Development and structure of the VariaNTS corpus

Arts, Floor; Baskent, Deniz; Tamati, Terrin N.

Published in:
Speech Communication

DOI:
10.1016/j.specom.2020.12.006

Speech Communication 127 (2021) 64–72

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Development and structure of the VariaNTS corpus: A spoken Dutch corpus containing talker and linguistic variability

Floor Arts[a,b], Deniz Başkent[a,b], Terrin N. Tamati[a,b,c,*]

[a] Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, Hanzeplein 1, Postbox 30.001, 9700 RB, Groningen, the Netherlands
[b] Research School of Behavioral and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, BCN-Office, FA 30, Postbox 196, 9700 AD, Groningen, the Netherlands
[c] Department of Otolaryngology – Head and Neck Surgery, The Ohio State University Wexner Medical Center, 915 Olentangy River Rd., Columbus, OH, USA

ARTICLE INFO

Keywords:
Spoken Dutch
Speech corpus
Talker variability
Linguistic variability

ABSTRACT

Speech perception and spoken word recognition are not only affected by *what* is being said, but also by *who* is speaking. Currently, publicly available corpora of spoken Dutch do not offer a wide variety of linguistic materials produced by multiple talkers. The VariaNTS (Variatie in Nederlandse Taal en Sprekers) corpus is a Dutch spoken corpus that was developed to maximize both linguistic and talker variability. It contains 1000 items from 11 linguistic subcategories, recorded by 8 male and 8 female native speakers of standard Dutch. The corpus contains audio recordings, orthographic transcriptions, item-specific details such as word frequencies, neighborhood densities and phonotactic probabilities, and talker details. The VariaNTS corpus aims to provide new materials to be used for broad assessment of speech perception and word recognition in Dutch clinical and academic settings.

1. Introduction

In real-world speech communication, outside the research lab or clinic, listeners must be able to deal with a wide range of adverse and challenging listening conditions. Real-world listening conditions involve the recognition of speech that varies in linguistic characteristics, often heard in the presence of noise or competing talkers, and produced by diverse talkers (Mattys et al., 2012; Gilbert et al., 2013). Thus, successful speech recognition depends on the linguistic content of the utterance as well as the characteristics of the talker, among other factors.

In everyday conversations, listeners use linguistic context to facilitate speech recognition. Grammatical and semantic context, within and across sentences, enable prediction and restoration of words or speech sounds that are inaccessible to the listener. For example, in the sentence *He ate his soup with a* __, the final word *spoon* is predictable from the preceding context. Highly predictable words, based on context, are easier to recognize than words that are less predictable from context (e.g., Duffy and Giolas, 1974; Kalikow et al., 1977), even when presented in anomalous (nonsense) sentences that lack semantic context information (Miller and Isard, 1963). Sentences with neutral verbs, such as *to say, to give*, or *to hold*, do not provide the listener with enough semantic context information to aid in the recognition of speech. Semantic context facilitates speech perception.

The linguistic characteristics of individual words also impact successful speech understanding. Listeners must match the speech input to the correct lexical entry in the mental lexicon, selecting the perceived word from several other, acoustically similar words. Lexical frequency and density characteristics have been found to have a large effect on spoken word recognition (Howes, 1957; Savin, 1963; Luce and Pisoni, 1998). High-frequency words are easier to identify than low-frequency words (Howes, 1957; Savin, 1963; Luce and Pisoni, 1998), and words that are phonetically unique with few neighbors (i.e., phonetically similar words) are easier to recognize than words with many neighbors (Luce and Pisoni, 1998). In contrast, nonwords are not represented in the lexicon, requiring adequate perception of speech sounds for accurate recognition. As such, accurate recognition depends greatly on the combination of sounds contained in the nonwords and their frequency of occurrence in the language (i.e., phonotactic probability) and their relation to typical phonological sequences in real words in the language (i.e., density) (Janse and Newman, 2012).

In addition to linguistic content, the speech signal also contains rich indexical information (Abercrombie, 1967), conveying information about the talker, such as his/her identity, gender, age, or regional origin. This talker information also affects speech perception and spoken word

* Corresponding author at: Department of Otolaryngology, The Ohio State University Wexner Medical Center, Columbus, OH, USA.
E-mail addresses: f.arts@umcg.nl (F. Arts), d.baskent@umcg.nl (D. Başkent), terrin.tamati@osumc.edu (T.N. Tamati).

https://doi.org/10.1016/j.specom.2020.12.006
Received 27 August 2020; Received in revised form 23 November 2020; Accepted 8 December 2020
Available online 28 December 2020
0167-6393/

recognition (e.g., Pisoni, 1997; Mattys et al., 2012). Normal-hearing (NH) listeners are highly sensitive to talker details and make use of those details to identify a familiar talker (e.g., Van Lancker et al., 1985a; Van Lancker et al., 1985b; Kreiman and Sidtis, 2011) or a regional accent (e.g., Van Bezooijen and Gooskens, 1999; Clopper and Pisoni, 2004). In addition, listeners benefit from learning talker-specific details in recognition, and demonstrate more accurate recognition of speech produced by a familiar talker (Nygaard et al, 1994). Talker variability can also present a challenge to successful speech understanding. Recognizing speech from multiple talkers and coping with different regional or foreign accents is more demanding than speech produced by a single talker with no discernable accent. Listeners are less accurate and slower to recognize speech produced by multiple talkers compared to a single talker (e.g., Peters, 1955; Creelman, 1957; Mullennix et al., 1989). Further, listeners are less accurate at understanding speech from nonnative talkers (Bradlow and Pisoni, 1999) and talkers with an unfamiliar regional accent (e.g., Mason, 1946; Clopper and Bradlow, 2008).

Thus, linguistic and talker variability significantly affect speech perception and word recognition, and are therefore important factors to consider in the development of speech perception experiments or clinical speech perception assessments. Many existing spoken Dutch corpora available for speech perception assessment were not designed to maximize linguistic and talker variability, limiting potential research or clinical uses of existing material. A general overview of available Dutch materials is presented first.

1.1. Linguistic variability

Studies on speech perception and spoken word recognition typically involve a wide variety of spoken materials, involving multiple linguistic levels (e.g., sound, word, sentence) with control and/or manipulation of the linguistic content of the materials. Speech materials are usually developed for individual research studies; therefore, preparation of a new corpus requires careful selection of the materials, based on the linguistic and/or usage properties. Two important sources for Dutch word materials are lexical databases and frequency lists. Lexical databases such as CELEX (Burnage, 1990; Baayen et al., 1993), SUBTLEX-NL (Keuleers et al., 2010) and Referentiebestand Nederlands (Martin et al., 2005), and frequency lists such as Frequentielijsten Corpora (Instituut Nederlandse Taal, 2014), are freely available. For researchers developing materials for individual studies, words can be filtered according to specific criteria, recorded by a new set of talkers, and further customized based on the particular goals of the study. While an effective and appropriate approach to conducting speech perception experiments, developing and recording unique word lists for different individual studies is highly time-consuming. In contrast, although more time-consuming in initial development, creating a corpus of spoken materials containing a large number of speech samples, a variety of linguistic content, and multiple talkers for wide use can facilitate future research studies and present a more practical approach in the long term.

There are a number of large sets of spoken Dutch materials that are available for speech perception research. However, existing corpora were often developed to contain and/or control for only a small number of linguistic factors. One widely used source of spoken Dutch words is the Nederlandse Vereniging voor Audiologie (NVA) corpus (Bosman and Smoorenburg, 1995). The NVA corpus consists of read speech and contains 12 lists of 12 Consonant-Vowel-Consonant (CVC) words, produced by one female talker. Words are balanced for phoneme occurrence across lists. The NVA words are frequently used in clinical and research settings to assess speech recognition, based on phonemes or words correctly identified in quiet or noise. However, the wider application of NVA words is limited, given that the corpus only contains words and does not control for lexical characteristics, such as word frequency, familiarity, and neighborhood characteristics.

Some widely used spoken Dutch sentence materials are also available. Plomp and Mimpen (1979) developed a sentence corpus that contains read speech and consists of 10 lists of 13 sentences, produced in a clear style by one female talker. For sentence consistency, four criteria were applied: simple wording describing everyday situations, 8-9 syllables per sentence, no words longer than 3 syllables, and balanced phoneme occurrence within and across sentence lists. Designed for speech-in-noise testing, these criteria were chosen to obtain a good compromise between having controlled materials to facilitate testing and having speech representative of everyday speech (Smoorenburg, 1992). More recently, Versfeld et al. (2000) developed the VU sentences, following criteria similar to Plomp and Mimpen, but extending the corpus to 39 lists of 13 sentences, produced by two male and two female speakers in a normal speaking style. The VU sentences are considered to be representative of daily speech, since sentences were taken from newspapers and have different topics and grammatical structures. The VU sentences are frequently used for speech recognition assessment in hearing-impaired (HI) populations and cochlear implant (CI) listeners. Although both the VU and Plomp and Mimpen sentence corpora contain a large number of sentences with similar grammatical structure, important linguistic factors such as the semantic context or predictability of the sentences as well as the characteristics of the key words within the sentences were not considered in the design or controlled within the materials. Thus, wider application of these materials in speech perception studies may be limited.

Similar limitations arise with the LIST (Leuven Intelligibility Sentence Test; Van Wieringen and Wouters, 2008). LIST is a Flemish sentence corpus that contains 35 lists of 10 sentences, developed for both Flemish and Dutch listeners, produced by a female (Van Wieringen and Wouters, 2008) and later a male talker (Jansen et al., 2014) in a clear speaking style. Sentences were selected based on similar criteria as Plomp and Mimpen (1979), and are simple, everyday sentences containing 2-5 key words. Additionally, the words appearing in the sentences must be common to Dutch spoken in both Belgium and the Netherlands. The LIST sentences were developed for speech recognition assessment in quiet and noise, but primarily intended for use in Belgium. However, as with the above sentences, the LIST materials were not designed to contain a variety of sentence or word types and semantic predictability, and lexical characteristics were not controlled.

1.2. Talker variability

Several Dutch corpora have also been developed to capture a great deal of talker variability, and have included multiple talkers and regional dialects. However, existing multi-talker corpora were not developed to include a variety of linguistic material and/or control for linguistic factors within the materials. The Nederlandse Dialectenbank (Van Oostendorp, 2014) contains over a thousand hours of audio recorded regional dialects of Dutch within and outside the Netherlands. In order to obtain a good representation of Dutch dialects for phonetic and phonological research, recordings in the corpus are made of spontaneous speech, including spontaneous conversations, monologues, and interviews, from several speakers in regions of the Netherlands and elsewhere. Although the corpus contains speech recordings from many diverse speakers, with over 2000 speakers just from the Netherlands alone, the linguistic content of each recording is not controlled. The structure and topic of the material varies greatly by speaker, and important linguistic features of general interest in speech perception studies may not be represented in the corpus (e.g., words with varying frequency).

The Corpus of Regional Dutch Speech (Nerbonne et al., 2013) is a corpus of words and nonwords read by Dutch speakers from different regions of the Netherlands and Belgium. Again, designed for study of regiolectal differences in Dutch, male radio announcers from 8 different regions of the Netherlands and Belgium were recorded using standard Dutch (Netherlandic or Belgian Dutch) and regional speech. In this corpus, materials were more controlled across speakers, consisting of different pronunciations of 200 words and 200 nonwords for each speaker. Since the corpus contains a large number of speech samples, a variety of linguistic content, and multiple talkers, it may be useful for future speech perception studies. However, the potential uses of the corpus materials may be limited due to the inclusion of only male talkers and only word-length materials.

The Corpus Gesproken Nederlands (CGN) (Oostdijk, 2000) is a corpus that maximizes talker variability. Designed for broad use in linguistics and language and speech technology, it consists of approximately a thousand hours of free speech (about 10 million words) in total, including speech produced in different communicative settings, such as spontaneous telephone and in-person dialogues, interviews, lectures, as well as read aloud passages. Overall, materials were produced by a large number of speakers of standard Dutch, with most speaking Netherlandic Dutch speakers and a few Belgian Dutch. Given that most of the materials consist of spontaneous speech, speakers do not consistently produce the same set (or subset) of materials. Additionally, the content and the linguistic characteristics of the materials were not controlled. Thus, although CGN is widely used for the linguistic analysis of spoken Dutch, it was not specifically designed for speech perception research and researchers would need to select recordings and compose their own balanced stimulus lists.

The Dutch Polyphone Corpus (Den Os et al., 1995) is another corpus containing a large number of speakers, but with more controlled materials. The corpus was originally designed to use in the development of speech recognition systems, and contains read and spontaneous speech from by 5050 different speakers of Netherlandic Dutch (2434 female/2616 male). Each speaker produced 50 items, including 32 read items, both sentences and numbers, as well as 4 spontaneous responses to probe questions. The sentence materials were designed to be phonetically rich. Sentences are at least four words long, have a maximum of 80 characters, and contain must Dutch phonemes at least once. Overall, 2500 sets of sentences containing all Dutch phonemes were constructed and distributed across speakers. However, in addition to the limited amount of speech materials for each speaker, recordings were done via telephone, resulting in poor sound quality. As such, the Dutch Polyphone Corpus is not ideally suited for speech perception research.

1.3. Combined linguistic and talker variability

Although the above corpora are centered mainly on either linguistic or talker variability, there are a small number of existing corpora for which both were considered in their development. The EUROM1 corpus (Chan et al., 1995) contains high-quality recordings of 64 different speakers of Dutch from the Netherlands. Developed to facilitate the phonetic comparison of languages, the corpus contains a variety of linguistic materials, including read passages, sentences, and words, including numbers, CVCs, and CVCs in carrier phrases (one word before/after). However, the materials are broken up into categories based on the number of talkers. A *Many Talker Corpus* consists of 3 passages, 1 block of 5 sentences, and 5 blocks of 20 numbers produced by 64 talkers (34 female/30 male). A *Few Talker Corpus* consists of 5×2 blocks of 33 CVCs, 5×5 blocks of 20 numbers, 5 blocks of 5 sentences, and 15 passages produced by a subset of 10 speakers (5 female/5male of the 64 appearing in the *Many Talker Corpus*). Finally, a *Very Few Talker Corpus* consists of 10 blocks of 33 CVC phrases produced by 4 speakers (2 female/2 male of the 10 appearing in the *Few Talker Corpus*). However, the wider application of this corpus may be limited since only a small subset of the materials was recorded by all 64 speakers, resulting in diminished talker variability and different sets of materials for different speakers.

Another corpus with potential wide application in speech perception research is the IFA Dutch Spoken Language Corpus (Van Son et al., 2001). The IFA corpus is a large corpus of words, sentences, and longer speech passages, with approximately 50,000 words in total. It contains speech representing different speaking styles, with informal stories, retold stories, as well as read speech. Therefore, IFA offers variety in both linguistic content and speaking styles. The read speech involves a longer text, sentences, nonsense sentences, words, syllables, and individual sounds. Materials were produced by 4 female and 4 male talkers; a part of the materials were produced by all talkers while the rest was variable across talkers (or subsets of talkers). Overall, the IFA corpus has several qualities that may allow for wider use, including a large number of speech samples, a variety of linguistic content, and multiple talkers. However, IFA was not developed to represent a large variety of linguistic content; it is limited in the types of sentences and words included, and does not control for semantic predictability and lexical characteristics within the sentence and word sets. For example, isolated words and sentences appear in the longer stories, and generally include only frequently occurring words in the Dutch language or short sequences designed to obtain sounds of the Dutch language (e.g., /hVd/, /VCV/). Further, the number of items repeated across all talkers is limited, thereby limiting the number of talkers that can be used if a study requires repeated sentences across talkers or conditions.

While several corpora were developed to include either linguistic and talker variability in the past twenty years, most corpora are not suited for speech perception research or are too limited for wider applications. As summarized above, the existing speech corpora were not designed to contain a large amount of speech samples, and were either developed to contain controlled linguistic materials or multiple talkers, but not both. The EUROM1 corpus (Chan et al., 1995) and IFA Dutch Spoken Language Corpus (Van Son et al., 2001) may be useful for some basic scientific or clinical applications. However, since linguistic factors and talker variability both contribute to speech perception, there is a need for a predefined set of materials, balanced for multiple linguistic factors and recorded by a substantial variety of speakers, to be used by clinicians and researchers for speech perception research.

1.4. The VariaNTS corpus

The newly developed VariaNTS corpus, Variatie in Nederlandse Taal en Sprekers (Variability in Dutch Language and Talkers), aims to fill the gap between linguistic and talker variability for the systematic assessment of spoken Dutch in academic or clinical settings. It is a spoken corpus of Dutch that was designed to maximize linguistic and talker variability. The corpus contains four types of linguistic materials, further subdivided into eleven linguistic categories. Materials are recorded by 16 different speakers, balanced for gender, and additional information is registered for each talker. The corpus provides materials for robust speech perception and recognition assessment, including real-life variation listeners need to deal with in everyday speech communication. Development, possible applications, and availability of the corpus are described.

2. Corpus development

2.1. General corpus content

The VariaNTS corpus contains four types of linguistic materials: words, nonwords, meaningful sentences, varying in predictability, and meaningless sentences, referred to as anomalous sentences. The word and nonword categories are subdivided into four subcategories, based on two linguistic features. Word subcategories are defined by word frequency, which refers to how frequently a word occurs in language, and neighborhood density (ND), which is defined as the number of existing words that can result from addition, deletion or substitution of a single phoneme (Luce and Pisoni, 1998). The nonwords are subdivided by ND, and phonotactic probability, which is a measure of similarity to real Dutch words, taking the probabilities of phoneme pairs in the nonword to occur in real Dutch words, and averaging those. Meaningful sentences are divided into two categories: sentences with high predictability and low predictability. Anomalous sentences had no subcategory. An overview of the corpus categories is presented in Table 1. Categories and linguistic features are explained in-depth in Section 2.3. All items were produced by 16 different speakers.

2.2. Linguistic materials

2.2.1. Words: corpus content

The VariaNTS corpus contains 300 Dutch words. Word forms were retrieved from the Dutch CELEX database, a multilingual lexical database developed for English, German, and Dutch, consisting of words from written texts. It contains extensive information on word orthography, phonology, morphology, syntax, and word frequency. CELEX was used as the major source for the VariaNTS corpus for three reasons. First, CELEX is an exhaustive database of 42 million words, making it a very complete representation of existing Dutch words. Second, CELEX words come from a range of text sources, i.e., dictionaries, fiction books, and non-fiction books, ensuring that estimates of lexical characteristics are more representative of the language experienced in everyday communication. Third, word criteria can be set strictly due to extensive search options, making CELEX a highly efficient database that performs high-precision word search.

The 300 words are subdivided into four categories of 75 words. The subcategories represent different lexical categories based on word frequency and ND: low frequency with low ND (LF-LD), low frequency with high ND (LF-HD), high frequency with low ND (HF-LD), and high frequency with high ND (HF-HD). To create the four lexical categories, an exhaustive list of words was retrieved from CELEX, and word frequency per million and phonological ND were determined for each word. Word frequencies were retrieved from CELEX. NDs were collected from the CLEARPOND database (Marian et al., 2012), an openly available database that provides phonological and orthographic NDs for Dutch, English, French, German, and Spanish, both within and across languages. The final 300 words for the VariaNTS corpus were selected by dividing word frequencies and NDs into three groups: high, medium, and low. Words with the lowest frequencies were selected to represent LF words, and words with the highest frequencies were selected to represent HF words. Finally, words within the HF and LF categories were further filtered based on ND (HD or LD). This way, only the most representative tokens were included in each category.

The words included in our final set have syllable structures of CVC ($N = 142$), CVCC ($N = 62$), and CCVC ($N = 96$), and word classes of noun ($N = 247$) and adjective ($N = 53$). No words ending with *w* or containing very rare phoneme sequences (e.g., *x* or *q* in initial position) are included. Word frequencies vary from 1 to 2288, and NDs range from 0 to 40. These ranges were not predefined, but were naturally established by selecting only the 300 most representative words. To make each lexical category reflect Dutch speech sounds, at least one token of each Dutch vowel and diphthong is included in each category. In the LF-HD category, the vowel *uu* and diphthong *eu* were initially not present. Two words with LF and medium ND (MD) were added, approximating the lower ND boundary of the LF-HD words (ND = 18) as closely as possible (ND = 17). This led to the inclusion of the words *buur* and *deuk* in the LF-HD category. Similarly, the word *paus* was added to the HF-LD category in order to include the vowel *au*. ND for this word approximated the upper ND boundary of the HF-LD words (ND = 8) as closely as possible (ND = 10). Within one lexical category, the maximum number of word forms containing the same vowel is 12.

To confirm that each category satisfied the word frequency and ND criteria, word frequencies and NDs were compared between subcategories. Mann-Whitney-Wilcoxon tests showed that CELEX word frequencies differed significantly between LF words (M = 1.34, SD = 0.59) and HF words (M = 164.81, SD = 268.77) (W = 22500, $p < 0.001$). This was confirmed by comparisons of individual LF and HF lexical categories (e.g., LF-LD and HF-LD) ($\chi^2(3) = 245.39$, $p < 0.001$), showing significant differences between individual LF and HF categories, but not between two LF or HF categories (i.e., LF-LD and LF-HD, and HF-LD and HF-HD).

Similar comparisons were also carried out based on word frequencies obtained from the semi-spoken SUBTLEX-NL database (Keuleers et al., 2010). One possible limitation of using CELEX as a source database is that it was developed 25 years ago, and may not be representative CELEX of contemporary Dutch language. SUBTLEX-NL database is a lexical database based on movie subtitle transcripts, written by native speakers of Dutch. SUBTLEX-NL may be more reflective of contemporary language than CELEX since it was developed more recently. In addition, movie subtitle transcripts may better reflect spoken Dutch language than written texts, because there is more room for spontaneity and improvisation in spoken language. Finally, the number of words in both databases is similar, i.e. 42 million words in CELEX and 43.7 million words in SUBTLEX-NL, so the words selected for the corpus should be present in both databases. Similar to the analyses above for CELEX word frequencies, SUBTLEX-NL frequencies differed significantly between LF words (M = 1.37, SD = 0.16) and HF words (M = 193.15, SD = 492.00) (W = 22486, $p < 0.001$). Again, comparing individual LF and HF lexical categories confirmed this finding ($\chi^2(3) = 235.23$, $p < 0.001$).

Finally, NDs calculated from the CLEARPOND database were compared between LD and HD categories. NDs differed significantly between LD words (M = 5.49, SD = 2.32) and HD words (M = 22.73, SD = 5.14) (W = 22500, $p < 0.001$). Comparisons of individual LD and HD lexical categories (e.g., LF-LD and LF-HD) confirmed this finding ($\chi^2(3) = 225.67$, $p < 0.001$). Taken together, the results validate the structure of the VariaNTS corpus, and the four lexical categories based on both CELEX and SUBTLEX-NL.

2.2.2. Words: familiarity ratings

Word frequencies have been found to relate to word familiarity, with a stronger relation for spoken word corpora than for text-based corpora (Tanaka-Ishii and Terada, 2011). To confirm this relation between word frequency and subjective familiarity, a control study was conducted, in

Table 1
Overview of the VariaNTS corpus: linguistic materials, subcategories, and total numbers of items per subcategory (N) and linguistic category (total N), per speaker and across all speakers. ND = Neighborhood Density.

Linguistic material	Subcategory	*N* (per speaker)	Total Category *N* (per speaker)	*N* (all speakers)	Total Category *N* (all speakers)
Words	Low frequency + low ND (LF-LD)	75	300	1200	4800
	Low frequency + high ND (LF-HD)	75		1200	
	High frequency + low ND (HF-LD)	75		1200	
	High frequency + high ND (HF-HD)	75		1200	
Nonwords	Low probability + low ND (LP-LD)	75	300	1200	4800
	Low probability + high ND (LP-HD)	75		1200	
	High probability + low ND (HP-LD)	75		1200	
	High probability + high ND (HP-HD)	75		1200	
Meaningful sentences	High predictability (HP)	125	250	2000	4000
	Low predictability (LP)	125		2000	
Anomalous sentences	—	150	150	2400	2400

which 10 native speakers of Dutch (7F; 3M) provided familiarity ratings for all 300 words. Participants' age ranged from 19 to 26 years. Each participant received a 10 euro voucher for participation.

Ratings were obtained for each word using a 7-point scale, ranging from *zeer onbekend* (*highly unfamiliar*) to *zeer bekend* (*highly familiar*). Familiarity was defined to the participants as knowing the meaning of the word, but also as recognizing the word in written or spoken form without knowing the exact meaning. Mean familiarity ratings per word were calculated, and Mann-Whitney-Wilcoxon tests were performed to compare the high and low frequency and ND groups.

Familiarity ratings differed significantly between HF words (M = 6.99, SD = 0.05) and LF words (M = 6.43, SD = 0.82) (W = 18520, $p < 0.001$). Comparison of individual LF and HF word categories, i.e. LF-LD with HF-LD and LF-HD with HF-HD, confirmed these findings, showing significant rating differences between the category pairs ($\chi^2(3) = 120.9$, $p < 0.001$). For ND, familiarity ratings did not differ significantly between HD words (M = 6.70, SD = 0.69) and LD words (M = 6.72, SD = 0.59). Comparison of individual LD and HD nonword categories, i.e. LD-LF with HD-LF, and LD-HF with HD-HF, confirmed the absence of significant rating differences. As expected, word familiarity differed across word frequency groups, but not across ND categories.

2.2.3. Nonwords: corpus content

The VariaNTS corpus contains 300 nonwords. The nonwords are divided into four categories of 75 nonwords, based on ND, and phonotactic probability, which is a measure of resemblance of a nonword to real words in a language. The four categories are low ND with low phonotactic probability (LD-LP), low ND with high phonotactic probability (LD-HP), high ND with low phonotactic probability (HD-LP), and high ND with high phonotactic probability (HD-HP).

To create the four nonword categories, an exhaustive list of nonwords was created manually by combining Dutch phonemes into phonotactically permissible Dutch sound sequences, and combining these sequences into CVC, CVCC and CCVC structures. For each nonword, ND and phonotactic probability were determined. Similar to word NDs, nonword NDs were retrieved from CLEARPOND. Phonotactic probabilities were calculated from biphone frequencies (BFs), which represent the frequency of two subsequent phonemes co-occurring in real Dutch words. BFs were also retrieved from CLEARPOND, which provides position-specific mean log BFs as described by Vitevitch and Luce (2004). BFs were determined for the CV and VC biphones, and in case of CVCC or CCVC structure, also for the CC biphone. The BFs were then summed into a single phonotactic probability for each nonword. For final nonword selection, a similar procedure was used as for the words, dividing the large set of nonwords into high, medium, and low ND and phonotactic probability groups. This way, only the most representative tokens were included in each category. Nonwords were first divided by ND, and second by phonotactic probability.

The nonwords included in our set have syllable structures of CVC (N = 91), CVCC (N = 131), and CCVC (N = 78). The nonwords do not look or sound like existing Dutch words or names, and all nonwords can be pronounced properly by native Dutch speakers. NDs range from 0 to 36, and phonotactic probabilities range from 0.0002 to 0.0375. Again, these ranges were not predefined, but were naturally established by selecting the 300 most representative nonwords.

To make sure the nonword categories properly represent Dutch phonotactics, each Dutch vowel and diphthong is represented at least once in each category. In the HD-HP category, the diphthong *au* was initially not present, so the nonword *plaut* was added, having a medium phonotactic probability (MP) that approximated the lower boundary of the HD-HP words (BF = 0.0122) as closely as possible (BF = 0.0075). The representation of vowels and diphthongs is well balanced within each nonword category, resulting in a maximum of 6 nonwords in each category that contain the same vowel.

Non-parametric tests were performed to confirm the representativeness of the high and low ND and phonotactic probability groups. Mann-Whitney-Wilcoxon tests revealed significantly higher phonotactic probabilities for the HP nonwords (M = 0.02, SD = 0.01) than for the LP nonwords (M = 0.00, SD = 0.00) (W = 22500, $p < 0.001$). Furthermore, NDs differed significantly between HD nonwords (M = 12.98, SD = 5.47) and LD nonwords (M = 0.38, SD = 0.56) (W = 22500, $p < 0.001$).

2.2.4. Nonwords: probability ratings

Phonotactic probability represents the frequency of phoneme sequences in existing language, and is therefore likely to be reflected in subjective ratings of nonword resemblance to real words, referred to as probability ratings (Frisch et al., 2000). To evaluate this relation, a control study was conducted in which the same 10 participants that rated word familiarity provided probability ratings for all 300 nonwords.

Probability was rated on a 7-point scale, ranging from *zeer onwaarschijnlijk* (*highly improbable*) to *zeer waarschijnlijk* (*highly probable*). Mean probability ratings per nonword were calculated, and Mann-Whitney-Wilcoxon tests were performed to compare the high and low probability and ND groups. Probability ratings did not differ significantly between HP nonwords (M=3.66, SD=0.88) and LP nonwords (M=3.58, SD=0.85). Comparison of individual HP and LP nonword categories, i.e. LP-LD with HP-LD and LP-HD with HP-HD, confirmed the absence of rating differences. This finding is not consistent with the expectation that HP nonwords would be rated as more probable than LP nonwords. For ND, probability ratings differed significantly between HD nonwords (M = 4.23, SD = 0.64) and LD nonwords (M = 3.02, SD = 0.60) (W = 20428, $p < 0.001$). Comparison of individual LD and HD nonword categories, i.e. LD-LP with HD-LP, and LD-HP with HD-HP, confirmed these findings, showing significant rating differences between the category pairs ($\chi^2(3) = 149.94$, $p < 0.001$). These findings suggest that nonword probability ratings may rely on resemblance of a nonword as a whole, rather than on the frequency of individual biphones in real language.

2.2.5. Meaningful sentences

A set of 250 meaningful sentences with variable predictability is included in the corpus, subdivided into 125 sentences with high predictability (HP) and 125 sentences with low predictability (LP). The sentences were created following the same procedure that Kalikow and colleagues (1977) used to develop the SPIN sentences for English language. In each sentence, the final word represents the *key word* of the sentence. For the VariaNTS set of sentences, 125 high-frequency words from our word set were used as key words. For each key word, one HP sentence and one LP sentence were created by hand.

The HP sentences contain three cue words per sentence. *Cue words* are nouns, adjectives, verbs, and adverbs which are highly related to the key word. As a consequence, HP sentences cue the listener towards the key word, resulting in high key word predictability. For example, in the sentence *De schoonmaakster dweilde de vieze [vloer]* (*The cleaner mopped the dirty [floor]*), the three cue words *schoonmaakster* (*cleaner*), *dweilde* (*mopped*) and *vieze* (*dirty*) serve as cue words for the key word *vloer* (*floor*). The LP sentences are neutral sentences that do not create an expectation of the key word. Thirty-one neutral Dutch verbs, i.e. verbs that do not put any constrain on the following object, were combined with singular and plural forms of neutral subjects, such as *de jongen* (*the boy*), *het meisje* (*the girl*), *de mannen* (*the men*), and *de vrouwen* (*the women*). An example is the sentence *Het meisje geeft haar vriend een [fiets]* (*The girl gives her friend a [bicycle]*), in which the neutral words *meisje (girl), geven* (*to give*) and *vriend (friend)* do not create a clear expectation of the key word, resulting in low predictability context for the final key word. Neutral subjects and verbs were randomly selected for each LP sentence.

All 250 sentences meet four criteria, based on the development of the SRT sentences by Plomp and Mimpen (1979), and the VU sentences by Versfeld et al. (2000). First, each sentence contains four content words in total, including the key word. Second, to create a good compromise between everyday speech and testing efficiency, sentence length is equalized at 5-8 words per sentence. Third, sentences are declarative sentences, i.e. no questions or exclamations are included. Finally, sentences contain no capitals other than the sentence-initial one, excluding names, and no punctuation characters are used, avoiding undesired pauses within the sentence.

To confirm high and low predictability in the sentences, the 10 participants mentioned earlier provided their best guess of each key word. They were presented with a larger set of 300 sentences, consisting of an HP and LP sentence designed specifically for each of the 150 high-frequency words in the corpus, with the key word replaced by a blank space. Participants filled in the first monosyllable that came to mind to complete the sentence. Scores were calculated for each sentence, where score was defined as the total number of correct target key words. For HP sentences, sentences with scores ranging from 5 to 10 were judged as predictable. For LP sentences, sentences with a score of 0 or 1 were judged to be unpredictable. HP and LP sentences for 125 key words met the inclusion scores and were included in the final corpus, resulting in a final set of 250 sentences, of which 125 are HP sentences (M = 8.47, SD = 1.63), and 125 are LP sentences (M = 0.10, SD = 0.30).

2.2.6. *Anomalous sentences*

A set of 150 anomalous sentences was created based on the high-predictability sentences described in Section 2.2.5. The key words were preserved in their original sentence-final position. The remaining three content words in each sentence were listed, resulting in an extensive list of different nouns, verbs, adjectives, and adverbs. All the content words in the list were shuffled randomly, so that three new content words were assigned to each sentence. Incorrect word classes were corrected manually, by interchanging the incorrect word with another word from a sentence that was also assigned an incorrect word class. Whenever the resulting sentence contained any form of meaning, as judged by the first author, one or more content words were interchanged between sentences until they became senseless.

The above mentioned 10 participants provided sensibility ratings for all 150 sentences. Participants were asked to rate each sentence on a 7-point scale, ranging from *zeer weinig betekenis* (*highly senseless*) to *zeer veel betekenis* (*highly sensible*). Mean sensibility ratings per sentence were calculated. Average mean sensibility rating across all sentences was 3.87 (SD = 0.70). Mean ratings per sentence were normally distributed and showed no outliers, indicating that all sentences were on average equally anomalous.

2.3. *Audio recordings*

2.3.1. *Speakers*

Sixteen native Dutch speakers (8M; 8F) were recorded producing the full set of linguistic materials. All speakers were judged by the first author to be speakers of standard Dutch, meaning they had no substantial audible accent. Speakers were university or higher-education students, or had graduated a maximum of one year prior to the recording session. Speakers' ages ranged from 20 to 28 years. All speakers had hearing thresholds ≤25 dB HL at frequencies between 250 and 8000 Hz, and normal or corrected vision. Participants provided signed informed consent prior to taking part in the recording session, and agreed to allow us to include the recordings in a database to be shared for research or clinical purposes. They received an hourly wage for their participation. The study was approved by the ethics committee of the University Medical Center Groningen (METc 2018/427).

Details for each individual speaker are listed in Appendix A. Speakers provided their height and weight, which indicate general speaker size. These were collected since speaker size is related to vocal characteristics, such as pitch and timbre. Fundamental frequency (F0), i.e. the rate of vocal fold vibration, is determined by the length, size, and tension of the vocal folds, and gives rise to the perception of voice pitch. The perception of timbre is related to the dispersion of formant frequencies, i.e. resonances of the vocal tract, associated with speaker height and vocal tract length (VTL). Mean F0 was calculated for each speaker after recording and postprocessing, by extracting and averaging the mean F0 across all items.

2.3.2. *Equipment*

Items were presented visually one by one on a MacBook laptop using PsyScope X Build 77 (Cohen et al., 1993). An external keyboard was attached to the Macbook via USB port, so that necessary repetitions could be indicated by the experimenter during recording. Speakers were seated in front of the Macbook screen, wearing a unidirectional SM10A head-mounted microphone (Shure Inc., Chicago, IL) positioned 1-2 cm from the left corner of the mouth. Use of a head-mounted microphone enabled speakers to move their heads in a natural way without affecting the recording settings. The microphone output was connected through a Tube Microphone Preamplifier (ART ProAudio, Niagara Falls, NY) to a MicroBook IIc digitizer (MOTU, Cambridge, MA). The digital output signal was transmitted via USB ports to the Macbook, where utterances were recorded in Audacity (Audacity Team, 2018) at a sampling rate of 44.1 kHz. Recordings of the complete sets of materials were stored in 16-bit digital WAV sound files.

2.3.3. *Procedure*

Recordings were done in a sound-attenuated booth, where both speaker and experimenter were seated. The experimenter was either the first author, or a trained intern who was supervised by the first author. One recording session took approximately 2 hours. Speakers were instructed to speak naturally, using their normal speaking volume and rate. They wore a head-mounted microphone, which was brought in place and adjusted by the experimenter.

First, a test recording was done, in which speakers read aloud '*De noordenwind en de zon*' ('*The north wind and the sun*', International Phonetic Association, 1949). This is a six-sentence text containing all Dutch phonemes, ensuring a complete evaluation of recorded speech quality. To ensure good signal quality, noise level and speech intensity of the test recording were assessed by the experimenter in Praat (Boersma and Weenink, 2018) before recording the corpus materials. In case of intensity levels under 70 dB or above 80 dB, or excessive noise as judged by the experimenter, microphone settings were re-adjusted, and the test recording and assessment were repeated.

All items were recorded in a single session. The order of the categories was the same for each speaker: words, meaningful sentences, nonwords, and finally anomalous sentences. The order of items within the categories was randomized for each speaker. Items were presented one-by-one as text on screen, and speakers read the presented item out loud. In case of pronunciation errors, dysfluencies, odd prosody, noise, or other disturbances, an item was repeated after presenting all other items of the corresponding category. Word and nonword presentation was self-paced, allowing speakers to determine the pace at which they wished to read the items. For the two sentence categories, a presentation duration was specified for each sentence: 5 seconds for each meaningful sentence, and 6 seconds for each anomalous sentence. This was done to maintain a medium speaking rate for each sentence, maximizing intelligibility and minimizing speaking errors. Each category was recorded only once. All items in one category were initially stored in a single file.

2.3.4. *Postprocessing*

Stored recordings, containing all items within one category, were cut and saved as separate WAV files. A Praat script automatically identified individual items, making use of intensity differences between speech and silence. For individual word and nonword audio files, the corresponding word or nonword was used as a file name. Sentences were coded to facilitate identification, and these codes were used as file names for individual sentence audio files. Text files were read into the Praat script, containing words, nonwords or sentence codes in the order in which they were recorded for each speaker, enabling recorded items and file names to be matched automatically. These matches were checked manually by the experimenter.

Individual files were trimmed in order to remove unnecessary silence and noises, e.g. swallowing, lip movements, breathing, or movement of the microphone cable. The onset of an utterance was defined as the onset of voice, noise, or aspiration for producing the item. The end of the utterance was a compromise between cutting off as much silence as possible, and keeping a natural sounding utterance without cutting off sound. The latter criterion was assessed auditorily by the first author.

Items that a speaker accidentally skipped, and cut files containing disturbances such as dysfluencies, odd intonation, laughing, or noise, were rerecorded in a second recording session. This session took place approximately three weeks after the first recording session, and was done with eight speakers. The same equipment, procedure, and post-processing method were used, except items were presented in a text file on the Macbook screen instead of PsyScope. This change was made because of the small numbers of items that needed to be rerecorded for each speaker. After adding the trimmed recordings from the second session, all recordings were levelled to each other by a Praat script, using an intended intensity level of 65 dB.

3. Potential applications

The VariaNTS corpus is primarily intended as a resource for the broad assessment of speech perception and spoken word recognition in Dutch, for both research and clinical purposes. In particular, the inclusion of linguistic and talker variability in the corpus will facilitate research on speech perception and spoken word recognition in conditions more reflective of real-life listening environments. The VariaNTS corpus contains a large variety of linguistic materials produced by several different male and female speakers. Although spoken corpora that incorporate a variety of linguistic materials produced by multiple speakers are more widely available in other languages, such as English (e.g., Bradlow and Bent, 2002; Smiljanic and Bradlow, 2005; Bradlow and Alexander, 2007), no such spoken corpus is yet available for spoken Dutch.

The materials in the VariaNTS corpus may be particularly useful for the development of new basic scientific or clinical speech perception studies in Dutch. In our lab, for example, we are already using the corpus to examine the processing of indexical and linguistic information in speech. We performed a voice perception study, in which easy and hard words and nonwords from the VariaNTS corpus were used to investigate NH listeners' perception of voice cues, including F0 and VTL, described above. The terms 'easy' and 'hard' refer to processing demands of the linguistic information: LF-HD words and LP-HD nonwords are expected to be more difficult to process than HF-LD words and HP-LD nonwords. By comparing perception thresholds for the three voice cues between the different linguistic categories, we investigated the effects of linguistic variation on the processing of voice cues.

Another possible application of the corpus is clinical word or sentence recognition testing. The VariaNTS corpus offers multiple types of sentences that may be used for meaningful sentence recognition assessment in Dutch. Clinical speech perception assessments that incorporate linguistic and talker variability are more widely available in other languages. In English, for example, sentence recognition assessments incorporating linguistic and talker variability have been developed, such as the Pediatric AzBio Sentence Test (Spahr et al., 2014), the AzBio Sentence Test (Spahr et al., 2012), and PRESTO (Gilbert et al., 2013). These tests have been used with clinical populations, including HI children and adults and CI users (e.g., Schafer et al., 2012; Moberly et al., 2017; Sladen et al., 2017; Hillyer et al., 2019). Both linguistic and talker variability have been shown to impact sentence recognition in adult CI users; in particular, higher-variability materials produced by many speakers and sentences with low semantic predictability or anomalous sentences present a challenge to successful sentence recognition (e.g., Moberly et al., 2018; O'Neill et al., 2019; Tamati et al., 2019).

The words and nonwords in the VariaNTS corpus may also be useful in both clinical and basic scientific research. Some clinical word recognition assessments in English control for linguistic variability, such as the Lexical Neighborhood Test (LNT; Kirk et al., 1999) and the Multisyllabic Lexical Neighborhood Test (MLNT; Kirk et al., 1999) for children. In conventional Dutch clinics, word recognition is frequently used to measure speech perception performance, but no distinction is made between easy and hard words in clinical assessments, such as in the NVA word lists (Bosman and Smoorenburg, 1995). Hard words require more fine-grained phonetic discrimination than easy words, and are more difficult to recognize. Therefore, if hard words are not considered, perception scores from easy words alone may be misleadingly high, and may not reflect a listener's true difficulties in recognizing speech.

Nonwords can be included in recognition tasks to assess perception of Dutch speech sounds. Whereas hard words still have a certain degree of predictability, because they are stored in memory and can therefore be guessed or restored, nonwords do not have this benefit. Bosman and Smoorenburg (1995) found that CVC nonwords are indeed harder to perceive than CVC real words, as evidenced by less accurate recognition of nonwords and individual phonemes within nonwords. Listeners can make estimates of subsequent speech sounds because they know the phonotactic rules of their language, but other than that, nonwords are unpredictable. Nonword repetition tasks in English clinical settings have been shown to be useful for assessing phonological processing and development in children (e.g., Dillon et al., 2004; Coady and Evans, 2008) and adults with hearing impairment. The materials in the VariaNTS corpus can be used to develop similar real word and nonword recognition tasks for Dutch.

The VariaNTS corpus includes multiple tokens of each vowel and diphthong in each word and nonword category, embedded in different consonants. This is an important factor in word and nonword perception, especially in HI and CI listeners. Vowels and diphthongs are characterized by different formant frequency distributions, and these distributions differ even for the same vowel or diphthong when surrounded by different consonants. In HI listeners, hearing loss in higher frequencies may result in weaker perception of vowels with high formant distributions, such as *ie*, and low-frequency hearing impairment may complicate perception of vowels with low formants, such as *oe* (Van Tassel and Yanz, 1987). Furthermore, HI listeners may not only use static formant frequencies to identify a vowels or diphthong, but also formant transitions, within a single vowel or diphthong, or in transition to consonants, and vice versa. Thus, VariaNTS words and nonwords can be used for speech perception assessment in listeners with varying types of hearing loss.

Everyday speech perception is not only reflected in the large variety of linguistic materials in the corpus, but also in the large number of different speakers. Additionally, natural speaking style is preserved in each talker by not providing any specific speaking instructions. Speaking style differences, such as speaking rate, F0 variability, and coarticulation effects, contribute even more to the talker variability for everyday speech communication. The VariaNTS corpus thus provides new materials for broad speech perception and spoken word recognition assessment in Dutch clinical and academic settings. Effects of linguistic and talker variability on speech perception, particularly in HI and CI listeners, can be investigated more thoroughly, contributing to a better understanding of speech perception and spoken word recognition in these listener groups.

4. Corpus availability

The VariaNTS corpus is property of the Speech Perception Lab of University Medical Center Groningen. The corpus is freely available for research purposes upon request on Zenodo: https://zenodo.org/record/3932039. Users will have access to audio recordings, orthographic transcriptions, item-specific information, and speaker information.

The trimmed audio files are stored in a separate folder for each speaker, which contain the materials organized by linguistic category (words, nonwords, meaningful sentences, and anomalous sentences). Each category folder is further subdivided into lexical or sentence subcategory folders, if applicable. Words and nonwords are listed alphabetically, and sentences are ordered by sentence code. Sentence codes and orthographic transcriptions of all items are listed in Appendix B.

For each linguistic category, orthographic transcriptions and relevant information are provided in text files. For words, lexical category, vowel class, word class, syllable structure, frequencies per million from CELEX and SUBTLEX-NL, ND, and mean familiarity rating are listed. For nonwords, nonword category, vowel class, syllable structure, ND, phonotactic probability, and mean probability rating are provided. Meaningful sentences are listed by sentence-specific code, and orthographic transcriptions, predictability category, and predictability scores are listed. Finally, anomalous sentences are listed by sentence-specific code, complemented with orthographic transcriptions and mean sensibility scores.

Speaker information is listed in separate text files for each speaker, which can be found in the individual speaker folders. Gender, age, height in cm, weight in kg, mean F0, place of birth, second language(s), regional accent, and daily occupation are provided for each speaker. One limitation that users should be aware of is that for five speakers, a limited set of one or more meaningful sentences was not recorded due to technical reasons. For these speakers, the speaker file also specifies which items are missing. Details on incomplete recordings and speaker details are provided in Appendix A.

5. Conclusion

The VariaNTS corpus is a new corpus of spoken Dutch containing a variety of speech materials for use in a wide range of scientific and clinical research studies. The VariaNTS corpus is the first Dutch corpus to offer different linguistic materials produced by multiple speakers. The corpus contains 1000 items from 4 types of linguistic materials, further subdivided into 11 subtypes, produced by 16 different speakers. It is intended to serve as a new resource for speech perception and spoken word recognition assessment, closely approximating everyday listening environments, and enabling reliable comparison of performance within and between listener groups.

Author contributions

Author Floor Arts participated in the conceptualization, data curation, formal analysis, methodology development, and writing of the manuscript – original draft and review & editing.

Author Dr. Deniz Başkent participated in the conceptualization, funding acquisition, and writing of the manuscript – original draft and review & editing

Author Dr. Terrin N. Tamati participated in the conceptualization, formal analysis, methodology development, funding acquisition, and writing of the manuscript – original draft and review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Sjors Bech for his assistance with the recruitment and recording of participants, and Roos van Doorn and Anne Nijman for their assistance with postprocessing of the audio files. This work was supported by a VICI grant (918-17-603) from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw) to Deniz Başkent, a VENI grant (275-89-035) from the NWO to Terrin N. Tamati, and funds from the Heinsius Houbolt Foundation, Amsterdam, The Netherlands. The study is part of the research program of the Otorhinolaryngology Department of the University Medical Center Groningen: Healthy Aging and Communication.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.specom.2020.12.006.

References

Abercrombie, D., 1967. Elements of General Phonetics. Edinburgh University Press, Edinburgh.
Audacity Team (2018). Audacity(R): free audio editor and recorder [Computer application]. Version 2.2.2, retrieved February 20th 2018 from https://audacityteam.org/.
Baayen, H., Piepenbrock, R., van Rijn, H., 1993. The CELEX lexical database. CD-ROM). Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
Bezooijen, R.van, Gooskens, C.S., 1999. Identification of language varieties: the contribution of different linguistic levels. J. Lang. Soc. Psychol. 18 (1), 31–48.
Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.41, retrieved 6 August 2018 from http://www.praat.org/.
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. Audiol 34 (5), 260–284.
Bradlow, A.R., Pisoni, D.B., 1999. Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors. J. Acoust. Soc. Am. 106 (4), 2074–2085.
Bradlow, A.R., Bent, T., 2002. The clear speech effect for non-native listeners. J. Acoust. Soc. Am. 112 (1), 272–284.
Bradlow, A.R., Alexander, J.A., 2007. Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. J. Acoust. Soc. Am. 121 (4), 2339–2349.
Burnage, G., 1990. CELEX. A Guide for Users. CELEX, Nijmegen.
Chan, D., Fourcin, A., Gibbon, D., Granstrom, B., Huckvale, M., Kokkinakis, G., Kvale, K., Lamel, L., Lindberg, B., Moreno, A., Mouropoulos, J., Senia, F., Trancoso, I., Veld, C., Zeiliger, J., 1995. EUROM- a spoken language resource for the EU. In: EUROSPEECH '95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology, 1, pp. 867–870.
Clopper, C.G., Pisoni, D.B., 2004. Some acoustic cues for the perceptual categorization of American english regional dialects. J. Phon. 32 (1), 111–140.
Clopper, C.G., Bradlow, A.R., 2008. Perception of dialect variation in noise: intelligibility and classification. Lang. Speech 51 (3), 175–198.
Coady, J.A., Evans, J.L., 2008. Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). Int. J. Lang. Commun. Disord. 43 (1), 1–40.
Cohen, J.D., MacWhinney, B., Flatt, M., Provost, J., 1993. PsyScope: a new graphic interactive environment for designing psychology experiments. Behav. Res. Methods Instrum. Comput. 25 (2), 257–271.
Creelman, C.D., 1957. Case of the unknown talker. J. Acoust. Soc. Am. 29 (5), 655.
Dillon, C.M., Burkholder, R.A., Cleary, M., Pisoni, D.B., 2004. Nonword repetition by children with cochlear implants: accuracy ratings from normal-hearing listeners. J Speech Lang. Hear Res. 47 (5), 1103–1116.
Duffy, J.R., Giolas, T.G., 1974. Sentence intelligibility as a function of key word selection. J. Speech Hear Res. 17 (4), 631–637.
Frisch, S.A., Large, N.R., Pisoni, D.B., 2000. Perception of Wordlikeness: effects of segment probability and length on the processing of nonwords. J. Mem. Lang. 42 (4), 481–496.
Gilbert, J.L., Tamati, T.N., Pisoni, D.B., 2013. Development, reliability, and validity of PRESTO: a new high-variability sentence recognition test. J. Am. Acad. Audiol. 24 (1), 26–36.
Hillyer, J., Elkins, E., Hazlewood, C., Watson, S.D., Arenberg, J.G., Parbery-Clark, A., 2019. Assessing cognitive abilities in high-performing cochlear implant users. Front. Neurosci. 12, 1056.
Howes, D., 1957. On the relation between the probability of a word as an association and in general linguistic usage. J. Abn. Soc. Psychol. 54 (1), 75–85.
Instituut Nederlandse Taal (2014). Frequentielijsten Corpora. Available from https://ivdnt.org/downloads/tstc-frequentielijsten-corpora.
International Phonetic Association, 1949. The Principles of the International Phonetic Association. University College, London: International Phonetic Association.
Janse, E., Newman, R.S., 2012. Identifying nonwords: effects of lexical neighborhoods, phonotactic probability, and listener characteristics. Lang. Speech 56 (4), 421–441.
Jansen, S., Koning, R., Wouters, J., van Wieringen, A., 2014. Development and validation of the Leuven intelligibility sentence test with male speaker (LIST-m). Int. J. Audiol. 53, 55–59.
Kalikow, D.N., Stevens, K.N., Elliott, L.L., 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J. Acoust. Soc. Am. 61 (5), 1337–1351.

Keuleers, E., Brysbaert, M., New, B., 2010. SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. Behav. Res. Methods 42 (3), 643–650.
Kirk, K.I., Eisenberg, L.S., Martinez, A.S., Hay-McCutcheon, M., 1999. The lexical neighborhood test: test–retest reliability and inter-list equivalency. J. Am. Acad. Audiol. 10, 113–123.
Kreiman, J., Sidtis, D., 2011. Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception. Wiley-Blackwell, Oxford.
Luce, P.A., Pisoni, D.B., 1998. Recognizing spoken words: the neighborhood activation model. Ear Hear 19 (1), 1–36.
Marian, V., Bartolotti, J., Chabal, S., Shook, A., 2012. CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. PLoS ONE 7 (8), e43230.
Martin, W., Maks, I., Bopp, S., Groot, M., 2005. Referentiebestand Nederlands (RBN). Den Haag: Taalunie.
Mason, H.M., 1946. Understandability of speech in noise as affected by region of origin of speaker and listener. Speech Monogr. 13 (2), 54–68.
Mattys, S.L., Davis, M.H., Bradlow, A.R., Scott, S.K., 2012. Speech recognition in adverse conditions: a review. Lang. Cogn. Process. 27 (7), 953–978.
Miller, G.A., Isard, S., 1963. Some perceptual consequences of linguistic rules. J. Verb. Learn. Verb. Behav. 2, 217–228.
Moberly, A.C., Pisoni, D.B., Harris, M.S., 2017. Visual working memory span in adults with cochlear implants: some preliminary findings. World J. Otorhinolaryngol. Head Neck Surg. 3 (4), 224–230.
Moberly, A.C., Vasil, K.J., Wucinich, T.L., Safdar, N., Boyce, L., Roup, C., Frush Holt, R., Adunka, O.F., Castellanos, I., Shafiro, V., Houston, D.M., Pisoni, D.B., 2018. How does aging affect recognition of spectrally degraded speech? Laryngoscope 128 (5). https://doi.org/10.1002/lary.27457. Suppl.
Mullennix, J.W., Pisoni, D.B., Martin, C.S., 1989. Some effects of talker variability on spoken word recognition. J. Acoust. Soc. Am. 85 (1), 365–378.
Nerbonne, J., Ommen, S.van, Gooskens, C., Impe, L., Kürschner, S., 2013. A corpus of regional Dutch speech. Phonet. Eur. 205–231.
Nygaard, L.C., Sommers, M.S., Pisoni, D.B., 1994. Speech perception as a talker-contingent process. Psychol. Sci. 5 (1), 42–46.
O'Neill, E.R., Kreft, H.A., Oxenham, A.J., 2019. Cognitive factors contribute to speech perception in cochlear-implant users and age-matched normal-hearing listeners under vocoded conditions. J. Acoust. Soc. Am. 146 (1), 195.
Oostdijk, N., 2000. The spoken Dutch corpus. Overview and first evaluation. In: Proc of LREC-2000, 2, pp. 887–894.
Oostendorp, M.van, 2014. Phonological and phonetic databases at the Meertens Institute. In: Durand, J., Kristoffersen, G. (Eds.), The Oxford Handbook of Corpus Phonology. Oxford: OUP, pp. 546–551.
Os, E.A.den, Boogaart, T.I., Boves, L., Klabbers, E., 1995. The Dutch polyphone corpus. In: EUROSPEECH '95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology, 1, pp. 825–828.
Peters, R.W. (1955). The relative intelligibility of single-voice and multiple-voice messages under various conditions of noise. In: Joint Project Report No. 56, U.S. Naval School of Aviation Medicine (pp. 1-9). Pensacola, FL.
Pisoni, D.B., 1997. Looking at the "Stars": a first report on the interrelations among measures of speech perception, intelligibility and language development in pediatric cochlear implant users. In: Fifth International Cochlear Implant Conference. New York, NY. May 1-3.
Plomp, R., Mimpen, A.M., 1979. Improving the reliability of testing the speech reception threshold for sentences. Audiol 18 (1), 43–52.
Savin, H.B., 1963. Wordfrequency effect and errors in the perception of speech. J. Acoust. Soc. Am. 35 (2), 200–206.
Schafer, E.C., Pogue, J., Milrany, T., 2012. List equivalency of the AzBio sentence test in noise for listeners with normal-hearing sensitivity or cochlear implants. J. Am. Acad. Audiol. 23 (7), 501–509.
Sladen, D.P., Gifford, R.H., Haynes, D., Kelsall, D., Benson, A., Lewis, K., Zwolan, T., Fu, Q., Gantz, B., Gilden, J., Westerberg, B., Gustin, C., O'Neil, L., Driscoll, C.L., 2017. Evaluation of a revised indication for determining adult cochlear implant candidacy. Laryngoscope 127 (10), 2368–2374.
Smiljanic, R., Bradlow, A.R., 2005. Production and perception of clear speech in Croatian and English. J. Acoust. Soc. Am. 118 (3), 1677–1688.
Smoorenburg, G.F., 1992. Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. J. Acoust. Soc. Am. 91 (1), 421–437.
Son, R.J.J.H.van, Binnenpoorte, D.M., Heuvel, H.van den, Pols, L., 2001. The IFA corpus: a phonemically segmented Dutch "open source" speech database. In: Proceedings EUROSPEECH, pp. 2051–2054, 2001.
Spahr, A.J., Dorman, M.F., Litvak, L.M., Van Wie, S., Gifford, R.H., Loizou, P.C., Loiselle, L.M., Oakes, T., Cook, S., 2012. Development and validation of the AzBio sentence lists. Ear Hear 33 (1), 112–117.
Spahr, A.J., Dorman, M.F., Litvak, L.M., Cook, S., Loiselle, L.M., De Jong, M.D., Hedley-Williams, A., Sunderhaus, L.S., Hayes, C.A., Gifford, R.H., 2014. Development and validation of the pediatric AzBio sentence lists. Ear Hear 35 (4), 418–422.
Tamati, T.N., Ray, C., Vasil, K.J., Pisoni, D.B., Moberly, A.C., 2019. High- and low-performing adult cochlear implant users on high-variability sentence recognition: differences in auditory spectral resolution and neurocognitive functioning. J. Am. Acad. Audiol. https://doi.org/10.3766/jaaa.18106.
Tanaka-Ishii, K., Terada, H., 2011. Word familiarity and frequency. Stud Linguist 65 (1), 96–116.
Tassel, D.J.van, Yanz, J.L., 1987. Speech recognition threshold in noise: Effects of hearing loss, frequency response, and speech materials. J. Speech Hear Res. 30, 377–386.
Van Lancker, D., Kreiman, J., Emmorey, K., 1985a. Familiar voice recognition: patterns and parameters. Part I: recognition of backward voices. J. Phon. 13 (1), 19–38.
Van Lancker, D., Kreiman, J., Wickens, T.D., 1985b. Familiar voice recognition: patterns and parameters. Part II: recognition of rate-altered voices. J. Phon. 13 (1), 39–52.
Versfeld, N.J., Daalder, L., Festen, J.M., Houtgast, T., 2000. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. J. Acoust. Soc. Am. 107 (3), 1671–1684.
Vitevitch, M.S., Luce, P.A., 2004. A Web-based interface to calculate phonotactic probability for words and nonwords in English. Behav. Res. Methods Instrum. Comput. 36 (3), 481–487.
Wieringen, A.van, Wouters, J., 2008. LIST and LINT: sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. Int. J. Audiol. 47 (6), 348–355.