Acoustic Cues to Lexical Stress in Spastic Dysarthria

Heejin Kim[1], Mark Hasegawa-Johnson[2], Adrienne Perlman[3]

[1] Beckman Institute, University of Illinois, Urbana-Champaign, USA
[2] Department of ECE, University of Illinois, Urbana-Champaign, USA
[3] Department of SHS, University of Illinois, Urbana-Champaign, USA
{hkim17, jhasegaw, aperlman}@illinois.edu

Abstract

The current study examined the acoustic cues to lexical stress produced by speakers with spastic dysarthria and healthy control speakers. Of particular interest was the effect of stress location, which represented whether lexical stress was on the first vs. second syllable of the word. Results suggest that speakers with dysarthria convey lexical stress differently than do control speakers. The difference was greater for second-syllable stressed words compared to first-syllable stressed words. In addition, for the first-syllable stressed words, speakers with dysarthria utilized the pitch and intensity cues to a greater degree compared to control speakers.

Index Terms: lexical stress, dysarthric speech, acoustic cues

1. Introduction

Dysarthria refers to a neuromotor speech disorder, which can result from a variety of causes such as congenital disorders (e.g., cerebral palsy), degenerative diseases (e.g., Parkinson's disease), or traumatic brain injuries (TBI). Specific patterns of dysarthria may vary, depending on the affected brain region; however, dysarthria is often characterized by segmental deficits such as imprecise consonants and vowels. A decreased range of pitch and loudness has been noted as prosodic characteristics of dysarthria [1, 2], affecting the prosodic structures of the language such as stress, rhythm, and intonation. Recent acoustic studies on prosodic features in dysarthria have demonstrated that, despite a limited range of pitch and loudness, speakers with severe dysarthria are able to mark prosodic contrast such as sentential stress (i.e., contrastive stress on different words in a sentence) and questions vs. statements, but they utilize the acoustic cues differently than control speakers [3, 4, 5, 6, 7, 8]. The purpose of the current study was to further enhance our understanding of prosodic deficits in dysarthric speech, by examining acoustic cues to lexical stress (i.e., relative prominence of a syllable within a word).

Lexical stress plays a critical role in English; specifically, it disambiguates the word's meaning (e.g., CONtract vs. conTRACT) and facilitates language acquisition [9]. Research also suggests that applications such as automatic speech recognition and text-to-speech synthesis can benefit from incorporating lexical stress to improve the recognition accuracy and the naturalness of speech [10, 11]. It is widely agreed that lexical stress in English is marked through variations of intensity, f_0, and duration (i.e., stressed syllables have greater intensity, higher f_0, and longer duration compared to unstressed syllables) [12, 13, 14]. However, studies on normal speech reported different acoustic findings, particularly with respect to the relative roles of acoustic cues. For example, in [15], duration and intensity were more relevant than pitch for disambiguating stress-minimal words of nouns vs. verbs, while in [16], an f_0 range was the most important property of lexical stress. In addition, the overlapping degree of acoustic measures between stressed vs. unstressed syllables was large to the extent that automatic classification accuracy of stress remained approximately at 67%-84% [17, 18].

Acoustic cues to lexical stress are further complicated in relation to the location of the stressed syllable in a word. Specifically, previous studies have suggested that the relative contributions of intensity, f_0, and duration vary depending on whether lexical stress is on the first syllable or second syllable of a word. Evidence for this finding is seen in studies that have compared lexical stress production between a non-native speaker and a native speaker [19, 20] and between disordered speech and normal speech [21]. For example, in normal speech produced by healthy native American English speakers, a greater number of cues such as intensity, f_0, and duration were utilized for marking the stressed first syllable of a noun, while only duration was used for the stressed second syllable of a verb [19, 21]. On the other hand, [20] reported that duration and amplitude, but not f_0, were used to a greater degree when stress was on the first syllable rather than the second syllable.

Acoustic studies on lexical stress in dysarthria are rare, and most studies have examined the acoustic correlates of sentential stress with varied findings. According to [6], speakers with dysarthria due to cerebral palsy marked sentential stress through increased intensity to a greater degree compared to a control group, while in [7], speakers with cerebral palsy exhibited more reliance on duration with reduced f_0 and intensity variations. In a study of sentential stress produced by speakers with TBI [22], the mild TBI group exhibited fewer differences in word duration, f_0 movement, and f_0 slope between stressed and unstressed words compared to control speakers, while no differences were found for the intensity range. In [23], on the other hand, subjects with severe TBI revealed a significant reduction in the difference between stressed and unstressed words mostly in duration, but not in mean f_0 or intensity. Different findings across studies may be due to differences in methods or the dysartrhia severity levels of participants; however, much remains to be learned about the abnormal patterns of acoustic cues to prosodic contrasts in dysarthria.

The current study investigated acoustic cues to lexical stress in American English produced by speakers with spastic cerebral palsy. The following questions were addressed: 1) do speakers with dysarthria convey lexical stress differently from normal speakers?, 2) if so, by which acoustic cue do speakers with dysarthria differ mostly from normal speakers?, and 3) what word type between the first- vs. second-syllable stressed words is more deviant from normal speech? Given the perceptual impression of a decreased range of pitch and loudness in dysarthria, duration would be modulated to a greater degree for conveying lexical stress compared to pitch and intensity. Additionally, given that spastic dysarthria is often characterized by abrupt onset, the acoustic difference between stressed vs. unstressed syllables would be larger for first-syllable stress words compared to second-syllable stress words.

This study will lead to better understand prosodic deficits in dysarthic speech. In addition, improving contrastive prosody in dysarthria is beneficial not only for reducing monotony and improving naturalness but also for increasing articulatory displacement, thereby enhancing segmental contrasts as well [7]. Therefore, quantifying lexical stress abnormalities will provide insights to therapeutic strategies to gain the most enhancements in intelligibility.

2. Method

2.1. Participants

Participants were 6 male speakers diagnosed with cerebral palsy and 5 male control speakers. Speech data used for the present study was a subset of the Universal Access (UA) database collected under a project developing automatic speech recognition systems for dysarthric speech [24]. Most speakers with dysarthria were clients of the Rehabilitation Education Center at the University of Illinois at Urbana-Champaign. One speaker was recruited from the Madison, Wisconsin area. The intelligibility of participants ranged from 28% to 62%. Intelligibility rating methods are described below. Control speakers reported no history of a language disorder.

2.2. Recording procedure and material

Recording was made with an 8-microphone array, which was previously developed for the AVICAR project [25]. The array was mounted on the top of the laptop computer screen. Speech was recorded through an 8-channel Firewire audio interface (MOTU 828mkII), with a sampling rate of 48 kHz. Speakers sat comfortably in front of a laptop computer, and were asked to read an isolated word displayed on a PowerPoint slide on the computer. An experimenter sat beside the speaker and advanced the slides after the subject produced the target word.

Each speaker produced three blocks of 255 real words in English, totaling 765 words. Each block contained five categories of words: 10 digits (e.g. *zero, one*), 19 computer commands (e.g. *enter, delete*), the 26 words of the international radio alphabet (e.g. *alpha, bravo*), 100 common words (e.g. *a, the*), and 100 uncommon words (e.g. *naturalization, exploit*). The 100 common words were the most frequently occurring words in the Brown corpus of written English [26]. The 100 uncommon words were selected from children's novels digitized by Project Gutenberg [27], using an algorithm that allowed for recording of all possible sequences of phones in American English. Speakers were always given a break between blocks and were allowed additional breaks as needed.

After recording, each speaker's productions were rated for intelligibility for the purpose of acquiring a dysarthria severity index. Five naive listeners were recruited per speaker. Listeners were between 18 - 40 years old, and native speakers of American English. They had no identified language disabilities, no more than incidental experience with persons having speech disorders, and no training in phonetic transcription. Each listener transcribed a total of 225 words comprised of words from the second block recording. Listeners were informed that they would be listening to real words in American English produced by an individual with a speech disorder. They were instructed to provide orthographic transcriptions of each word. They listened to one word at a time through headphones in a quiet room. After the task was completed, the percentage of correct responses was calculated per listener, and averaged across five listeners who transcribed the same speaker's data. Based on the average percent accuracy, each speaker's intelligibility was classified into one of four categories: very low (0-25%), low (26%-50%), mid (51%-75%) and high (76%-100%).

2.3. Acoustic and statistic analysis

A total of 30 words were selected from the recording of each speaker's speech. All words were either tri- or quadra-syllabic. Bisyllabic words were excluded to control for the influence of boundary tones. Eleven words contained lexical stress on the first syllable (e.g., *anybody, episode, epithet*), while 19 words contained lexical stress on the second syllable (e.g., *banana, exaggerate, adjacent*).

The beginning and end points of the nucleus vowel in the first and second syllables were manually marked using Praat [28]. Vowels were identified by the presence of the second formant and strong glottal pulses. The following six measures were extracted from each vowel using Praat.

- Peak intensity
- Average intensity
- Integral intensity: RMS energy over the vowel duration
- Peak f_0
- Average f_0
- Duration

The f_0 measures were extracted on the 70 - 500Hz analysis range, amplitude on the 20-100 dB display range. For each measure, a ratio was obtained by dividing the value of the stressed syllable by the corresponding value of unstressed syllable, yielding one ratio for each measure per word. A ratio value greater than 1 indicated that the stressed syllable had a certain degree of prominence compared to the unstressed syllable.

Acoustic measures were subject to two-way ANOVA analyses, with the independent factors of Stress Location (first syllable vs. second syllable) and Group (speakers with dysarthria (DS) vs. control speakers (CS)). When a significant interaction was found, a one-way ANOVA was performed. First, the effect of Stress Location was examined separately for each speaker group. Second, the effect of Group was examined separately for each condition of stress location.

3. Results

3.1. Peak intensity

The peak intensity ratio was greater than 1 for both types of stress locations for both groups (Figure 1a). However, the DS group exhibited a greater peak intensity ratio when stress was on the first syllable of the word compared to second-syllable, while an opposite pattern was observed for the CS group. According to a two-way ANOVA analysis, a Stress Location x Group interaction was significant ($F(1, 325) = 32.955$, $p=.000$).

Subsequent one-way ANOVAs for the effect of Stress Location revealed that, for both groups, peak intensity ratios differed between the first vs. second-syllable stressed words (DS: $F(1, 148) = 8.853$, $p<.01$; CS: ($F(1,177) = 26.854$, $p<.001$). One-way ANOVAs for the effect of Group indicated that, for both stress locations, the DS and CS groups differed significantly (when stress was on the first syllable: $F(1, 118) = 4.819$, $p<.05$; when stress was on the second syllable: $F(1,207) = 42.787$, $p<.001$).

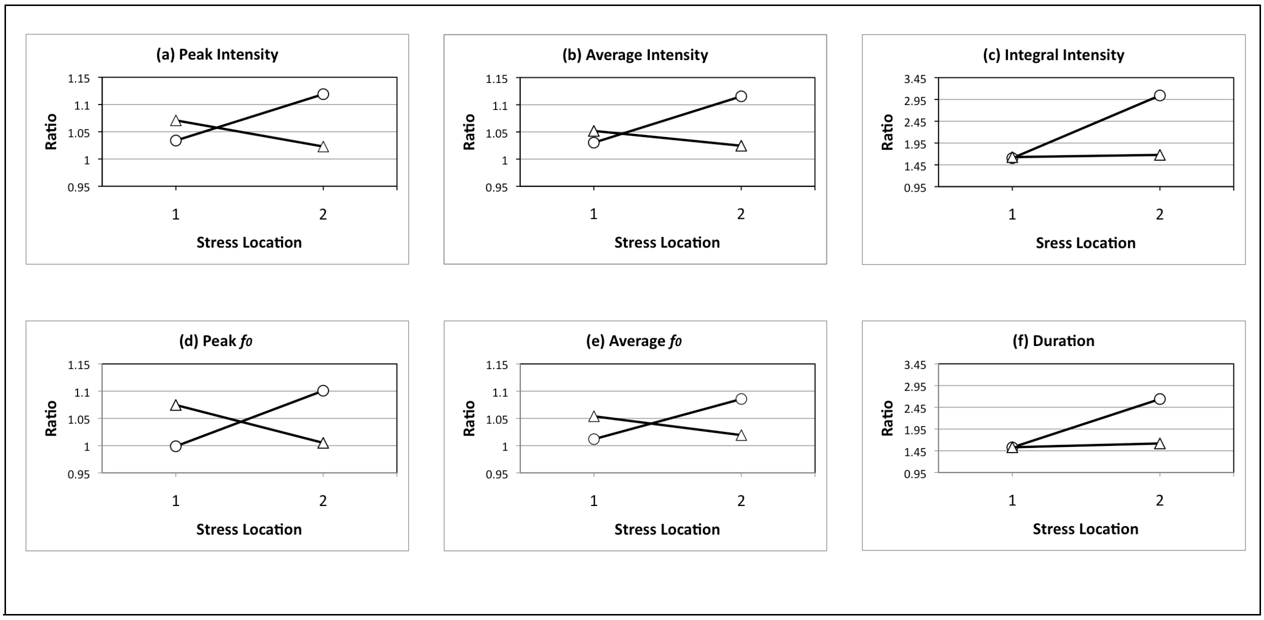

Figure 1: *Bar graphs showing the effect of Stress location and Group on each measure: -△-indicates the DS group, and-○- the CS group.*

3.2. Average intensity

Average intensity ratios were greater than 1 for all conditions (Figure 1b). A two-way ANOVA showed a significant interaction Stress Location x Group ($F(1, 325) = 30.082$, $p<.001$): the DS group indicated a greater average intensity ratio when stress was on the first syllable of the word compared to second-syllable stressed words, while an opposite pattern was observed for the CS group (Figure 1b).

One-way ANOVAs on the effect of Stress Location revealed that, for both groups, the average intensity ratio differed significantly between the first- vs. second-syllable stressed words (DS: $F(1, 148) = 4.313$, $p<.05$; CS: $(1,177) = 31.101$, $p<.001$). Regarding the effect of Group, the average intensity ratio differed significantly between the DS and CS groups only when stress was on the second syllable ($F(1,207) = 48.917$, $p<.001$).

3.3. Integral intensity

Integral intensity ratios were greater than 1 for all conditions (Figure 3c) and a Stress Location x Group interaction was significant ($F(1, 325)=14.584$, $p=.000$). One-way ANOVAs on the effect of Stress Location revealed that, only for the CS group, integral intensity ratios differed significantly between the first vs. second syllable stressed words ($F(1, 177) = 22.268$, $p<.001$). Regarding the effect of Group, the ratios differed significantly between the DS and CS groups only when stress was on the second syllable ($F(1,207) = 28.813$, $p<.001$).

3.4. Peak f_0

A peak f0 ratio was close to 1 for the DS group when stress was on the second syllable (Figure 1d). On the other hands, the ratio was lower than 1 for the CS group when stress was on the first syllable of the word, indicating that f_0 was rarely modulated for the first-syllable stress words. A two-way ANOVA confirmed that a Stress Location x Group interaction was significant ($F(1, 319) = 22.582$, $p<.001$).

One-way ANOVAs on the effect of Stress Location indicated that, for both groups, peak f_0 ratios differed significantly depending on the location of stress (DS: $F(1, 146 = 5.586$, $p<.05$: CS; $F(1, 173) = 21.576$, $p<.001$). Regarding the effect of Group, the ratios differed significantly between the DS and CS groups for both stress locations (when stress was on the first syllable: $F(1, 117) = 6.925$, $p<.05$; when stress was on the second syllable: $F(1, 202) = 19.106$, $p<.001$).

3.5. Average f_0

Similar to peak f_0 measures, average f_0 ratios were close to 1 for the CS group when stress was on the first syllable, and for the DS group when stress was on the second syllable (Figure 3e). A Stress Location x Group interaction was significant ($F(1, 319) = 9.822$, $p<.01$).

One-way ANOVAs on the effect of Stress Location revealed that, only for the CS group, the average f_0 ratio was significantly different depending on the location of stress ($F(1, 173) = 11.737$, $p<.01$). Regarding the effect of Group, the ratios differed significantly between the DS and CS groups only when stress was placed on the second syllable ($F(1, 202) = 8.684$, $p<.01$).

3.6. Duration

Duration ratios were greater than 1 for all conditions, and a Stress Location x Group interaction was significant ($F(1, 325) = 12.947$, $p<.001$). A one-way ANOVA on the effect of Stress Location indicated that, only for the CS group, the duration ratio was significantly different depending on the location of the stress ($F(1,177) = 23.339$, $p<.001$): when stress was on the second syllable, the difference in duration between stressed vs. unstressed syllables was greater compared to when stress was placed on the first syllable. Regarding the effect of Group, the duration ratios differed significantly between the DS and CS groups only for the second-syllable stressed words ($F(1,207) = 27.452$, $p<.001$).

4. Conclusions

This study provided a quantitative analysis of acoustic cues to lexical stress in individuals with spastic dysarthria. Our results indicate that speakers with CP-associated dysarthria can convey lexical stress through prosodic modulations, as seen in increased intensity, pitch and duration values on stressed syllables compared to unstressed syllables. In addition, an interaction between Stress Location and Group was significant for all measures, suggesting that variations of intensity, pitch and duration for the first- vs. second-syllable stressed words were different depending on Group (DS vs. CS). As expected from the abrupt onset in spastic dysathria, the DS group utilized the pitch and intensity cues to a greater degree for the first-syllable stressed words than did the CS group.

Previous studies have suggested that variations of intensity, f_0 and duration would differ depending on whether stress is on the first vs. second syllable of words. In the current study, evidence was seen for both groups, but the CS group exhibited a greater difference than the DS group. Specifically, for the DS group, only three measures (peak intensity, average intensity, peak f_0) were significantly different as a function of Stress Location, while for the CS group, all six measures were significantly different.

Regarding the effect of Group, the results of the current study suggest that DS and CS groups differ to a greater degree for second-syllable stressed words. Specifically, when stress was on the first syllable of the word, the DS and CS groups were significantly different only for the peak intensity and peak f_0 ratios. When stress was on the second syllable, on the contrary, the groups exhibited significant differences for all six measures.

In comparing the ratios of different measures, the lowest values were found in the pitch measures for both groups and the highest values in the duration measure. Our results suggest that, for marking lexical stress, pitch is modulated to a smaller degree compared to other cues. In the DS group, a larger ratio of duration compared to intensity and pitch measures supports a decreased range of pitch and loudness in dysarthric speech. However, given that the CS group also appears to rely on duration more than other cues, it is possible that the characteristics of the data (i.e., isolated word production) are also responsible for the relatively small values of intensity and pitch measures. A greater contribution of intensity and pitch would be found in spontaneous speech; this is left for future studies. Our findings suggest that second-syllable stress words require more therapeutic efforts than first-syllable stress words to improve lexical stress marking in dysarthric speech. Future studies with perceptual data are needed to corroborate the acoustic findings in this study and to address whether a particular cue is more relevant to the perception of lexical stress than others.

5. Acknowledgements

This study has been supported by NIH (R21-DC008090-A) grant. The authors thank the subjects for their participation.

6. References

[1] Darley, F., Aronson, A. and Brown, J., “Differential diagostic patterns of dysarthria”, JSHR 12:246-269, 1969.

[2] Darley, F., Aronson, A. and Brown, J., “Motor speech disprders”, Philadephia, PA:W.B. Saunders., 1975.

[3] Liss, J.M. and Weismer, G. Selected acoustic characteristics of contrastive stress in control geriatric, apraxic and ataxic dysarthric speakers”, Clinical Linguistics and Phonetics, 8:43-66, 1994.

[4] Patel, R., “The acoustics of contrastive prosody in adults with cerebral palsy”, Journal of Medical Speech Pathology, 12(4): 189-193, 2004.

[5] Patel, R. and Campellone, P, “Acoustic and perceptual cues to contrastive stress in dysarthria, Journal of Speech Language and Hearing Research, 52:206-222, 2009.

[6] Le Dorze, G., Ouellet, L. and Ryalls, J., “Intonation and speech rate in dysarthric speech”, Journal of communication disorders, 27:1-18, 1994.

[7] Patel, R., “Prosodic control in severe dysarthia: Preserved ability to mark the question-statement contrast”, Journal of Speech and Language, and Hearing Research, 45:858-870, 2002.

[8] Patel, R., “Acoustic characteristics of the question-statement contrast in severe dysarthria due to cerebral palsy”, Journal of Speech and Language, and Hearing Research, 46:1401-1415, 2003.

[9] Mehler et al., 1996. Cutler, A. and Norris, D., “The role of strong syllables in segmentation for lexical access”, Journal of Experimental Psychology: Human Perception and Performance, 14:113-121, 1988.

[10] Wang, C. and Seneff, S., “Lexical stress modeling for improved speech recognition of spontaneous telephone speech in the JUPITE domain”, Eurospeech Proc., 2001.

[11] Klatt, D., “Review of text-to-speech conversion for English”, JASA, 82(3):737-793, 1987.

[12] Lieberman P., “Some acoustic correlates of word stress in American English”, Journal of the Acoustical Society of America, 32: 451-454, 1960.

[13] Klatt, D.J., “Linguistic uses of segmental duration in English: acoustic and perceptual evidence”, Journal of the Acoustical Society of America, 59:1208-21, 1976.

[14] Sluijter, A. and Van Heuven, V., “Spectral balance as an acoustic correlate of linguistic stress”, JAS, 100(4):2471-2485, 1996.

[15] Sereno, J. and Jongman A., “Acoustic correlates of grammatical class”, Language and Speech, 38 (a): 57-76, 1995.

[16] Howell, P., “Cue trading in the production and perception of vowel stress’, Journal of the Acoustical Society of America, 94(4): 2063-1073, 1993.

[17] Jenkin, L. and Scordilis, M., “Development and comparison of three syllable stress classifers”, ICSLP Proc., 1996.

[18] Van Kuijk, D. and Boves, L., “Acoustic characteristics of lexical stress in continuous telephone speech”, Speech Communication, 27: 95-111, 1999.

[19] Lai, Y., Sereno, J. and Jongman, A., “Acoustic realization and perception of English lexical stress by Mandarin learners”, paper presented at the acoustic week in Canada, Vancouver, BC Canada, 2008.

[20] Zuraiq. W. and Sereno, J., “English lexical stress cues in native English and non-native Arabic speakers”, ICPhS Proc., 2007.

[21] Walker, P., Joseph, L. and Goodman, J., “The production of linguistic prosody in subjects with aphasia”, Clinical Linguistics and Phonetics, 23(7):529-549, 2009.

[22] Wang, Y., Kent, R., Duffy, J. and Thomas, J., “Dysarthria associated with traumatic brain injury: speaking rate and emphatic stress”, Journal of Communication Disorders 38:231-260, 2005.

[23] McHenry, M., “The ability to effect intended stress following traumatic brain injury”, Brain Injury, 12:495-503, 1998.

[24] Kim, H., Hasegawa-Johnson M, Perlman A., Gunderson J., Huang T., Watkin K. and Frame, S. “Dysarthric speech database for universal access research”, Interspeech Proc., 2008

[25] Lee, B., Hasegawa-Johnson, M., Goudeseune, C., Kamdar S., Borys, S., Liu, M. and Huang, T., “Avicar: Audio-visual speech corpus in a car envirnemnt”, Interspeech Proc., 2004.

[26] Francis, W., Kucera, H., “Frequency analysis of English usage: lexicon and grammar”, Houghton Mifflin Company, Boston, 1982.

[27] Hart, M., Project Gutenberg, Online: http://www.gutenberg.org/, accessed on 2005.

[28] Praat, Online: http://citeseer.ist.psu.edu/176038.html, accessed on 19 Mar 2008.