Clinical Linguistics & Phonetics, Early Online, 2010, 1–8

informa
healthcare

An acoustic analysis of the vowel space in young and old cochlear-implant speakers

VERONIKA NEUMEYER, JONATHAN HARRINGTON,
& CHRISTOPH DRAXLER

Institut für Phonetik und Sprachverarbeitung, Ludwig-Maximilans Universität München, Germany

(Received 26 March 2010; Accepted 3 May 2010)

Abstract
The main purpose of this study was to compare acoustically the vowel spaces of two groups of cochlear implantees (CI) with two age-matched normal hearing groups. Five young test persons (15–25 years) and five older test persons (55–70 years) with CI and two control groups of the same age with normal hearing were recorded. The speech material consisted of five German vowels V = /aː, eː, iː, oː, uː/ in bilabial and alveolar contexts. The results showed no differences between the two groups on Euclidean distances for the first formant frequency. In contrast, Euclidean distances for F2 of the CI group were shorter than those of the control group, causing their overall vowel space to be compressed. The main differences between the groups are interpreted in terms of the extent to which the formants are associated with visual cues to the vowels. Further results were partially longer vowel durations for the CI speakers.

Keywords: *cochlear implant, CI, articulation, vowel space, vowel durations*

Introduction

Cochlear implantation has been possible for ∼ 25 years, during which time it has become a common treatment option in many countries for individuals with significant hearing impairment. Despite the technical progress that has been made to Cochlear Implants (CI), there are still many limitations which affect both CI users' speech perception and speech production.

According to Horga and Liker (2006), acoustic analyses show a significantly reduced vowel space for profoundly deaf speakers without cochlear implants compared with CI users and hearing controls. Immediately after cochlear implantation, the vowel space of a CI user is similar to that of a hearing impaired speaker. About 1 year after implantation, the formants shift closer to those of listeners with normal hearing who tend to have more expanded vowel spaces than hearing impaired listeners with hearing aids. After this formant shift, Horga and Liker (2006) showed that the CI users also produced more intelligible vowels except for the vowel /a/. This result was supported by an acoustic analysis and a vowel intelligibility test.

Correspondence: Veronika Neumeyer, Institut für Phonetik und Sprachverarbeitung, Ludwig-Maximilans Universität München, Schellingstraße 3, D 80799, München, Germany. Tel: +49 (0)89 2180-2758. E-mail: veronika.neumeyer@phonetik.uni-muenchen.de

ISSN 0269-9206 print/ISSN 1464-5076 online © 2010 Informa UK Ltd.
DOI: 10.3109/02699206.2010.491173

In another study concerned with vowel production, Uchanski and Geers (2003) compared F2 of English speaking CI users with those of a control group with normal hearing. They investigated the vowels /i/ and /ɔ/ which had the highest and lowest F2 values, respectively, in the variety of English they investigated. The results of their study were that 87% of the formant values for /i/ and 88% of the formant values for /ɔ/ of the CI group were in the range of the formant values of the group with normal hearing. These results suggest that CI users and normal hearing listeners may have broadly similar vowel spaces.

These results of Uchanski and Geers (2003) are, however, not consistent with those reported in Liker et al. (2007), who compared the formants of CI children of five Croatian vowels (/i, e, a, o, u/) with those of a control group without hearing impairment. The participants were recorded three times within 1 year. Liker et al. (2007) hypothesized that the CI users would have a smaller and more fronted vowel space than the controls. Although their results showed F2-differences between the groups, Liker et al.'s hypothesis that the vowel space of CI users would increase and become less fronted from the first to the third recording session could not be unequivocally demonstrated.

In an investigation of Swedish children, Ibertsson et al. (2008) provided some evidence in support of Liker et al.'s (2007) finding of a smaller vowel space for a CI group. They investigated nine Swedish long vowels and measured the Euclidean distance in the F1–F2 plane between each vowel and the mean first and second formant frequencies of all the vowels. The vowel space of the children with cochlear implant was more compressed and differed significantly from that of the children with normal hearing.

The aim of the present study is to investigate the vowel production of two different groups of CI users, an adolescent and an adult one, and compare them with two age-matched control groups. Our aim in this study was to test four hypotheses. First, the vowel-space of the CI speakers was predicted to be smaller and more centralized than that of normal listeners. Second, we predicted a greater deviation between the groups in F2 than in F1. This is because the relationship between acoustics and articulation is visually less transparent for the second, compared with the first, formant frequency. More specifically, it is difficult to attribute visually an F2-change unambiguously to a change in the vocal tract configuration: an increase in F2 could, for example, result either from tongue fronting or from lip-unrounding, or quite possibly both. Moreover, the visual interpretation of the articulatory activity that gives rise to F2 shifts is further complicated by the fact that the tongue is itself largely hidden from view. In contrast, there is a much more transparent relationship between F1 changes and jaw-height: in general, closing the mouth and therefore raising the jaw—an articulatory activity that is clearly visible—results in a lowering of F1 and vice-versa (Lindblom and Sundberg, 1971). Consequently, we predict that the relationship between jaw-height and F1 is more easily learnable by the CI group than the relationship between lip-rounding/constriction location and F2. Third, we tested whether, as shown in other studies (e.g., Whitehead and Jones, 1976; Lane et al., 1995; Uchanski and Geers, 2003), the duration of vowels for CI speakers is greater than for unimpaired speakers. Finally, we predicted greater differences between the CI and normal-hearing speakers for younger than for older speakers: this was because, whereas the younger CI speakers were hearing impaired from birth, the participants in our older CI group acquired deafness in adulthood.

Method

Speakers and materials

The speakers were 10 adults with profound deafness and cochlear implants (CI) and 10 adults with normal hearing (NH) as a control group. All speakers were first language speakers of the same Southern Germany variety. Each group consisted of five younger women (CI: age range 16–27 years, mean age 21 years, 2 months; NH: age range 15–24 years, mean age 21 years) and five older women (CI: age range 58–68 years, mean age 63 years, 9 months; NH: age range 59–68 years, mean age 62 years, 9 months). Only three of the CI users had bilateral cochlear implants. The mean duration for which the two groups had worn their cochlear implant was 3.9 years for the old and 9.5 years for the young group. A two sample *t*-test showed that this difference was not significantly different ($t(5.53) = 1.83$, $p > 0.1$), presumably because of the small number of participants ($n = 5$) in each age group. The older participants with cochlear implants all had had normal hearing for the most part of their life and were deafened because of an accident or progressive amblyacousia.

The speech material consisted of the five German long vowels /aː, eː, iː, oː, uː/ (which occur in the first syllable of e.g., *baten*, *beten*, *bieten*, *boten*, *Bude*). All of these vowels are quite close to Cardinal Vowels in quality, with the exception of /aː/, which is a low central vowel and intermediate between Cardinal Vowels 4 and 5. Each of these vowels occurred syllable-initially in symmetrical alveolar and bilabial contexts in lexical trochaic target words. We chose not to use the same carrier phrase for each context but to embed the target word medially in what we considered to be more naturally occurring sentences. In all cases, the sentences were produced as a single prosodic phrase and the target word was prosodically accented, that means produced with sentence stress (Ladd, 1996). There were two kinds of sentences (Table I): in the first kind the focus was broad, whereas the second elicited a production with narrow focus on the target word. Following Hirschberg (2005) and Beckman and Venditti (2010), we defined an utterance to have broad focus if a dialogue does not require any of its constituents to be produced with a much higher degree of prominence than its other constituents. For example, the sentence *I'm*

Table I. List of sentences in broad (above) and narrow focus (below) sentences showing the vowel (V) and symmetrical consonant (C) contexts of the target words (in bold).

V	C	Sentence
Broad focus		
iː	b	Der Baum wurde von einem **Biber** gefällt.
	d	Ich habe **Dieter** vergessen.
eː	b	Es wurde ein **Beben** gemessen.
	d	Ich habe den **Teetisch** gedeckt.
aː	b	Wir haben den Turm von **Babel** gemalt.
	d	Ich habe die **Daten** mitgebracht.
oː	b	Sie ist mit dem **Moped** da.
	d	Er hat seine **Note**n dabei.
uː	b	Schlagen Sie das Wort im **Duden** nach.
	d	Er hat einen **Buben** gesehen.
Narrow focus		
iː	b	Das Wort '**Biber**' ist ganz kurz.
	d	Hier ist '**Dieter**' vergessen.
eː	b	Hier habe ich '**Beben**' gelesen.
	d	Hier steht '**Teetisch**' geschrieben.
aː	b	Sie hat '**Babel**' gesagt.
	d	Das Wort "**Daten**' ist dran.
oː	b	Das heißt '**Moped**', was da steht.
	d	Sie hat '**Noten**' gehört.
uː	b	Hier steht '**Buben**' drauf.
	d	Dann ist '**Duden**' dran.

going to London tomorrow has broad focus in response to a question *what are your plans?* but narrow focus on the last word (which is produced with a high degree of prominence) in response to the question *I thought you're going to London next week?*.

Every sentence was read five times. The total number of stimuli available was thus: (10 CI + 10 NH) = 20 speakers × 5 (vowels) × 2 (labial/aveolar) × 2 kinds of sentences (broad/narrow focus) × 5 five repetitions = 2000 items (100 per speaker).

Parameters

The recordings were all carried out in quiet rooms in Munich, either at the University Hospital 'rechts der Isar', or at the Institute of Phonetics and Speech Processing, or at the homes of the participants. The same investigator was present in all recordings. The speech data were recorded with a Sennheiser USB 36 headset, and digitized directly to a battery-powered laptop at a sampling frequency of 22.05 kHz. The sentences were randomized and presented individually on the screen of the laptop using the software system SpeechRecorder (Draxler and Jänsch, 2004). We instructed our participants to read the sentences at a normal rate. They could take a break at any time they wanted.

The digitized speech data was automatically segmented and labelled using the HMM-based Munich-Automatic-Segmentation system MAUS (Schiel, 2004). The segmentation and annotations were subsequently checked and manually corrected using the Praat and Emu (Harrington, 2010) software systems. The formant values of F1 and F2 were computed with the default settings in Emu using a Blackman window, a window size of 25 ms, and a frame shift of 5 ms. Formant tracking errors such as when F2 was mis-tracked as F3 (which sometimes occurs in high back vowels in which F1 and F2 are close together) were subsequently manually corrected. We also converted the formant frequency values in Hz to the auditory Bark scale according to the formulae given in Traunmüller (1990). There is evidence to suggest that a transformation to the Bark scale is more in accordance with the way that the acoustic signal is auditorily processed (Schroeder et al., 1979; Zwicker and Feldtkeller, 1967). Moreover, a Bark transformation has been shown to remove a certain amount of non-phonetic, speaker-specific information from the signal (e.g., Bladon and Lindblom, 1981) as a result of which vowels can be more clearly separated when formants are represented on the Bark scale (Syrdal and Gopal, 1986).

Instead of representing vowels in the formant plane with a single static slice extracted at the vowel target, we followed the methodology in Watson and Harrington (1999) and Harrington et al. (2008) by parameterizing the entire shape of the (Bark-scaled) vowel formant as a function of time, thereby preserving dynamic information. To do this, we reduced each formant trajectory to a point in a three-parameter space using the discrete cosine transformation (DCT): this technique decomposes any digital signal into a set of 1/2 cycle cosine waves which, if summed, reconstruct entirely the original signal. The three-parameter space was formed from the amplitudes of the first three DCT-coefficients (at frequencies of 0, 1/2, and 1 cycle) which encode major properties of the formant's shape: specifically, these first three DCT-coefficients are proportional to the formant's mean, linear slope, and curvature, respectively.

We then calculated in this three-dimensional DCT space separately for each of the four groups (young/old, CI/normal) and separately for F1 and for F2 the Euclidean distance from each vowel token to the centre of the space. This measure has been shown to be related to vowel reduction (e.g., Wright, 2003): in general, the greater the compression of the vowel space, the shorter the Euclidean distances from the tokens to the space's centre.

We also measured vowel length differences by calculating the ratio of vowel-to-word duration (which normalizes for variation in speech rate).

Results

Figures 1 and 2 show boxplots of the Euclidean distances to the centre of the F1 (Figure 1) and F2 (Figure 2) spaces separately for the four groups and by vowel category. As far as F1 is concerned, there are some indications of shorter Euclidean distances in both young and old CI compared with normal groups in /aː, iː, uː/, whereas the data for the mid-vowels /eː, oː/ are inconsistent. The F2-data, in contrast, show a more consistent trend: as Figure 2 shows, the Euclidean distances were shorter for CI than for normal speakers in both age groups for all vowels except /aː/. The same figure also shows a greater divergence on this measure between CI and normal speakers for the younger speakers.

Figure 3 shows the distribution of the mean values of the five vowels by age group and separately for the CI (black dotted) and NH speakers (grey). For both age groups, there is some evidence that the distance in F2 between the front vowels /iː, eː/ and the back vowels /oː,

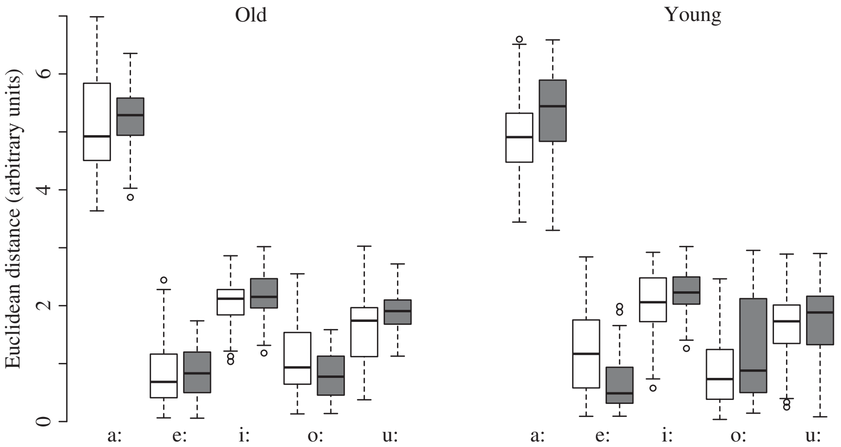

Figure 1. Boxplots of the Euclidean distances (arbitrary units) to the centre in the F1-DCT space pooled separately for the older participnts (left) and for the younger participants (right). The white boxes represent the cochlear implant users, the filled grey boxes the normal hearing.

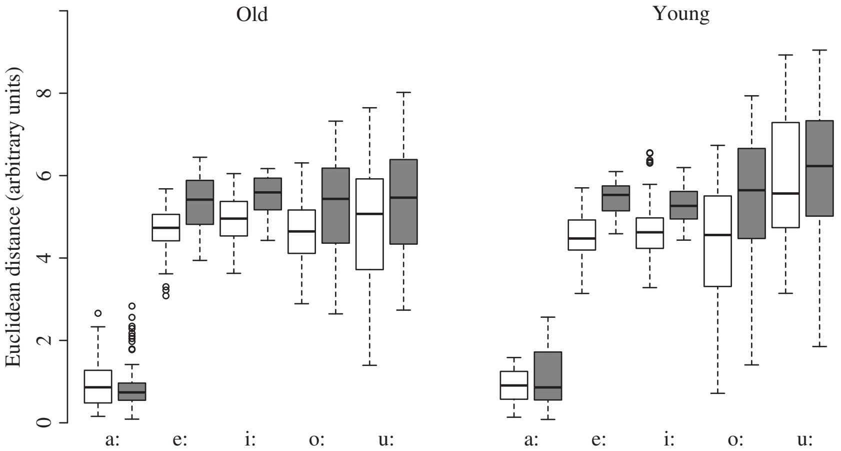

Figure 2. Boxplots of the Euclidean distances (arbitrary units) to the centre in the F2-DCT space pooled separately for the older participnts (left) and for the younger participants (right). The white boxes represent the cochlear implant users, the filled grey boxes the normal hearing.

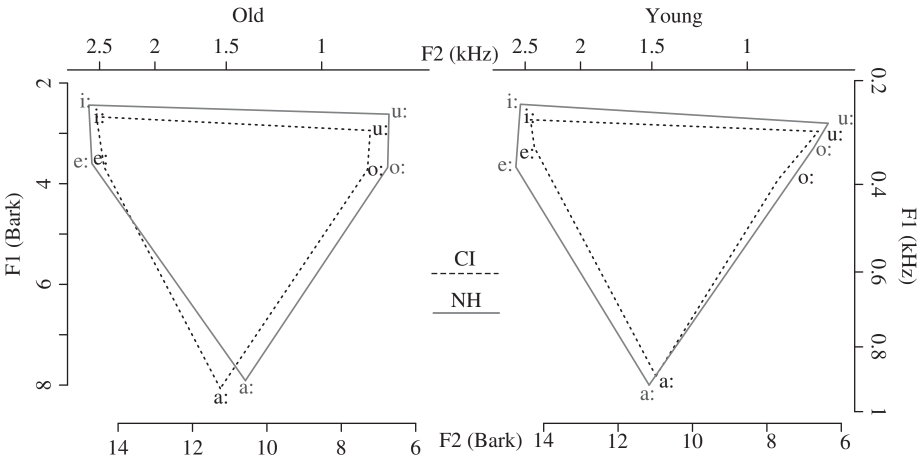

Figure 3. Vowel spaces (in Hz) for CI- (black dotted line) and NH- (grey solid line) speakers shown for the older (left) and younger (right) speakers.

uː/ is shorter for the CI than for NH speakers as a result of which the vowel space is horizontally compressed. The data in Figure 3 are consistent with our hypotheses that the CI speakers have a smaller vowel-space than the control-group and the differences between the groups in F2 are greater than in F1.

The results of a repeated-measures ANOVA with dependent-variable F1-Euclidean distance and independent variables Vowel class (five levels), Speaker group (two levels) and Age (two levels) showed no significant differences between CI and normal hearing speakers on this measure. On the other hand, the results with the same independent variables showed significant differences between CI and normal hearing groups on F2-Euclidean distances ($F[1,16] = 8.99$, $p < 0.01$).

The results of our duration analysis show a broadly similar pattern between the groups with some age-dependent differences. A repeated measures-ANOVA with the same independent variables as before showed that there were no significant main effects for duration in Age and Speaker-group and predictably (given the well-known relationship between phonetic height and duration e.g., Lehiste, 1961) a significant effect for Vowel ($F[1,64] = 161.5874$, $p < 0.001$). There were also significant interactions between Vowel and Speaker-group ($F[1,64] = 4.25$, $p < 0.01$) and between Vowel and Age ($F[1,64] = 7.21$, $p < 0.001$). *Post-hoc* Bonferroni tests showed that the durations of /eː/ and /iː/ were greater for the CI compared with NH speakers.

Discussion

The main aim of this study was to investigate whether CI and normal speakers differ in the way that they produce vowels. The results showed that Euclidean distances to the centre of the vowel space were shorter for CI speakers, from which we infer that their vowel space was horizontally compressed.

Following other investigations of CI-speech (e.g., Svirsky and Tobey, 1991; Perkell et al., 2001; Vick et al., 2001; Horga and Liker, 2006; Liker et al., 2007; Ibertsson et al., 2008), we hypothesized that the CI's reduced auditory feedback may be the source both of their more compressed vowel space in speech production and (although we did not test this in this study) under-differentiation of vowels in perception. Our analysis of F1 showed no clear differences between the two speaker groups. One of the reasons for the lack of differences may be because of the clear visibility of jaw height changes on the one hand, and the correlation between jaw position and F1 on the other (Lindblom and Sundberg, 1971): that is, CI listeners can abstract phonetic height (i.e., the relative position of a vowel between CV1 and CV4) from jaw position, as a result of which they and normal listeners may differ minimally in F1. In contrast, we would suggest that the more compressed F2 range in our CI speakers comes about because they do not have the opportunity to infer visually tongue backing and hence phonetic backness in the same way. These conclusions are supported by other findings in the literature showing F2, but not F1, differences between CI and normal speakers (Liker et al., 2007) and those of Lachs et al. (2001) who showed a greater intelligibility of vowels for CI speakers using an audiovisual speech input compared with those who made use of auditory speech input alone.

Although we hypothesized a greater divergence in vowel production between CI and normal hearing speakers for the younger than the older group (because the younger CI speakers had been hearing-impaired since early childhood whereas the older CI users had had normal hearing and therefore unimpaired auditory feedback for most part of their lives), our results showed no significant age-dependent differences. One possible explanation for these negative findings is that the older CI speakers mostly had progressive hearing loss and so were not fitted with cochlear implants at the same stage relative to the onset of deafness. For this reason they spent years with restricted auditory feedback until they got their implants. During this time articulatory skills can diminish because of the impaired monitoring of a participant's own speech.

Compatibly with other studies (e.g., Whitehead and Jones, 1976; Lane et al. 1995; Uchanski and Geers, 2003), we found vowel duration of the CI group to be partially greater than that of the controls. We are currently unsure of the source of these vowel duration differences, but note that the Lombard effect induced by normal hearing listeners through binaural masking also produces an increase in vowel duration (Garnier et al., 2006; Junqua, 1993; Lane and Tranel, 1971; van Summers et al., 1988).

In summary, our main finding from this study is that the vowel space of CI speakers is compressed in phonetic backness, but not in phonetic height, which we attribute to the fact that tongue movements are less easily seen than jaw height differences. We are currently investigating tongue movements of CI speakers using electromagnetic articulometry; and we are also investigating the role of audiovisual input for speech perception in CI-listeners (Bergeson et al., 2005).

Acknowledgements

We thank Yuki Era and Dominique Vornwald for help with labelling and Florian Schiel for his comments. We are also very grateful to all the participants for participating in this study.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Beckman, M., & Venditti, J. (2010). *Tone and intonation. A Handbook of Phonetics*. Chichester: Wiley-Blackwell.

Bergeson, T.R., Pisoni, D.B., & Davis, R.A.O. (2005). Development of audiovisual comprehensions in prelingually deaf children with cochlear implants. *Ear & Hearing*, *26*, 149–164.

Bladon, R., & Lindblom, B. (1981) Modelling the judgement of vowel quality differences. *Journal of the Acoustical Society of America*, *69*, 1414–1422.

Draxler, C., & Jänsch, K. (2004). *SpeechRecorder - A universal platform independent multi-channel audio recording software*. Proceedings of the IVth International Conference on Language Resources and Evaluation (LREC), Lisbon, Portugal, pp. 559–562.

Garnier, M., Bailly, L., Dohen, M., Welby, P., & Lœvenbruck, H. (2006). An acoustic and articulatory study of Lombard speech: Global effects on the utterance. *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, 2246–2249.

Harrington, J. (2010). *The Phonetic Analysis of Speech Corpora*. Chichester: Wiley-Blackwell.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: an acoustic and perceptual study. *Journal of the Acoustical Society of America*, *123*, 2825–2835.

Hirschberg, J. (2005) *Pragmatics and intonation. The Handbook of Pragmatics*. Chichester: Wiley-Blackwell.

Horga, D., & Liker, M. (2006). Voice and pronunciation of cochlear implant speakers. *Clinical Linguistics & Phonetics*, *20*, 211–217.

Ibertsson, T., Sahlén, B., & Löfqvist, A. (2008). Vowel spaces in Swedish children with cochlear implants. *Journal of the Acoustical Society of America*, *123*, 3330.

Junqa, J.C. (1993). The Lombard reflex and its role on human listener and automatic speech recognisers. *Journal of the Acoustical Society of America*, 93, 510–524.

Lachs L., Pisoni, D.B., & Kirk, K.I. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear & Hearing*, *22*, 236–251.

Ladd, R.D. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Ladefoged, P. (1967). *Three areas of experimental phonetics*. London: Oxford University Press.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *The Journal of Speech and Hearing Research*, *14*, 677–709.

Lane, H., Wozniak, J., Matthies, M., Svirsky, M., & Perkell, J. (1995). Phonemic resetting versus postural adjustments in the speech of cochlear implant users: an exploration of voice-onset time. *Journal of the Acoustical Society of America*, *98*, 3096–3106.

Lehiste, I., & Peterson, G.E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America*, *33*, 419–425.

Liker, M., Mildner, V., & Sindija, B. (2007). Acoustic analysis of the speech of children with cochlear implants: A longitudinal study. *Clinical Linguistics & Phonetics*, *21*, 1–11.

Lindblom, B., & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America*, *50*, 1166–1179.

Perkell, J., Numa, W., Vick, J., Lane, H., Balkany, T., & Gould, J. (2001). Language-specific, hearing-related changes in vowel spaces: A preliminary study of English- and Spanish-speaking cochlear implant users. *Ear and Hearing*, *22*, 461–470.

Schiel, F. (2004). *MAUS goes iterative*. Proceedings of the IVth International Conference on Language Resources and Evaluation (LREC), Lisbon, Portugal, pp. 1015–1018.

Schroeder, M., Atal, B., & Hall, J. (1979) Auditory analysis and timbre perception. In B. Lindblom and S. Öhman (eds.): *Frontiers of Speech Communication Research* (pp. 217–229). London: Academic Press.

Svirsky, M., & Tobey, E. (1991). Effect of different types of auditory stimulation on vowel formant frequencies in multichannel cochlear implant users. *Journal of the Acoustical Society of America*, *89*, 2895–2904.

Syrdal, A., & Gopal, H. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, *79*, 1086–1100.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, *88*, 97–100.

Uchanski, R., & Geers, A. (2003). Acoustic characteristics of the speech of young cochlear implant users: A comparison with normal-hearing age-mates. *Ear & Hearing*, *24*, 90–105.

Van Summers, W., Pisoni, D.B., Bernacki, R.H., Pedlow, R.I., & Stokes, M.A. (1988). Effect of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America*, *84*, 912–928.

Vick, J.C., Lane, H., Perkell, J.S., Matthies, M.L., Gould, J., & Zandipour, M. (2001). Covariation of cochlear implant users' perception and production of vowel contrasts and their identification by listeners with normal hearing. *Journal of Speech, Language, and Hearing Research*, *44*, 1257–1267.

Watson, C.I., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English Vowels. *Journal of the Acoustical Society of America*, 106, 458–468.

Whitehead, R., & Jones, K. (1976). Influence of consonant environment on duration of vowels produced by normal hearing, hearing impaired and deaf adult speakers. *Journal of the Acoustical Society of America*, *60*, 513–515.

Wright, R. (2003). Lexical competition and reduction in speech. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI.* (pp. 75–87). Cambridge: Cambridge University Press.

Zwicker, E. & Feldtkeller, R. (1967) *Das Ohr als Nachrichtenempfänger*. Stuttgart: Hirzel Verlag.