How Dysarthric Prosody Impacts Naïve Listeners' Recognition

Verkhodanova, Vass; Timmermans, Sanne; Coler, Matt; Jonkers, Roel; de Jong, Bauke; Lowie, Wander

Published in:
Speech and Computer

DOI:
10.1007/978-3-030-26061-3_52

Publication date:
2019

How Dysarthric Prosody Impacts Naïve Listeners' Recognition

Vass Verkhodanova[1,2(✉)], Sanne Timmermans[3], Matt Coler[1], Roel Jonkers[4], Bauke de Jong[2,3], and Wander Lowie[2,4]

[1] Campus Fryslân, University of Groningen, Groningen, The Netherlands
{v.verkhodanova,m.coler}@rug.nl
[2] Research School of Behavioural and Cognitive Neurosciences, University of Groningen, Groningen, The Netherlands
w.m.lowie@rug.nl
[3] University Medical Center Groningen, Groningen, The Netherlands
{s.h.timmermans,b.m.de.jong}@umcg.nl
[4] Center for Language and Cognition Groningen, University of Groningen, Groningen, The Netherlands
r.jonkers@rug.nl

Abstract. The class of speech disorders known as dysarthria arise from disturbances in muscular control over the speech mechanism caused by damage of the central or peripheral nervous system. Dysarthria is typically classified into one of six classes, each corresponding to a different neurological disorder with distinct prosodic cues [3]. The assumption in this classification is that dysarthric speech can be classified implicit on the basis of perception. In this study, we investigate how accurately naïve listeners can recognize stress and intonation in dysarthric speech, and if different neurological disorders impact the ability to convey meaning with these same two cues. To those ends, we collected speech data from Dutch speakers diagnosed with cerebellar lesions (ataxic dysarthria), Parkinson's Disease (hypokinetic dysarthria), Multiple Sclerosis (mixed classes of dysarthria) and from a healthy control group. Thirteen naïve Dutch listeners participated in the perceptual experiment which targeted recognition of intended realization of four prosodic functions: lexical stress, sentence type, boundary marking and focus. We analyzed recognition accuracy for different groups and performed acoustic analyses to check for fundamental frequency trajectories. Results attest to different accuracy recognition results for different disease groups. The sentence type recognition task was the most sensitive of all tasks for differentiating different diseases both on perceptual and acoustic levels of analysis.

Keywords: Dysarthria · Prosody · Parkinson's disease · Multiple sclerosis · Spinocerebellar ataxia · Speech perception · Speech recognition

A. A. Salah et al. (Eds.): SPECOM 2019, LNAI 11658, pp. 510–519, 2019.
https://doi.org/10.1007/978-3-030-26061-3_52

1 Introduction

Dysarthria is a condition which is caused both by weakness of muscles used in speech and by difficulties in the control over them. The most common and simple description of this speech disorder is "slow speech that can be difficult to understand" [1]. Common causes of dysarthria arise from cerebral dysfunction at the level of brainstem nuclei, supra nuclear brain dysfunction or neuromuscular impairment. Neurological conditions that may lead to dysarthria include Parkinson's Disease (PD), Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS), head injury, Spinocerebellar Ataxia (SCA) and a number of others. Since dysarthria causes communication difficulties, it may lead to social deprivation and depression [1].

The seminal contribution to understanding dysarthria was made by Darley et al. [2,3], who introduced the classification system of dysarthrias. Since then the system (hereafter, the *Mayo System*) has been widely used for research and clinical purposes. The Mayo System links brain pathology based on the lesion site and perceptual speech characteristics, united in clusters of deviant speech dimensions. Despite the wide use of the system, there are doubts in its suitability for clinical purposes. For example, two independent studies tested the classification accuracy for neurologists and neurology trainees [4], and for neurologists, residents in neurology, and speech therapists [5]. Both had reported accuracy of correct classification to be from about 35% to 40%, concluding that perceptual judgement alone is not reliable, and other sources of information should be taken into consideration by clinicians.

Since then, researchers have tried to classify dysarthrias using acoustic cues to support the Mayo System. In the study by Guerra et al. [6], authors matched the acoustic measurements to the perceptual cues used by clinicians, and compared performance of two different classifiers to the clinicians' judgements on the speech corpus of different dysarthrias linked to eight neurological disorders in their corpus. The combined feature set of perceptual judgments and objective measurements provided more accurate information about the speech disturbances, while the best classifier proved to be self-organising maps (SOM), which improved the accuracy of clinicians' judgements by nearly 20% [6]. These findings indicate the value of acoustic analysis as an additional tool for clinical purposes.

There has been research dedicated to purely acoustic metrics to reliably differentiate dysarthrias. In the study by Liss et al. [12] the rhythm metrics are assessed, addressing dimension of prosody on the corpus of five different dysarthrias with different prosodic profiles. The results showed the ability of rhythm metrics to distinguish healthy speech from moderate and severe dysarthric speech as well as to discriminate dysarthria subtypes with accuracy up to 80%. The follow up study [11] investigated whether speech envelope modulation spectra, which quantifies the rhythmicity of speech within specified frequency bands, could be used for automatic analysis. Discriminant function analysis showed 84%–100% accuracy for different dysarthrias compared to all others, with hypokinetic dysarthria scoring at 100% [12].

Another study by Lansford and Liss [10] explored the dimension of articulation, focusing on the vowel metrics. They investigated whether such metrics could be used to distinguish healthy from dysarthric speech and among three different classes of dysarthria (ataxic, hypokinetic dysarthria, hyperkinetic and mixed flaccid-spastic dysarthria). All explored vowel metrics, particularly metrics that capture vowel distinctiveness, showed significant differences between dysarthric and healthy control speakers to be more sensitive and specific predictors of dysarthria. However, only the slope of the second formant (F2) demonstrated between-group differences across the dysarthrias. The second study by Lansford and Liss [9] investigated whether vowel metrics reflect the human perceptual performance, namely judging intelligibility of dysarthric speech, showing the correlation between classification by disordered vowels metrics and intelligibility judgements.

The study by Kim et al. [8] explored both dimensions of articulation and prosody simultaneously, using eight acoustic features as predictors for classification of different classes of dysarthria occurring from PD, stroke, multiple system atrophy or traumatic brain injury. Interestingly, the reported results have shown that classification accuracy by dysarthria type was typically worse than by disease type or severity, while the best classification was achieved when disease type was the grouping variable. Regarding intelligibility, F2 slope showed significance for each disease group, serving as the universal predictor. Articulation rate however was not a significant predictor of speech intelligibility for speakers with Parkinson's Disease, while it showed significance in the pooled analysis [8].

In this study, we further investigate the perceptual side of dysarthria classes. We explore if different dysarthias affect the ability of speakers to convey intended prosody. We have collected recordings of three groups of diseases - Parkinson's Disease (PD), SpinoCerebellar Ataxia group (SCA) and Multiple Sclerosis (MS) that are frequent causes of different dysarthrias, namely hypokinetic dysarthria, ataxic dysarthria and either spatic, flaccid or mixed dysarthria. Many studies have indicated that such dysarthrias have different prosodic deficit profiles [2,11,15], which, among other cues, is reflected by different disturbances of fundamental frequency (f_0).

To determine if naïve listeners could recognise intended intonation and stress patterns produced by speakers of different disease groups, we approached the question from two perspectives: first related to prosody recognition and second related to acoustics. For the former we hypothesized, that if there is a correlation between disease groups and accuracy of recognition, PD would be most prominent. For the latter, we hypothesised that f_0 would hinder the listeners' accuracy of recognition at least for PD group. To test these hypotheses we collected data (Sect. 2.1), designed a perception experiment (Sects. 2.2–2.4), and performed an acoustic and recognition accuracy analyses (Sect. 2.5).

2 Methods

2.1 Data Collection

Speech recordings were collected from 32 Dutch native speakers, 24 patients (eight per disease group) and eight control speakers. The demographics can be seen in Table 1.

Table 1. Participants demographics. Age and duration of disease are given in years

Group name	Mean age	Gender (F:M)	Diagnoses	Disease duration mean, range
PD	53.9	4:4	Idiopathic PD	mean: 11.5, range: 20
SCA	55.3	5:3	Spinocerebellar ataxia, adult form of Alexander disease, idiopathic late onset cerebellar ataxia, multiple system atrophy with cerebellar ataxia	mean:6, range: 10
MS	51.9	4:4	Primary progressive MS, secondary progressive MS, relapsing-remitting MS	mean: 13.5, range: 21
HC	56.2	4:4	–	–

Every participant except for the healthy control speakers (HC) exhibited dysarthric symptoms due to neurological disorder according to the neurologist. Speakers reported (corrected-to) normal vision and hearing and signed informed consent. Exclusion criteria for patients were cognitive problems assessed by Minimal Mental State Examination (MMSE $<$ 26), brain damage caused by stroke that inflicted aphasia and/or apraxia of speech, and language and/or (motor) speech disorders other than dysarthria. Exclusion criteria for healthy controls were cognitive problems (MMSE $<$ 26), brain damage, language and/or (motor) speech disorders. The recording sessions took place in quiet rooms at the University Medical Centre Groningen or at participants' homes with the TASCAM-DR100 recorder and an external Senheiser e865 microphone placed at around a 40 cm distance from a participant.

The collection and analysis of the material was approved by the Medical Ethics Committee of the University Medical Center Groningen.

2.2 Participants for Perceptual Experiment

Thirteen native Dutch listeners participated in the prosody recognition experiment (mean age 29). All 13 were "naïve" listeners with no prior experience with speech disorders. All participants reported normal hearing.

2.3 Stimuli

Stimuli for this study were created from a prosody task, that included exercises on four Dutch prosody functions: lexical stress, sentence type, boundary marking, and focus intonation [13]. Table 3 summarizes Dutch prosody functions and their perceptual correlates based on [13,18].

Table 2. Prosodic functions and their perceptual correlates based on [13,18]. Perceptually prominent correlates according to Rietveld and Heuven [18] are marked bold.

Function name	Description	Perceptual correlates (for undisturbed speech)	Name used in the current study
Lexical function	Discriminates between words	f_0 **change**, (vowel) duration, intensity	Lexical stress
Phrasing function	Segments the speech stream in information units	**preboundary lengthening pauses**, f_0 change	Boundary Marking
Attentional marking	Highlights the most important elements in a unit	f_0 **change**, (vowel) duration, intensity	Focus
Intentional marking	Nuances meaning	f_0 change	–
Sentence type	Discriminates between questions and statements	**general** f_0 **rise (question)**, high initial f_0 (question)	Sentence Typing
Emotional prosody	Discriminates between different emotional states	**general** f_0, f_0 **span, speech rate**	–

Four exercises included sentence completion (to elicit lexical stress and boundary intonation), repetition (boundary intonation) and the production of negative/affirmative questions and statements (sentence type). As the result, from these exercises we had created pairs of stimuli for every prosody function:

- Words segmented from the completed sentences that differ by stress placement: first or second syllable (e.g., *dóórlopen* - 'continue' and *doorlópen* - 'complete');
- Phrases syntactically identical but different in question or statement intonation (e.g.*de toets gehaald?* - '<he> passed the test?' and *de toets gehaald.* - '<he> passed the test');
- Phrases syntactically identical but different in complete/statement or incomplete/iteration intonation (e.g., *Andre houdt van honden*, <...> - 'Andre likes dogs, <...>' and *Andre houdt van honden.* - 'Andre likes dogs');

– Phrases syntactically identical but different in prosodically emphasised words - focus intonation (e.g., *ik ken haar van* ***dansles***. - 'I know her from the **dancing class**.' (as opposed to another class) and *ik ken haar van dansles.* - 'I know her from the dancing class.'.

The total amount of stimuli was 1233, 320 for the stress and for sentence type, and 310 and 283 for boundary marking and focus. Fewer stimuli for two latter functions was due to patients quitting during the last part of the protocol and due to their incorrect execution of exercise parts.

2.4 Procedure

Participants of the recognition experiment completed a recognition task in which they listened to the stimuli in four blocks corresponding to four prosody functions. Participants were told that they would hear words and phrases that were different either in stress or intonation and were asked to answer a simple question by picking one option from a list (e.g., "was the phrase question or statement?" – "(1) question, (2) statement, (3) impossible to decide"), there were always three options with one being "impossible to decide". The experiment was built within the OpenSesame program [14]. For every block, procedure consisted of a short practice session and a main part. In the practice session, to get participants acquainted with the task, they were asked to assess two stimuli of two different voices. For the main part there were 192 stimuli randomly pooled from the set representing current prosody function in such a way, that there would be six stimuli per speaker in each block. The speech samples were intensity normalized and presented over headphones (Koss Pro4S). Participants could listen to each sample only once.

2.5 Analysis

To analyse listeners' accuracy of dysarthric prosody recognition we calculated percentage of correct, incorrect and unspecified ("impossible to decide") answers along with the confidence interval (CI) estimation for the particular answers using Normal Approximation Method of the Binomial Confidence Interval.

To analyse pitch trajectories of different disease groups and healthy speakers, we assessed f_0 slopes within each stimulus. To do so, we divided each stimulus recording in two parts (the ratio between parts was 1:1 for stimuli of the lexical stress function, for other stimuli it was 7:3). For each part we calculated f_0 average derivative and calculated the difference between the parts of the recording. Pitch tracking was performed with the Talkin's RAPT algorithm [19] implemented in the SPTK toolkit for Python [17]. The RAPT algorithm identifies pitch candidates with the cross-correlation function and then attempts to select the "best fit" at each frame by dynamic programming [16,19]. RAPT results have been shown to be informative for Dutch dysarthric speech [20].

3 Results

General accuracy calculation for different disease groups did not show any striking differences, though predictably the HC group were recognized best of all, and the PD group performed worst with the highest percentage of unspecified answers. The percentage of unspecified answers was also very small for the HC group compared to other groups (see Table 3).

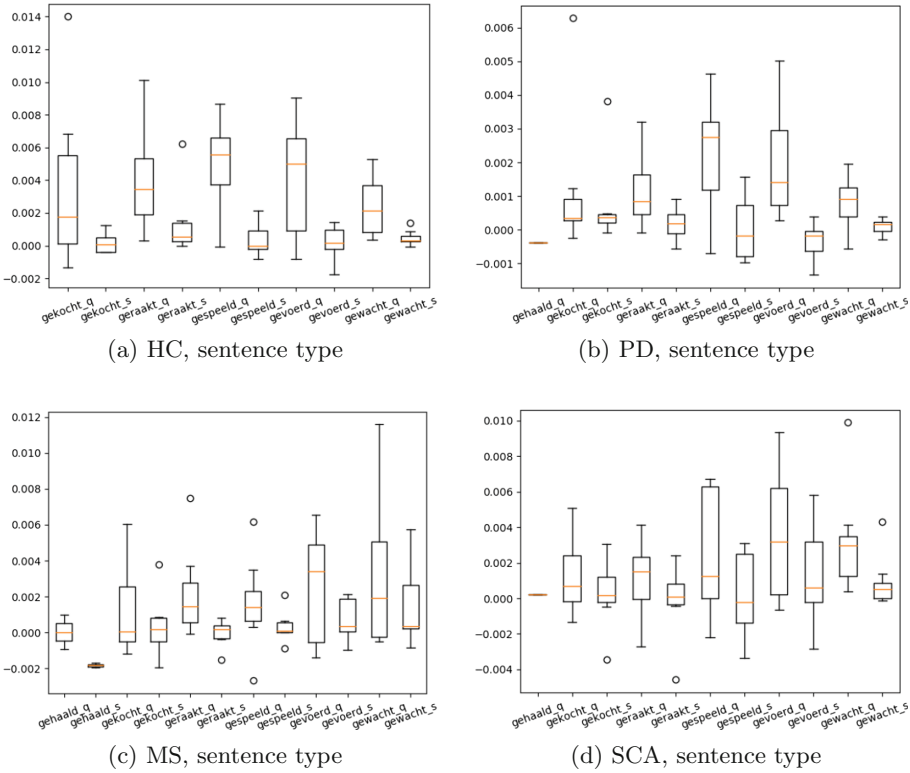
Table 3. Recognition accuracy for different disease groups and healthy speakers

Disease group	Percentage of correct answers with CI	Percentage of incorrect answers with CI	Percentage of unspecified answers with CI
HC	67 ± 1.8	27 ± 1.8	4 ± 0.8
MS	60 ± 2.0	28 ± 1.8	11 ± 1.3
SCA	56 ± 2.0	28 ± 1.8	14 ± 1.4
PD	55 ± 2.0	25 ± 1.7	18 ± 1.5

When assessing the differences for listeners' performance depending on the target prosodic function, disease groups yielded different accuracy results. Overall, boundary and focus tasks were the most difficult prosodic functions for listeners to recognise intended prosody, especially the focus where the percentage of the unspecified answers was the highest (up to 23 ± 3.4), but even those functions showed difference between healthy and dysarthric speech. Lexical stress was relatively successful for HC and MS, while SCA and PD showed lower accuracy results. Sentence type was the best recognised function for every disease group, with the smallest numbers of unspecified answers. It was also the only function where PD group did not score the worst.

Further analysis of accuracy targetted specific prosody patterns, that is first or second syllables for the lexical stress, question or statement for the sentence type, finished or unfinished intonation for boundary marking, presence or absence of focus intonation for the focus. Except for the focus, the difference between accuracy for two specific prosody patterns was very clear within each group. Questions were better recognised than statements, stressed first syllable - better than the stressed second syllable, finished intonation - better than unfinished.

To determine, if f_0 trajectories would reflect perceptual aspect, we conducted Kruskal-Wallis rank sum tests for non-parametric data to determine f_0 trajectory differences across the data. We compared differences between the f_0 derivatives for stimuli pairs. For all but one stimuli pair, significant results were found in sentence type task for two disease groups: HC and PD. Other prosodic functions did not exhibit any stable significant results within any disease group. The box plot of f_0 trajectories for sentence type function in different diseases is shown on Fig. 1.



(a) HC, sentence type

(b) PD, sentence type

(c) MS, sentence type

(d) SCA, sentence type

Fig. 1. HC and PD f_0 derivative differences in sentence typing. Difference between derivatives are placed on the y-axis, stimuli tags are placed on the x-axis: 'q' after each word means question, 's' - statement.

We also checked if there was a correlation between accuracy of listeners' recognition and speakers' disease duration, and found that there was none.

4 Discussion

In this study we explored the abilities of naïve listeners to recognitize intonation and stress patterns produced by speakers of different disease groups. We indeed found that different diagnoses, that cause different dysarthria types, affect the intelligibility of prosodic patterns differently. HC group was always distinguishable from any dysarthria groups based on the listeners' recognition accuracy. As we hypothesised, listeners performed poorest on stimuli produced by PD group (three out of four prosody function tasks). Sentence type was the function that listeners were more successful at recognising in the PD group than in the SCA group. This might be because the SCA speaker's tendency towards equalized vowel/syllable durations within utterances and unusually large f_0 range across utterances [7] interfered with their ability to mark sentence types.

Moreover, not all the tasks were found to be reliable to assess prosody deficiency. The focus recognition task was very difficult for listeners in general, causing high numbers of unspecified ("impossible to decide") answers. The sentence type recognition proved to be the clearest task, and was the only one that showed correlation with f_0 trajectories estimation. However, it is obvious that f_0 trajectories cannot act as a single predictor for different dysarthria classes or for the accuracy of listeners recognition, but it is obviously a meaningful cue for differentiating healthy and dysarthric speech.

Despite the small number of speakers and participants, and the lack of information about severity of dysarthria, we showed that assessing the naïve listeners' speech perception is potentially informative for further exploring the link between acoustic and perceptual cues for classifying different dysarthrias. Further research will target other acoustic cues such as duration, temporal cues and formant measurements that might affect listeners' prosody recognition of different dysarthric speech.

Acknowledgements. We thank Lea Busweiler, the student research assistant, for the help in the data collection and Vladimir Shapranov for assistance and advice with Python scripting. We also thank all the participants who volunteered to participate in our experiment.

References

1. Mayo Clinic: dysarthria overview. https://www.mayoclinic.org/diseases-conditions/dysarthria/symptoms-causes/syc-20371994. Accessed 10 Apr 2019
2. Darley, F.L., Aronson, A.E., Brown, J.R.: Clusters of deviant speech dimensions in the dysarthrias. J. Speech Lang. Hear. Res. **12**(3), 462–496 (1969)
3. Darley, F.L., Aronson, A.E., Brown, J.R.: Differential diagnostic patterns of dysarthria. J. Speech Lang. Hear. Res. **12**(2), 246–269 (1969)
4. Fonville, S., Van Der Worp, H., Maat, P., Aldenhoven, M., Algra, A., Van Gijn, J.: Accuracy and inter-observer variation in the classification of dysarthria from speech recordings. J. Neurol. **255**(10), 1545–1548 (2008)
5. Van der Graaff, M., et al.: Clinical identification of dysarthria types among neurologists, residents in neurology and speech therapists. Eur. Neurol. **61**(5), 295–300 (2009)
6. Guerra, E.C., Lovey, D.F.: A modern approach to dysarthria classification. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439), vol. 3, pp. 2257–2260. IEEE (2003)
7. Kent, R.D., Kent, J.F., Duffy, J.R., Thomas, J.E., Weismer, G., Stuntebeck, S.: Ataxic dysarthria. J. Speech Lang. Hear. Res. **43**(5), 1275–1289 (2000)
8. Kim, Y., Kent, R.D., Weismer, G.: An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. J. Speech Lang. Hear. Res. 417–429 (2011)
9. Lansford, K.L., Liss, J.M.: Vowel acoustics in dysarthria: mapping to perception. J. Speech Lang. Hear. Res. 68–80 (2014)
10. Lansford, K.L., Liss, J.M.: Vowel acoustics in dysarthria: speech disorder diagnosis and classification. J. Speech Lang. Hear. Res. 57–67 (2014)

11. Liss, J.M., LeGendre, S., Lotto, A.J.: Discriminating dysarthria type from envelope modulation spectra. J. Speech Lang. Hear. Res. 1246–1255 (2010)
12. Liss, J.M., et al.: Quantifying speech rhythm abnormalities in the dysarthrias. J. Speech Lang. Hear. Res. 1334–1352 (2009)
13. Martens, H., Van Nuffelen, G., Cras, P., Pickut, B., De Letter, M., De Bodt, M.: Assessment of prosodic communicative efficiency in Parkinson's disease as judged by professional listeners. Parkinson's Dis. **2011** (2011). https://doi.org/10.4061/2011/129310
14. Mathôt, S., Schreij, D., Theeuwes, J.: OpenSesame: an open-source, graphical experiment builder for the social sciences. Behav. Res. Methods **44**(2), 314–324 (2012)
15. Miller, P.H.: Dysarthria in multiple sclerosis: clinical bulletin, information for health professionals. Accessed 29 Mar 2019
16. Morrison, D., Wang, R., De Silva, L.C.: Ensemble methods for spoken emotion recognition in call-centres. Speech Commun. **49**(2), 98–112 (2007)
17. Research Group of Professor Satoshi Imai, Kobayashi, P.T.: SPTK: the speech signal processing toolkit (version 3.11). http://sp-tk.sourceforge.net/
18. Rietveld, T., Van Heuven, V.J.: Algemene Fonetiek (3e geheel herziene druk). Coutinho, Bussum (2009)
19. Talkin, D.: A robust algorithm for pitch tracking (RAPT). In: Speech Coding and Synthesis, pp. 495–518 (1995)
20. Verkhodanova, V., Coler, M.: Prosodic and segmental correlates of spontaneous dutch speech in patients with Parkinson's disease: a pilot study. In: Speech Prosody 9th International Conference, pp. 163–166 (2018)