A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. *Scientific Reports*, 9(1), 1–12.
- Deroche, M. L. D., Lu, H.-P., Limb, C. J., Lin, Y.-S., & Chatterjee, M. (2014). Deficits in the pitch sensitivity of cochlear-implanted children speaking English or Mandarin. *Frontiers in Neuroscience*, 8, 1–13.
- DiNino, M., & Arenberg, J. G. (2018). Age-related performance on vowel identification and the spectral-temporally modulated ripple test in children with normal hearing and with cochlear implants. *Trends in Hearing*, 22, 2331216518770959.
- El Boghdady, N., Gaudrain, E., & Başkent, D. (2019). Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users? *The Journal of the Acoustical Society of America*, 145(1), 417–439.
- Everhardt, M. K., Sarampalis, A., Coler, M., Baskent, D., & Lowie, W. (2020). *Meta-Analysis on the Identification of Linguistic and Emotional Prosody in Cochlear Implant Users and Vocoder Simulations*. LWW.
- Fagan, M. K., & Pisoni, D. B. (2010). Hearing experience and receptive vocabulary development in deaf children with cochlear implants. *The Journal of Deaf Studies and Deaf Education*, 15(2), 149–161.
- Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106(3), 1511–1522.
- Flaherty, M. M., Buss, E., & Leibold, L. J. (2019). Developmental effects in children’s ability to benefit from F0 differences between target and masker speech. *Ear and Hearing*, 40(4), 927–937.
- Flaherty, M. M., Buss, E., & Leibold, L. J. (2021). Independent and combined effects of fundamental frequency and vocal tract length differences for school-age children’s sentence recognition in a two-talker masker. *Journal of Speech, Language, and Hearing Research*, 64(1), 206–217.

- Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., & Graves, S. (2012). *Package 'car'* (R package version 3.0-8) [Computer software].
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 115(5), 2246–2256.
- Fu, Q.-J., Chinchilla, S., Nogaki, G., & Galvin, J. J. (2005). Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *The Journal of the Acoustical Society of America*, 118(3), 1711–1718.
- Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., & Başkent, D. (2014). Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology*, 15(6), 1037–1048.
- Gaudrain, E., & Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear and Hearing*, 39(2), 226–237.
- Geers, A. E., Nicholas, J. G., & Sedey, A. L. (2003). Language skills of children with early cochlear implantation. *Ear and Hearing*, 24(1), 46S–58S.
- Geurts, L., & Wouters, J. (2001). Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *The Journal of the Acoustical Society of America*, 109(2), 713–726.
- Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., & Olszewski, C. (2007). Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear and Hearing*, 28(3), 412–423.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton-Century-Crofts.
- Gifford, R. H., Noble, J. H., Camarata, S. M., Sunderhaus, L. W., Dwyer, R. T., Dawant, B. M., Dietrich, M. S., & Labadie, R. F. (2018). The relationship between spectral modulation detection and speech recognition: Adult versus pediatric cochlear implant recipients. *Trends in Hearing*, 22, 2331216518771176.
- Gordon, K. A., Wong, D. D. E., Valero, J., Jewell, S. F., Yoo, P., & Papsin, B. C. (2011). Use it or lose it? Lessons learned from the developing brains of children who are deaf and use cochlear implants to hear. *Brain Topography*, 24(3), 204–219.
- Green, T., Faulkner, A., & Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *The Journal of the Acoustical Society of America*, 116(4), 2298–2310.
- Green, T., Faulkner, A., Rosen, S., & Macherey, O. (2005). Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification. *The Journal of the Acoustical Society of America*, 118(1), 375–385.
- Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & Shakespeare, B. (2009). Speech perception abilities of adults with dyslexia: Is there any evidence for a true deficit? *Journal of Speech, Language, and Hearing Research*, 52(6), 1510–1529.

- Hilkhuyzen, G., Gaubitch, N., Brookes, M., & Huckvale, M. (2012). Effects of noise suppression on intelligibility: Dependency on signal-to-noise ratios. *The Journal of the Acoustical Society of America*, 131(1), 531–539.
- Hopyan-Misakyan, T. M., Gordon, K. A., Dennis, M., & Papsin, B. C. (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychology*, 15(2), 136–146.
- Iverson, P., Smith, C. A., & Evans, B. G. (2006). Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration. *The Journal of the Acoustical Society of America*, 120(6), 3998–4006.
- Johnson, J., McGettigan, C., & Lavan, N. (2020). Comparing unfamiliar voice and face identity perception using identity sorting tasks. *Quarterly Journal of Experimental Psychology*, 73(10), 1537–1545.
- Jung, K. H., Won, J. H., Drennan, W. R., Jameyson, E., Miyasaki, G., Norton, S. J., & Rubinstein, J. T. (2012). Psychoacoustic Performance and Music and Speech Perception in Prelingually Deafened Children with Cochlear Implants. *Audiology and Neurotology*, 17(3), 189–197.
- Kawahara, H., & Irino, T. (2005). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In *Speech Separation by Humans and Machines* (pp. 167–180). Springer. https://doi.org/10.1007/0-387-22794-6_11
- Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational Masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Red.), *Auditory Perception of Sound Sources* (pp. 143–189). Springer US.
- Kirk, K. I., Miyamoto, R. T., Lento, C. L., Ying, E., O'Neill, T., & Fears, B. (2002). Effects of age at implantation in young children. *The Annals of Otology, Rhinology & Laryngology*, 111(5), 69–73.
- Klein, K. E., Walker, E. A., Kirby, B., & McCreery, R. W. (2017). Vocabulary facilitates speech perception in children with hearing aids. *Journal of Speech, Language, and Hearing Research*, 60(8), 2281–2296.
- Kopelovich, J. C., Eisen, M. D., & Franck, K. H. (2010). Frequency and electrode discrimination in children with cochlear implants. *Hearing Research*, 268(1), 105–113.
- Kovačić, D., & Balaban, E. (2009). Voice gender perception by cochlear implantees. *The Journal of the Acoustical Society of America*, 126(2), 762–775.
- Kovačić, D., & Balaban, E. (2010). Hearing history influences voice gender perceptual performance in cochlear implant users. *Ear and Hearing*, 31(6), 806–814.
- Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons.
- Kronenberger, W. G., Pisoni, D. B., Henning, S. C., & Colson, B. G. (2013). Executive functioning skills in long-term users of cochlear implants: A case control study. *Journal of Pediatric Psychology*, 38(8), 902–914.

- Landsberger, D. M., Padilla, M., Martinez, A. S., & Eisenberg, L. S. (2018). Spectral-temporal modulated ripple discrimination by children with cochlear implants. *Ear and Hearing*, 39(1), 60–68.
- Laneau, J., Wouters, J., & Moonen, M. (2004). Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *The Journal of the Acoustical Society of America*, 116(6), 3606–3619.
- Lavan, N., Merriman, S. E., Ladwa, P., Burston, L. F. K., Knight, S., & McGettigan, C. (2020). ‘Please sort these voice recordings into 2 identities’: Effects of task instructions on performance in voice sorting studies. *British Journal of Psychology*, 111(3), 556–569.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, 49(2B), 467–477.
- Litovsky, R. Y. (2005). Speech intelligibility and spatial release from masking in young children. *The Journal of the Acoustical Society of America*, 117(5), 3091–3099.
- Litovsky, R. Y., Goupell, M. J., Misurelli, S. M., & Kan, A. (2017). Hearing with cochlear implants and hearing aids in complex auditory scenes. In J. Middlebrooks, J. Simon, A. Popper, & R. Fay (Eds.), *The Auditory System at the Cocktail Party* (Vol. 60, pp. 261–291). Springer.
- Lund, E. (2016). Vocabulary knowledge of children with cochlear implants: A meta-analysis. *Journal of Deaf Studies and Deaf Education*, 21(2), 107–121.
- Luo, X., Fu, Q.-J., & Galvin, J. J. (2007). Cochlear implants special issue article: Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification*, 11(4), 301–315.
- Manrique, M., Cervera-Paz, F. J., Huarte, A., Perez, N., Molina, M., & García-Tapia, R. (1999). Cerebral auditory plasticity and cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 49, S193–S197.
- Massida, Z., Marx, M., Belin, P., James, C., Fraysse, B., Barone, P., & Deguine, O. (2013). Gender categorization in cochlear implant users. *Journal of Speech, Language, and Hearing Research*, 56(5), 1389–1401.
- May-Mederake, B. (2012). Early intervention and assessment of speech and language development in young children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 76(7), 939–946.
- McCreery, R. W., Spratford, M., Kirby, B., & Brennan, M. (2017). Individual differences in language and working memory affect children’s speech recognition in noise. *International Journal of Audiology*, 56(5), 306–315.
- McCreery, R. W., Walker, E. A., Spratford, M., Lewis, D., & Brennan, M. (2019). Auditory, cognitive, and linguistic factors predict speech recognition in adverse listening conditions for children with hearing loss. *Frontiers in Neuroscience*, 13, 1–11.
- McGettigan, C., Rosen, S., & Scott, S. K. (2014). Lexico-semantic and acoustic-phonetic processes in the perception of noise-vocoded speech: Implications for cochlear implantation. *Frontiers in Systems Neuroscience*, 8.

- Meister, H., Fürsen, K., Streicher, B., Lang-Roth, R., & Walger, M. (2016). The use of voice cues for speaker gender recognition in cochlear implant recipients. *Journal of Speech, Language, and Hearing Research*, 59(3), 546–556.
- Meister, H., Landwehr, M., Pyschny, V., Walger, M., & Wedel, H. von. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *International Journal of Audiology*, 48(1), 38–48.
- Meister, H., Walger, M., Lang-Roth, R., & Müller, V. (2020). Voice fundamental frequency differences and speech recognition with noise and speech maskers in cochlear implant recipients. *The Journal of the Acoustical Society of America*, 147(1), EL19–EL24.
- Misurelli, S. M., & Litovsky, R. Y. (2015). Spatial release from masking in children with bilateral cochlear implants and with normal hearing: Effect of target-interferer similarity. *The Journal of the Acoustical Society of America*, 138(1), 319–331.
- Moore, B. C. J., & Carlyon, R. P. (2005). Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Red.), *Pitch* (Vol. 24, pp. 234–277). Springer-Verlag.
- Moore, T. J. (1981). Voice communication jamming research. *AGARD Conference Proceedings 331: Aural Communication in Aviation*, 2:1-2:6.
- Most, T., & Aviner, C. (2009). Auditory, visual, and auditory-visual perception of emotions by individuals with cochlear implants, hearing aids, and normal hearing. *The Journal of Deaf Studies and Deaf Education*, 14(4), 449–464.
- Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2020a). Development of voice perception is dissociated across gender cues in school-age children. *Scientific Reports*, 10(1), 1–11.
- Nagels, L., Gaudrain, E., Vickers, D., Matos Lopes, M., Hendriks, P., & Başkent, D. (2020b). Development of vocal emotion recognition in school-age children: The EmoHI test for hearing-impaired populations. *PeerJ*, 8, 1–14.
- Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2021). School-age children benefit from voice gender cue differences for the perception of speech in competing speech. *The Journal of the Acoustical Society of America*, 149(5), 3328–3344.
- Nakata, T., Trehub, S. E., & Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *The Journal of the Acoustical Society of America*, 131(2), 1307–1314.
- Nicholas, J. G., & Geers, A. E. (2007). Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to profound hearing loss. *Journal of Speech, Language, and Hearing Research*, 50(4), 1048–1062.
- Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N.-Y., Quittner, A. L., Fink, N. E., & Team, Cd. I. (2010). Spoken language development in children following cochlear implantation. *JAMA*, 303(15), 1498–1506.

- Osberger, M. J., Miyamoto, R. T., Zimmerman-Phillips, S., Kemink, J. L., Stroer, B. S., Firszt, J. S., & Novak, M. A. (1991). Independent evaluation of the speech perception abilities of children with the Nucleus 22-channel cochlear implant system. *Ear and Hearing*, 12(4), 66S-80S.
- Peng, S.-C., Chatterjee, M., & Lu, N. (2012). Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends in Amplification*, 16(2), 67–82.
- Peng, S.-C., Tomblin, J. B., Cheung, H., Lin, Y.-S., & Wang, L.-S. (2004). Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. *Ear and Hearing*, 25(3), 251–264.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253–260.
- Pyman, B., Blamey, P., Lacy, P., Clark, G., & Dowell, R. (2000). The development of speech perception in children using cochlear implants: Effects of etiologic factors and delayed milestones. *Otology & Neurotology*, 21(1), 57–61.
- Pyschny, V., Landwehr, M., Hahn, M., Walger, M., von Wedel, H., & Meister, H. (2011). Bimodal hearing and speech perception with a competing talker. *Journal of Speech, Language, and Hearing Research*, 54(5), 1400–1415.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Renfrew, C. E. (1995). *Word Finding Vocabulary Test*. Speechmark Publishing.
- Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, 133(4), 2431–2443.
- Roser, M., Appel, C., & Ritchie, H. (2020). *Human Height*. Our World in Data. <https://ourworldindata.org/human-height>
- Saleh, S. M., Saeed, S. R., Meerton, L., Moore, D. R., & Vickers, D. A. (2013). Clinical use of electrode differentiation to enhance programming of cochlear implants. *Cochlear Implants International*, 14(S4), 16–18.
- Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, 18(7), 559–572.
- Sjoberg, K. M., Driscoll, V. D., Gfeller, K., Welhaven, A. E., Kirk, K. I., & Prusick, L. (2017). The impact of electric hearing on children’s timbre and pitch perception and talker discrimination. *Cochlear Implants International*, 18(1), 36–48.
- Skuk, V. G., Kirchen, L., Oberhoffner, T., Guntinas-Lichius, O., Dobel, C., & Schweinberger, S. R. (2020). Parameter-specific morphing reveals contributions of timbre and fundamental frequency cues to the perception of voice gender and age in cochlear implant users. *Journal of Speech, Language, and Hearing Research*, 63(9), 3155–3175.

- Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language & Hearing Research*, 57(1), 285–296.
- Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, 118(5), 3177–3186.
- Smith, D. R. R., Walters, T. C., & Patterson, R. D. (2007). Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. *The Journal of the Acoustical Society of America*, 122(6), 3628–3639.
- Spencer, P. E. (2004). Individual differences in language performance after cochlear implantation at one to three years of age: Child, family, and linguistic factors. *Journal of Deaf Studies and Deaf Education*, 9(4), 395–412.
- Staller, S. J., Dowell, R. C., Beiter, A. L., & Brimacombe, J. A. (1991). Perceptual abilities of children with the Nucleus 22-channel cochlear implant. *Ear and Hearing*, 12(4), 34S–47S.
- Stickney, G. S., Assmann, P. F., Chang, J., & Zeng, F.-G. (2007). Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *The Journal of the Acoustical Society of America*, 122(2), 1069–1078.
- Stickney, G. S., Zeng, F.-G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America*, 116(2), 1081–1091.
- Stordahl, J. (2002). Song recognition and appraisal: A comparison of children who use cochlear implants and normally hearing children. *Journal of Music Therapy*, 39(1), 2–19.
- Sussman, E., & Steinschneider, M. (2009). Attention effects on auditory scene analysis in children. *Neuropsychologia*, 47(3), 771–785.
- Sussman, E., Wong, R., Horváth, J., Winkler, I., & Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. *Hearing Research*, 225(1), 117–127.
- Svirsky, M. A., Chin, S. B., & Jester, A. (2007). The effects of age at implantation on speech intelligibility in pediatric cochlear implant users: Clinical outcomes and sensitive periods. *Audiological Medicine*, 5(4), 293–306.
- Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., & Miyamoto, R. T. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science*, 11(2), 153–158.
- Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, 85(4), 1699–1707.
- Torppa, R., Faulkner, A., Huottilainen, M., Järvikivi, J., Lipsanen, J., Laasonen, M., & Vainio, M. (2014). The perception of prosody and associated auditory cues in early-implanted children: The role of auditory working memory and musical activities. *International Journal of Audiology*, 53(3), 182–191.

- Visram, A. S., Kluk, K., & McKay, C. M. (2012). Voice gender differences and separation of simultaneous talkers in cochlear implant users with residual hearing. *The Journal of the Acoustical Society of America*, 132(2), EL135–EL141.
- Wei, C.-G., Cao, K., & Zeng, F.-G. (2004). Mandarin tone recognition in cochlear-implant subjects. *Hearing Research*, 197(1–2), 87–95.
- Welch, G. F., Saunders, J., Edwards, S., Palmer, Z., Himonides, E., Knight, J., Mahon, M., Griffin, S., & Vickers, D. A. (2015). Using singing to nurture children’s hearing? A pilot study. *Cochlear Implants International*, 16(S3), 63–70.
- Wouters, J., McDermott, H. J., & Francart, T. (2015). Sound coding in cochlear implants: From electric pulses to hearing. *IEEE Signal Processing Magazine*, 32(2), 67–80.
- Zaltz, Y., Goldsworthy, R. L., Kishon-Rabin, L., & Eisenberg, L. S. (2018). Voice discrimination by adults with cochlear implants: The benefits of early implantation for vocal-tract length perception. *Journal of the Association for Research in Otolaryngology*, 19(2), 193–209.
- Zeng, F.-G. (2002). Temporal pitch in electric hearing. *Hearing Research*, 174(1), 101–106.
- Zeng, F.-G., Tang, Q., & Lu, T. (2014). Abnormal Pitch Perception Produced by Cochlear Implant Stimulation. *PLoS ONE*, 9(2).

5 | Development of vocal emotion categorization in school-age children

Published as:

L. Nagels, E. Gaudrain, D. Vickers, M. Matos Lopes, P. Hendriks, and D. Başkent (2020). Development of vocal emotion recognition in school-age children: The EmoHI test for hearing-impaired populations. *PeerJ*, 8.

Abstract

Traditionally, emotion recognition research has primarily used pictures and videos, while speech materials are not always readily available or are not of good quality, which may be particularly important for studies with hearing-impaired listeners. Here we present a vocal emotion categorization test with pseudospeech productions from multiple speakers expressing three core emotions (happy, angry, and sad): the EmoHI test. The high sound quality recordings make the test suitable for use with populations of children and adults with normal or impaired hearing. Here we present normative data for the development of vocal emotion categorization, the ability to classify speakers' vocal emotions into different categories, in normal-hearing (NH) school-age children using the EmoHI test. Furthermore, we investigated cross-language effects by testing NH Dutch and English children, and the suitability of the EmoHI test for hearing-impaired populations, specifically for prelingually deaf Dutch children with cochlear implants (CIs). Our results show that NH children's performance improved significantly with age from the youngest group tested onwards (4-6 years: 48.9%, on average). However, NH children's performance did not reach adult-like values (adults: 94.1%) even for the oldest age group tested (10-12 years: 81.1%). Additionally, the effect of age on NH children's development did not differ across languages. All except one CI child performed at or above chance-level showing the suitability of the EmoHI test. In addition, seven out of fourteen CI children performed within the NH age-appropriate range, and even nine out of fourteen CI children did so when performance was adjusted for hearing age, measured from their age at CI implantation. However, CI children showed great variability in their performance, ranging from ceiling (97.2%) to below chance-level performance (27.8%), which could not be explained by chronological age alone. The strong and consistent development in performance with age, the lack of significant differences across the tested languages for NH children, and the above-chance performance of most CI children affirm the usability and versatility of the EmoHI test.

5.1 Introduction

Development of emotion recognition in children has been studied extensively using visual stimuli, such as pictures or sketches of facial expressions (e.g., Rodger et al., 2015), or audiovisual materials (e.g., Nelson & Russell, 2011), and particularly in some clinical groups, such as autistic children (e.g., Harms, Martin & Wallace, 2010). However, not much is known about the development of vocal emotion recognition,

even in typically developing children (Sauter et al., 2013; Scherer, 1986). While children can recognize facial and vocal emotions reliably and associate them with external causes already from the age of 5 years on (Pons et al., 2004), it seems to take until late childhood for this ability to develop to adult-like levels (Sauter et al., 2013; Tonks et al., 2007). The recognition of vocal emotions relies heavily on the perception of related vocal acoustic cues, such as mean fundamental frequency (F0) and intensity, as well as fluctuations in these cues, and speaking rate (Scherer, 1986). Based on earlier research on the development of voice perception (Mann et al., 1979; Nittrouer & Miller, 1997), children's performance may be lower compared to adults due to differences in their weighting of acoustic cues and a lack of robust representations of auditory categories for vocal emotions. For instance, Morton and Trehub (2001) showed that when acoustic cues and linguistic content convey contradictory emotions, children mostly rely on linguistic content to judge emotions, whereas adults mostly rely on affective prosody. In addition, children and adults both are better at facial emotion recognition than vocal emotion recognition (Chronaki et al., 2015; Nelson & Russell, 2011). All these observations combined indicate that the formation of robust representations for vocal emotions is highly complex and possibly a long-lasting process even in typically developing children.

Research with hearing-impaired children has shown that they do not perform as well on vocal emotion recognition compared to their NH peers (Chatterjee et al., 2015; Dyck et al., 2004; Hopyan-Misakyan et al., 2009; Nakata et al., 2012). Hopyan-Misakyan (2009) showed that 7-year-old children with CIs performed as well as their NH peers on visual emotion recognition but scored significantly lower on vocal emotion recognition. Visual emotion recognition generally seems to develop faster than vocal emotion recognition (Nelson & Russell, 2011; Nowicki & Duke, 1994), particularly in hearing-impaired children (Hopyan-Misakyan et al., 2009), which may indicate that visual emotions are perceptually more prominent or easier to categorize than vocal emotions. For hearing-impaired children, a higher reliance on visual emotion cues as compensation for spectrotemporally degraded auditory input may be an effective strategy, as emotion recognition in daily life is usually multimodal. However, it may lead to less robust auditory representations of vocal emotions and knowledge about their acoustic properties. Luo et al. (2018) also showed that CI users' ability to recognize vocal emotions was related to their self-reported quality of life, which emphasizes the importance of recognizing vocal emotions in addition to visual emotions. Wiefferink et al. (2013) suggested that reduced auditory exposure and language delays may also lead to delayed social-emotional development and reduced conceptual knowledge about emotions, which

in turn result in a negative impact on emotion recognition. This is also evidenced by CI children's reduced differences in mean F0 cues and F0 variations in emotion production compared to their NH age peers (Chatterjee et al., 2019). Effects of conceptual knowledge on children's discrimination abilities have also been observed earlier, for instance, in research on pitch discrimination (Costa-Giomi & Descombes, 1996). Costa-Giomi and Descombes (1996) found that French children showed better pitch discrimination when they had to use the single meaning terms 'aigu' and 'grave' to denote pitch than the multiple meaning words 'haut' [high] and 'bas' [low], which besides pitch also can be used to denote differences in space and loudness. This finding demonstrates that the label to refer to a concept may affect the labeling process itself. Thus, if CI children do not have clear conceptual knowledge about emotions, this will also similarly affect their ability to label them correctly. Finally, perceptual limitations, such as increased F0 discrimination thresholds (Deroche et al., 2014), may also play a role in CI children's abilities to recognize vocal emotions. Nakata et al. (2012) found that children with CIs especially had difficulties with differentiating happy from angry vocal emotions. This finding suggests that CI children primarily use speaking rate to categorize vocal emotions, as this cue differentiates sad from happy and angry vocal emotions but is similar for the latter two emotions. Therefore, hearing loss also seems to influence the weighting of different acoustic cues, and hence likely also affects the formation of representations of vocal emotions.

Vocal emotion recognition also differs from visual emotion recognition due to the potential influence of linguistic factors. Research regarding cross-language effects on emotion recognition has also demonstrated the importance of auditory exposure for vocal emotion recognition. Most studies have demonstrated a so-called 'native language benefit' showing that listeners are better at recognizing vocal emotions produced by speakers from their own native language than from another language. (Bryant & Barrett, 2008; Scherer et al., 2001; Van Bezooijen et al., 1983). This effect has mainly been attributed to cultural differences (Van Bezooijen et al., 1983), but also effects of language distance have been reported (Scherer et al., 2001), i.e., differences in performance were larger when the linguistic distance (the extent to which the features of two languages differ from each other) between the speakers' and listeners' native languages was larger. Interestingly, Bryant and Barrett (2008) did not find a native language benefit for low-pass filtered vocal emotion stimuli, which filtered out both the linguistic message and the language-specific phonological information. Fleming et al. (2014) also demonstrated a similar native-language benefit for voice recognition based on differences in phonological

familiarity. For CI children, reduced auditory exposure may also lead to reduced phonological familiarity, and could therefore also contribute to their difficulties with the recognition of vocal emotions.

As most research on the development of the general ability to recognize emotions has used visual or audiovisual materials such as pictures or videos, good-quality speech materials are scarce. While the audio quality may only have a small effect on NH listeners' performance, it may be imperative for hearing-impaired listeners' ability to recognize different vocal emotions. Hence, we recorded high sound quality vocal emotion stimuli produced by multiple speakers with three basic emotions (happy, angry, and sad) that are suitable to test hearing-impaired children and adults: the EmoHI test. We aimed to investigate how NH school-age children's ability to categorize vocal emotions develops with age and to obtain normative data for the EmoHI test for future applications, for instance, with clinical populations. In addition, we tested children of two different native languages, namely Dutch and English, to investigate potential cross-language effects, and we collected data from Dutch prelingually deaf children with CIs, to investigate the applicability of the EmoHI test to hearing-impaired children.

5.2 Materials & Methods

5.2.1 Participants

We collected normative data from fifty-eight Dutch and twenty-five English children between 4 and 12 years of age, and fifteen Dutch and fifteen English adults between 20 and 30 years of age with normal hearing. All NH participants were monolingual speakers of Dutch or English and reported no hearing or language disorders. Normal hearing (hearing thresholds at 20 dB HL) was screened with pure-tone audiometry at octave-frequencies between 500 and 4000 Hz. In addition, we collected data from fourteen prelingually deaf Dutch children with CIs between 4 and 16 years of age. The study was approved by the Medical Ethical Review Committee of the University Medical Center Groningen (METc 2016.689). A written informed consent form was signed by adult participants and the parents or legal guardians of children before data collection.

5.2.2 Stimuli and Apparatus

We made recordings of six native Dutch speakers producing two non-language specific pseudospeech sentences using three core emotions (happy, sad, and angry), and a neutral emotion (not used in the current study). All speakers were native monolingual speakers of Dutch and did not have any speech, language, or hearing disorders. Speakers gave written informed consent for the distribution and sharing of the recorded materials. To keep our stimuli relevant to the general emotion recognition literature and suitable for usage across different languages, the pseudospeech sentences that we used, *Koun se mina lod belam* [kʌʊn sə miːnəː lɔd beːlɑm] and *Nekal ibam soud molen* [neːkɑl iːbɑm sɑʊt moːlən], were based on the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus materials by Bänziger et al. (2012). These pseudosentences are meaningful neither in Dutch nor in English, nor any other Indo-European languages. Speakers were instructed to produce the sentences in a happy, sad, angry, or neutral manner using emotional scripts that were also used for the GEMEP corpus stimuli (Scherer & Bänziger, 2010). We chose these three core emotions as previous studies have reported that children first learn to identify happy, angry, and sad emotions, respectively, followed by fear, surprise, and disgust (Widen & Russell, 2003), and hence we could test children from very young ages. The stimuli were recorded in an anechoic room at a sampling rate of 44.1 kHz.

We pre-selected 96 productions, including neutral productions, ($2 \text{ productions} \times 2 \text{ sentences} \times 4 \text{ emotions} \times 6 \text{ speakers}$), and performed a short online survey with Dutch and English adults to confirm that the stimuli were categorized reliably and to select the four speakers whose productions were categorized best. Table 5.1 shows an overview of these four selected speakers' demographic information and voice characteristics. The neutral productions and the productions of the other two speakers were part of the online survey, and are available with the stimulus set, but were not used in the current study to simplify the task for children. Our final set of stimuli consisted of 36 stimuli with three productions (*one sentence repeated once + the other sentence*) per emotion and per speaker ($3 \text{ productions} \times 3 \text{ emotions} \times 4 \text{ speakers}$) as well as 4 practice stimuli, one production per speaker, that were used for the training session.

Table 5.1: Overview of the EmoHI test speakers' voice characteristics.

Speaker	Age (years)	Sex	Height (m)	Mean F0 (Hz)	F0 range (Hz)
T2	36	F	1.68	302.23	200.71 – 437.38
T3	27	M	1.85	166.92	100.99 – 296.47
T5	25	F	1.63	282.89	199.49 – 429.38
T6	24	M	1.75	167.76	87.46 – 285.79

5.2.3 Procedure

NH and CI children were tested in a quiet room at their home, and NH adults were tested in a quiet testing room at the two universities. Since the present experiment was part of a larger project on voice and speech perception (Perception of Indexical Cues in Kids and Adults (PICKA)), data were collected from the same population of children and adults in multiple experiments (Nagels et al., 2020, 2021). The experiment was preceded by a training session consisting of 4 practice stimuli. The experiment consisted of 36 stimuli and the total duration of the experiment was approximately 6 to 8 minutes. All stimuli were presented to participants in a randomized order.

The experiment was conducted on a laptop with a touchscreen using a child-friendly interface that was developed in Matlab (Figure 5.1). The auditory stimuli were presented via Sennheiser HD 380 Pro headphones for NH children and adults, and via Logitech Z200 loudspeakers for CI children. The presentation level of the stimuli was calibrated to a sound level of 65 dBA. CI children were instructed to use the settings they most commonly use in daily life and to keep the settings consistent throughout the experiment. In each trial, participants heard a stimulus and then had to indicate which emotion was conveyed by clicking on one of three clowns with corresponding facial expressions on the screen. We also checked if children were able to correctly identify the emotions of each portrayed clown before the start of the experiment. Visual feedback on the accuracy of responses was provided to motivate participants. Participants saw confetti falling down the screen after a correct response, and the parrot shaking its head after an incorrect response. After every two trials, one of the clowns in the back went one step up the ladder until the experiment was finished to keep children engaged and to give an indication of the progress of the experiment.

Figure 5.1: The experiment interface of the EmoHI test.

5.2.4 Data analysis

NH children's accuracy scores were analyzed using the *lme4* package (Bates et al., 2014) in R. A mixed-effects logistic regression model with a three-way interaction between *language* (Dutch and English), *emotion* (happy, angry, and sad), and *age* in years, rounded to two decimal places, and random intercepts per participant and per stimulus was computed to determine the effects of language, emotion, and age on NH children's ability to categorize vocal emotions. We used backward stepwise selection with ANOVA Chi-Square tests to select the best fitting model, starting with the full factorial model, in *lme4* syntax: `accuracy ~ language * emotion * age + (1 | participant) + (1 | stimulus)`, and deleting one fixed factor at a time based on its significance. In addition, we performed Dunnett's Tests on the NH Dutch and English data with *accuracy* as an outcome variable and *age group* as a predictor variable using the *DescTools* package (Signorell et al., 2018) to investigate at what age NH Dutch and English children show adult-like performance. Finally, we examined the data of CI children to investigate if they could reliably perform the task.

5.3 Results

5.3.1 NH Dutch and English data

Figure 5.2 shows the accuracy scores of NH Dutch (left panel) and English (right panel) participants as a function of their age (dots) and age group (boxplots). The model comparison analysis showed that the full model with random intercepts per participant and per stimulus was significantly better than the full models with only random intercepts per participant ($\chi^2(1) = 393$, $P < 0.001$) or only random intercepts per stimulus ($\chi^2(1) = 51.9$, $P < 0.001$). Backward stepwise selection showed that the best fitting and most parsimonious model was the model with only a fixed effect of *age*, in lme4 syntax: `accuracy ~ age + (1 | participant) + (1 | stimulus)`. This model did not significantly differ from the full model ($\chi^2(10) = 12.90$, $P = 0.23$) or any of the other models while being the most parsimonious. Figure 5.2 shows the data of individual participants and the median accuracy scores per age group for the NH Dutch and English participants. NH children's ability to categorize vocal emotions significantly increased as a function of age ($\beta = 8.91$, $Estimate = 0.30$, $SE = 0.034$, $P < 0.001$). We did not find any significant effects of language or emotion on children's accuracy scores. Finally, the results of the Dunnett's Tests showed that the accuracy scores of Dutch NH children of all tested age groups differed from Dutch NH adults (*4-6 years difference* = -0.47, $P < 0.001$; *6-8 years difference* = -0.31, $P < 0.001$; *8-10 years difference* = -0.19, $P < 0.001$; *10-12 years difference* = -0.15, $P < 0.001$), and the accuracy scores of English NH children of all tested age groups differed from English NH adults (*4-6 years difference* = -0.43, $P < 0.001$; *6-8 years difference* = -0.27, $P < 0.001$; *8-10 years difference* = -0.20, $P < 0.001$; *10-12 years difference* = -0.12, $P < 0.01$). The mean accuracy scores per age group and per participant group are shown in Table 5.2.

Figure 5.2: Vocal emotion categorization in NH children and adults. Accuracy scores of NH Dutch and English children and adults for the EmoHI test per age group and per language. The dotted line shows the chance-level performance of 33.3% correct. The dots show individual data points at participants' age (Netherlands (NL): $N_{\text{children}} = 58$, $N_{\text{adults}} = 15$; United Kingdom (UK): $N_{\text{children}} = 25$, $N_{\text{adults}} = 15$). The boxplots show the median accuracy scores per age group, and the lower and upper quartiles. The whiskers indicate the lowest and highest data points within plus or minus 1.5 times the interquartile range.

Table 5.2: Overview of the mean accuracy scores per participant and age group.

Age groups	Participant groups		
	Dutch NH	English NH	Dutch CI
4-6 years	48.7%	49.3%	34.7%
6-8 years	65.2%	64.8%	48.6%
8-10 years	76.7%	71.8%	37.5%
10-12 years	81.2%	80.6%	57.6%
12-14 years	N/A	N/A	50.0%
14-16 years	N/A	N/A	76.9%
Adults	96.1%	92.0%	N/A

5.3.2 Data of CI children

Figure 5.3 shows the accuracy scores of Dutch CI children as a function of their chronological age (left panel) and hearing age (right panel), i.e., the number of years after children received their CI. The mean accuracy scores per age group are shown in Table 5.2. All except one CI child performed at or above chance-level. Based on Figure 5.3, we can see that seven out of fourteen CI children (50%) performed within the NH age-appropriate range. Considering CI children’s hearing age instead of their chronological age, nine out of fourteen CI children (64.3%) perform within the NH age-appropriate range. However, there is a large variability in CI children’s performance, which varies from ceiling (97.2%) to below chance-level performance (27.8%). The development in CI children’s performance with age does not seem to be as consistent as we found for NH children, which suggests that their performance is not merely due to age-related development.

Figure 5.3: Vocal emotion categorization in Dutch CI children. Accuracy scores of Dutch CI children (N=14) for the EmoHI test per age group. The dotted line shows the chance-level performance of 33.3% correct. The dots show individual data points at Dutch CI children's chronological age (a) and at their hearing age (b). The boxplots show NH Dutch children's median accuracy scores per age group, and the lower and upper quartiles, reproduced from Figure 5.2. The whiskers indicate the lowest and highest data points of NH Dutch children within plus or minus 1.5 times the interquartile range.

5.4 Discussion

5.4.1 Age effects

As shown by our results and the data displayed in Figure 5.2, NH children's ability to categorize vocal emotions improved gradually as a function of age. In addition, we found that, on average, even the oldest age group of 10- to 12-year-old Dutch and English children did not show adult-like performance yet. The 4-year-old NH children that were tested performed at or above chance level while adults generally showed near ceiling performance, indicating that our test covers a wide range of age-related performances. Our results are in line with previous findings that NH children's ability to categorize vocal emotions improves as a function of age (Sauter et al., 2013; Tonks et al., 2007). It may be that children require more auditory experience to form robust representations of vocal emotions or rely on different acoustic cues than adults, as was shown in research on the development of sensitivity to voice cues (Mann et al., 1979; Nitttrouer & Miller, 1997). It is still unclear on which specific acoustic cues children are basing their decisions and how this differs from adults. Future research using machine-learning approaches may be able to further explore such aspects. Finally, the visual feedback may have caused some learning effects, although the correct response was not shown after an error, and learning would pose relatively high demands on auditory working memory since there were only three productions per speaker and emotion presented in a randomized order.

5.4.2 Language effects

Comparing data from NH children from two different native languages, we did not find any cross-language effects between Dutch and English children's development of vocal emotion categorization, even though the materials were produced by Dutch native speakers. Earlier research has demonstrated that although adults are able to categorize vocal emotions across languages, there still seems to be a native language benefit (Bryant & Barrett, 2008; Scherer et al., 2001; Van Bezooijen et al., 1983). Listeners seem better at categorizing vocal emotions that are produced by speakers of their native language than another language. However, it should be noted that five (Bryant & Barrett, 2008; Scherer et al., 2001) and nine (Van Bezooijen et al., 1983) different and more complex emotions were used in these studies which likely poses a considerably more difficult task than differentiating three basic emotions. In addition, the lack of a native language benefit in our results may also be due to the fact that Dutch and English are phonologically closely related

languages. This idea is also in line with the language distance (Scherer et al., 2001) and phonological familiarity effects (Bryant & Barrett, 2008). We are currently collecting data from Turkish children and adults on the EmoHI test to investigate whether there are any detectable cross-language effects for typologically and phonologically more distinct languages.

5.4.3 CI children

The data of CI children show that only one CI child performed below chance-level, indicating that almost all CI children could reliably perform the task and that the task seems sufficiently easy to capture their vocal emotion categorization abilities. In addition, seven out of fourteen CI children performed within the NH age-appropriate range, and if we consider CI children's hearing age instead of their chronological age, nine out of fourteen CI children still fell within that range. Vocal emotion categorization performance was generally lower in CI children compared to NH children and did not seem to follow the same consistent developmental trajectory that we found for NH children. The general lower performance of CI children and the lack of a strong relation between CI children's performance and chronological or hearing age are in line with findings from previous studies (Chatterjee et al., 2015; Hopyan-Misakyan et al., 2009; Nakata et al., 2012). The variability in CI children's performance was large and covered the entire performance range, which also demonstrates that the EmoHI test can capture a wide range of performances. Besides age, CI children's performance could be affected by differences in social-emotional development causing reduced conceptual knowledge on emotions and their properties (Chatterjee et al., 2019; Wiefferink et al., 2013), and differences in their hearing abilities causing perceptual limitations (Nakata et al., 2012). For instance, individual differences in CI children's vocal emotion categorization abilities may also rely on their F0 discrimination thresholds, which are generally higher and more variable in CI children compared to NH children (Deroche et al., 2014). We are currently working on an in-depth analysis of CI children's data, as their performance seems to also be largely related to their hearing abilities (Nakata et al., 2012), a perceptual effect, and social-emotional interaction and development (Wiefferink et al., 2013), a cognitive effect, in addition to age.

5.5 Conclusions

The results of the current study provide baseline normative data for the development of vocal emotion categorization in typically developing, school-age children with normal hearing using the EmoHI test. Our results show that there is a large but relatively slow and consistent development in children's ability to categorize vocal emotions. Furthermore, the data from the CI children show that they seem to be able to carry out the EmoHI test reliably, but the improvement in their performance as a function of age was not as consistent as for NH children. The clear development observed in NH children's performance as a function of age and the generalizability of performance across the tested languages show the EmoHI test's suitability for different ages and potentially also across different languages. Additionally, the above-chance performance of most CI children and the high sound quality stimuli also demonstrate that the EmoHI test is suitable for testing hearing-impaired populations.

Bibliography

- Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion*, 12(5), 1161–1179.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., & Grothendieck, G. (2014). Package ‘lme4’. R Foundation for Statistical Computing, Vienna, 12.
- Bryant, G., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, 8(1–2), 135–148.
- Chatterjee, M., Kulkarni, A. M., Siddiqui, R. M., Christensen, J. A., Hozan, M., Sis, J. L., & Damm, S. A. (2019). Acoustics of emotional prosody produced by prelingually deaf children with cochlear implants. *Frontiers in Psychology*, 10, 1–15.
- Chatterjee, M., Zion, D. J., Deroche, M. L. D., Burianek, B. A., Limb, C. J., Goren, A. P., Kulkarni, A. M., & Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research*, 322, 151–162.
- Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. S. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology*, 33(2), 218–236.
- Costa-Giomi, E., & Descombes, V. (1996). Pitch labels with single and multiple meanings: A study with French-speaking children. *Journal of Research in Music Education*, 44(3), 204–214.
- Deroche, M. L. D., Lu, H.-P., Limb, C. J., Lin, Y.-S., & Chatterjee, M. (2014). Deficits in the pitch sensitivity of cochlear-implanted children speaking English or Mandarin. *Frontiers in Neuroscience*, 8, 1–13.
- Dyck, M. J., Farrugia, C., Shochet, I. M., & Holmes-Brown, M. (2004). Emotion recognition/understanding ability in hearing or vision-impaired children: Do sounds, sights, or words make the difference? *Journal of Child Psychology and Psychiatry*, 45(4), 789–800.
- Fleming, D., Giordano, B. L., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences*, 111(38), 13795–13798.
- Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, 20(3), 290–322.
- Hopyan-Misakyan, T. M., Gordon, K. A., Dennis, M., & Papsin, B. C. (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychology*, 15(2), 136–146.

- Luo, X., Kern, A., & Pulling, K. R. (2018). Vocal emotion recognition performance predicts the quality of life in adult cochlear implant users. *The Journal of the Acoustical Society of America*, 144(5), EL429–EL435.
- Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, 27(1), 153–165.
- Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, 72(3), 834–843.
- Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2020). Development of voice perception is dissociated across gender cues in school-age children. *Scientific Reports*, 10(1), 1–11.
- Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2021). School-age children benefit from voice gender cue differences for the perception of speech in competing speech. *The Journal of the Acoustical Society of America*, 149(5), 3328–3344.
- Nakata, T., Trehub, S. E., & Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *The Journal of the Acoustical Society of America*, 131(2), 1307–1314.
- Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology*, 110(1), 52–61.
- Nitttrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *The Journal of the Acoustical Society of America*, 101(4), 2253–2266.
- Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, 18(1), 9–35.
- Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European journal of developmental psychology*, 1(2), 127–152.
- Rodger, H., Vizioli, L., Ouyang, X., & Caldara, R. (2015). Mapping the development of facial expression recognition. *Developmental Science*, 18(6), 926–939.
- Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, 31(1), 97–113.
- Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99(2), 143–165.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32(1), 76–92.
- Scherer, K. R., & Bänziger, T. (2010). On the use of actor portrayals in research on emotional expression. In K. R. Scherer, T. Bänziger, & E. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 271–294). Oxford University Press.

- Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arppe, A., Baddeley, A., Barton, K., Bolker, B., & Borchers, H. W. (2018). DescTools: Tools for descriptive statistics (R package version 0.99.25) [Computer software].
- Tonks, J., Williams, W. H., Frampton, I., Yates, P., & Slater, A. (2007). Assessing emotion recognition in 9–15-years olds: Preliminary analysis of abilities in reading emotion from faces, voices and eyes. *Brain Injury*, 21(6), 623–629.
- Van Bezooijen, R., Otto, S. A., & Heenan, T. A. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology*, 14(4), 387–406.
- Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, 39(1), 114–128.
- Wiefferink, C. H., Rieffe, C., Ketelaar, L., De Raeve, L., & Frijns, J. H. M. (2013). Emotion understanding in deaf children with a cochlear Implant. *Journal of Deaf Studies and Deaf Education*, 18(2), 175–186.

6 | Discussion

When listeners process speech, they attend not only to the linguistic message but also to the characteristics of the speaker's voice which provide information about the speaker, such as their age, affective state, regional accent, or gender (Scott & McGettigan, 2016). Listeners can use these voice characteristics to recognize and distinguish different speakers, which is particularly important for segregating different speech streams in the presence of multiple speakers (Bregman, 1994; Cherry, 1953). The primary aim of this dissertation was to investigate the development of the perception of voice characteristics, specifically voice gender cues and vocal emotions, in NH and CI children as a function of age, taken as a proxy for cognitive development. Four main research questions were addressed. Firstly, how do the perception of voice gender cues (voice discrimination, voice categorization, and the benefit from target-masker voice differences for perceiving speech in competing speech) and vocal emotions (voice categorization) develop in NH school-age children as a function of age? Secondly, how does NH children's ability to discriminate differences in voice gender cues relate to their perceptual weighting of voice gender cues for categorization and their ability to benefit from target-masker differences in voice gender for perceiving speech in competing speech? Thirdly, how does the development of an auditory system based on the spectrotemporally degraded CI signal affect the development of the perception of voice gender cues and vocal emotions in CI children? Finally, how does CI children's ability to discriminate differences in voice gender cues relate to their perceptual weighting of voice gender cues for categorization and their ability to benefit from voice gender cue differences for the perception of speech in competing speech?

To answer these research questions, we conducted four experimental studies related to the perception of voice characteristics. First, we investigated the development of the ability to discriminate differences in voice gender cues and the perceptual weighting of voice gender cues for categorization in NH children during the school-age years (**Chapter 2**). Secondly, we examined if NH children benefitted from target-masker differences in voice gender cues for perceiving speech in competing speech, and how this benefit developed during the school-age years (**Chapter 3**). Thirdly, we examined how the development of an auditory system based on the spectrotemporally degraded CI signal affects the development of the discrimination and categorization of voice gender cues, and the benefit from voice gender cue differences for perceiving speech in competing speech in CI children (**Chapter 4**). Finally, we investigated the development of the categorization of vocal emotions in NH children during the school-age years, and how the development of

an auditory system based on the spectrotemporally degraded CI signal affected this development in CI children (**Chapter 5**). The current chapter discusses the implications of the results of these experimental studies in light of the four main research questions of this dissertation.

6.1 The development of voice characteristics perception in NH children

The first research question of this dissertation focused on the development of the perception of voice characteristics in NH children during the school-age years. This research question was primarily addressed in Chapters 2, 3, and 5 of this dissertation. Studying the development of the perception of voice characteristics in NH children is a prerequisite for understanding these abilities in CI children, for whom the limitations imposed by the spectrotemporally degraded CI signal not only affect the perception of voice characteristics but also its development. Furthermore, we addressed the second research question by examining how NH children's discrimination of differences in voice gender cues affects their perceptual weighting of voice gender cues for categorization in Chapter 2 and their benefit from voice gender cue differences for perceiving speech in competing speech in Chapter 3.

6.1.1 Development of discrimination and categorization of voice gender cues in NH children

In Chapter 2, we examined the development of the discrimination and categorization of voice gender cues, specifically speakers' mean F0 and their VTL, in NH children during the school-age years. In addition, we studied the relationship between these two abilities, as discrimination is a lower-order perception task that involves the direct comparison of stimuli based on their acoustic features, while categorization is a higher-order cognitive task that requires the comparison of stimuli to stored representations of auditory categories (Gibson, 1969; Pisoni, 1973). NH children may be more sensitive to differences in speakers' mean F0 than their VTL, as F0 is a dynamic cue that varies widely among and within speakers. Listeners are therefore exposed to large variations in F0 relative to VTL, which can only be modified to a very limited extent within the same speaker (Fitch and Giedd, 1999). F0 differences can also help listeners to recognize syllabic structures (Rosen, 1992; Soderstrom et al., 2003), while VTL differences provide information about the phonetic structure of speech segments which are particularly necessary for the

correct interpretation of vowels (Irino and Patterson, 2002; Ladefoged and Broadbent, 1957). Following the Developmental Weighting Shift (DWS) model for the categorical perception of phonemes by Nittrouer and Miller (1997), children may primarily rely on dynamic acoustic cues, such as F0, at first, and learn to increase their weighting of static acoustic cues, such as VTL, gradually via auditory and language experience. The data obtained from the two experiments described in Chapter 2 partly contradicted this hypothesis. We found that the discrimination of VTL differences was adult-like in children around the age of 8, while the discrimination of F0 differences was still not adult-like around the age of 12. Thus, children's discrimination abilities were, in fact, adult-like earlier for VTL than for F0, unlike our prediction based on the DWS model. However, we found the opposite developmental pattern across voice cues for voice gender categorization. Children's perceptual weight attributed to differences in F0 and VTL cues for voice gender categorization was adult-like around the age of 6 for F0 and around the age of 10 for VTL, which was in line with our prediction based on the DWS model. Besides these different developmental patterns for discrimination and categorization, there was only a significant correlation between both abilities for the 4- to 6-year-old age group. Hence, adult-like performance for F0 and VTL is acquired at different rates for the two abilities, and the developmental patterns across voice cues for discrimination and categorization are not closely related. The dissociations in the developmental patterns for voice discrimination and voice categorization across voice cues and abilities imply that these abilities inherently rely on different auditory processes and rely on perceptual cues in a different way.

Our hypothesis that children are more sensitive to differences in speakers' mean F0 than their VTL was partially inspired by the DWS model for the categorical perception of phonemes (Nittrouer and Miller, 1997). This model posits that children initially rely on dynamic acoustic cues, e.g., formant transitions, for phoneme categorization and need to learn to increase their weighting of static acoustic cues, e.g., fricative noise, when these are more informative than dynamic cues. While our results for the categorization experiment seem in line with the DWS model, i.e., earlier acquisition of adult-like weighting of differences in F0 than VTL, we found the opposite developmental pattern for discrimination, i.e., earlier acquisition of adult-like discrimination of VTL cues than F0 cues. However, the DWS model is based on categorization and not discrimination. Hence, although we found that children's discrimination of F0 and VTL cues and weighting of differences in F0 and VTL for voice gender categorization both take a relatively long time to mature, similarly as reported for the discrimination and categorization

of phonemic contrasts (e.g., Hazan & Barrett, 2000; Nittrouer & Miller, 1997), these abilities do not seem to be closely related. An alternative explanation for the earlier acquisition of adult-like discrimination of VTL cues than F0 cues could also be that these voice cues are processed differently based on their acoustic traits. Results from neuroimaging studies also indicate that there may be a discrepancy in the cortical processing of differences in speakers' F0 and VTL (Kreitewolf et al., 2014; Lattner et al., 2005; Von Kriegstein et al., 2010). Kreitewolf et al. (2014) found that Heschl's gyrus, i.e., the transverse temporal gyrus, is mainly involved in the processing of changes in speakers' F0, while the superior temporal gyrus and the superior temporal sulcus were mainly activated when changes in speakers' VTL occurred. Differences in the maturation of these cortical areas could also contribute to the observed discrepancy that we found in the acquisition of adult-like discrimination and sensitivity to differences in F0 and VTL cues in the current study.

Interestingly, Mirman et al. (2004) found a similar trade-off effect between discrimination and categorization based on the acoustic dynamics of phonemic cues in NH adult listeners. They concluded that dynamic acoustic cues may be more perceptually salient than static cues, which makes listeners rely on them more for categorization. However, due to their rapidly changing and dynamic nature, subtle differences in dynamic cues may be more difficult to perceive for listeners than static cues. Neural dissociations have also been reported between the areas that are involved in different auditory perception tasks, which seem to be related to the cognitive and memory demands of the task. For instance, Johnsrude et al. (2000) found that patients with right temporal lobe excisions that included Heschl's gyrus showed similar pitch discrimination thresholds as normal control subjects, but they performed poorer at determining the direction of pitch changes, i.e., increasing or decreasing. These results show that there is a discrepancy in the cortical areas that are involved in simple sensory pitch discrimination and those involved in high-order pitch perception tasks that require sequential analysis and the perception of relative pitch, i.e., maintaining pitch information in working memory. Zatorre et al. (1994) similarly found differences between discrimination of two successive pure tones and discrimination of the first and last pure tone in a sequence of pure tones. Although both tasks require listeners to maintain pitch information in working memory, the demands to maintain this information in memory are larger in the latter task, which increased cortical activity. The underlying mechanisms for the processing of F0 and VTL cues and the involved neural substrates therefore seem to depend on voice cues and task demands, which is also in line with the different developmental patterns across abilities and voice cues that we observed.

An alternative explanation for the differences in the perceptual weighting of F0 and VTL cues for voice gender categorization between children and adults is that children have different voice gender categories. As there are no apparent differences in the F0 and VTL parameters between male and female children's voices before puberty (Perry et al., 2001; Vorperian and Kent, 2007), children may attend to other acoustic cues that are more informative for distinguishing the voice gender of child voices, such as differences in speaking rate or prosody. If the voice gender categories that children have are not so much based on differences in F0 and VTL, this may also explain why they are less sensitive to differences in these voice cues, as auditory categories can affect listeners' sensitivity to variability within and between categories (Goldstone, 1994; Kuhl, 1991).

6.1.2 Development of a benefit from voice gender cue differences for perceiving speech in competing speech in NH children

In Chapter 3, we extended the study described in Chapter 2 by again looking at the perception of F0 and VTL cues, but now to see how the benefit from target-masker differences in F0 and VTL cues for perceiving speech in competing speech developed in NH children during the school-age years. We also examined how children's benefit from target-masker differences in F0 and VTL was related to their F0 and VTL discrimination thresholds that were presented in Chapter 2. Previous research has shown that it takes children considerably longer developmentally to reach an adult-like performance level for the perception of speech in competing speech than in competing non-speech noise (Bonino et al., 2013; Corbin et al., 2016; Hall et al., 2002; Leibold and Buss, 2013). This discrepancy seems to be related to their developing cognitive abilities, such as attention and inhibition, but also their developing voice discrimination abilities may play a role (Buss et al., 2017a; Flaherty et al., 2019; Nagels et al., 2020). Based on previous research, children seem to benefit similarly to adults when the masker speech is produced by speakers of a different sex than the target speaker (Leibold et al., 2018; Newman and Morini, 2017; Wightman and Kistler, 2005), but they do not seem to benefit from subtle differences in mean F0 only or VTL only between the target and the masker (Flaherty et al., 2019, 2021), unlike adults (Başkent and Gaudrain, 2016; Darwin et al., 2003). The results from our speech in competing speech perception experiment demonstrated that children benefitted from target-masker differences in F0, VTL, or both voice cues at all tested ages. Although the overall accuracy scores of children continued to differ from those of adults, their benefit from voice cue differences remained proportionally the same across age once we interpolated their accuracy

scores to the same performance level. The benefit from target-masker differences in F0 and VTL and the overall accuracy scores of children were not related to their F0 and VTL discrimination thresholds. Hence, the benefit from voice differences for perceiving speech in competing speech does not seem to be limited by the poorer voice discrimination thresholds of children compared to those of adults. Nevertheless, children's overall speech in competing speech perception abilities remained poorer compared to those of adults at all tested ages. The continued development of children's overall speech in competing speech perception abilities seems to be primarily related to other cognitive-developmental factors than those involved in voice discrimination, such as those related to selective attention mechanisms.

The lack of age effects on children's benefit from voice differences when participants' accuracy scores were interpolated, and the continued development of their overall speech in competing speech perception abilities suggest that factors other than their not fully developed voice discrimination abilities are more closely related. Nevertheless, it is important to consider the size of the voice differences that were used with respect to the voice discrimination thresholds of children. Similar to the voice gender categorization experiment in Chapter 2, the target-masker differences in F0 and VTL that were used in the current study were above the voice discrimination thresholds of most children. The proportionally large voice differences may have prevented a closer relationship between children's benefit from F0 and VTL differences and their respective discrimination thresholds. However, Flaherty et al. (2019) did not find a significant correlation between children's F0 discrimination thresholds and their benefit from smaller target-masker differences of 3 st, 6 st, or 9 st in F0 either. Research on pitch discrimination has also reported a discrepancy between children's pitch discrimination abilities and their ability to use pitch differences to segregate two streams of pure tones (Sussman et al., 2007; Sussman and Steinschneider, 2009). Hence, it is still unclear if voice discrimination thresholds affect children's benefit if more subtle voice differences were to be used that are closer to the discrimination thresholds of children.

Our results are partially in line with those of Flaherty et al. (2019), who also did not find a correlation between children's benefit from target-masker differences in F0 and their F0 discrimination thresholds. However, our observations for children's benefit from target-masker voice differences contradict the findings of Flaherty et al. (2019, 2021). While we found a clear benefit from target-masker differences in F0 or VTL for children at all tested ages, Flaherty et al. (2019, 2021) found that young children only benefitted from combined differences in F0 and VTL, although

this benefit was still lower compared to older children and adults. However, there are some important differences in the experiment design of our study and that of Flaherty et al. (2019, 2021) that complicate a direct comparison of results. While we used sentences with a carrier phrase and sentence chunks produced by a single-talker masker, Flaherty et al. (2019, 2021) used disyllabic words and sentences produced by a two-talker masker. The target-masker differences in F0 and VTL were also much larger in the current study compared to those used by Flaherty et al. (2019, 2021). The discrepancy between the results of the current study and those of Flaherty et al. (2019, 2021) concerning a benefit from voice differences in school-age children for perceiving speech in competing speech mainly seems to be caused by these differences in the experiment design of our studies. Particularly the use of target sentences with the same consistent carrier phrase in the current study may have made it easier for children to use the voice differences for speech stream segregation and to keep track of the target speech stream (Bonino et al., 2013; Freyman et al., 2004; Sussman-Fort and Sussman, 2014). Using a single-talker speech masker may have also provided more opportunity for “glimpsing” acoustic information because of temporary reductions in masking (Buss et al., 2017b; Litovsky, 2005; Rosen et al., 2013), although children may not be able to use such low-level glimpses as efficiently as adults (Buss et al., 2017b). Future research is required to determine the exact contribution of these different design parameters to children’s benefit from voice differences for perceiving speech in competing speech.

6.1.3 Development of categorization of vocal emotions in NH children

In Chapter 5, we extended the results from the earlier studies on the perception of voice gender cues by studying the perception of other voice characteristics, namely vocal emotions. For NH children, we hypothesized that their ability to categorize vocal emotions would continue to develop during the school-age years (e.g., Sauter et al., 2013; Tonks et al., 2007), similar as for voice gender categorization in Chapter 2. Previous research has mainly been focused on the perception of visual emotion cues using pictures or video materials (e.g., Nelson & Russell, 2011; Rodger et al., 2015). However, both children and adults seem to perform better at categorizing visual emotions compared to vocal emotions (Chronaki et al., 2015; Nelson and Russell, 2011). Some studies have also reported a so-called *native language benefit* for the recognition of vocal emotions, indicating that listeners show better recognition of vocal emotions that are produced by speakers from their native language instead of speakers from another language (Bryant and Barrett, 2008; Scherer et al., 2001;

Van Bezooijen et al., 1983). These effects could be attributed to cultural differences (Van Bezooijen et al., 1983). Furthermore, the genetic relationships between languages and the extent to which the phonological systems of languages differ from each other seem to affect the size of the differences in performance (Bryant and Barrett, 2008; Scherer et al., 2001). For the current study, we developed a new vocal emotion test with high sound quality recordings to make the test suitable to use with hearing-impaired children and adults: the EmoHI test. The materials consisted of pseudospeech sentence productions that were produced by multiple Dutch native speakers conveying three basic emotions (happy, angry, and sad) to also enable the testing of listeners across different languages (see Chapter 5 for a more elaborate description). Our results for NH Dutch and English children demonstrated that their abilities to categorize the three tested vocal emotions correctly gradually improved as a function of age and did not differ across the tested languages. In addition, we found that even the oldest tested age group of 10- to 12-year-old children did not show adult-like performance yet on average. While the 4-year-old children performed at or slightly above chance, adults showed near ceiling performance, indicating that our newly developed test, the EmoHI test, covered a wide range of performance levels.

The gradual improvement in NH children's performance during the school-age years is in agreement with the findings of earlier studies on vocal emotion recognition (e.g., Sauter et al., 2013; Tonks et al., 2007). A potential explanation for this observed development is that, similar to voice gender categorization in Chapter 2, children need to gain more auditory and language experience to form robust auditory categories and learn which voice cues they should rely on for which emotion. For instance, children may have to learn which F0 contour pattern is associated with each emotion, in the same way that they need to learn which acoustic features can be used to recognize different phonemes (e.g., Hazan & Barrett, 2000; Nitttrouer & Miller, 1997). The relatively extensive development in children's ability to categorize vocal emotions during the school-age years may be caused partially by the multimodality of emotions. Listeners can reliably recognize a person's emotions based on visual or auditory cues only, although performance considerably increases when both types of cues are available (e.g., Bänziger et al., 2012; de Boer et al., 2021). However, both children and adults seem to perform better at categorizing emotions based on visual cues than auditory cues (Chronaki et al., 2015; Nelson and Russell, 2011). As visual emotion categorization seems to be acquired earlier and seems to be relatively easier, children may rely primarily on visual emotion cues early in life, and hence require some time to form robust

categories of vocal emotions based on subtle acoustic cues. Compared to the voice gender categorization task in Chapter 2, which required listeners to categorize voices into two categories based on differences across two acoustic cues, the vocal emotion recognition task in Chapter 5 was also considerably more complex. Not only were there more voice categories, but there were also substantially more acoustic cues that contributed to the vocal emotion that was produced, e.g., differences in F0 contour, intensity, or speaking rate. Integrating all available acoustic cues to categorize vocal emotions correctly may also partially explain why even the oldest tested children did not show adult-like performance yet.

Furthermore, we did not find any significant differences between the performance of Dutch and English NH children and adults. However, the previous studies that mentioned such differences across languages used tasks that were considerably more complex using five or nine different emotions (Bryant and Barrett, 2008; Scherer et al., 2001). We may also not have found any cross-language effects because Dutch and English are both Germanic languages, which are also phonologically relatively closely related. Hence, cross-language effects may be present when a more complex task is used or when the phonological systems of the tested languages are more different (Van Bezooijen et al., 1983). A native language benefit has also been reported for voice discrimination in NH children and adults (Fleming et al., 2014; Johnson et al., 2011). Listeners are better at discriminating the voices of speakers in their native language than in another language. According to work by Johnson et al. (2011) with seven-month-old infants, this ‘native language benefit’ is mainly a consequence of phonological familiarity with one’s native language, which develops as a function of language experience. Instead of applying specifically to voice discrimination, the effects of auditory and language experience, i.e., the earlier mentioned phonological familiarity effects, may apply to the perception of voice cues in general and explain the observed cross-language effects in the categorization of vocal emotions in previous studies (Bryant and Barrett, 2008; Scherer et al., 2001; Van Bezooijen et al., 1983). Supporting this idea, Bryant and Barrett (2008) did not find a native language benefit for low-pass filtered vocal emotion stimuli, which filtered out the language-specific phonological information in productions. The extent to which such phenomena as a native language benefit or phonological familiarity effects apply to voice perception abilities in general should be explored in future studies.

6.1.4 Summary of findings for NH children

The four experimental studies that we conducted with NH children showed that their voice perception abilities continue to develop during the school-age years, even though their peripheral auditory system is already almost fully developed at birth (Litovsky, 2015; Werner, 2007). Even for the discrimination of voice cues, a lower-order perception task, a clear development was observed during the school-age years. Auditory peripheral factors do not seem to explain the relatively long development of voice perception abilities in NH children. Instead, cognitive-developmental factors, such as immature processing efficiency, or different processing strategies and auditory representations than adults (Hazan & Barrett, 2000; Nitttrouer & Miller, 1997; Werner & Leibold, 2011) seem to play a large role. This idea is also supported by the observed dissociations in developmental patterns across abilities and voice cues in Chapter 2, which are most likely explained by differences in the cognitive demands of the tasks and potentially also the cortical areas that are involved (Johnsrude et al., 2000; Zatorre et al., 1994). Compared to voice discrimination, voice categorization is a higher-order cognitive task that relies on the retrieval of stored representations of auditory categories from long-term memory (Gibson, 1969; Pisoni, 1973). Perceiving speech in competing speech also relies heavily on higher-order cognitive abilities, such as selective auditory attention and inhibition of the interference caused by masking (Arbogast et al., 2002; Brungart et al., 2001, 2006; Cooke et al., 2008; Evans et al., 2016; Mattys et al., 2009; Ruggles et al., 2011; Scott et al., 2004; Swaminathan et al., 2015). The different voice perception abilities of NH children that were tested in Chapters 2, 3, and 5 all seem to develop during the school-age years. However, the not fully developed voice discrimination thresholds of NH children were not closely related to their categorization abilities or their ability to perceive speech in competing speech. In general, our results imply that the processing of voice cues like speech cues (e.g., Hazan & Barrett, 2000; Nitttrouer & Miller, 1997), seems to develop towards adult-like performance levels as children gain more auditory and language experience. Children seem to require time and experience to learn to which acoustic cues they should attend and to develop the same cognitive abilities to use voice differences in a similarly effective manner as adults.

6.2 The development of voice characteristics perception in CI children

The third research question of this dissertation was focused on the development of the perception of voice characteristics in prelingually deaf CI children. This research question was primarily addressed in Chapters 4 and 5 of this dissertation, in which the results of CI children for the four experimental studies are discussed. As most research on the perception of voice characteristics in CI users has tested postlingually deaf CI adults, not much is known yet about developmental effects. As discussed in Chapters 2, 3, and 5, we found considerably large developmental effects on NH children's perception of voice gender cues and vocal emotions during the school-age years. Developmental effects may also contribute similar to the perception of voice characteristics in CI children, but the development in CI children may also differ in some aspects from NH children due to the development of an auditory system based on the spectrotemporally degraded CI speech signal. Furthermore, the fourth research question was addressed by investigating if CI children's discrimination of differences in voice gender cues was related to their perceptual weighting of voice gender cues and their benefit from voice gender cue differences for perceiving speech in competing speech.

6.2.1 Development of discrimination and categorization of voice gender cues and a benefit from voice gender cue differences for perceiving speech in competing speech in CI children

In Chapter 4, we investigated the perception of voice gender cues, specifically speakers' mean F0 and VTL, in CI children. We examined how their abilities to discriminate and categorize voice gender cues, and their benefit from target-masker voice differences for perceiving speech in competing speech developed as a function of age. We also studied the relationship between these different voice perception abilities. Previous research has shown that the ability to discriminate differences in speakers' mean F0 (El Boghdady et al., 2019; Gaudrain & Başkent, 2018; Geurts & Wouters, 2001; Green et al., 2004, 2005) and their VTL (El Boghdady et al., 2019; Gaudrain & Başkent, 2018) is poorer in postlingually deaf CI adults compared to NH listeners. The higher F0 and VTL discrimination thresholds of postlingually deaf CI adults also seem to lead to a deviant weighting of differences in F0 and particularly VTL for voice gender categorization (Fuller et al., 2014; Meister et al., 2016; Skuk et al., 2020). In addition, CI adults' higher voice discrimination thresholds have been hypothesized to relate to their difficulties with

perceiving speech in competing speech (Auinger et al., 2017; El Boghdady et al., 2019; Meister et al., 2020; Pyschny et al., 2011; Stickney et al., 2007). Similar to postlingually deaf CI adults, prelingually deaf CI children may also have less access to differences in F0 and VTL due to the spectrotemporally degraded CI speech signal. However, unlike postlingually deaf CI adults, prelingually deaf CI children acquired language primarily via the spectrotemporally degraded CI speech signal, and hence developed language differently. CI children also show higher brain plasticity compared to CI adults (Manrique et al., 1999), which may make them better at learning to capitalize on reduced F0 and VTL cues that are available in the CI signal than CI adults. The effects of developing an auditory system based on the spectrotemporally degraded CI speech signal and potential perceptual limitations on the voice perception abilities in CI children remain unclear. Our results show that the F0 and VTL discrimination thresholds of CI children were overall higher, i.e., their discrimination abilities were poorer, than those of their NH age peers, but their discrimination abilities improved as a function of chronological age. The F0 and VTL discrimination thresholds of CI children were also better on average than those reported in postlingually deaf CI adults (El Boghdady et al., 2019; Gaudrain & Başkent, 2018). Furthermore, the developmental pattern across voice cues, i.e., earlier acquisition of adult-like discrimination of VTL cues than F0 cues, did not significantly differ from that observed in NH children. Even though the F0 and VTL discrimination thresholds of CI children were generally higher than those of NH children, this did not lead to a deviant weighting of differences in F0 and VTL for voice gender categorization. Similar to NH children, the perceptual weight that CI children attributed to F0 and VTL differences also became more adult-like as a function of chronological age. Nevertheless, CI children with high F0 discrimination thresholds attributed significantly less perceptual weight to differences in F0 for voice gender categorization. Finally, the ability to perceive speech in competing speech was overall lower in CI children compared to their NH age peers, but their performance improved as a function of age and they benefitted to a similar degree from target-masker differences in F0 and VTL. To summarize, although the perceptual abilities of CI children were overall poorer than those of their NH age peers, we found clear developmental effects on their performance related to their chronological age in all three experiments. The observed developmental patterns across voice cues for all three tested abilities (voice discrimination, voice categorization, and the benefit from target-masker voice differences for perceiving speech in competing speech) were also more similar to those observed in NH children than those reported for postlingually deaf CI adults

(Fuller et al., 2014; Gaudrain & Başkent, 2018; Meister et al., 2016; Skuk et al., 2020). Hence, these results imply that high brain plasticity in CI children may lead to more effective use of reduced acoustic cues that are available in the spectrotemporally degraded CI signal compared to CI adults.

Although voice discrimination and speech in competing speech perception were generally poorer in CI children, this did not result in a deviant weighting of F0 and VTL cues for voice gender categorization or prevent a benefit from target-masker voice differences in F0 and VTL. Moreover, the perception of voice gender cues also improved in CI children as a function of chronological age. The developmental effects on F0 and VTL discrimination in CI children are in line with earlier studies (speaker discrimination using parallel changes in F0 and VTL: Cleary et al., 2005; pitch discrimination using harmonic complexes: Deroche et al., 2014; pitch discrimination using pulse trains: Kopelovich et al., 2010). However, as is reported in these studies, chronological age could not account for the large variability in the discrimination abilities of CI children. Deroche et al. (2014) reported that chronological age could only explain around 10% of the variability in the F0 discrimination thresholds of CI children. Some other factors that could explain the large variability among the performance of CI children are related to the used CI processing strategy (Wouters et al., 2015), the underlying etiology (Pyman et al., 2000), the degree of auditory nerve survival, differences in the placement of the electrode array, or the adaptation to mismatches in the nerve-electrode array (Gordon et al., 2011). All CI children reported in this dissertation had devices from Cochlear and were using the ACE strategy or the MP3000 strategy, which is based on the ACE strategy (Wouters et al., 2015). Hence, the contribution of differences in CI processing strategy is expected to be limited in the current study, but these differences may be important for comparing these results to those of postlingually deaf CI adults, who use a wider range of devices and processing strategies (El Boghdady et al., 2019; Gaudrain & Başkent, 2018). Pyman et al. (2000) also only found a higher incidence of motoric and cognitive delays for children who had the cytomegalovirus (CMV) as the underlying etiology for deafness. The degree of auditory nerve survival, electrode placement, and adaptation to mismatches in the nerve-electrode array are difficult to quantify in CI children, apart from the fact that these factors partly depend on the underlying etiology of deafness and the duration of the period of auditory deprivation (Gordon et al., 2011). Environmental factors, such as CI children's daily CI use, socioeconomic status, and parental involvement in the rehabilitation process (Boons et al., 2012; Busch et al., 2020; Geers et al., 2003; Niparko et al., 2010; Spencer, 2004) may also play a role in explaining the

observed variability. These factors should be kept in mind for future research with larger sample sizes and more heterogeneous groups of CI children than in the current study.

There are some clear discrepancies between the results from the current study with prelingually deaf CI children and earlier work from our research group with postlingually deaf CI adults using experiments with similar methods (El Boghdady et al., 2019; Fuller et al., 2014; Gaudrain & Başkent, 2018). First, the discrimination thresholds of CI children were better on average than those of CI adults (El Boghdady et al., 2019; Gaudrain & Başkent, 2018). Secondly, CI children attributed perceptual weight to differences in both F0 and VTL for voice gender categorization, unlike CI adults, who mainly seem to attribute weight to F0 differences but not so much to VTL differences for voice gender categorization for similar word stimuli (Fuller et al., 2014; Meister et al., 2016; Skuk et al., 2020). CI children's weighting of F0 and VTL cues also did not differ from those of their NH age peers. CI children may have better access to F0 and VTL cues on average or make better use of the acoustic cues that are available via the spectrotemporally degraded CI signal than CI adults. Some evidence for the fact that CI adults may not optimally use acoustic cues is provided by research on vowel recognition in CI adults and NH adults that were exposed to noise-vocoded CI-simulated speech (Iverson et al., 2006; McGettigan et al., 2014). Both CI adults and NH adults made suboptimal use of the acoustic cues that are relatively well preserved in the noise-vocoded and the spectrotemporally degraded CI signals for vowel recognition, such as vowel duration and voicing cues. These results indicate that CI adults may have difficulties with adapting their already existing acoustic representations and adjusting them to the spectrotemporally degraded CI signal. Furthermore, CI children may also have formed more robust representations based on the reduced acoustic cues that are related to speakers' voice gender than CI adults, as they acquired language primarily via the spectrotemporally degraded CI signal. This explanation also partly fits with the finding of Meister et al. (2016) that CI adults could use VTL cues for voice gender categorization when sentence stimuli instead of word stimuli were used. Thus, CI adults may after all be able to use VTL cues for voice gender categorization to some extent but could need more processing time and acoustic information to do so, while word stimuli may already be sufficient for CI children to extract this information. Future research is required to explore the potential effects of training and auditory rehabilitation on the voice perception abilities of CI users. Based on our results with CI children, these acoustic cues seem to be available in the spectrotemporally degraded CI speech signal. However,

training effects could be limited in postlingually deaf CI users, as most of them are older adults and have lower brain plasticity than young CI children (Manrique et al., 1999), which could impede the adaptation of acoustic representations.

6.2.2 Development of categorization of vocal emotions in CI children

In Chapter 5, besides the categorization of voice gender cues, we also investigated the categorization of vocal emotions in CI children. As discussed earlier in this chapter, the ability to categorize vocal emotions develops during the school-age years in NH children (e.g., Sauter et al., 2013; Tonks et al., 2007). This development mainly seems to be related to the fact that children require time to learn which acoustic cues are associated with which vocal emotion categories. For CI children, the ability to recognize vocal emotions is poorer on average compared to their NH age peers (Chatterjee et al., 2015; Hopyan-Misakyan et al., 2009; Nakata et al., 2012). Their performance may be poorer because of reduced auditory and language exposure, which can also lead to delayed social-emotional development (Wiefferink et al., 2013) and potentially reduced conceptual knowledge of vocal emotions (Chatterjee et al., 2019). Furthermore, perceptual limitations imposed by the CI likely play a role by reducing their perception of, for instance, F0 contours (Deroche et al., 2016, 2019). Our results demonstrated that although the ability to categorize vocal emotions was poorer in CI children compared to their NH age peers, their abilities improved as a function of chronological age. However, similar to their ability to categorize speakers' voice gender described in Chapter 4, the development of CI children's ability to categorize vocal emotions was less consistent than in NH children. All except one CI child performed above chance-level (27.8%) and 50% of CI children performed within the NH age-appropriate range. When we considered CI children's hearing age instead of their chronological age, even 64.3% had scores within the NH age-appropriate range, although most of these scores were still at the lower ends of this range. To summarize, the vocal emotion categorization abilities of CI children were overall poorer than those of their NH age peers, yet their abilities did not plateau and improved as a function of age.

The less consistent development of the vocal emotion categorization abilities of CI children as a function of age is in line with reports from previous studies (Chatterjee et al., 2015; Hopyan-Misakyan et al., 2009; Nakata et al., 2012). The results suggest that, as described for CI children's ability to categorize speakers' voice gender in Chapter 4, other factors besides regular cognitive development seem to play an important role, such as their CI processing strategy or their adaptation to mismatches in the nerve-electrode array. However, apart from factors related to CI

children's hearing abilities (Nakata et al., 2012), an additional factor may play a large role in vocal emotion categorization, namely their social-emotional development (Wiefferink et al., 2013). Wiefferink et al. (2013) found that preschool children with CIs between 2.5 and 5 years of age performed poorer on many tasks testing emotion understanding compared to their NH age peers, including nonverbal emotion understanding tasks. These findings suggest that reduced auditory exposure may also negatively impact CI children's social-emotional development and their conceptual knowledge of emotions, although this relationship is likely bidirectional. Some evidence for poorer conceptual knowledge of emotions in CI children is provided by the study of Chatterjee et al. (2019) on the production of vocal emotions in CI children. The vocal emotion productions of CI children were more monotonous and varied less in F0 and mean F0 than those produced by NH children and adults. The lack of clear distinctions between different emotions in CI children's productions suggests that they also may have less clear representations of vocal emotions. CI children may also attend more to other acoustic cues, such as differences in speaking rate (Nakata et al., 2012), which are represented better in the spectrotemporally degraded CI signal.

6.2.3 Summary of findings for CI children

The results of CI children demonstrated that the development of an auditory system based on the spectrotemporally degraded CI signal mainly affects the discrimination of differences in voice characteristics. While the F0 and VTL discrimination thresholds of CI children were generally poorer than those of their NH age peers, most CI children showed a similar weighting of differences in F0 and VTL for voice gender categorization. CI children were also able to benefit from target-masker differences in F0 and VTL for perceiving speech in competing speech. CI children still had difficulties with the categorization of vocal emotions and the perception of speech in competing speech, but also these abilities did not seem to plateau and improve during the school-age years. In general, these findings show that the voice perception abilities of CI children develop to a certain extent during the school-age years, which was also observed in NH children. Furthermore, our results suggest that CI children are better able to capitalize on the reduced F0 and VTL cues that are available in the spectrotemporally degraded CI speech signal than postlingually deaf CI adults. These observations indicate that F0 and VTL cues are represented in the CI signal to a certain degree, and CI children are able to acquire the ability to effectively use the reduced F0 and VTL cues for voice gender categorization and to better perceive speech in competing speech, unlike most CI adults.

6.3 Future perspectives

In Chapters 2, 3, and 4 of this dissertation, we used the DWS model for phoneme categorization as a framework to explain the development of the perceptual weighting of acoustic cues in children during the school-age years. We made an analogy between the dynamic formant transitions and the static fricative noise cues that are used in this model and the dynamic F0 and the static VTL cues that were used in the current study. As explained in Chapter 1 of this dissertation, this distinction was mainly based on the informativeness of these cues at a linguistic level. Dynamic cues primarily seem to help with the parsing of speech into syllables, similar to F0 differences, while static cues are more informative for phonetic structures, similar to VTL differences. Besides these functions, speakers' mean F0 and VTL primarily define a speakers' voice gender (Skuk & Schweinberger, 2014; Smith & Patterson, 2005; Titze, 1989), and differences in both types of acoustic cues result in spectral changes. However, there are many spectrotemporal acoustic cues, such as durational cues that signal other phonemic contrasts (Hazan & Barrett, 2000), but also those related to vocal emotions as described in Chapter 5, for which similar distinctions could be made. For instance, CI children's ability to categorize vocal emotions may be related to their ability to discriminate differences in speaking rate or F0 contours, as was suggested by Nakata et al. (2012). We took some first steps here to show that the same perceptual mechanisms that have been found in the literature to affect the development of speech perception, particularly phoneme categorization, could also apply to voice perception. Additional research is required to investigate how domain-general or domain-specific these auditory perceptual mechanisms of voice and speech perception are. Related to this, there may also be parallels between the development of vocal emotion categorization and the development of similar abilities in other modalities, such as the development of visual emotion categorization, as was suggested in Chapter 5.

In Chapters 3 and 4 of this dissertation, we also investigated if NH and CI children could benefit from target-masker differences in F0 and VTL for perceiving speech in competing speech. Based on earlier results by Flaherty et al. (2019, 2021) that showed that young children did not benefit from target-masker voice differences in F0 or VTL, and considering NH and CI children's elevated F0 and VTL discrimination thresholds, we expected this may not be the case. Nevertheless, we found a significant benefit from differences in F0 and VTL for both groups of children and no significant correlation with their discrimination thresholds. The target-masker differences of -12 st in F0 and 3.8 st in VTL in the current study were, however, well above most children's discrimination thresholds. A closer

relationship with their discrimination threshold could be present when more subtle voice differences that are nearer to children's discrimination thresholds are used.

Finally, in the experimental studies that are presented in this dissertation, we used chronological age and hearing age as a proxy for cognitive development. Although chronological age proved to be an adequate predictor of the tested voice perception abilities in NH children, future research collecting longitudinal data should investigate to what degree these effects are attributable to age. Using cognitive tasks, which test abilities such as selective attention, inhibition, and working memory, could also help to explain the individual differences in the perceptual abilities of NH and CI children (Kronenberger et al., 2013). Particularly for CI children, for whom the development as a function of age was much less consistent than for NH children, it is fundamental to further study these effects. These results could provide more information about what performance level can be expected of individual CI children and give a better indication of whether additional speech and language therapy could be beneficial. Also, the potential effects of auditory training and rehabilitation on voice perception could be investigated. This information may provide a clearer explanation of why most prelingually deaf CI children performed better than the postlingually deaf CI adults in earlier studies that used similar experiment designs, and whether auditory training could be beneficial. Future studies could also explore the effects of brain plasticity in NH and CI children, for instance by exposing NH children to noise-vocoded CI-simulated speech and measuring if they make more optimal use of the available acoustic cues than NH adults (Iverson et al., 2006; McGettigan et al., 2014).

6.4 Conclusions

This dissertation investigated how the perception of voice gender cues and vocal emotions develops in NH children as a function of age during the school-age years. We also investigated how NH children's discrimination of differences in voice gender cues affects other voice perception abilities by using multiple perceptual tasks. Furthermore, we studied how the acquisition of an auditory system based on the spectrotemporally degraded CI signal affects the development of these perceptual abilities in CI children. Finally, we examined how CI children's ability to discriminate differences in voice gender cues affects other voice perception abilities. Combined, the results of these experimental studies show strong developmental effects related to age for NH and CI children on all tested abilities, namely (1) Discrimination of differences in voice gender cues; (2) Perceptual weighting of

voice gender cues for categorization; (3) Benefit from voice gender cue differences for perceiving speech in competing speech; and (4) Categorization of vocal emotions. These findings not only have implications for theories of voice perception but also for those of speech perception and language processing in general, for which voice-related indexical cues often receive little attention.

For NH children, the development of their voice perception abilities during the school-age years seems to be primarily related to cognitive-developmental factors, which improve as more auditory and language experience is gained. For CI children, we found that their voice and speech perception abilities were overall poorer than those of their NH age peers, but their performance also improved with age and was even similar in some aspects, i.e., voice gender categorization and benefit from target-masker voice differences in F0 and VTL. The development in the voice and speech perception abilities of CI children as a function of age was less consistent than for NH children. This less consistent development seems partially caused by perceptual limitations, i.e., CI children with higher F0 discrimination thresholds showed a lower weighting of differences in F0 for voice gender categorization. However, the large individual differences among CI children also give some indication that cognitive-developmental factors, such as immature processing efficiency, or different processing strategies and auditory representations than adults (Hazan and Barrett, 2000; Nitttrouer and Miller, 1997; Werner and Leibold, 2011), play a large role. Compared to postlingually deaf CI adults, the results from the current study suggest that prelingually deaf CI children may use the spectrotemporally degraded voice and speech cues provided by the CI signal more effectively. The early exposure to spectrotemporally degraded speech and higher brain plasticity may give CI children an advantage over CI adults, who have to adapt their existing acoustic representations that are based on a normal acoustic hearing. These results emphasize that what is true for postlingually deaf CI adults, who are more often studied, may not always apply to prelingually deaf CI children.

To summarize, the findings from the experimental studies that are presented in this dissertation show strong developmental effects on different aspects of voice and speech perception in both NH and CI children. For CI children, the spectrotemporally degraded CI signal makes the observed developmental effects much less consistent than for NH children. The spectrotemporally degraded CI signal particularly reduces their voice discrimination abilities, but this does not seem to prevent them from forming representations of auditory categories that are broadly similar to those of their NH age peers.

Bibliography

- Arbogast, T. L., Mason, C. R., & Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America*, 112(5), 2086–2098.
- Auinger, A. B., Riss, D., Liepins, R., Rader, T., Keck, T., Keintzel, T., Kaider, A., Baumgartner, W.-D., Gstoettner, W., & Arnoldner, C. (2017). Masking release with changing fundamental frequency: Electric acoustic stimulation resembles normal hearing subjects. *Hearing Research*, 350, 226–234.
- Boons, T., Brokx, J. P. L., Dhooge, I., Frijns, J. H. M., Peeraer, L., Vermeulen, A., Wouters, J., & van Wieringen, A. (2012). Predictors of spoken language development following pediatric cochlear implantation. *Ear and Hearing*, 33(5), 617–639.
- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press.
- Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America*, 120(6), 4007–4018.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, 110(5), 2527–2538.
- Busch, T., Vermeulen, A., Langereis, M., Vanpoucke, F., & van Wieringen, A. (2020). Cochlear implant data logs predict children’s receptive vocabulary. *Ear and Hearing*, 41(4), 733–746.
- Buss, E., Flaherty, M. M., & Leibold, L. J. (2017). Development of frequency discrimination at 250 Hz is similar for tone and /ba/ stimuli. *The Journal of the Acoustical Society of America*, 142(1), EL150–EL154.
- Chatterjee, M., Kulkarni, A. M., Siddiqui, R. M., Christensen, J. A., Hozan, M., Sis, J. L., & Damm, S. A. (2019). Acoustics of emotional prosody produced by prelingually deaf children with cochlear implants. *Frontiers in Psychology*, 10, 1–15.
- Chatterjee, M., Zion, D. J., Deroche, M. L. D., Burianek, B. A., Limb, C. J., Goren, A. P., Kulkarni, A. M., & Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research*, 322, 151–162.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5), 975–979.
- Cleary, M., Pisoni, D. B., & Kirk, K. I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 48(1), 204–223.
- Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, 123(1), 414–427.

- Deroche, M. L. D., Kulkarni, A. M., Christensen, J. A., Limb, C. J., & Chatterjee, M. (2016). Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants. *Frontiers in Neuroscience*, 10, 1–15.
- Deroche, M. L. D., Lu, H.-P., Kulkarni, A. M., Caldwell, M., Barrett, K. C., Peng, S.-C., Limb, C. J., Lin, Y.-S., & Chatterjee, M. (2019). A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. *Scientific Reports*, 9(1), 1–12.
- Deroche, M. L. D., Lu, H.-P., Limb, C. J., Lin, Y.-S., & Chatterjee, M. (2014). Deficits in the pitch sensitivity of cochlear-implanted children speaking English or Mandarin. *Frontiers in Neuroscience*, 8, 1–13.
- El Boghdady, N., Gaudrain, E., & Başkent, D. (2019). Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users? *The Journal of the Acoustical Society of America*, 145(1), 417–439.
- Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., & Scott, S. K. (2016). Getting the Cocktail Party Started: Masking Effects in Speech Perception. *Journal of Cognitive Neuroscience*, 28(3), 483–500.
- Flaherty, M. M., Buss, E., & Leibold, L. J. (2019). Developmental effects in children’s ability to benefit from F0 differences between target and masker speech. *Ear and Hearing*, 40(4), 927–937.
- Flaherty, M. M., Buss, E., & Leibold, L. J. (2021). Independent and combined effects of fundamental frequency and vocal tract length differences for school-age children’s sentence recognition in a two-talker masker. *Journal of Speech, Language, and Hearing Research*, 64(1), 206–217.
- Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., & Başkent, D. (2014). Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology*, 15(6), 1037–1048.
- Gaudrain, E., & Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear and Hearing*, 39(2), 226–237.
- Geers, A. E., Nicholas, J. G., & Sedey, A. L. (2003). Language skills of children with early cochlear implantation. *Ear and Hearing*, 24(1), 46S–58S.
- Geurts, L., & Wouters, J. (2001). Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *The Journal of the Acoustical Society of America*, 109(2), 713–726.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton-Century-Crofts.
- Gordon, K. A., Wong, D. D. E., Valero, J., Jewell, S. F., Yoo, P., & Papsin, B. C. (2011). Use it or lose it? Lessons learned from the developing brains of children who are deaf and use cochlear implants to hear. *Brain Topography*, 24(3), 204–219.
- Green, T., Faulkner, A., & Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *The Journal of the Acoustical Society of America*, 116(4), 2298–2310.

- Green, T., Faulkner, A., Rosen, S., & Macherey, O. (2005). Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification. *The Journal of the Acoustical Society of America*, 118(1), 375–385.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4), 377–396.
- Hopyan-Misakyan, T. M., Gordon, K. A., Dennis, M., & Papsin, B. C. (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychology*, 15(2), 136–146.
- Iverson, P., Smith, C. A., & Evans, B. G. (2006). Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration. *The Journal of the Acoustical Society of America*, 120(6), 3998–4006.
- Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, 123(1), 155–163.
- Kopelovich, J. C., Eisen, M. D., & Franck, K. H. (2010). Frequency and electrode discrimination in children with cochlear implants. *Hearing Research*, 268(1), 105–113.
- Kronenberger, W. G., Pisoni, D. B., Henning, S. C., & Colson, B. G. (2013). Executive functioning skills in long-term users of cochlear implants: A case control study. *Journal of Pediatric Psychology*, 38(8), 902–914.
- Litovsky, R. (2015). Development of the auditory system. In *Handbook of Clinical Neurology* (Vol. 129, pp. 55–72). Elsevier.
- Manrique, M., Cervera-Paz, F. J., Huarte, A., Perez, N., Molina, M., & García-Tapia, R. (1999). Cerebral auditory plasticity and cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 49, S193–S197.
- Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59(3), 203–243.
- McCullough, E. A., Clopper, C. G., & Wagner, L. (2019). Regional dialect perception across the lifespan: Identification and discrimination. *Language and Speech*, 62(1), 115–136.
- McGettigan, C., Rosen, S., & Scott, S. K. (2014). Lexico-semantic and acoustic-phonetic processes in the perception of noise-vocoded speech: Implications for cochlear implantation. *Frontiers in Systems Neuroscience*, 8.
- Meister, H., Fürsen, K., Streicher, B., Lang-Roth, R., & Walger, M. (2016). The use of voice cues for speaker gender recognition in cochlear implant recipients. *Journal of Speech, Language, and Hearing Research*, 59(3), 546–556.
- Meister, H., Walger, M., Lang-Roth, R., & Müller, V. (2020). Voice fundamental frequency differences and speech recognition with noise and speech maskers in cochlear implant recipients. *The Journal of the Acoustical Society of America*, 147(1), EL19–EL24.
- Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2020). Development of voice perception is dissociated across gender cues in school-age children. *Scientific Reports*, 10(1), 1–11.

- Nakata, T., Trehub, S. E., & Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *The Journal of the Acoustical Society of America*, 131(2), 1307–1314.
- Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N.-Y., Quittner, A. L., Fink, N. E., & Team, Cd. I. (2010). Spoken language development in children following cochlear implantation. *JAMA*, 303(15), 1498–1506.
- Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *The Journal of the Acoustical Society of America*, 101(4), 2253–2266.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253–260.
- Pyman, B., Blamey, P., Lacy, P., Clark, G., & Dowell, R. (2000). The development of speech perception in children using cochlear implants: Effects of etiologic factors and delayed milestones. *Otology & Neurotology*, 21(1), 57–61.
- Pyschny, V., Landwehr, M., Hahn, M., Walger, M., von Wedel, H., & Meister, H. (2011). Bimodal hearing and speech perception with a competing talker. *Journal of Speech, Language, and Hearing Research*, 54(5), 1400–1415.
- Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proceedings of the National Academy of Sciences*, 108(37), 15516–15521.
- Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, 31(1), 97–113.
- Scott, S. K., & McGettigan, C. (2016). The voice: From identity to interactions. In D. Matsumoto, H. C. Hwang, & M. G. Frank (Eds.), *APA handbook of nonverbal communication*. (2015-02851-011; pp. 289–305). American Psychological Association.
- Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *The Journal of the Acoustical Society of America*, 115(2), 813–821.
- Skuk, V. G., Kirchen, L., Oberhoffner, T., Guntinas-Lichius, O., Döbel, C., & Schweinberger, S. R. (2020). Parameter-specific morphing reveals contributions of timbre and fundamental frequency cues to the perception of voice gender and age in cochlear implant users. *Journal of Speech, Language, and Hearing Research*, 63(9), 3155–3175.
- Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language & Hearing Research*, 57(1), 285–296.
- Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, 118(5), 3177–3186.
- Spencer, P. E. (2004). Individual differences in language performance after cochlear implantation at one to three years of age: Child, family, and linguistic factors. *Journal of Deaf Studies and Deaf Education*, 9(4), 395–412.

- Stickney, G. S., Assmann, P. F., Chang, J., & Zeng, F.-G. (2007). Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *The Journal of the Acoustical Society of America*, 122(2), 1069–1078.
- Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, J. G., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports*, 5(1), 11628.
- Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, 85(4), 1699–1707.
- Tonks, J., Williams, W. H., Frampton, I., Yates, P., & Slater, A. (2007). Assessing emotion recognition in 9–15-years olds: Preliminary analysis of abilities in reading emotion from faces, voices and eyes. *Brain Injury*, 21(6), 623–629.
- Werner, L. A. (2007). Issues in human auditory development. *Journal of Communication Disorders*, 40(4), 275–283.
- Werner, L. A., & Leibold, L. J. (2011). Auditory development in normal-hearing children. In J. S. Gravel, R. Sewald, & A. M. Tharpe (Red.), *Handbook of Pediatric Audiology*. Sage Publications.
- Wiefferink, C. H., Rieffe, C., Ketelaar, L., De Raeye, L., & Frijns, J. H. M. (2013). Emotion understanding in deaf children with a cochlear Implant. *Journal of Deaf Studies and Deaf Education*, 18(2), 175–186.
- Wouters, J., McDermott, H. J., & Francart, T. (2015). Sound coding in cochlear implants: From electric pulses to hearing. *IEEE Signal Processing Magazine*, 32(2), 67–80.
- Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*, 14(4), 1908–1919.

Nederlandse samenvatting

Wanneer we luisteren naar spraak, dan letten we met name op wat een spreker precies probeert te zeggen. Tegelijkertijd geeft de stem van de spreker echter ook veel informatie over de spreker zelf. Vaak kunnen we bijvoorbeeld horen waar iemand vandaan komt op basis van hun accent, of hoe oud en gezond iemand is door te luisteren naar hoe krakerig of hees hun stem klinkt. Deze informatie is niet alleen belangrijk om de spreker beter te leren kennen, maar ook om de spreker beter te kunnen herkennen en verstaan. In een druk café moeten we bijvoorbeeld iemands stem kunnen onderscheiden van allerlei andere stemmen en achtergrondgeluiden om te verstaan wat er wordt gezegd. Vooral voor doven en slechthorenden zijn dit soort situaties vaak een uitdaging. In het geval van ernstige slechthorendheid of doofheid kan een *cochleair implantaat* (CI) luisteraars in staat stellen om alsnog enigszins spraak te verstaan. De kwaliteit van het *spraaksignaal* dat kan worden waargenomen via een CI is echter ernstig verminderd. Het verstaan van spraak, met name in rumoerige omgevingen, en het achterhalen van de *stemkenmerken* van de spreker leveren daarom nog steeds vaak problemen op voor CI-gebruikers. Tegenwoordig krijgen veel dove en slechthorende kinderen op jonge leeftijd al één of twee CI's en leren ze gesproken taal en spraak bijna volledig op basis van het verslechterde CI-spraaksignaal. De gevolgen hiervan op de ontwikkeling van het verstaan van stemkenmerken en spraak voor CI-kinderen zijn echter nog onduidelijk. Daarnaast is het onbekend in hoeverre deze ontwikkeling bij CI-kinderen verschilt van normaalhorende kinderen. In dit proefschrift heb ik daarom de ontwikkeling onderzocht in het verstaan van stemkenmerken en spraak in achtergrondspraak bij normaalhorende kinderen en kinderen met een CI.

1.1. Cochleaire implantaten

Een CI is een elektronische gehoorprothese voor doven en ernstig slechthorenden. Bij een regulier gehoorapparaat wordt het normale *akoestische* spraaksignaal versterkt, maar spraak wordt nog steeds op dezelfde fysieke manier overgebracht via het trommelmvlies, middenoor en slakkenhuis (*cochlea*) naar de gehoorzenuw. Bij een CI is er sprake van een geheel andere manier van horen. Via een externe microfoon boven het buitenoor wordt spraak omgezet in een *elektrisch* spraaksignaal dat direct via de geïmplanteerde elektrodes in het slakkenhuis naar de gehoorzenuw wordt gestuurd. Het signaal dat door de elektrodes wordt afgegeven aan de gehoorzenuw is vrij onnauwkeurig, waardoor de kwaliteit van de akoestische

informatie die CI-gebruikers ontvangen ernstig verminderd is ten opzichte van de informatie die normaalhorende mensen ontvangen. Het verstaan van spraak blijft voor veel CI-gebruikers daarom vaak vermoeiend en uitdagend. Naast het minder goed verstaan van spraak zorgt het verslechterde CI-spraaksignaal er ook voor dat subtiele akoestische informatie, zoals verschillen in stemkenmerken, minder goed te horen zijn voor CI-gebruikers. Uit eerder onderzoek blijkt dat volwassen CI-gebruikers minder goed de verschillen tussen onder andere een mannen- en vrouwenstem en de emoties van sprekers kunnen horen. Het slechter verstaan van stemkenmerken lijkt er ook toe te leiden dat CI-gebruikers verschillende sprekers minder goed uit elkaar kunnen houden, waardoor het verstaan van spraak in achtergrondspraak wordt bemoeilijkt.

In onderzoek naar het verstaan van spraak door CI-gebruikers wordt vaak een onderscheid gemaakt tussen *postlinguaal* dove en *prelinguaal* dove CI-gebruikers. Postlinguaal dove CI-gebruikers zijn op latere leeftijd doof of ernstig slechthorend geworden. Meestal hebben postlinguaal dove CI-gebruikers hierdoor gesproken taal en spraak ontwikkeld op basis van een normaal akoestisch spraaksignaal. Prelinguaal dove CI-gebruikers zijn meestal doof vanaf de geboorte of zijn op vroege leeftijd doof of slechthorend geworden tijdens de periode van taalontwikkeling. Daarom hebben veel prelinguaal dove CI-kinderen gesproken taal en spraak ontwikkeld via het verslechterde CI-spraaksignaal. Postlinguaal dove CI-volwassenen en prelinguaal dove CI-kinderen hebben gesproken taal en spraak dus op een fundamenteel andere manier ontwikkeld. Terwijl CI-volwassenen hun *akoestische representaties*, zoals het verschil tussen hoe een mannen- of vrouwenstem klinkt, moeten aanpassen aan het verslechterde CI-spraaksignaal, zijn de akoestische representaties van CI-kinderen alleen op het verslechterde CI-spraaksignaal gebaseerd. Enerzijds is het mogelijk dat CI-volwassenen hierdoor een betere basis hebben gevormd voor akoestische representaties, doordat deze gebaseerd zijn op het normale akoestische spraaksignaal dat rijk is aan informatie. Anderzijds zou de aanpassing aan het verslechterde CI-spraaksignaal hinderlijk kunnen zijn voor CI-volwassenen, omdat ze moeten leren om andere akoestische informatie te gebruiken dan de informatie die beschikbaar was via het normale akoestische spraaksignaal. Bovendien is de *hersenplasticiteit* van jonge kinderen vaak groter dan die van oudere volwassenen. Hersenplasticiteit zorgt ervoor dat we de organisatie van onze hersenen kunnen aanpassen. Dit vergemakkelijkt het leren van nieuwe vaardigheden en informatie. CI-kinderen zouden daarom misschien beter in staat zijn om de akoestische informatie die beschikbaar is in het CI-spraaksignaal te leren gebruiken dan CI-volwassenen.

1.2. Stemkenmerken

Zoals eerder genoemd kunnen we meer te weten komen over sprekers op basis van hun stemkenmerken. Twee van de meest invloedrijke stemkenmerken zijn de gemiddelde toonhoogte en het zogenaamde timbre ofwel de diepte van iemands stem. De gemiddelde toonhoogte, de *fundamentele frequentie* (F_0), van iemands stem wordt bepaald door de snelheid waarmee de stembanden openen en sluiten. Mannen hebben over het algemeen een lagere gemiddelde toonhoogte dan vrouwen, omdat hun stembanden vaak langer zijn en daardoor minder snel kunnen openen en sluiten. Het timbre van iemands stem wordt voornamelijk bepaald door de lengte van het spraakkanaal van de spreker ofwel *vocal tract length* (VTL). Deze bestaat uit de afstand van de stembanden tot aan de mond- en neusholte en hangt sterk samen met de algehele lengte van de spreker. Aangezien mannen over het algemeen langer zijn dan vrouwen, hebben ze daarom meestal niet alleen een lagere, maar ook een diepere stem dan vrouwen. Als we een vrouwenstem kunstmatig willen manipuleren zodat deze als een duidelijke mannenstem gaat klinken, dan moeten we dus zowel de F_0 van de stem verlagen als de gesimuleerde VTL verlengen. Daarnaast zijn er stemkenmerken die niet zozeer fysiek bepaald zijn, maar informatie geven over de gemoedstoestand of intenties van de spreker. We kunnen de emoties van een spreker bijvoorbeeld bepalen door te luisteren naar variaties in toonhoogte, luidheid of spreek snelheid.

Hoewel de meeste normaalhorende kinderen geen problemen hebben met het horen van spraak en andere geluiden, zien we toch een grote ontwikkeling in hun vermogen om verschillende stemmen van elkaar te kunnen onderscheiden en te categoriseren. Eerder onderzoek heeft bijvoorbeeld aangetoond dat kinderen een groter verschil in de gemiddelde F_0 van twee sprekers nodig hebben dan volwassenen om de stemmen uit elkaar te houden. Voor *stemdiscriminatie* moet de luisteraar een directe vergelijking maken tussen twee of meer stemmen. Hiervoor moeten luisteraars goed onthouden hoe beide stemmen klinken en hun aandacht specifiek richten op een soms miniem akoestisch verschil tussen de stemmen. Stemdiscriminatie is daarom een perceptuele vaardigheid die berust op zogenaamde *cognitieve vaardigheden*, zoals het tijdelijk onthouden van informatie om een bepaalde taak uit te voeren (*werkgeheugen*), de aandacht richten op relevante informatie (*selectieve aandacht*), of het onderdrukken van irrelevante informatie (*inhibitie*). Ook voor de categorisatie van stemmen in verschillende categorieën, bijvoorbeeld als een mannenstem of vrouwenstem of als een boze of blij stem, blijkt uit eerder onderzoek dat het een tijd duurt voordat kinderen hierin even goed zijn als volwassenen. *Stemcategorisatie* houdt in dat luisteraars bepaalde stemkenmerken

kunnen relateren aan een akoestische representatie die ze al hebben. De luisteraar moet daarom een duidelijk beeld hebben van welke akoestische informatie relevant is, bijvoorbeeld F0 en VTL, en weten wat de bijbehorende waardes zijn. Zo zijn een lage F0 en grote VTL kenmerkend voor een mannenstem. Stemcategorisatie is in tegenstelling tot stemdiscriminatie afhankelijk van de voorkennis die de luisteraar heeft en kan daarom als een cognitief complexere perceptuele vaardigheid worden gezien. Tijdens de schooljaren vindt er ook een ontwikkeling plaats van de vaardigheid om spraak in achtergrondspraak te kunnen verstaan. Voor het verstaan van spraak in achtergrondspraak moet de luisteraar alle aandacht richten op wat wordt gezegd door de *doelspreker*, en daarbij de *achtergrondspreekers* negeren. Cognitieve vaardigheden, zoals selectieve aandacht en inhibitie, lijken daarom ook voor deze perceptuele vaardigheid erg belangrijk te zijn.

Voorgaand onderzoek naar het waarnemen van verschillen in F0 en VTL bij CI-volwassenen wijst uit dat er een sterk verband is tussen hoe goed CI-gebruikers verschillen kunnen horen in de F0 en VTL van sprekers (stemdiscriminatie) en hoe goed ze deze informatie kunnen gebruiken om te bepalen of ze een mannen- of vrouwenstem horen (stemcategorisatie). Vooral verschillen in de VTL van sprekers lijken niet goed te worden doorgegeven via het verslechterde CI signaal, waardoor CI-volwassenen met name verschillen in F0 gebruiken om mannen- en vrouwenstemmen te categoriseren. Dit lijkt er toe te leiden dat CI-volwassenen geen voordeel hebben van verschillen in de F0 en VTL van sprekers om spraak in achtergrondspraak beter te kunnen verstaan, zoals normaalhorende volwassenen. Veruit de meeste onderzoeken naar het verstaan van stemkenmerken zijn gericht geweest op CI-volwassenen. Hierdoor is er nog maar weinig bekend over het verstaan van stemkenmerken door CI-kinderen. Aangezien normaalhorende kinderen een duidelijke ontwikkeling laten zien in stemdiscriminatie, stemcategorisatie en het verstaan van spraak in achtergrondspraak, is het de vraag hoe de ontwikkeling van deze perceptuele vaardigheden precies verloopt in CI-kinderen. Enerzijds is het denkbaar dat het verstaan van stemkenmerken bij CI-kinderen net als bij CI-volwassenen ernstig beperkt wordt door het verslechterde CI-spraaksignaal, en dat de ontwikkeling van hun perceptuele vaardigheden hierdoor op een bepaald moment afvlakt en een plafondniveau bereikt. Anderzijds zou het kunnen dat we eenzelfde soort ontwikkeling zien in het verstaan van stemkenmerken bij CI-kinderen als bij normaalhorende kinderen, omdat CI-kinderen beter leren omgaan met de informatie die beschikbaar is in het verslechterde CI-spraaksignaal vanwege de plasticiteit van hun hersenen. Tenslotte is het onduidelijk of er een verband is tussen de verschillende perceptuele vaardigheden in normaalhorende kinderen en CI-kinderen.

1.3. Samenvatting van bevindingen

Het voornaamste doel van dit proefschrift was om de ontwikkeling in het verstaan van stemkenmerken en spraak in achtergrondspraak te onderzoeken in normaalhorende kinderen en kinderen met een CI. Specifiek is er gekeken naar het waarnemen van verschillen in de F0 en VTL van sprekers, dat wil zeggen de akoestische informatie die grotendeels het verschil tussen een mannen- en vrouwenstem bepaalt, en vocale emoties. Daarnaast was het tweede doel om de relatie tussen verschillende perceptuele vaardigheden te bestuderen. Hiervoor heb ik vier experimenten uitgevoerd met zowel normaalhorende kinderen als kinderen met een CI, waarin de ontwikkeling van stemdiscriminatie, stemcategorisatie en het verstaan van spraak in achtergrondspraak zijn onderzocht. Het eerste gedeelte van dit proefschrift is gefocust op de resultaten van normaalhorende kinderen, om meer inzicht te krijgen in hoe de ontwikkeling in het verstaan van stemkenmerken normaliter verloopt. Deze inzichten zijn noodzakelijk om de complexe combinatie van factoren op de ontwikkeling in CI-kinderen te begrijpen. In het tweede gedeelte worden de resultaten van CI-kinderen en de vergelijking tussen beide groepen besproken, om een beter beeld te krijgen van de effecten die het verslechterde CI-spraaksignaal heeft op de ontwikkeling van perceptuele vaardigheden.

In **Hoofdstuk 2** is de eerste studie beschreven waarin de ontwikkeling van de discriminatie en categorisatie van stemkenmerken, in het bijzonder verschillen in de F0 en VTL van sprekers, in normaalhorende kinderen in de schoolgaande leeftijd (4-12 jaar) is onderzocht. Eerder onderzoek heeft aangetoond dat kinderen minder goed zijn in stemdiscriminatie en stemcategorisatie dan volwassenen. Waarschijnlijk wordt dit veroorzaakt doordat deze perceptuele vaardigheden op cognitieve vaardigheden berusten die nog in ontwikkeling zijn, zoals werkgeheugen, selectieve aandacht en inhibitie. Voor deze studie zijn twee verschillende experimenten uitgevoerd. In het eerste experiment werd stemdiscriminatie onderzocht door het kleinst waarneembare verschil in F0 en VTL tussen stemmen te meten dat kinderen en volwassenen nog konden horen. De F0 en VTL van dezelfde spreker zijn hiervoor gemanipuleerd, zodat de stemmen alleen in dit opzicht van elkaar verschillen. In het begin van het experiment was het verschil in F0 of VTL duidelijk hoorbaar voor de luisteraars, maar het verschil werd steeds kleiner gemaakt, totdat de luisteraar geen verschil meer kon horen tussen de stemmen. Het tweede experiment naar stemcategorisatie bestond uit een taak waarin kinderen en volwassenen stemmen met verschillende F0 en VTL waardes moesten categoriseren als een mannenstem of een vrouwenstem. Hiervoor zijn wederom de F0 en VTL waardes van dezelfde spreker gemanipuleerd. De stem klonk hierdoor als een

duidelijke vrouwenstem (hoge F0 en korte gesimuleerde VTL), een duidelijke mannenstem (lage F0 en grote gesimuleerde VTL), of een enigszins ambigue stem. Uit de resultaten bleek dat kinderen in de basisschoolleeftijd voor zowel stemdiscriminatie als stemcategorisatie een sterke ontwikkeling lieten zien. Voor stemdiscriminatie bleek dat kinderen rond de leeftijd van 8 jaar even goed verschillen in de VTL van sprekers konden horen als volwassenen, maar dat zelfs de oudste kinderen van 12 jaar nog niet zo goed waren als volwassenen in het waarnemen van verschillen in de F0 van sprekers. Voor stemcategorisatie toonden de resultaten aan dat kinderen rond de leeftijd van 6 jaar even goed mannen- en vrouwenstemmen konden categoriseren als volwassenen op basis van verschillen in F0, maar pas rond de leeftijd van 10 jaar oud op basis van verschillen in VTL. Voor beide perceptuele vaardigheden was er dus een duidelijke ontwikkeling te zien tijdens de basisschoolleeftijd, maar er lijkt geen directe onderlinge relatie te zijn tussen stemdiscriminatie en stemcategorisatie. Terwijl kinderen beter waren in het waarnemen van verschillen in de VTL dan de F0 van sprekers, waren ze slechter in het gebruik van verschillen in de VTL dan F0 van sprekers voor het categoriseren van mannen- en vrouwenstemmen. Een van de verklaringen hiervoor komt voort uit eerder onderzoek naar de categorisatie van verschillende klanken. Volgens een bekende theorie hierover, de *'Developing Weighting-Shift Hypothesis'* (DWS-hypothese), letten kinderen voor categorisatie vooral op verschillen in dynamische, snel veranderende stemkenmerken, zoals F0-verschillen, in plaats van op statische, langzaam veranderende stemkenmerken, zoals VTL-verschillen. De oorzaak hiervan is dat dynamische stemkenmerken vaak een belangrijke rol spelen in het bepalen van woordgrenzen en lettergrepen. Dit is vooral van belang in de beginfase van de taalontwikkeling. Statische stemkenmerken zijn vaak belangrijker voor het onderscheiden van verschillende klanken, en kinderen hebben daarom misschien meer tijd nodig om te leren dat deze informatie belangrijk is. Een alternatieve verklaring is dat kinderen andere akoestische representaties hebben dan volwassenen omdat er nauwelijks verschillen in de F0 en VTL zijn tussen de stemmen van jongens en meisjes tijdens de basisschoolleeftijd. De stembanden en algehele lengte van kinderen verschillen namelijk nauwelijks tot aan de puberteit. Kinderen hebben daarom misschien nog niet geleerd dat dit belangrijke informatie is om te bepalen of een spreker een man of vrouw is.

De studie in **Hoofdstuk 3** is een vervolg op de studie in Hoofdstuk 2, waarin het gebruik van verschillen in F0 en VTL voor het verstaan van spraak in achtergrondspraak werd onderzocht. Om spraak in achtergrondspraak te kunnen verstaan, moeten we de stem van een spreker, de doelspreker, kunnen onderscheiden van andere sprekers, de achtergrondsprekers. De verwachting was

dat kinderen nog niet zo goed zouden zijn als volwassenen in het verstaan van spraak in achtergrondspraak, doordat belangrijke cognitieve vaardigheden hiervoor, zoals selectieve aandacht en inhibitie, nog in ontwikkeling zijn. Daarnaast heeft eerder onderzoek aangetoond dat het voor normaalhorende volwassenen makkelijker is om spraak te verstaan wanneer de stemmen van de doelspreker en een achtergrondspreeker verschillen in F0 en VTL. Aangezien uit voorgaand onderzoek en de resultaten van de vorige studie is gebleken dat kinderen minder goed verschillen in de F0 en VTL van sprekers kunnen horen, profiteren ze misschien niet in dezelfde mate als volwassenen van deze verschillen. Het verband tussen stemdiscriminatie en het profiteren van verschillen in stemkenmerken voor het verstaan van spraak in achtergrondspraak heb ik onderzocht door een derde experiment uit te voeren met dezelfde normaalhorende kinderen en volwassenen die hebben deelgenomen aan de vorige twee experimenten. Voor het experiment moesten deelnemers aangeven welke twee kleur- en nummerwoorden werden genoemd door de doelspreker in een zin, bijvoorbeeld *Laat de hond zien waar de rode twee is*, terwijl ze tegelijkertijd een achtergrondspreeker hoorden praten. De achtergrondspraak werd geproduceerd door dezelfde spreker als de doelspraak, maar de F0 en VTL waardes van de spreker werden gemanipuleerd zodat er soms wel en soms geen verschil was tussen de stemmen van de doelspreker en achtergrondspreeker. De resultaten lieten een duidelijke ontwikkeling zien in de vaardigheid van kinderen om spraak in achtergrondspraak te kunnen verstaan. Echter, zelfs de oudste kinderen van 12 jaar waren nog niet zo goed in het verstaan van de doelspreker als volwassenen. Tegen de verwachting in hadden kinderen verhoudingsgewijs evenveel voordeel van verschillen in de F0 en VTL van sprekers als volwassenen. Kinderen en volwassenen konden de doelspreker allebei beter verstaan wanneer de F0 en VTL van de sprekers verschilden. Het minder goed kunnen horen van verschillen in de F0 en VTL van sprekers lijkt kinderen dus niet te belemmeren bij het gebruik van deze stemverschillen voor het verstaan van achtergrondspraak. Wellicht zijn kinderen met name minder goed in het verstaan van spraak in achtergrondspraak doordat de relevante cognitieve vaardigheden nog in ontwikkeling zijn, in plaats van hun ontwikkeling in stemdiscriminatie.

In **Hoofdstuk 4** wordt voortgebouwd op de eerdere twee studies door de ontwikkeling van de waarneming van stemkenmerken en het verstaan van spraak in achtergrondspraak in CI-kinderen te onderzoeken. Dezelfde experimenten als in de eerdere twee studies heb ik hiervoor uitgevoerd met CI-kinderen. Eerder onderzoek heeft aangetoond dat CI-volwassenen verschillen in de F0 en met name de VTL van sprekers minder goed kunnen horen dan normaalhorende volwassenen, en dat dit de categorisatie van mannen- en vrouwenstemmen op basis van verschillen in

F0 en VTL lijkt te bemoeilijken. Ook lijken CI-volwassenen geen voordeel te hebben van verschillen in de F0 en VTL van sprekers voor het verstaan van spraak in achtergrondspraak. De ontwikkeling van stemdiscriminatie, stemcategorisatie en het verstaan van spraak in achtergrondspraak bij CI-kinderen zou daarom op een bepaald moment in hun taal- en spraakontwikkeling kunnen afvlakken door de beperkingen van het verslechterde CI-spraaksignaal. Daarentegen zou het ook kunnen dat CI-kinderen beter leren omgaan met de informatie die beschikbaar is in het verslechterde CI-spraaksignaal dan CI-volwassenen door de vroege blootstelling aan het signaal en doordat ze een grote mate van hersenplasticiteit hebben. De resultaten voor stemdiscriminatie toonden aan dat CI-kinderen minder goed waren in het waarnemen van verschillen in de F0 en VTL van sprekers dan normaalhorende kinderen, maar dat deze vaardigheid wel lijkt te verbeteren naarmate CI-kinderen ouder worden. Bovendien waren CI-kinderen gemiddeld gezien beter in het horen van verschillen in F0 en VTL dan de CI-volwassenen die eerder zijn getest via eenzelfde soort experiment. Uit de resultaten voor stemcategorisatie bleek dat CI-kinderen mannen- en vrouwenstemmen categoriseerden op basis van zowel verschillen in F0 als VTL, net als normaalhorende kinderen. Ook in dit opzicht verschilden de CI-kinderen van CI-volwassenen, die hiervoor voornamelijk F0-verschillen lijken te gebruiken. Tot slot hadden CI-kinderen meer moeite met het verstaan van spraak in achtergrondspraak dan normaalhorende kinderen, echter, ook deze vaardigheid verbeterde met leeftijd. Bovendien hadden CI-kinderen verhoudingsgewijs evenveel voordeel van stemverschillen als normaalhorende kinderen en volwassenen. Samengevat kunnen we stellen dat CI-kinderen meer moeite hebben met het waarnemen van subtiele verschillen in stemkenmerken en het verstaan van spraak in achtergrondspraak dan normaalhorende kinderen, maar dit verbetert naarmate ze ouder worden. CI-kinderen zijn zelfs beter hierin dan CI-volwassenen. Een grote mate van hersenplasticiteit en de vroege blootstelling aan het verslechterde CI-spraaksignaal lijken er dus voor te zorgen dat CI-kinderen de informatie van het CI-spraaksignaal beter leren gebruiken dan CI-volwassenen.

In **Hoofdstuk 5** worden de resultaten van de eerdere studies uitgebreid door de ontwikkeling in de categorisatie van een ander stemkenmerk, namelijk vocale emoties, te onderzoeken bij normaalhorende en CI-kinderen. Net als voor de categorisatie van mannen- en vrouwenstemmen, hangt de categorisatie van vocale emoties af van akoestische representaties, bijvoorbeeld hoe een blij of boze stem klinkt. De categorisatie van vocale emoties is echter complexer dan de categorisatie van mannen- en vrouwenstemmen, omdat emoties over een groter aantal dimensies kunnen verschillen en bepaald worden door diverse soorten akoestische informatie,

zoals variaties in F0, gemiddelde F0, luidheid en spreesnelheid. Bovendien speelt ook de sociaal-emotionele ontwikkeling van kinderen hoogstwaarschijnlijk een rol in de categorisatie van vocale emoties. De ontwikkeling van de categorisatie van emoties heb ik onderzocht door een vierde experiment uit te voeren met dezelfde groep van normaalhorende kinderen en volwassenen en CI-kinderen als beschreven in de eerdere studies. Hiervoor heb ik een nieuwe test ontwikkeld, de EmoHI test. Deze test bestaat uit opnames van spraak met een goede geluidskwaliteit, zodat ze geschikt zijn om mensen met een gehoorbeperking te testen. Aangezien we voor de opnames gebruik hebben gemaakt van zinnen die bestonden uit zogenaamde niet-bestaande *pseudowoorden*, bijvoorbeeld *Koun se mina lod belam*, heb ik ook sprekers met een andere moedertaal getest, namelijk Engelse kinderen en volwassenen. Voorgaand onderzoek laat namelijk zien dat verschillen in de uitspraak van klanken een effect kunnen hebben op hoe goed luisteraars de emoties van sprekers kunnen categoriseren. Dit zou een voordeel kunnen geven aan Nederlandse kinderen en volwassenen in vergelijking met de Engelse kinderen en volwassenen, aangezien de opnames door Nederlandse sprekers zijn ingesproken. De resultaten toonden aan dat de categorisatie van emoties zich in normaalhorende Nederlandse kinderen ontwikkelt gedurende de basisschoolleeftijd, en zelfs rond de leeftijd van 12 nog verschilt van volwassenen. CI-kinderen hadden over het algemeen meer moeite met het correct categoriseren van emoties dan normaalhorende kinderen, maar er was veel variatie binnen deze groep. Tenslotte waren er, anders dan verwacht, geen verschillen tussen Nederlandse en Engelse normaalhorende kinderen en volwassenen. Het zou kunnen dat dit komt doordat Nederlands en Engels qua klanken sterk op elkaar lijken. Voor sprekers van een taal die meer verschilt van het Nederlands, zoals Turks, zullen er misschien wel verschillen zijn.

De resultaten uit dit proefschrift laten zien dat er een grote ontwikkeling plaatsvindt in het verstaan van stemkenmerken en spraak in achtergrondspraak bij normaalhorende kinderen en kinderen met een CI tijdens de basisschoolleeftijd en zelfs nog daarna. De resultaten van normaalhorende kinderen impliceren dat deze ontwikkeling niet zozeer gerelateerd is aan het gehoor van kinderen, maar eerder aan hun cognitieve vaardigheden, zoals selectieve aandacht en inhibitie, die nog niet zo goed zijn als die van volwassenen. De cognitieve vaardigheden van CI-kinderen lijken daarom van belang voor de beoordeling van hun perceptuele vaardigheden in het kader van wetenschappelijk onderzoek en in de klinische praktijk. Ten slotte tonen onze resultaten aan dat CI-kinderen beter dan CI-volwassenen in staat lijken te zijn om de akoestische informatie die beschikbaar is in het verslechterde CI-spraaksignaal te gebruiken. Een grote mate van hersenplasticiteit en vroege blootstelling aan het verslechterde spraaksignaal lijken hierbij een grote rol in te

spelen. Mogelijk zijn er ook nog andere factoren, zoals het feit dat CI-kinderen meestal twee CI's hebben en CI-volwassenen vaak maar één CI en mogelijke verschillen in de mate van restgehoor van CI-gebruikers, die het verstaan van stemkenmerken en spraak in achtergrondspraak kunnen beïnvloeden.

List of publications

2021

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., and Başkent, D. (2021). School-age children benefit from voice gender cue differences for the perception of speech in competing speech. *The Journal of the Acoustical Society of America*, 149(5).

2020

Nagels, L., Bastiaanse, R., Başkent, D., & Wagner, A. (2020). Individual differences in lexical access among cochlear implant users. *Journal of Speech, Language, and Hearing Research*, 63 (1), 286-304.

Nagels L., Gaudrain, E., Vickers, D., Matos Lopes, M., Hendriks, P., Başkent, D. (2020). Development of vocal emotion recognition in school-age children: the EmoHI test for hearing-impaired populations. *PeerJ*, 8.

Nagels L., Gaudrain, E., Vickers, D., Hendriks, P., Başkent, D. (2020). Dissociated development of voice perception across gender cues in school-age children. *Scientific Reports*, 10 (1).

2019

Wagner, A. E., Nagels, L., Toffanin, P., Opie, J. M., & Başkent, D. (2019). Individual Variations in Effort: Assessing Pupillometry for the Hearing Impaired. *Trends in Hearing*, 23, 1–18.

Groningen dissertations in linguistics (GRODIL)

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach*.
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure*.
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation*.
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation*.
5. Gosse Bouma (1993). *Nonmonotonicity and Categorical Unification Grammar*.
6. Peter I. Blok (1993). *The Interpretation of Focus*.
7. Roelien Bastiaanse (1993). *Studies in Aphasia*.
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist*.
9. Wim Kosmeijer (1993). *Barriers and Licensing*.
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach*.
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity*.
12. Ton van der Wouden (1994). *Negative Contexts*.
13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorical Grammar*.
14. Petra Hendriks (1995). *Comparatives and Categorical Grammar*.
15. Maarten de Wind (1995). *Inversion in French*.
16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance*.
17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition*.
18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items*.
19. Karen Lattewitz (1997). *Adjacency in Dutch and German*.
20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch*.
21. Henny Klein (1997). *Adverbs of Degree in Dutch*.
22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The case of French 'des'/'du'-NPs*.
23. Rita Landeweerd (1998). *Discourse semantics of perspective and temporal structure*.
24. Mettina Veenstra (1998). *Formalizing the Minimalist Program*.
25. Roel Jonkers (1998). *Comprehension and Production of Verbs in aphasic Speakers*.
26. Erik F. Tjong Kim Sang (1998). *Machine Learning of Phonotactics*.
27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses*.
28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure*.
29. H. Wee (1999). *Definite Focus*.
30. Eun-Hee Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean tense and aspect in discourse*.
31. Ivilin P. Stoianov (2001). *Connectionist Lexical Processing*.
32. Klarien van der Linde (2001). *Sonority substitutions*.
33. Monique Lamers (2001). *Sentence processing: using syntactic, semantic, and thematic information*.

34. Shalom Zuckerman (2001). *The Acquisition of "Optional" Movement.*
35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding.*
36. Esther Ruigendijk (2002). *Case assignment in Agrammatism: a cross-linguistic study.*
37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection.*
38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren.*
39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and segments in level-specific deficits.*
40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension.*
41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition.*
42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study.*
43. Hein van Schie (2003). *Visual Semantics.*
44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian.*
45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic Structures.*
46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance.*
47. Wouter Jansen (2004). *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology.*
48. Judith Rispens (2004). *Syntactic and phonological processing in developmental dyslexia.*
49. Danielle Bougaïré (2004). *L'approche communicative des campagnes de sensibilisation en santé publique au Burkina Faso: Les cas de la planification familiale, du sida et de l'excision.*
50. Tanja Gaustad (2004). *Linguistic Knowledge and Word Sense Disambiguation.*
51. Susanne Schoof (2004). *An HPSG Account of Nonfinite Verbal Complements in Latin.*
52. M. Begoña Villada Moirón (2005). *Data-driven identification of fixed expressions and their modifiability.*
53. Robbert Prins (2005). *Finite-State Pre-Processing for Natural Language Analysis.*
54. Leonoor van der Beek (2005) *Topics in Corpus-Based Dutch Syntax.*
55. Keiko Yoshioka (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse.*
56. Sible Andringa (2005). *Form-focused instruction and the development of second language proficiency.*
57. Joanneke Prenger (2005). *Taal telt! Een onderzoek naar de rol van taalvaardigheid en tekstbegrip in het realistisch wiskundeonderwijs.*
58. Neslihan Kansu-Yetkiner (2006). *Blood, Shame and Fear: Self-Presentation Strategies of Turkish Women's Talk about their Health and Sexuality.*
59. Mónika Z. Zempléni (2006). *Functional imaging of the hemispheric contribution to language processing.*

-
60. Maartje Schreuder (2006). *Prosodic Processes in Language and Music*.
  61. Hidetoshi Shiraishi (2006). *Topics in Nivkh Phonology*.
  62. Tamás Biró (2006). *Finding the Right Words: Implementing Optimality Theory with Simulated Annealing*.
  63. Dieuwke de Goede (2006). *Verbs in Spoken Sentence Processing: Unraveling the Activation Pattern of the Matrix Verb*.
  64. Eleonora Rossi (2007). *Clitic production in Italian agrammatism*.
  65. Holger Hopp (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing*.
  66. Gerlof Bouma (2008). *Starting a Sentence in Dutch: A corpus study of subject- and object-fronting*.
  67. Julia Klitsch (2008). *Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia*.
  68. Janneke ter Beek (2008). *Restructuring and Infinitival Complements in Dutch*.
  69. Jori Mur (2008). *Off-line Answer Extraction for Question Answering*.
  70. Lonneke van der Plas (2008). *Automatic Lexico-Semantic Acquisition for Question Answering*.
  71. Arjen Versloot (2008). *Mechanisms of Language Change: Vowel reduction in 15th century West Frisian*.
  72. Ismail Fahmi (2009). *Automatic term and Relation Extraction for Medical Question Answering System*.
  73. Tuba Yarbay Duman (2009). *Turkish Agrammatic Aphasia: Word Order, Time Reference and Case*.
  74. Maria Trofimova (2009). *Case Assignment by Prepositions in Russian Aphasia*.
  75. Rasmus Steinkrauss (2009). *Frequency and Function in WH Question Acquisition. A Usage-Based Case Study of German L1 Acquisition*.
  76. Marjolein Deunk (2009). *Discourse Practices in Preschool. Young Children's Participation in Everyday Classroom Activities*.
  77. Sake Jager (2009). *Towards ICT-Integrated Language Learning: Developing an Implementation Framework in terms of Pedagogy, Technology and Environment*.
  78. Francisco Dellatorre Borges (2010). *Parse Selection with Support Vector Machines*.
  79. Geoffrey Andogah (2010). *Geographically Constrained Information Retrieval*.
  80. Jacqueline van Kruiningen (2010). *Ondervijzontwerp als conversatie. Probleemoplossing in interprofessioneel overleg*.
  81. Robert G. Shackleton (2010). *Quantitative Assessment of English-American Speech Relationships*.
  82. Tim Van de Cruys (2010). *Mining for Meaning: The Extraction of Lexico-semantic Knowledge from Text*.

83. Therese Leinonen (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects*.
84. Erik-Jan Smits (2010). *Acquiring Quantification. How Children Use Semantics and Pragmatics to Constrain Meaning*.
85. Tal Caspi (2010). *A Dynamic Perspective on Second Language Development*.
86. Teodora Mehotcheva (2010). *After the fiesta is over. Foreign language attrition of Spanish in Dutch and German Erasmus Student*.
87. Xiaoyan Xu (2010). *English language attrition and retention in Chinese and Dutch university students*.
88. Jelena Prokić (2010). *Families and Resemblances*.
89. Radek Šimík (2011). *Modal existential wh-constructions*.
90. Katrien Colman (2011). *Behavioral and neuroimaging studies on language processing in Dutch speakers with Parkinson's disease*.
91. Siti Mina Tamah (2011). *A Study on Student Interaction in the Implementation of the Jigsaw Technique in Language Teaching*.
92. Aletta Kwant (2011). *Geraakt door prentenboeken. Effecten van het gebruik van prentenboeken op de sociaal-emotionele ontwikkeling van kleuters*.
93. Marlies Kluck (2011). *Sentence amalgamation*.
94. Anja Schüppert (2011). *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish*.
95. Peter Nabende (2011). *Applying Dynamic Bayesian Networks in Transliteration Detection and Generation*.
96. Barbara Plank (2011). *Domain Adaptation for Parsing*.
97. Cagri Coltekin (2011). *Catching Words in a Stream of Speech: Computational simulations of segmenting transcribed child-directed speech*.
98. Dörte Hessler (2011). *Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners: The Whole is More than the Sum of its Parts*.
99. Herman Heringa (2012). *Appositional constructions*.
100. Diana Dimitrova (2012). *Neural Correlates of Prosody and Information Structure*.
101. Harwintha Anjarningsih (2012). *Time Reference in Standard Indonesian Agrammatic Aphasia*.
102. Myrte Gosen (2012). *Tracing learning in interaction. An analysis of shared reading of picture books at kindergarten*.
103. Martijn Wieling (2012). *A Quantitative Approach to Social and Geographical Dialect Variation*.
104. Gisi Cannizzaro (2012). *Early word order and animacy*.
105. Kostadin Cholakov (2012). *Lexical Acquisition for Computational Grammars. A Unified Model*.

106. Karin Beijering (2012). *Expressions of epistemic modality in Mainland Scandinavian. A study into the lexicalization-grammaticalization-pragmaticalization interface.*
107. Veerle Baaijen (2012). *The development of understanding through writing.*
108. Jacolien van Rij (2012). *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults.*
109. Ankelien Schippers (2012). *Variation and change in Germanic long-distance dependencies.*
110. Hanneke Loerts (2012). *Uncommon gender: Eyes and brains, native and second language learners, & grammatical gender.*
111. Marjoleine Sloos (2013). *Frequency and phonological grammar: An integrated approach. Evidence from German, Indonesian, and Japanese.*
112. Aysa Arylova. (2013) *Possession in the Russian clause. Towards dynamicity in syntax.*
113. Daniël de Kok (2013). *Reversible Stochastic Attribute-Value Grammars.*
114. Gideon Kotzé (2013). *Complementary approaches to tree alignment: Combining statistical and rule-based methods.*
115. Fridah Katushemererwe (2013). *Computational Morphology and Bantu Language Learning: an Implementation for Runyakitara.*
116. Ryan C. Taylor (2013). *Tracking Referents: Markedness, World Knowledge and Pronoun Resolution.*
117. Hana Smiskova-Gustafsson (2013). *Chunks in L2 Development: A Usage-based Perspective.*
118. Milada Walková (2013). *The aspectual function of particles in phrasal verbs.*
119. Tom O. Abuom (2013). *Verb and Word Order Deficits in Swahili-English bilingual agrammatic speakers.*
120. Gülsen Yılmaz (2013). *Bilingual Language Development among the First Generation Turkish Immigrants in the Netherlands.*
121. Trevor Benjamin (2013). *Signaling Trouble: On the linguistic design of other-initiation of repair in English conversation.*
122. Nguyen Hong Thi Phuong (2013). *A Dynamic Usage-based Approach to Second Language Teaching.*
123. Harm Brouwer (2014). *The Electrophysiology of Language Comprehension: A Neurocomputational Model.*
124. Kendall Decker (2014). *Orthography Development for Creole Languages.*
125. Laura S. Bos (2015). *The Brain, Verbs, and the Past: Neurolinguistic Studies on Time Reference.*
126. Rimke Groenewold (2015). *Direct and indirect speech in aphasia: Studies of spoken discourse production and comprehension.*
127. Huiping Chan (2015). *A Dynamic Approach to the Development of Lexicon and Syntax in a Second Language.*
128. James Griffiths (2015). *On appositives.*
129. Pavel Rudnev (2015). *Dependency and discourse-configurationality: A study of Avar.*

130. Kirsten Kolstrup (2015). *Opportunities to speak. A qualitative study of a second language in use.*
131. Güliz Güneş (2015). *Deriving Prosodic structures.*
132. Cornelia Lahmann (2015). *Beyond barriers. Complexity, accuracy, and fluency in long-term L2 speakers' speech.*
133. Sri Wachyunni (2015). *Scaffolding and Cooperative Learning: Effects on Reading Comprehension and Vocabulary Knowledge in English as a Foreign Language.*
134. Albert Walsweer (2015). *Ruimte voor leren. Een etnogafisch onderzoek naar het verloop van een interventie gericht op versterking van het taalgebruik in een knowledge building environment op kleine Friese basisscholen.*
135. Aleyda Lizeth Linares Calix (2015). *Raising Metacognitive Genre Awareness in L2 Academic Readers and Writers.*
136. Fathima Mufeeda Irshad (2015). *Second Language Development through the Lens of a Dynamic Usage-Based Approach.*
137. Oscar Strik (2015). *Modelling analogical change. A history of Swedish and Frisian verb inflection.*
138. He Sun (2015). *Predictors and stages of very young child EFL learners' English development in China.*
139. Marieke Haan (2015). *Mode Matters. Effects of survey modes on participation and answering behavior.*
140. Nienke Houtzager (2015). *Bilingual advantages in middle-aged and elderly populations.*
141. Noortje Joost Venhuizen (2015). *Projection in Discourse: A data-driven formal semantic analysis.*
142. Valerio Basile (2015). *From Logic to Language: Natural Language Generation from Logical Forms.*
143. Jinxing Yue (2016). *Tone-word Recognition in Mandarin Chinese: Influences of lexical-level representations.*
144. Seçkin Arslan (2016). *Neurolinguistic and Psycholinguistic Investigations on Evidentiality in Turkish.*
145. Rui Qin (2016). *Neurophysiological Studies of Reading Fluency. Towards Visual and Auditory Markers of Developmental Dyslexia.*
146. Kashmiri Stec (2016). *Visible Quotation: The Multimodal Expression of Vieuipoint.*
147. Yinxing Jin (2016). *Foreign language classroom anxiety: A study of Chinese university students of Japanese and English over time.*
148. Joost Hurkmans (2016). *The Treatment of Apraxia of Speech. Speech and Music Therapy, an Innovative Joint Effort.*
149. Franziska Köder (2016). *Between direct and indirect speech: The acquisition of pronouns in reported speech.*

150. Femke Swarte (2016). *Predicting the mutual intelligibility of Germanic languages from linguistic and extra-linguistic factors.*
151. Sanne Kuijper (2016). *Communication abilities of children with ASD and ADHD. Production, comprehension, and cognitive mechanisms.*
152. Jelena Golubović (2016). *Mutual intelligibility in the Slavic language area.*
153. Nynke van der Schaaf (2016). *"Kijk eens wat ik kan!" Sociale praktijken in de interactie tussen kinderen van 4 tot 8 jaar in de buitenschoolse opvang.*
154. Simon Šuster (2016). *Empirical studies on word representations.*
155. Kilian Evang (2016). *Cross-lingual Semantic Parsing with Categorical Grammars.*
156. Miren Arantzeta Pérez (2017). *Sentence comprehension in monolingual and bilingual aphasia: Evidence from behavioral and eye-tracking methods.*
157. Sana-e-Zehra Haidry (2017). *Assessment of Dyslexia in the Urdu Language.*
158. Srđan Popov (2017). *Auditory and Visual ERP Correlates of Gender Agreement Processing in Dutch and Italian.*
159. Molood Sadat Safavi (2017). *The Competition of Memory and Expectation in Resolving Long-Distance Dependencies: Psycholinguistic Evidence from Persian Complex Predicates.*
160. Christopher Bergmann (2017). *Facets of native-likeness: First-language attrition among German emigrants to Anglophone North America.*
161. Stefanie Keulen (2017). *Foreign Accent Syndrome: A Neurolinguistic Analysis.*
162. Franz Manni (2017). *Linguistic Probes into Human History.*
163. Margreet Vogelzang (2017). *Reference and cognition: Experimental and computational cognitive modeling studies on reference processing in Dutch and Italian.*
164. Johannes Bjerva (2017). *One Model to Rule them all. Multitask and Multilingual Modelling for Lexical Analysis: Multitask and Multilingual Modelling for Lexical Analysis.*
165. Dieke Oele (2018). *Automated translation with interlingual word representations.*
166. Lucas Seuren (2018). *The interactional accomplishment of action.*
167. Elisabeth Borleffs (2018). *Cracking the code - Towards understanding, diagnosing and remediating dyslexia in Standard Indonesian.*
168. Mirjam Günther-van der Meij (2018). *The impact of degree of bilingualism on L3 development English language development in early and later bilinguals in the Frisian context.*
169. Ruth Koops van 't Jagt (2018). *Show, don't just tell: Photo stories to support people with limited health literacy.*
170. Bernat Bardagil-Mas (2018). *Case and agreement in Panará.*
171. Jessica Overweg (2018). *Taking an alternative perspective on language in autism.*
172. Lennie Donné (2018). *Convincing through conversation: Unraveling the role of interpersonal health communication in health campaign effectiveness.*
173. Toivo Glatz (2018). *Serious games as a level playing field for early literacy: A behavioural and neurophysiological evaluation.*

174. Ellie van Setten (2019). *Neurolinguistic Profiles of Advanced Readers with Developmental Dyslexia*.
175. Anna Pot (2019). *Aging in multilingual Netherlands: Effects on cognition, wellbeing and health*.
176. Audrey Rousse-Malpat (2019). *Effectiveness of explicit vs. implicit L2 instruction: a longitudinal classroom study on oral and written skills*.
177. Rob van der Goot (2019). *Normalization and Parsing Algorithms for Uncertain Input*.
178. Azadeh Elmianvari (2019). *Multilingualism, Facebook and the Iranian diaspora*.
179. Joëlle Ooms (2019). *"Don't make my mistake": Narrative fear appeals in health communication*.
180. Annerose Willemsen (2019). *The floor is yours: A conversation analytic study of teachers' conduct facilitating whole-class discussions around texts*.
181. Frans Hiddink (2019). *Early childhood problem-solving interaction: Young children's discourse during small-group work in primary school*.
182. Hessel Haagsma (2020). *A Bigger Fish to Fry: Scaling up the Automatic Understanding of Idiomatic Expressions*.
183. Juliana Andrade Feiden (2020). *The Influence of Conceptual Number in Coreference Establishing: An ERP Study on Brazilian and European Portuguese*.
184. Sirkku Lesonen (2020). *Valuing variability: Dynamic usage-based principles in the L2 development of four Finnish language learners*.
185. Nathaniel Lartey (2020). *A neurolinguistic approach to the processing of resumption in Akan focus constructions*.
186. Bernard Amadeus Jaya Jap (2020). *Syntactic Frequency and Sentence Processing in Standard Indonesian*.
187. Ting Huang (2020). *Learning an L2 and L3 at the same time: help or hinder?*.
188. Anke Herder (2020). *Peer talk in collaborative writing of primary school students: A conversation analytic study of student interaction in the context of inquiry learning*.
189. Ellen Schep (2020). *Attachment in interaction: A conversation analytic study on dinner conversations with adolescents in family-style group care*.
190. Yulia Akinina (2020). *Individual behavioural patterns and neural underpinnings of verb processing in aphasia*.
191. Camila Martinez Rebolledo (2020). *Comprehending the development of reading difficulties in children with SLI*.
192. Jakolien den Hollander (2021). *Distinguishing a phonological encoding disorder from Apraxia of Speech in individuals with aphasia by using EEG*.
193. Rik van Noord (2021). *Character-based Neural Semantic Parsing*.
194. Anna de Koster (2021). *Acting Individually or Together? An Investigation of Children's Development of Distributivity*.
195. Frank Tsiwah (2021). *Time, tone and the brain: Behavioral and neurophysiological studies on time reference and grammatical tone in Akan*.

-
196. Amélie la Roi (2021). *Idioms in the Aging Brain*.
  197. Nienke Wolthuis (2021). *Language impairments and resting-state EEG in brain tumour patients: Revealing connections*.
  198. Nienke Smit (2021). *Get it together: Exploring the dynamics of teacher-student interaction in English as a foreign language lessons*.
  199. Svetlana Averina (2021). *Bilateral neural correlates of treatment-induced changes in chronic aphasia*.
  200. Wilasinee Siriboonpipattana (2021). *Neurolinguistic studies on the linguistic expression of time reference in Thai*.
  201. Irene Graafsma (2021). *Computer programming skills: a cognitive perspective*.
  202. Pouran Seifi (2021). *Processing and comprehension of L2 English relative clauses by Farsi speakers*.
  203. Hongying Peng (2021). *A Holistic Person-Centred Approach to Mobile-Assisted Language Learning*.
  204. Nermina Cordalija (2021). *Neurolinguistic and psycholinguistic approaches to studying tense, aspect, and unaccusativity*.
  205. Aida Salčić (2021). *Agreement processing in Dutch adults with dyslexia*.
  206. Eabele Tjepkema (2021). *Exploring content-based language teaching practices to stimulate language use in grades 7 and 8 of Frisian trilingual primary education*.
  207. Liefke Reitsma (2021). *Bilingualism and Contact-Induced Language Change: Exploring variation in the Frisian verbal complex*.
  208. Steven Gilbers (2021). *Ambitionz az a Ridah: 2Pac's Changing Accent and Flow in Light of Regional Variation in African-American English Speech and Hip-Hop Music*.
  - 209. Leanne Nagels (2021). *From Voice to Speech: The perception of voice characteristics and speech in children with cochlear implants*.**

GRODIL

Center for Language and Cognition Groningen (CLCG)

P.O. Box 716

9700 AS Groningen

The Netherlands