Acoustic and perceptual speech characteristics of native Mandarin speakers with Parkinson's disease

Sih-Chiao Hsu, Yishan Jiao, Megan J. McAuliffe, Visar Berisha, Ruey-Meei Wu, and Erika S. Levy

Citation: The Journal of the Acoustical Society of America **141**, EL293 (2017); doi: 10.1121/1.4978342
View online: https://doi.org/10.1121/1.4978342
View Table of Contents: https://asa.scitation.org/toc/jas/141/3
Published by the Acoustical Society of America

Acoustic and perceptual speech characteristics of native Mandarin speakers with Parkinson's disease

Sih-Chiao Hsu
Department of Biobehavioral Sciences, Teachers College, Columbia University, 525 West 120th Street, New York, New York 10027, USA
sh3014@tc.columbia.edu

Yishan Jiao
Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287, USA
yjiao16@asu.edu

Megan J. McAuliffe
Department of Communication Disorders & New Zealand Institute of Language, Brain & Behaviour, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
megan.mcauliffe@canterbury.ac.nz

Visar Berisha
Department of Speech and Hearing Science and School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe Arizona 85287, USA
visar@asu.edu

Ruey-Meei Wu
Centre of Parkinson and Movement Disorders, Department of Neurology, National Taiwan University Hospital, College of Medicine, National Taiwan University, No.7, Chung-Shan South Road, Taipei, 100, Taiwan
robinwu@ntu.edu.tw

Erika S. Levy
Department of Biobehavioral Sciences, Teachers College, Columbia University, 525 West 120th Street, New York, New York 10027, USA
elevy@tc.columbia.edu

Abstract: This study examines acoustic features of speech production in speakers of Mandarin with Parkinson's disease (PD) and relates them to intelligibility outcomes. Data from 11 participants with PD and 7 controls are compared on several acoustic measures. In speakers with PD, the strength of association between these measures and intelligibility is investigated. Speakers with PD exhibited significant differences in fundamental frequency, pitch variation, vowel space, and rate relative to controls. However, in contrast to the English studies, speech rate was consistently slow and most strongly correlated with intelligibility. Thus, acoustic cues that strongly influence intelligibility in PD may vary cross-linguistically.

[MC]
Date Received: September 30, 2016 **Date Accepted:** February 21, 2017

1. Introduction

Individuals with Parkinson's disease (PD) commonly exhibit the motor speech disorder of hypokinetic dysarthria. To date, the majority of studies examining the key characteristics of hypokinetic dysarthria and their relationship to speech intelligibility have been conducted with speakers of English. However, extending this research to languages other than English is important for both theoretical and clinical reasons. Theoretically, an understanding of the key features influencing intelligibility across languages is required in the development of robust models of dysarthric speech—in particular when determining whether certain speech features are language-universal or language-specific.[1] For speech-language pathologists working with speakers of languages other than English, an understanding of key speech features and their language-specific consequences for intelligibility is vital in assessment and treatment planning.

In English speakers, key features of hypokinetic dysarthria include monoloudness, monotone, reduced stress, articulatory imprecision, breathy or hoarse voice, variable rate, and short rushes of speech.[2] Studies have also shown these speech features to be salient in speakers with PD who speak languages other than English, such as Spanish,[3] Korean,[4] and Japanese.[5,6] These perceptual features are often represented acoustically by vowel centralization,[7,8] a flattened F0 contour,[9] and normal or fast articulatory rate.[8] In English, the occurrence of these acoustic features has been linked to intelligibility deficits experienced by speakers with PD. In particular, measures of vowel articulation have been most closely associated with reductions in intelligibility in individuals with PD.[7,8] However, there has been limited investigation of how specific characteristics of hypokinetic dysarthria are linked to speech intelligibility in speakers of languages other than English. This is particularly apparent for Chinese.[1]

Approximately 20% of the world's languages are tonal, including Mandarin and Cantonese.[1] Tonal languages use variation in fundamental frequency (F0) to signal lexical contrasts. Hence in tonal languages, the ability to modify F0 in a language-appropriate manner is necessary to maintain intelligibility.[1,10] Given that hypokinetic dysarthria is generally characterized by impairments in F0 variation, it seems reasonable to suspect that this speech feature may have a larger influence on intelligibility in tonal languages than in non-tonal languages. To date, examination of the key speech features of hypokinetic dysarthria in tonal languages has focused predominately on Cantonese.[1] Similarly to speakers of English, Cantonese speakers exhibit monoloudness, monotone, and imprecise articulation.[1] Whitehill and Wong[10] reported reduced F0 variation at the sentence level, but relatively intact F0 variation at the syllable level in Cantonese speakers with PD. Furthermore, reduced pitch variation at the sentence level was associated with reduced speech intelligibility. To our knowledge, there have been few studies investigating the speech features of hypokinetic dysarthria in other Chinese languages[1]—such as Mandarin—or how salient speech features link to the intelligibility deficits experienced by this population. It is particularly important that studies consider Mandarin, as it differs from Cantonese in certain ways. For example, Cantonese has six tones, whereas Mandarin has four. Furthermore, Mandarin has a falling-rising tone that requires more complex pitch transitions than those seen in Cantonese (i.e., flat, falling, and rising tones only[10]). Thus, it is unclear that findings of studies focusing on Cantonese can be generalized to speakers of Mandarin. Although a study on young Mandarin adults with dysarthria due to cerebral palsy (CP) reported reduced intelligibility associated with a flattened F0 contour in their lexical tone,[11] it is not known whether lexical tone is affected in Mandarin speakers with PD and dysarthria.

The present study addressed two questions pertaining to Mandarin speakers with PD. (1) How does hypokinetic dysarthria affect measures of mean F0, F0 variation, vowel centralization, speech rate, and rate variance? (2) Is there a significant relationship between these measures and intelligibility outcomes? It should be noted that the current study focused on vowel acoustics because the Mandarin literature has documented that vowels make a greater contribution to sentence intelligibility than consonants do.[12] Moreover, studies on English speakers with dysarthria have reported a strong association between vowel distinctiveness and speech intelligibility.[7,8] It was predicted that, like English speakers with PD, Mandarin speakers with PD would exhibit vowel centralization, reductions in F0 variation, and a normal to fast speech rate in relation to healthy controls; however, it was anticipated that F0 variation would exhibit the strongest relationship with intelligibility outcomes.

2. Method

2.1 Speakers

Eleven Mandarin-speaking participants with PD (diagnosed by a neurologist) and hypokinetic dysarthria (henceforth, the PD group), and seven age- and gender-matched healthy controls (four males, three females), ranging in age from 64 to 86 years [mean (M) = 72, standard deviation (SD) = 7] (henceforth, the HC group), participated in the study. In addition, we calculated the ratio of males to females and mean age for the PD group and the results were used to match healthy controls with the same ratio and mean age. Biographical details of the participants with PD are presented in Table 1. All individuals provided written consent to participate.

2.2 Speech stimuli and recording procedure

All participants with PD and healthy controls were audio recorded in a sound-treated booth, while reading a Mandarin translation of the Rainbow Passage[13] in three

Table 1. Biographical details of the 11 Mandarin-speaking patients with Parkinson's disease.

Participant	Age	Sex	Years post diagnosis	Dysarthria severity
P1	76	M	1	Moderate
P2	62	M	10	Mild
P3	73	F	23	Severe
P4	79	F	14	Moderate-severe
P5	80	M	16	Moderate
P6	75	M	6	Moderate-severe
P7	83	M	10	Mild
P8	79	M	10	Moderate
P9	65	M	2	Moderate
P10	68	F	12	Moderate-severe
P11	68	M	18	Mild

speaking conditions (i.e., habitual speech and speech in response to prompts to speak loudly or slowly). The same recordings were used for intelligibility testing and acoustic analysis. Speech samples were recorded on a Shure SM10A head-mounted microphone placed 8 cm away from speakers' mouths and connected to a Tascam Dr-100 portable digital recorder with a sample rate of 44.1 kHz, and 16-bit resolution on a mono channel.

2.3 Listeners and measurement of EOU

Fifteen native Mandarin-speaking listeners with no prior experience with dysarthric speech rated ease of understanding (EOU) on a visual analog scale (VAS) marked with the endpoints "difficult" and "easy." This measure describes the effort required, on the part of the listener, to comprehend each speaker. EOU scores range from 0 to 100, with higher scores indicating that speech is easier to understand and lower scores indicating speech that is more difficult to understand. EOU has been used as a general index of speech intelligibility or speech severity in a number of studies of dysarthric speech[8]; thus, in the current study, EOU is also referred to as "intelligibility." As this study forms part of a larger investigation of the effect of various treatment manipulations on speech outcomes, the perceptual data were collected as part of that larger study. In the current experiment, we were interested in speech intelligibility of participants with PD only. Hence, listeners heard seven sentences produced by each of the participants with PD across three speaking conditions—habitual, loud, and slow. Seven sentences (sentences 1, 2, 3, 6, 8, 9, and 16) were extracted from the 19-sentence passage produced by each speaker. Each listener heard a total of 231 sentences (7 sentence stimuli $\times$ 3 conditions $\times$ 11 speakers with PD). The experiment was conducted free-field in a quiet room using customized computer software,[14] with all stimuli played via loudspeakers at a consistent distance from the listener. A calibration procedure was conducted, and speech stimuli were played at their actual recorded sound pressure level. For the purpose of this study on speech characteristics in Mandarin PD, listener ratings of the sentences in the habitual condition only are reported.

2.4 Automated acoustic analysis

Five acoustic features were extracted from participants' reading of the passage by means of an automatic feature extraction system described below. The pitch contour was estimated by averaging the results of three state-of-the-art pitch estimation methods only when the estimates of the three algorithms were within 10 Hz of each other.[15–17] The algorithms estimate the F0 on a 10-ms windows with 50% overlap. Pitch variation was measured by estimating the SD of the pitch contour using a 1-s window with 50% overlap. In other words, we processed each speaker's pitch contour using a 1-s sliding window and obtained multiple "local" estimates of pitch variation for each speaker. We discarded 1-s windows for which there were fewer than ten estimates of F0 and averaged the F0 SD of the remaining windows into a single pitch variation value for a speaker. Larger values of this value indicate greater local variation in F0. Vowel space area (VSA) was estimated using an automated estimation algorithm.[18,19] In brief, this method automatically extracts the values of the first and second formants from the speech stream for all voiced segments and estimates the area of the bounding polygon. The approach does not require manual segmentation and correlates strongly with the more cumbersome approach that requires hand-labeling of

Table 2. Descriptive statistics for mean fundamental frequency (F0), F0 variation, vowel space area, speech rate (SR), and standard deviation of SR for healthy controls (HC) and individuals with Parkinson's disease.

Acoustic measure	HC (n = 7)	PD (n = 11)
Mean F0 (Hz)	153.97 ± 16.14	160.71 ± 43.12
F0 variation (Hz)[a]	16.94 ± 4.43	10.28 ± 3.18
VSA (kHz^2)[a]	0.16 ± 0.08	0.04 ± 0.08
Average SR across the passage (syll/s)[a]	3.44 ± 0.23	2.58 ± 0.24
Standard deviation of SR within the passage (syll/s)[a]	0.76 ± 0.09	0.63 ± 0.05

[a]$p < 0.01$.

corner vowels.[19] Speech rate was measured using the feature-based approach.[20,21] Furthermore, the rate was estimated for every 3-s segment in the speech signal with a 2-s overlap. This allowed an estimation of the average rate for the entire passage and the variation in speech rate.

2.5 Statistical analysis

Unpaired *t*-tests were used to compare the means of the acoustic feature values between the HC and the PD groups. This allowed us to accept or reject the null hypothesis that there was no difference in the mean for a particular feature value between the two groups. We further fit multiple regression models to examine the best linear combination of the acoustic variables. We also examined the correlation between EOU score and year post diagnosis in participants with PD. It should be noted that the conventions of Cohen[21] guided interpretations regarding the strength of the correlations.

3. Results and discussion

3.1 Speakers with PD versus healthy controls

Table 2 shows the mean values for each of the two groups across the five acoustic measures. Results of two-sample *t*-tests with unequal variances revealed a number of significant differences between the two groups. Participants with PD exhibited a significant reduction in pitch variation relative to the HC group ($t = 3.45$, $p < 0.01$), indicating that participants with PD exhibited reduced local pitch variation compared to HCs. As predicted, the group with PD exhibited significant vowel centralization relative to the HC group ($t = 2.96$, $p < 0.01$), suggesting that vowel space in the PD group was constrained. The difference in speech rate between the two groups was also statistically significant; speakers with PD showed a slower speech rate than the healthy controls ($t = 7.69$, $p < 0.01$). Standard deviation of speech rate (SDSR) was significantly lower in the PD group ($t = 3.02$, $p < 0.01$), indicating that speech rate was less variable in speakers with PD. While the group with PD exhibited a trend towards higher F0 than the HC group ($t = -0.41$, $p < 0.1$), the difference was not statistically significant.

As predicted, Mandarin speakers with PD exhibited a smaller pitch range and reduced VSA compared to the HC group. These findings are consistent with previous studies, which have found that reduced intelligibility is associated with flattened F0 contour at the word level in Mandarin speakers of CP with dysarthria[11] and, furthermore, that compressed VSA is associated with less distinct vowel articulation in speakers of Spanish[3] and English.[7,8] Hence, it is reasonable to conclude that these speech feature might be language-universal features of dysarthria.

The Mandarin speakers with PD revealed a consistently slow speech rate and low SDSR relative to healthy controls. A slower speech rate has been reported for Japanese speakers with PD, relative to healthy Japanese speakers.[6] In contrast, speech rate has been reported as variable in English speakers with PD (Ref. 22) and approximately normal in German speakers with PD.[23] The discrepancy in speech rate between tonal and non-tonal languages may be associated with linguistic differences. In the tonal language of Mandarin, speakers use pitch variation to distinguish word meaning. Although Japanese is not a tonal language, Japanese speakers can signal lexical differences by means of pitch patterns along with segmental duration.[24] According to Xu,[25] achieving pitch changes in Mandarin requires time (e.g., approximately 124 ms for healthy speakers to raise or lower a pitch by four semitones); thus, the relationship between duration and pitch changes in tone languages is complex. Further investigation is needed to determine whether the differences in speech rates between speakers with PD of tonal and non-tonal languages might be related to such linguistic differences.

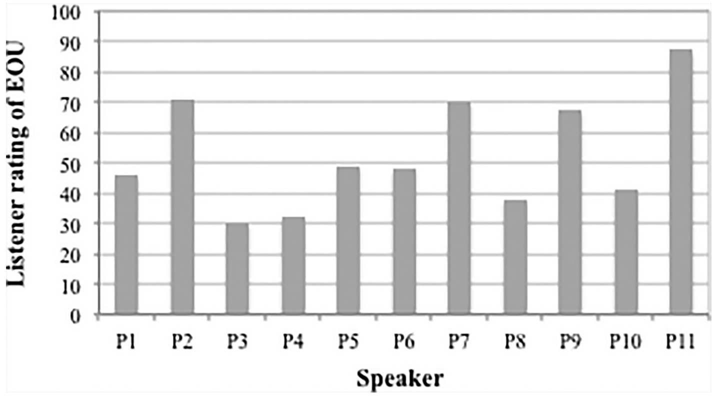

Fig. 1. Listener ratings of EOU on a scale of 0 to100 of speech produced by 11 speakers with Parkinson's disease. (Higher scores = easier to understand; lower scores = more difficult to understand.)

3.2 Speech intelligibility and multiple regression outcomes

Figure 1 demonstrates the range of EOU values exhibited by the speakers with PD. As can be seen, levels of intelligibility varied, ranging from mild to more severe. These data were subsequently used to examine the relationship between the acoustic measures and intelligibility in multiple regression analyses (see Tables 3 and 4).

We conducted a multiple regression analysis that began with a model that included all of the acoustic variables (mean pitch, pitch variation, VSA, speech rate, and rate variance) as predictors of EOU. The result of this model revealed that speech rate and rate variance were strongly correlated. Thus, we implemented a second multiple regression model that included mean pitch, pitch variation, VSA, and speech rate. We elected to use speech rate instead of rate variance because it was highly correlated with EOU ($p = 0.004$). The result of the second multiple regression model revealed that a combination of the four acoustic variables significantly predicted EOU, $F(4, 6) = 5.01$, $p = 0.041$ ($p < 0.05$), with speech rate significantly contributing to the prediction. The adjusted R squared value was 0.62, indicating that 62% of the variance in EOU was explained by the model and the effect was large.[21]

We also examined the correlation between EOU and year post diagnosis. The results of the analysis reveal that year post diagnosis was moderately, but not significantly, correlated with EOU ($r = -0.42$, $p = 0.66$), suggesting that disease progress of PD did not affect intelligibility.

The finding that speech rate most strongly correlated with intelligibility outcomes was unexpected. However, it is perhaps not surprising considering the slow rate of speech evidenced by the current cohort of participants with PD. That is, as speech rate "normalized," intelligibility improved. The literature shows that speech rate has differential effects on pitch variability. In English, as individuals with dysarthria reduce their rate of speech, reductions in pitch range at the sentence level co-occur.[26] In contrast, Xu's study on neurologically healthy Mandarin speakers[27] found that pitch peak and slope at the word level did not vary at different speech rates. Our finding is in partial agreement with Xu. That is, pitch variation at the word level did not significantly increase as rate was slowed, but mean pitch did. The findings from the previous and current studies are particularly important for speakers of Mandarin, given the importance of pitch variation to intelligibility in tonal languages.[1,10]

The finding of only a weak association between vowel space and intelligibility was of interest. In speakers of English, enhanced vowel space has consistently been

Table 3. M, SD, and intercorrelations for EOU and mean F0, F0 variation, VSA, and SR (N = 11).

Variable	M	SD	Mean F0	F0 variation	VSA	SR
EOU	52.64	18.42	−0.66[a]	0.27	0.30	0.75[b]
Mean F0	160.71	43.12	—	0.11	0.10	−0.52[a]
F0 variation	10.28	3.18	—	—	−0.33	0.14
VSA	44542.20	83949.47	—	—	—	0.10
SR	2.58	0.24	—	—	—	—

[a] $p < 0.05$.
[b] $p < 0.01$.

Table 4. Summary of multiple regression analysis for mean F0, F0 variation, VSA, and SR predicting ease of understanding. Note that $R^2 = 0.77$, $F (4, 6) = 5.01$, $p < 0.05$.

Variable	B	SEB	β
Mean F0	−0.16	0.10	−0.37
F0 variation	1.89	1.23	0.33
VSA	<0.001	<0.001	0.25
SR	37.35	18.16	0.49[a]
Constant	−40.43	56.74	

[a] $p < 0.1$.

associated with intelligibility gain[7,8]—and it appears that the same strength of association did not co-occur, at least for the cohort of participants studied in the current investigation. Therefore, the specific acoustic cues that most strongly influence intelligibility in speakers with PD appear to vary cross-linguistically. Future studies that compare these relationships across speakers of different languages but with similar levels of intelligibility deficit would be of interest.

4. Conclusion

Generally consistent with Mandarin CP[11] and English PD[7,8] studies, Mandarin speakers with PD exhibited a smaller local pitch range and reduced vowel space area compared to healthy controls. However, in contrast to the English studies, speech rate was consistently slow and most strongly correlated with intelligibility—vowel space area played a lesser role. Future studies that investigate these relationships across multiple languages are required to ascertain the language-specific effects of salient speech features in the dysarthrias.

Acknowledgments

We thank all of the participants, the Center for Parkinson and Movement Disorders at National Taiwan University Hospital, and the Ear, Nose, and Throat Department at Mackay Memorial Hospital, Taipei, Taiwan. This research was also supported by the National Institute of Deafness and Other Communication Disorders, National Institutes of Health Grant No. 1R21DC013812-01A1 from V.B. and M.J.M.

References and links

1. T. L. Whitehill and J. K-Y. Ma, "Motor speech disorders in Chinese," in *Motor Speech Disorders: A Cross-Language Perspective*, edited by N. Miller and A. Lowit (Multilingual Matters, Bristol, 2009), p. 143–153.
2. F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," J. Speech Lang. Hear. Res. **12**(2), 246–269 (1969).
3. J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Spectral and cepstral analyses for Parkinson's disease detection in Spanish vowels and words," Expert Syst. **32**(6), 688–697 (2015).
4. H. J. Shim, W. K. Park, and D. H. Ko, "파킨슨병환자의말 명료도와 모음 공간 특성" ["Characteristics of speech intelligibility and the vowel space in patients with Parkinson's disease"], Phonet. Speech Sci. **4**, 161–169 (2012).
5. H. Mori, Y. Kobayashi, H. Kasuya, N. Kobayashi, and H. Hirose, "Evaluation of fundamental frequency (F0) characteristics of speech in dysarthrias: A comparative study," Acoust. Sci. Tech. **26**(6), 540–543 (2005).
6. M. Nishio and S. Niimi, "Comparison of speaking rate, articulation rate and alternating motion rate in dysarthric speakers," Folia Phoniatr. Logop. **58**(2), 114–131 (2006).
7. M. J. McAuliffe, A. Fletcher, S. Kerr, G. O'Beirne, and T. Anderson, "Effect of dysarthria type, speaking condition and listener age on speech intelligibility," Am. J. Speech-Lang. Pathol. **26**(1), 113–123 (2017).
8. K. Tjaden and G. E. Wilding, "Rate and loudness manipulations in dysarthria acoustic and perceptual findings," J. Speech Lang. Hear. Res. **47**(4), 766–783 (2004).
9. J. S. Laures and G. Weismer, "The effects of a flattened fundamental frequency on intelligibility at the sentence level," J. Speech Lang. Hear. Res. **42**(5), 1148–1156 (1999).
10. T. L. Whitehill and L. L. N. Wong, "Effect of intensive voice treatment on tone-language speakers with Parkinson's disease," Clin. Linguist. Phonetics **21**(11-12), 919–925 (2007).
11. J. Y. Jeng, G. Weismer, and R. D. Kent, "Production and perception of mandarin tone in adults with cerebral palsy," Clin. Linguist. Phon. **20**(**1**), 67–87 (2006).
12. F. Chen, L. L. Wong, and E. Y. Wong, "Assessing the perceptual contributions of vowels and consonants to Mandarin sentence intelligibility," J. Acoust. Soc. Am. **134**(2), EL178–EL184 (2013).
13. G. Fairbanks, *Voice and Articulation Drill Book*, 2nd ed. (Harper, New York, 1960), pp. 124–139.
14. E. W. Chang and Y. M. Chang, "Word Transcription VAS & Sentence VAS (Version 1.0) [Computer software]," 2015.

[15]A. Camacho and J. G. Harris, "A sawtooth waveform inspired pitch estimator for speech and music," J. Acoust. Soc. Am. **124**(3), 1638–1652 (2008).
[16]K. Kasi and S. A. Zahorian, "Yet another algorithm for pitch tracking," in *2002 IEEE International Conference on Acoustics, Speech and Signal Processing* (2002), Vol. 1, p. I-361.
[17]L. N. Tan and A. Abeer, "Multi-band summary correlogram-based pitch detection for noisy speech," Speech Commun. **55**(7), 841–856 (2013).
[18]V. Berisha, S. Sandoval, R. Utianski, J. Liss, and A. Spanias, "Characterizing the distribution of the quadrilateral vowel space area," J. Acoust. Soc. Am. **135**(1), 421–427 (2014).
[19]Y. Jiao, V. Berisha, M. Tu, and J. Liss, "Convex weighting criteria for speaking rate estimation," IEEE Trans. Speech Audio Process. **23**(9), 1421–1430 (2015).
[20]Y. Jiao, M. Tu, V. Berisha, and J. Liss, "Online speaking rate estimation using recurrent neural networks," *IEEE International Conference on Acoustics, Speech and Signal Processing* (2016), pp. 5245–5249.
[21]J. Cohen, "*Statistical Power Analysis for the Behavioral Sciences*," 2nd ed. (Lawrence Erlbaum, Mahwah, NJ, 1988).
[22]M. P. Caligiuri, "The influence of speaking rate on articulatory hypokinesia in Parkinsonian dysarthria," Brain Lang. **36**(3), 493–502 (1989).
[23]S. Skodda and U. Schlegel, "Speech rate and rhythm in Parkinson's disease," Move. Disord. **23**(7), 985–992 (2008).
[24]Y. Hirata, "Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts," Comput. Assist. Lang. Learn. **17**(3-4), 357–376 (2004).
[25]Y. Xu, "Tone in connected discourse," in *Encyclopedia of Language and Linguistics*, edited by K. Brown (Elsevier, Oxford, UK, 2006), pp. 742–750.
[26]K. Tjaden and G. Wilding, "The impact of rate reduction and increased loudness on fundamental frequency characteristics in dysarthria," Folia Phoniatr. Logop. **63**(4), 178–186 (2010).
[27]Y. Xu, "Consistency of tone-syllable alignment across different syllable structures and speaking rates," Phonetica **55**(4), 179–203 (1998).