The use of fundamental frequency for lexical segmentation in listeners with cochlear implants

Stephanie Spitzer, Julie Liss, Tony Spahr, Michael Dorman, and Kaitlin Lansford
Department of Speech and Hearing Science, Arizona State University, P.O. Box 870102, Tempe, Arizona 85287-0102
spitzer@asu.edu, julie.liss@asu.edu, tony.spahr@asu.edu, michael.dorman@asu.edu, kaitlin.lansford@asu.edu

Abstract: Fundamental frequency (F0) variation is one of a number of acoustic cues normal hearing listeners use for guiding lexical segmentation of degraded speech. This study examined whether F0 contour facilitates lexical segmentation by listeners fitted with cochlear implants (CIs). Lexical boundary error patterns elicited under unaltered and flattened F0 conditions were compared across three groups: listeners with conventional CI, listeners with CI and preserved low-frequency acoustic hearing, and normal hearing listeners subjected to CI simulations. Results indicate that all groups attended to syllabic stress cues to guide lexical segmentation, and that F0 contours facilitated performance for listeners with low-frequency hearing.

PACS numbers: 43.71.Ky, 43.71.An [JH]
Date Received: January 30, 2009 **Date Accepted:** March 23, 2009

1. Introduction

Many studies have demonstrated a benefit to speech understanding when the electric stimulation provided by a cochlear implant (CI) is complemented by preserved low-frequency acoustic hearing (e.g., Ching *et al.*, 2004; Dorman *et al.*, 2008; Gantz and Turner, 2004; Kiefer *et al.*, 2004; Tyler *et al.*, 2002; von Ilberg *et al.*, 1999). It has been hypothesized that the availability of this low-frequency information allows for, among other things, higher quality fundamental frequency (F0) representation than when stimulation is wholly electric (Brown and Bacon, 2009; Chang *et al.*, 2006; Kong *et al.*, 2005; Qin and Oxenham, 2006; Turner *et al.*, 2004). On this view, access to the F0 contour may facilitate tracking of the speech signal, which leads to better speech intelligibility, especially in noise.

The present investigation examined the hypothesis that the presence of the F0 contour improves performance specifically by allowing a listener to better segment the acoustic stream into its component words. This task of lexical segmentation is central to speech understanding. Although it is conducted seemingly effortlessly in most listening situations, parsing speech that is degraded or unreliable requires attention to prosodic contrasts, including variations in vowel and syllable durations, F0, amplitude, and vowel quality (Cutler and Butterfield, 1992). In English and other so-called "stress-timed" languages, listeners tend to treat strong (stressed) syllables as word-onsets, a strategy that is effective because of the statistical probabilities of English in which most words begin with stressed syllables (Cutler and Carter, 1987). This strategy is particularly important when the speech signal is impoverished and the words cannot be immediately recognized in the acoustic stream (Liss *et al.*, 1998; Liss *et al.*, 2002; Mattys, 2004; Mattys *et al.*, 2005). Further, it has been demonstrated that normal hearing (NH) listeners flexibly use all available acoustic cues to syllabic stress, but that reductions in F0 variation are especially detrimental to the task of lexical segmentation of degraded speech (Spitzer *et al.*, 2007). If the performance benefit accrued from the presence of low-frequency information is due to improved F0 representation as a cue for lexical segmentation decisions, this will be apparent in the patterns of lexical boundary errors (LBEs) that are elicited when F0 is present and when it is not. Specifically, the proportion of predicted error types (e.g., erroneously inserting a lexical boundary before a strong syllable than before a weak syllable) should be greater when F0 contour information is available than when it is not.

The purpose of the present study was to determine whether acoustic F0 contour information is used to facilitate lexical segmentation by listeners fitted with CIs and a group of NH controls subjected to CI simulations. We examined two groups of CI listeners, with different capacities for representing low-frequency information: those with CIs, and those with implants and residual low-frequency acoustic hearing in the ear opposite the implant (EAS). All participants transcribed utterances in which the F0 contour was either unaltered (normal) or flattened. Participants with capabilities for low-frequency hearing (NH and EAS) transcribed these unaltered and flattened F0 phrases in an electric-only condition (a proxy for traditional CI, in which they were not provided with low-frequency acoustic information) and in an electric-acoustic condition (in which low-frequency acoustic information also was provided). It was hypothesized that F0 flattening would reduce the listeners' ability to discern strong and weak syllables for the purposes of lexical segmentation in all conditions in which F0 could be sufficiently represented. We predicted that, because F0 representation in the electric-only condition would be poor, F0 flattening would have little effect on the proportion of predicted versus nonpredicted LBEs in this condition for all three listener groups. However, EAS and NH listeners should show detrimental effects of F0 flattening in the electric-acoustic condition, with lower predicted-to-nonpredicted LBE ratios than would be elicited in the unaltered F0 condition. This finding would confirm the listeners' use of F0 variation for lexical segmentation, and provide a *source* of the intelligibility benefit associated with the presence of low-frequency hearing.

2. Method

2.1 Participants

Twelve listeners with CIs and ten NH listeners participated in this study. Six of the CI patients had conventional cochlear implants (CI). The other six possessed residual low-frequency hearing in the non-implanted ear (EAS). The averaged hearing thresholds obtained from the non-implanted ears of EAS subjects revealed a steeply sloping sensori-neural hearing loss, with thresholds of 30 dB hearing loss (HL) through 250 Hz, and 70 dB HL by 750 Hz. Only two conventional CI listeners had measurable auditory thresholds from the non-implanted ear of less than 115 dB sound pressure level (SPL). One had audiometric thresholds of 85 dB SPL at 250 Hz, and 115 dB SPL at 500 Hz. The other had a flat, severe sensori-neural hearing loss from 250 to 8000 Hz, and an earplug was used to attenuate any audible low-frequency information for this listener.

2.2 Speech stimuli

Eighty phrases were developed for LBE analysis in previous investigations (see Liss *et al.*, 1998 for details). Briefly, they each consisted of six syllables (three to five words) that alternated in syllabic stress (iambic or trochaic). A male speaker produced high fidelity digital recordings of the phrases. These phrases were presented to all listeners in normal (unaltered) F0 and F0-flattened conditions. F0 flattening was achieved by calculating the mean F0 (Hz) of each digitized phrase using the *pitch contour* function in Computerized Speech Laboratory (Computerized-Speech Lab (CSL), 2004). The F0 contour was then flattened to this mean value using the numerical editor function in Analysis Synthesis Laboratory (2004); see Spitzer *et al.*, 2007 for details. The F0-flattened phrases were perceptually monotonous, with no perceptible distortion.

CI simulations for the NH listeners were achieved by processing the phrases through a 15-channel vocoder (Litvak *et al.*, 2007). Acoustic information from 350–5600 Hz was divided into 15 logarithmically spaced analysis bands. The energy in each band was used to modulate the amplitude of 15 spectrally-shaped noise bands. The center frequency of each output was equal to the center frequency of the corresponding analysis band. All outputs had a spectral peak at the center frequency and rolled off at a rate of 30 dB/octave from center. The manipulation was intended to simulate the perceptual experience of listeners fitted with CIs.

Table 1. Examples of the four possible categories of LBEs, based on type (insertion or deletion) and location (before strong or weak syllables). The bolded text [insertions before strong (IS) and deletions before weak (DW)] are the categories predicted by the metrical segmentation strategy hypothesis (Cutler and Butterfield, 1992).

LBE	Before a strong syllable	Before a weak syllable
Insertion	**Target: "account for who could knock"** **Response: "he got so he could knock"**	"its harmful note abounds" "its hard to know abounds"
Deletion	Target: "and spoke behind her sin" Response: "and spoke behind person"	**"push her equal culture"** **"pressure equal culture"**

2.3 Procedures

Because of the different hearing modes among the three participant groups, data collection procedures varied accordingly. The phrases were presented in a free field or under headphone (distinctions described below) in a sound attenuated booth. Presentation amplitudes, regardless of mode, were set at a comfortable listening level (65 dB SPL). Phrases were presented one time each, without the opportunity for replay, using the ALVIN software program (Hillenbrand and Gayvert, 2005). Listeners were instructed to use the computer keyboard to type exactly what they heard, and they did not receive feedback about their response accuracy. Phrases were presented in blocks of 20.

All listeners transcribed blocks of unaltered and F0-flattened phrases. For the NH and EAS listeners, these were transcribed in both an electric-only condition and in an electric-acoustic condition. Obviously, it was not possible for the conventional CI participants to receive the latter condition.

In the electric-only condition, phrases were presented via loudspeaker to the implanted ear. A standard foam earplug was used to attenuate the signal in the non-implanted ear for the EAS participants (see Gifford *et al.*, 2008). CI simulated signals were presented to the NH listeners at the left ear via headphones.

In the electric-acoustic condition, phrases were presented free field to the EAS listeners. The NH listeners received the CI simulated phrases to the left ear to simulate electric hearing. The right ear received phrases that had been passed through a low-pass filter with a cutoff frequency of 500 Hz (sixth order Butterworth) to simulate residual low-frequency hearing.

2.4 Analysis

Listener transcripts were coded independently by two judges proficient in LBE analysis. As per standard consensus procedures, codes were compared and discrepancies were either resolved or omitted from the analysis (see Liss *et al.*, 2002 for details). Dependent variables included the total number of LBEs committed, and the type (insertion or deletion) and location (before strong or weak syllables) of LBEs. Thus, errors could be of one of four types: lexical boundary insertion or deletion occurring either before a strong or weak syllable (IS, IW, DS, and DW). Examples of each type are provided in Table 1. Based on predictions generated from the metrical segmentation strategy hypothesis (Cutler and Carter, 1987), if listeners attend to syllabic stress to segment the speech stream, they will be most likely to erroneously insert lexical boundaries before strong syllables (i.e., treat a strong syllable as a word-onset), and erroneously delete lexical boundaries before weak syllables (i.e., treat a weak syllable as the second, or greater, syllable of a word). Thus, IS and DW errors are of the predicted type. Ratios of predicted error types also were calculated (IS+DW/total errors); higher ratios indicate greater reliance on—and quality of—syllabic stress information to segment the phrases.

Pearson chi-square analyses were conducted to determine if LBEs were distributed differently across type (insertion/deletion) and location (before strong/weak syllables) within each of the pools of LBEs. Additional chi-square analyses were conducted to identify differences in distributions between unaltered and F0-flattened LBE pools within condition, whereby

Table 2. Performance values for all listener groups across listening conditions (electric-only or electric-acoustic) with unaltered or F0-flattened phrases. Data columns contain mean intelligibility (% words correct and standard error), the proportions of types of LBEs (insertions or deletions, before strong or before weak syllables), and the total number of LBEs per condition. The final two columns contain *p*-values for the χ^2 analyses conducted on each condition LBE pool, and differences between distributions derived from F0-flattened stimuli relative to unaltered for each condition.

Groups	Condition	Phrases	% words correct (SE)	% IS	% IW	% DS	% DW	Total LBE	($\chi^2 df=1$) *p*-value	($\chi^2 df=3$) *p*-value
EAS *N*=6	E-only	F0-flat	32 (4.3)	33	21	16	30	82	0.003	0.000
	E-only	Unaltered	4 (3.1)	51	9	14	26	78	0.000	
	E-A	F0-flat	50 (2.4)	52	25	10	13	63	0.03	0.000
	E-A	Unaltered	56 (3.2)	57	9	16	18	56	0.002	
CI *N*=6	E-only	F0-flat	25 (6.5)	49	20	10	21	72	0.000	0.08
	E-only	Unaltered	34 (6.5)	59	13	7	21	55	0.000	
NHL *N*=10	E-only	F0-flat	40 (3.7)	38	17	13	32	121	0.000	0.03
	E-only	Unaltered	43 (2.3)	46	9	15	30	134	0.000	
	E-A	F0-flat	48 (2.4)	45	28	10	17	109	0.03	0.000
	E-A	Unaltered	68 (3.5)	60	7	5	28	60	0.000	

the proportions of errors elicited by F0-*unaltered* were treated as the expected cell values, and the F0-*flattened* were the observed cell values. Significant differences were regarded as a rejection of the null hypothesis that the values were sampled from the same distribution. A *p*-value of 0.05 was selected for all analyses.

3. Results and discussion

Table 2 contains the group performance data across conditions, as well as the results of the chi-square analyses. The first novel observation is that all groups in all conditions produced patterns of LBEs that align with the metrical segmentation strategy for predicted errors, with ten significant chi-square tests confirming differential distributions across LBE type and location. In all cases, insertions errors occurred more frequently before strong than before weak syllables; and deletion errors occurred more often before weak than before strong. This is further exemplified in Fig. 1, which shows the total proportions of predicted LBEs across groups and conditions. The ratio values on the ordinate range from 0.5 (reflecting instances in which predicted and nonpredicted errors occur equally often) to a value of 1 (in which all errors are of the predicted type). This figure is important in showing that all values exceeded 0.5, which is consistent with the explanation that listeners attended to available syllabic stress cues to guide lexical segmentation, even when F0 was not among those cues. These cues likely include variations in syllable durations, dispersion of vowel formant frequencies, and amplitude envelope modulations. Thus, as for NH listeners, those fitted with CI appear to rely on syllabic stress cues for lexical segmentation and to flexibly use the cues that are available and reliable to accomplish this task.

All groups showed evidence of attending to syllabic stress, but the primary question of this study is whether the flattening of F0 made a difference in LBE distributions. The relative heights of the bars in Fig. 1 suggest this to be the case for all conditions. Recall that we expected to see differences *only* in instances in which F0 was adequately represented. Returning to the final column of Table 2, we see the chi-square results of pitting the F0-flattened LBE distributions against those generated by the unaltered phrases. As predicted, these distributions were not different for the CI group, and they were different for the EAS and NH in the electric-acoustic condition. However, we also see that the F0 flattening served to significantly reduce adherence

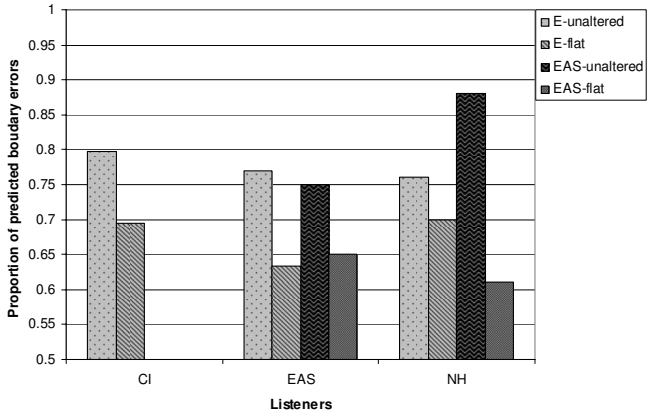

Fig. 1. Proportion of total LBEs that are of the predicted type (IS+DW/total LBE). The data are presented by listener group (CI, EAS, and NH listeners) and condition. The conditions include electric-only and electric-acoustic (EAS), in both the unaltered and F0-flattened conditions.

to the predicted error patterns for the EAS and NH groups in the electric-only condition as well. A likely explanation is that the electric-only condition did not effectively eliminate low-frequency and F0 information for these listeners who were capable of processing it. However, it also is of interest that, even though insignificant, the CI listeners showed the tendency for lower adherence to the predicted LBE patterns in the F0-flattened condition. It will be important to explore the possibility that even poorly represented F0 contour information can facilitate lexical segmentation.

These patterns of results support the hypothesis that syllabic stress information is exploited by listeners with CIs for the purposes of lexical segmentation, in much the same way that NH listeners deal with the CI simulation. Further, the presence of an F0 contour was shown to facilitate lexical segmentation decisions in EAS listeners, and this thereby points to a specific source of perceptual benefit afforded by the presence of low-frequency hearing.

4. Conclusion

Listeners fitted with CI appear to behave much like NH listeners in the deciphering of degraded speech in that they attend to syllabic stress cues to guide lexical segmentation. The present investigation provides evidence for the informative role of F0 as one of these cues. Specifically, F0 variation appears to be used, to the extent possible, to mark strong and weak syllables for the purposes of lexical segmentation. The data also provide preliminary support for the hypothesis that at least some of the performance benefit enjoyed by EAS listeners may be traceable to the availability of a high resolution F0 contour for facilitating lexical segmentation.

Acknowledgments

This research was supported by NIH Grant Nos. 5R01 DC6859 (J.L.) and R01 DC00654-18 (M.D.) from the National Institute on Deafness and Other Communication Disorders, and by a grant from Advanced Bionics Corp. (T.S.).

References and links

Analysis Synthesis Laboratory (**2004**). Software (KayPentax, NJ).

Brown, C. A., and Bacon, S. P. (**2009**). "Low-frequency speech cues and simulated electric-acoustic hearing," J. Acoust. Soc. Am. **125**, 1658–1665.

Chang, J. E., Bai, J. Y., and Zeng, F. G. (**2006**). "Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise," IEEE Trans. Biomed. Eng. **53**, 2598–2601.

Ching, T. Y., Incerti, P., and Hill, M. (**2004**). "Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears," Ear Hear. **25**, 9–21.

Computerized-Speech Lab (CSL) (**2004**). computer hardware (KayPentax, NJ).

Cutler, A., and Butterfield, S. (**1992**). "Rhythmic cues to speech segmentation: Evidence from juncture

misperception," J. Mem. Lang. **31**, 218–236.
Cutler, A., and Carter, D. M. (**1987**). "The predominance of strong syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142.
Dorman, M. F., Gifford, R., Spahr, A. J., and McKarns, S. (**2008**). "The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies," Audiol. Neuro-Otol. **13**, 105–112.
Gantz, B. J., and Turner, C. (**2004**). "Combining acoustic and electrical speech processing: Iowa/nucleus hybrid implant," Acta Oto-Laryngol. **124**, 344–347.
Gifford, R., Dorman, M., Spahr, A., Bacon, S., Skarzynski, H., and Lorens, A. (**2008**). "Hearing preservation surgery: Psychophysical estimates of cochlear damage in recipients of a short electrode array," J. Acoust. Soc. Am. **124**, 2164–2173.
Hillenbrand, J. M., and Gayvert, R. T. (**2005**). "Open source software for experimental design and control," J. Speech Lang. Hear. Res. **48**, 45–60.
Kiefer, J., Gstöettner, W., Baumgartner, W., Pok, S. M., Tillein, J., Ye, Q., and Von Ilberg, C. (**2004**). "Conservation of low-frequency hearing in cochlear implantation," Acta Oto-Laryngol. **124**, 272–280.
Kong, Y. Y., Stickney, G. S., and Zeng, F. G. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**, 1351–1361.
Liss, J. M., Spitzer, S. M., Caviness, J. N., and Adler, C. (**2002**). "The effects of familiarization on intelligibility and lexical segmentation of hypokinetic and ataxic dysarthria," J. Acoust. Soc. Am. **112**, 3022–3031.
Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., and Edwards, B. W. (**1998**). "Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech," J. Acoust. Soc. Am. **104**, 2457–2466.
Litvak, L. M., Spahr, A. J., Saoji, A. A., and Fridman, G. Y. (**2007**). "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners," J. Acoust. Soc. Am. **122**, 982–991.
Mattys, S. L. (**2004**). "Stress versus coarticulation: Towards an integrated approach to explicit speech segmentation," J. Exp. Psychol. Hum. Percept. Perform. **30**, 397–408.
Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple segmentation cues: A hierarchical framework," J. Exp. Psychol. Gen. **134**, 477–500.
Qin, M. K., and Oxenham, A. J. (**2006**). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," J. Acoust. Soc. Am. **119**, 2417–2426.
Spitzer, S. M., Liss, J. L., and Mattys, S. L. (**2007**). "Acoustic cues to lexical segmentation: A study of resynthesized speech," J. Acoust. Soc. Am. **122**, 3678–3687.
Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (**2004**). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," J. Acoust. Soc. Am. **115**, 1729–1735.
Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (**2002**). "Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization," Ear Hear. **23**, 98–105.
von Ilberg, C., Kiefer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Sturzebecker, E., and Klinke, R. (**1999**). "Electric-acoustic stimulation of the auditory system," ORL **61**, 334–340.