This is the published version of:


Copyright:

Copyright (2003) Acoustical Society of America. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the Acoustical Society of America. The following article appeared in *J. Acoust. Soc. Am.* Volume 114, Issue 2, pp. 1055-1068 (2003) and may be found at

http://dx.doi.org/10.1121/1.1593059.
Acoustic analysis of monophthong and diphthong production in acquired severe to profound hearing loss

Sallyanne Palethorpe and Catherine I. Watson

Speech, Hearing and Language Research Centre, Department of Linguistics, Macquarie University, Sydney, Australia 2109 and Macquarie Centre for Cognitive Science, Macquarie University, Sydney, Australia 2109

Rosalind Barker

Speech, Hearing and Language Research Centre, Department of Linguistics, Macquarie University, Sydney, Australia 2109

(Received 23 September 2002; revised 31 May 2003; accepted 2 June 2003)

The effect of diminished auditory feedback on monophthong and diphthong production was examined in postlingually deafened Australian-English speaking adults. The participants were 4 female and 3 male speakers with severe to profound hearing loss, who were compared to 11 age- and accent-matched normally hearing speakers. The test materials were 5 repetitions of hVd words containing 18 vowels. Acoustic measures that were studied included \( F_1, \) \( F_2, \) discrete cosine transform coefficients (DCTs), and vowel duration information. The durational analyses revealed increased total vowel durations with a maintenance of the tense/lax vowel distinctions in the deafened speakers. The deafened speakers preserved a differentiated vowel space, although there were some gender-specific differences seen. For example, there was a retraction of \( F_2 \) in the front vowels for the female speakers that did not occur in the males. However, all deafened speakers showed a close correspondence between the monophthong and diphthong formant movements that did occur. Gaussian classification highlighted vowel confusions resulting from changes in the deafened vowel space. The results support the view that postlingually deafened speakers maintain reasonably good speech intelligibility, in part by employing production strategies designed to bolster auditory feedback. © 2003 Acoustical Society of America. [DOI: 10.1121/1.1593059]

PACS numbers: 43.70.Dn [AL]

I. INTRODUCTION

This study was designed to investigate the effects of severe to profound postlingual deafness on monophthong and diphthong vowel production. Postlingual deafness is a hearing deficit that has occurred after language has been firmly established, usually after age 5, and can also be referred to as adventitious or acquired deafness. Speakers who have postlingual deafness can be called deafened. When the hearing deficit has occurred early enough to have a major effect on language acquisition, the speakers are termed (con-genitally) deaf. Numerous studies have investigated vowel production in deaf speakers (e.g., Bakkum et al., 1993; Ertmer et al., 1997; Fourakis et al., 1993; McCaffrey and Sussman, 1994; Monsen, 1976). The majority of these researchers have studied vowels in children and adolescents and report reduced vowels spaces, overlap between vowel categories, a reduction in \( F_1 - F_2 \) range, tense–lax substitution, diphthongization, and neutralization. However, there is a vast difference between investigating how vowel production has developed in the absence of hearing, and establishing what happens to production when hearing is lost or distorted over a prolonged period of time.

The investigation of vowel production in severe to profound deafness can assist in an understanding of the long-term role of auditory feedback in speech production. Monophthongs and diphthongs represent a valuable field of investigation as they rely less on tactile feedback than on motor-kinesthethic feedback that may be more vulnerable to the effects of long-term hearing deprivation. Waldstein (1990) noted that vowel production was affected by acquired deafness, stating that “postlingual deafness affects the production of all classes of speech sounds, suggesting that auditory feedback is implicated in regulating the phonetic precision of consonants, [and] vowels” (p. 2099). Although researchers are in agreement about the importance of auditory feedback in the acquisition of speech production skills (Tobey, 1993), there has been considerable debate in the literature regarding its role in speech maintenance. Clarification of the role of auditory feedback in speech maintenance is important for the construction of a comprehensive theory of the speech production system (Perkell et al., 1992).

Studies examining the role of auditory feedback can be divided into several categories. The first investigates the effect of altering feedback in normally hearing adults, thus providing evidence regarding the short-term effects of feedback deprivation. For example, effects such as increased intensity, segment duration, and spectral changes have been shown to occur when speaking in increased background noise (Lane and Tranel, 1971; van Summers et al., 1988). Compensatory production mechanisms have been shown to result from alterations in the auditory feedback frequencies (Natke and Kalverum, 2001; Houde and Jordan, 1998, 2002). Similar studies have been carried out in deaf and deafened speakers by observing the changes in speech production that occur as a consequence of removal of auditory aids for short
implants, or that occur pre- and postimplantation of cochlear implants (e.g., Economou et al., 1992; Lane et al., 2001; Kishon-Rabin et al., 1999; Perkell et al., 1992). By illustrating the long-term effects of feedback deprivation after skilled speech has been acquired, studies of the differential deterioration in the speech production skills of adults with profound acquired deafness (deafened speakers) “can illuminate the underlying control mechanism of speech” (Lane and Webster, 1991; p. 860). In addition, research into the effects of acquired profound deafness in early childhood and adolescence provides additional information about the developmental role of auditory feedback.

The results of investigations into the role of auditory feedback have led to a number of theories regarding the nature of the speech production system. Older theories advocated that speech production is a closed-loop system, relying on auditory feedback for continuous monitoring of output to prevent production errors (Fairbanks, 1954; Siegel and Pick, 1974). However, Perkell et al. (2000) argue that “it is unlikely that auditory feedback is used for closed-loop error correction in the intra-segmental control of individual articulatory movements, because the feedback delay is too large” (p. 238). Rather, speech motor control is predominantly an open-loop system, where auditory feedback, in conjunction with proprioceptive and visual feedback, enables the learning of a robust internal model that is hypothesized to be a neural representation of spatial, kinematic, tactile and/or proprioceptive features of movements used to predict a desired acoustic consequence. Once speech has been acquired there is less need for peripheral feedback, and the role of auditory feedback is more in tuning the settings of this robust internal model (Perkell et al., 1992). However, recent studies of fundamental frequency perturbation have shown that auditory feedback can be used in a closed-loop fashion in conjunction with internal representations (Donath et al., 2002; Larson et al., 2000) leading to suggestions that the mechanisms involved in suprasegmental control may be more labile than those involved in segmental control (Lane et al., 1997; Perkell et al., 2000; Svirsky et al., 1992).

The evidence for a robust mapping of vocal gestures to their acoustic consequences can be seen in maintenance of reasonably good intelligibility in the speech of people who have been deafened for a number of years (Lane and Webster, 1991). While discernible changes in suprasegmental features of deafened speech have been reported (Cowie and Douglas-Cowie, 1992; Leder and Spitzer, 1993), the segmental features appear to be less affected by the reduction of auditory feedback. However, alterations in vowel production with a reduction in auditory feedback have been shown to occur in acoustic studies of deafened speakers. The results show a pattern of distortion from normal production that varies between and within studies, often as a consequence of differences in auditory capability between speakers. Several studies have shown some reduction of vowel space either along the F1 and/or F2 dimensions (Economou et al., 1992; Smyth et al., 1991; Waldstein, 1990), although this effect is not always present (Tartter et al., 1989; Lane et al., 2001). For example, while Svirsky and Tobey (1991) reported no significant changes for the American English point vowels /i, a, u/, Langereis et al. (1999) found small changes to these vowels, and Barker (1995) and Cowie and Douglas-Cowie (1992) reported the centralization of /i/. Richardson et al. (1993) found that two of the five Australian English (AE) speakers they studied showed reduction in /i/ and /ɛ/, while F1 changed according to speaker. Langereis et al. (1999) and Waters (1986) reported that monophthongs were replaced with diphthongs, while Plant and Hammarberg (1983) found that diphthongs were reduced to monophthongs. Both Waldstein (1990) and Economou et al. (1992) noted that their deafened speakers maintained tense lax vowel distinctions, as did two of the three speakers in the study by Svirsky and Tobey (1991). Waldstein (1990) also showed that her deafened speakers showed increased within-speaker variability, particularly in the increased standard deviations for F2 of vowel means.

Vowel production has a complex dependency on audition: speakers need to use auditory feedback to achieve vowel contrasts that are perceptible by a listener. Little tactile feedback is available in vowel production, as, with the exception of high vowels (Fitzpatrick and Ni Chasaidhe, 2002), the tongue does not generally make direct contact with other articulators. However, other proprioceptive information is available: for example, cutaneous receptors in the vocal tract, sensors transmitting information about joint location and sensors in the muscles providing information on force and movement of articulators (Gracco, 1995). McCaffrey and Sussman (1994) felt that the important question was whether severely and profoundly deaf speakers can produce vowels with significant discriminability and with critical distances between formants for contrasting vowel height and place. Deafened speakers have diminished audibility despite the assistance of clarification (via hearing aids or cochlear implants). For speakers with significant losses, the fundamental frequency (F0) and vowel F1 are most likely to be audible as they are lower in frequency (less than 2 kHz) than F2 and F3. “Thus, the resulting vowel inventory may achieve contrast in auditory space along fewer formant dimensions than those employed by speakers with normal hearing (McCaffrey and Sussman, 1994; p. 939).

The aim of this study is to look at vowel production in deafened speakers. In contrast to the majority of research on deafened speech that generally focuses on monophthongs, we shall be extending our acoustic studies to look in depth at diphthong production, on the premise that the moving formant tracks of diphthongs may be harder to produce in deafened speakers due to the absence of much tactile feedback. In addition, we shall look at male and female speakers separately given the influence of gender on formant frequency ranges coupled with the relatively restricted range of audible sounds in deafened speakers. The acoustic measures to be examined are vowel duration, target timing, formant position, and formant trajectories. In the discussion of our results we shall interpret our findings in the light of the various auditory feedback issues that have been raised in other research.
II. METHOD

All participants were native Australian English speakers, who used speech as their main means of communication. Seven deafened adults (four females and three males) and 11 normally hearing (NH) adults (six females and five males) took part in the study. All of the seven deafened (DF) speakers were deafened postlingually. Duration of deafness ranged from 6 to 48 years. All of the deafened speakers were classified as severely to profoundly deaf (three-frequency pure-tone average minimum of 80 dB HL in better ear). All of the normally hearing speakers were aged from 32 to 68 years and satisfied our criteria for normal hearing.

Duration of deafness ranged from 6 to 48 years. All of the deafened speakers were classified as severely to profoundly deaf (three-frequency pure-tone average minimum of 80 dB HL in better ear). Table I provides more detailed information on the deafened speakers. All deafened speakers used hearing aids during the task, except one male speaker who was awaiting a cochlear implant; however, this speaker’s vowel space was not significantly different from that of the other male speakers. All NH speakers were aged from 32 to 68 years and satisfied our criteria for normal hearing (passing an unaided three-frequency pure-tone average of 25 dB HL or less bilaterally).

It was important to select NH speakers who matched the age range and accent characteristics of the deafened speaker. Australian English has traditionally been described as having three accent types (“broad,” “general,” and “cultivated”) based on the pronunciation of certain vowels (Bernard, 1970a; Harrington et al., 1997). Although it is convenient to refer to these three varieties as separate entities, there is considerable phonetic overlap between them, and they are perhaps better described as socio-phonetic variation along a broadness continuum. In addition, accent changes occur over the last 30 years in Australian English has been documented (Cox, 1999), particularly in the monophthongs /æ/, /æ/, /æ/, and /æ/ and the second targets of /oʊ/ and /ʌ/, and also seen in the monophthongisation of the diphthongs /aʊ/ and /ɛʊ/.

Two trained phoneticians listened to the data from all speakers to determine accent type in order to appropriately match speakers between the deafened and normally hearing groups.

In addition, to ensure that the vowel space of the normally hearing speakers was representative of a larger population of Australian speakers, all monophthong tokens were checked acoustically against appropriate tokens from the Australian National Database of Spoken English (ANDOSL) (Millar et al., 1994). This database contains both citation form and sentence material from over 250 speakers of ages ranging from 18 to 45+, representing the Australian accent continuum from broad to cultivated. The tokens of the normally hearing speakers were found to fall within the 95%-confidence intervals of the means of the citation form vowel tokens from the ANDOSL database.

The speech material consisted of five repetitions of 18 monophthongs and diphthongs presented in a citation-form /hVd/ context (Table II). The IPA representation of the Australian English vowels is phonemic (Mitchell and Delbridge, 1965). The stimulus words were presented in a different randomized order for each of the five repetitions. All speakers were recorded in the Macquarie University Speech Pathology Clinic. The acoustic profile of these rooms conforms to the Australian Hearing Services (AHS) requirements for

TABLE I. Deafened speaker characteristics.

<table>
<thead>
<tr>
<th>Speaker number</th>
<th>Sex</th>
<th>Age</th>
<th>Accent type</th>
<th>Type of hearing loss</th>
<th>Severity (3-frequency average loss in better ear -dB)</th>
<th>Duration of deafness (years)</th>
<th>Hearing aid use at time of recording</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>71</td>
<td>General</td>
<td>Acquired—gradual loss</td>
<td>83</td>
<td>40</td>
<td>Bilateral</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>72</td>
<td>General</td>
<td>Acquired—gradual loss</td>
<td>80</td>
<td>32</td>
<td>Left ear</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>60</td>
<td>Broad</td>
<td>Acquired—gradual then viral infection</td>
<td>105+</td>
<td>18</td>
<td>None—awaiting cochlear implant</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>52</td>
<td>Upper</td>
<td>Sudden onset (viral) then gradual deterioration</td>
<td>103</td>
<td>26</td>
<td>Bilateral</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>65</td>
<td>General</td>
<td>Acquired—gradual loss</td>
<td>85</td>
<td>28</td>
<td>Bilateral</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>61</td>
<td>General</td>
<td>Acquired—sudden onset (viral and drug reaction)</td>
<td>83</td>
<td>48</td>
<td>Bilateral</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>50</td>
<td>General</td>
<td>Acquired—gradual then sudden</td>
<td>90</td>
<td>4</td>
<td>Bilateral</td>
</tr>
</tbody>
</table>

TABLE II. Target words and IPA symbols.

<table>
<thead>
<tr>
<th>Word</th>
<th>IPA</th>
<th>Word</th>
<th>IPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEED</td>
<td>/i/</td>
<td>WHO’D</td>
<td>/u/</td>
</tr>
<tr>
<td>HID</td>
<td>/ɪ/</td>
<td>HERD</td>
<td>/ɜ/</td>
</tr>
<tr>
<td>HEAD</td>
<td>/æ/</td>
<td>HAYED</td>
<td>/ɑ/</td>
</tr>
<tr>
<td>HAD</td>
<td>/æ/</td>
<td>HIDE</td>
<td>/ɛ/</td>
</tr>
<tr>
<td>HARD</td>
<td>/ɑ/</td>
<td>HOID</td>
<td>/ɔ/</td>
</tr>
<tr>
<td>HUD</td>
<td>/ʌ/</td>
<td>HOED</td>
<td>/oʊ/</td>
</tr>
<tr>
<td>HOD</td>
<td>/ɒ/</td>
<td>HOWED</td>
<td>/aʊ/</td>
</tr>
<tr>
<td>HOARD</td>
<td>/ə/</td>
<td>HEARED</td>
<td>/ɑ/</td>
</tr>
<tr>
<td>HOOD</td>
<td>/ʊ/</td>
<td>HARED</td>
<td>/ɛ/</td>
</tr>
</tbody>
</table>
sound sensitive spaces. The recordings were made onto digital audio tape (DAT) using a Sennheiser ME80 condenser microphone placed at approximately 45 cm from the speaker’s face. The majority of the speakers also wore a Nasometer mask during the recording, as data were also being collected for another experiment; however, speech samples used from four of the male normally hearing speakers were recorded without a Nasometer mask. In order to test whether the mask compromised articulatory movement, in particular jaw movement, a comparison was made between the vowel formant data from two NH speakers (one male and one female) recorded both with and without the Nasometer mask. Statistical analysis showed no significant difference in the first two formant values at the targets of all relevant vowels.

The speech data were digitized at 20 kHz with a 16-bit resolution, and the first two formants and their bandwidths were automatically tracked using ESPS/WAVES (12th-order LPC analysis, cosine window, 49-ms frame size, 5-ms frame shift). All automatically tracked formants were checked for accuracy and, where the formant tracking was unreliable due to discontinuities, the formants were smoothed interactively using the spectrogram and DFT spectrum. All the labeling was done in EMU, a hierarchical speech data management system (Harrington and Cassidy, 1999). Using both the speech waveform and spectrogram as guides, the acoustic onset of the vowel was marked at the first complete pitch period and the offset marked at the closure for the /d/ (see Fig. 1). A single acoustic vowel target was marked in the monophthongs and two targets were marked in the diphthongs. The targets of the monophthongs and the first targets of the diphthongs were marked at a point where there was the least movement in the formant tracks. For the high vowels, this point occurred where \( F_2 \) reached a peak (see Fig. 1 showing the target placement for the vowel /i/); for open vowels the target was marked where \( F_1 \) was at a maximum; for back vowels the target was marked where \( F_2 \) reached a trough. If none of the above criteria was satisfied for a given vowel, the target was marked at the point of maximum amplitude in the waveform. The second target of the rising and falling diphthongs was marked according to the same criteria as the first target. The second target of the centring diphthongs was marked at the right boundary of the vowel because a second target cannot be measured reliably.

The formant frequencies of the vowels were normalized using the Lobanov method (Lobanov, 1971). This was done to reduce speaker-specific aspects of the acoustic signal in order to examine more effectively the group characteristics of the deafened and normally hearing speakers, in particular, the size of the vowel space. In addition, as vowel classification studies will be carried out, normalization is useful in lowering classification errors by reducing intervowel variability while maintaining the appropriate distances between the vowel centroids (Disner, 1980). The normalization was carried out separately for each of the four populations of speakers (NH male/female and DF male/female). For each speaker, the Lobanov normalization of a vowel target formant frequency was carried out by subtracting the mean of the five tokens from each target point and dividing the result by the standard deviation, using the formula

\[
F_i = \frac{(F_i - \bar{F}_i)}{SD_i},
\]

where \( F_i \) is a given formant, \( \bar{F}_i \) is the mean of \( F \) across all vowels, and \( SD_i \) is the standard deviation of \( \bar{F}_i \) about its mean for all vowels. In order to present the data graphically in Hz values comparable to unnormalized values, the data were rescaled using the mean and standard deviation (Disner, 1980: p. 260).

The time-varying nature of the formant tracks was modeled with the coefficients of the discrete cosine transform (DCT) (Zahorian and Jaggard, 1993; for the mathematical modeling equations, see Watson and Harrington, 1999). Pilot vowel classification studies were carried out on the present vowel data to see which DCT coefficients should be used to encode the formant tracks. As in Watson and Harrington (1999), we found that only the first and second DCT coefficients played a significant role in separating the vowels: the first coefficient is proportional to the mean of the original formant track; the second models both the direction and tilt of the original formant track. Therefore, in the results reported herein, we have represented the first two formant tracks with the first two discrete cosine transformation (DCT) coefficients.

Statistical analysis was carried out in SPSS using two-way analyses of variance (ANOVA) with vowel type (VT) and hearing type (HT) as main factors using a \( p \) value of 0.05. Due to the inherent nature of the individual vowels studied, VT was always a significant main effect and will not be mentioned further except when there is a significant VT–HT interaction. Post hoc two-sample \( t \)-test analyses had a significance level of 0.01. In carrying out these statistical tests, we recognize that while there is generally thought to be less homogeneity in the speech parameters of deafened speakers, it is important to establish group patterns of behavior. In general, the group statistics reported reflect the performance of the individual members of the group; however, if statistical results were to be biased by a part of the group of speakers, it would be noted in the text.
III. RESULTS

A. Timing

For the monophthongs, analysis of variance shows a significant main effect for deafness in female speakers \[F(1,21) = 48.23, p = 0.000\] and, in male speakers, a significant two-way interaction between HT and VT \[F(10,21) = 2.51, p = 0.006\]; post hoc t-tests show a significantly longer duration for all DF monophthongs except /n/, where the p value (0.013) is slightly above the 0.01 cutoff limit. The total mean durations (standard deviations) averaged over all vowels for female speakers are: NH: 262 (96) ms.; DF: 285 (101) ms.; those for male speakers are: NH: 221 (83) ms.; DF: 278 (106) ms, bearing in mind that the duration for /n/ is longer for DF compared to NH speakers, but not significantly so.

In addition, for all speakers the tense vowels are always longer than the lax vowels, and there is little difference in the lax/tense duration ratio between NH and DF speakers [female: NH (0.53), DF (0.52); male: NH (0.52), DF (0.52)]. This value is very similar to other tense/lax ratios of vowels in similar phonetic contexts found in Australian English; 0.56 in a study of male adolescents (Bernard, 1970b), and 0.53 (females) and 0.55 (males) from the ANDOSL database (Millar et al., 1994). Therefore, the tense/lax distinction is preserved in the deafened speakers, as was found by Waldstein (1990). The preservation of the tense/lax distinction is important for phonemic contrast in Australian English. Australian English has two tense/lax pairs /a/ and /a/, and /i/ and /i/ (although the former can also be distinguished by an onglide) (Harrington et al., 1997).

In order to examine whether there are differences between NH and DF speakers in the time to attain a monophthong vowel target, the more accurate onset-target timing information provided by the /l/ onglide that is generally present in Australian English was examined. We calculated the ratio of the duration between the vowel onset and vowel target to the total vowel duration and then arc-sine transformed the result. However, there was no significant difference between DF and NH male and female speakers for this target-time duration.

In examining the total duration of the diphthongs, analysis of variance shows a significant main effect for deafness with no significant interaction [female: \(F(1,13) = 56.896, p = 0.000\) male: \(F(1,13) = 79.601, p = 0.000\)]. The mean total duration averaged over all diphthongs is longer for the deafened speakers compared with the NH speakers for females [NH: 370 (56) ms.; DF: 410 (51) ms] and males [NH: 315 (56) ms.; DF: 377 (59) ms]. To investigate whether the DF speakers took longer to reach the second target, we looked at the arc-sined ratio of the time between the first and second targets to the total duration; however, no consistent significant differences between the NH and DF speakers are seen.

The gender difference noted here in total vowel duration, where female speakers generally have longer duration than male speakers, has been commented on previously (Hillenbrand et al., 1995; Simpson, 2001, 2002). To test the significance of such a difference, analyses of variance were carried out on each population of NH and DF speaker separately. For monophthong total durations, there was no significant gender difference for the DF speakers and a significant gender/vowel interaction for the NH speakers \([F(10,583) = 2.404, p = 0.008]\); post hoc t-tests showed that the total duration for female speakers’ monophthong productions was longer than that for males, but not significantly so for /n/ and /n/. The lack of a significant gender difference for the deafened speakers was due primarily to the increased duration of some of the tense monophthongs in one male speaker. The total duration for diphthongs was significantly longer for females than for males in both groups of speakers [NH: \(F(1,371) = 102.283, p = 0.000\); DF: \(F(1,231) = 24.324, p = 0.000\)].

Waldstein (1990) observed an increase in the standard deviations for total duration and formant frequencies in his study of monophthongs in deafened speakers, but did not carry out any statistical analyses of these differences in variability. In order to see whether this increased variability was also found in the present data, we examined the variability in total duration values across all vowels for male and female DF and NH speakers using a Levene’s test for equality of variance (Levene, 1960). Contrary to expectations, there are no consistently significant increases in total duration variability for the deafened speakers for either monophthongs or for diphthongs.

B. Formants

1. Monophthongs

Table III shows the means and standard deviations of the first two formant values for the male and female speakers. Figure 2(a) shows the F1 and F2 centroids at the vowel targets for each of the monophthongs from the female speakers: the individual deafened speakers are superimposed on the average for the NH speakers. For both formants in the female data there are significant two-way VT–HT interactions [formant 1: \(F(10,21) = 10.17, p = 0.000\); formant 2: \(F(10,21) = 20.50, p = 0.000\)]. Post hoc t-tests show significantly lower formant values in the deafened speakers for the F1 of /a/ and /a/ (there is a trend only \(p = 0.017\) in /e/ due to the lowered F1 in three speakers only). There is also significantly retracted F2 in /l/, /l/, /l/, /l/, /l/, /l/, and /l/. This retraction can be seen in Fig. 2(a), where the deafened vowel space is retracted in the F2 plane for the front vowels, and reduced in the F1 mainly in the open vowels. Figure 2(b) shows the data for the deafened male speakers and the averaged normally hearing speakers. Analysis of variance shows a significant VT–HT interaction for both [formant 1: \(F(10,21) = 5.49, p = 0.000\); formant 2: \(F(10,21) = 2.91, p = 0.002\)]. Post hoc t-tests reveal a significant decrease in F1 of the vowels /a/, /a/, and /a/ and retraction in F2 of /l/ for the deafened speakers: resulting in a reduction in vowel space due to a raising of the open vowels and retraction of /l/ with no change in the other front vowels.

We investigated the possibility of an increase of variability around the vowel mean in the deafened speakers as might be expected if there was instability in vowel production. Three different methods were used: Levene’s test for homogeneity of variance (Levene, 1960), coefficient of variation...
Diphthongs

2. Diphthongs

The diphthong formant track data are shown for both male and female speakers in Figs. 3(a) and (b). The figures depict the movement of the formant tracks from vowel onset to offset. Each formant track has been time normalized to 20 equidistant values; then, each of these 20 values is frequency normalized as described in Sec. II (Method). The normally hearing data are a solid line and each deafened speaker a dashed line. Although overall the F1 and F2 formant tracks for the NH and DF speakers are fairly similar, there are differences nonetheless. In particular, when the F2 tracks are above about 2 kHz in the NH speakers, they are noticeable lower in the DF speakers.

Table IV gives the means and standard deviations for F1 and F2 at the first and second target for the normally hearing and deafened female speakers and Table V show the equivalent information for the male speakers. Analysis of variance reveals that there is a significant VT–HT interaction for F1 and F2 of target one in the female speakers [T1P1:F(6,13) = 4.86, \( p = 0.000 \); T1P2:F(6,13) = 24.82, \( p = 0.000 \)]. The results of post hoc t-tests indicate that for target one, F1 is significantly reduced for the deafened female speakers for /æ/, /æ/, /oʊ/, /oul/, and /aʊ/, and F2 is significantly retracted in /æ/, /æ/, /oʊ/, and /aʊ/. There is also a significant VT–HT interaction for the first and second formants of target one in the male data [T1P1:F(6,13) = 3.11, \( p = 0.006 \); T1P2:F(6,13) = 4.57, \( p = 0.000 \)]. Post hoc t-tests show that F1 at target one is significantly reduced for /æ/, and F2 is significantly reduced for /æ/ and /æ/.

There is a significant VT–HT interaction for F2 of target two in the male data [F(6,13) = 15.56, \( p = 0.000 \); post hoc t-tests show that all words spoken by the deafened group except for /aʊ/ have a significantly lower F2 than the normally hearing group. There is a similar significant interaction for the second target F2 in the male data [F(6,13) = 3.02, \( p = 0.007 \)], and post hoc t-tests reveal that only for /æ/ and /oul/ is there a significantly lower F2 in the deafened compared with the normally hearing group.

Figures 3(a) and (b) also suggest that for the NH and DF there may be differences between the slope of the formant track between T1 and T2. In order to investigate any differences between the deafened and normally hearing speakers in the dynamic nature of the diphthongs, we calculated the DCT coefficients of the formant tracks. We modeled the formant track slope with the second coefficient of the DCT as it models both the direction and tilt of the original trajectory. The second coefficient indicated that the slopes of the formant trajectories for the NH and DF speakers were in the same direction, as can be seen in Figs. 3(a) and (b); however, there were some differences in the degree of slope. Analyses of variance were carried out on values of DCTs of the F1 and F2 formant tracks for each of the diphthongs for the NH and DF speakers. In the male speakers there is no significant difference in the slope of the formant tracks between the DF and NH speakers. In the female speakers there are significant interactions both for F1 and F2 [F1:F(6,13) = 2.84, \( p = 0.010 \); F2:F(6,13) = 6.99, \( p = 0.000 \)]. Post hoc analyses show that in the DF speakers there is significantly less slope for the F1 of /æ/, /æ/, /oul/, and /aʊ/ due to a lower F1 at the first target; there is also a significantly reduced slope for the F2 of /æ/, /æ/, /oul/, with a similar trend for /æ/ and /aʊ/ \( p = 0.013 \), caused by a retraction of F2 at the second tar-

| Table III. Average monophthong F1 and F2 values in Hz (standard deviation) for female (left) and male (right) normally hearing (NH) and deafened (DF) speakers. |
|---------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|
|                                | NH  | DF  | NH  | DF  | NH  | DF  | NH  | DF  |
|                                |     |     |     |     |     |     |     |     |
| Formant one                    |     |     |     |     |     |     |     |     |
|                                |     |     |     |     |     |     |     |     |
| HEED                           | 314(41) | 321(32) | 2850(105) | 2512(138) | 288(25) | 306(27) | 2282(58) | 2247(55) |
| HID                            | 367(42) | 377(29) | 2777(92) | 2417(133) | 324(22) | 332(22) | 2163(46) | 2158(67) |
| HEAD                           | 476(64) | 435(47) | 2623(76) | 2212(215) | 409(32) | 412(34) | 2020(51) | 1982(96) |
| HAD                            | 789(112) | 817(55) | 2118(229) | 1750(112) | 667(59) | 635(73) | 1726(61) | 1649(81) |
| HARD                           | 1044(67) | 869(70) | 1533(85) | 1465(91) | 764(42) | 711(53) | 1287(67) | 1301(63) |
| HOD                            | 1015(96) | 918(43) | 1565(86) | 1544(72) | 770(51) | 721(40) | 1317(41) | 1344(87) |
| HOARD                          | 742(84) | 712(85) | 1125(98) | 1092(87) | 633(56) | 579(52) | 991(46) | 991(52) |
| HOOD                           | 426(56) | 414(47) | 792(90)  | 758(66)  | 382(26) | 401(42) | 765(61) | 808(43) |
| WHO’D                          | 383(68) | 387(53) | 869(90)  | 798(100) | 352(25) | 366(27) | 922(129) | 914(49) |
| HERD                           | 454(47) | 424(57) | 1800(102) | 1650(79) | 433(34) | 434(45) | 1493(80) | 1517(53) |

(except for /æ/ due to a lower

\( p = 0.000 \)); post hoc t-tests show that all words spoken by the deafened group except for /aʊ/ have a significantly lower F2 than the normally hearing group. There is a similar significant interaction for the second target F2 in the male data [F(6,13) = 3.02, \( p = 0.007 \)], and post hoc t-tests reveal that only for /æ/ and /oul/ is there a significantly lower F2 in the deafened compared with the normally hearing group.

Figures 3(a) and (b) also suggest that for the NH and DF there may be differences between the slope of the formant track between T1 and T2. In order to investigate any differences between the deafened and normally hearing speakers in the dynamic nature of the diphthongs, we calculated the DCT coefficients of the formant tracks. We modeled the formant track slope with the second coefficient of the DCT as it models both the direction and tilt of the original trajectory. The second coefficient indicated that the slopes of the formant trajectories for the NH and DF speakers were in the same direction, as can be seen in Figs. 3(a) and (b); however, there were some differences in the degree of slope. Analyses of variance were carried out on values of DCTs of the F1 and F2 formant tracks for each of the diphthongs for the NH and DF speakers. In the male speakers there is no significant difference in the slope of the formant tracks between the DF and NH speakers. In the female speakers there are significant interactions both for F1 and F2 [F1:F(6,13) = 2.84, \( p = 0.010 \); F2:F(6,13) = 6.99, \( p = 0.000 \)]. Post hoc analyses show that in the DF speakers there is significantly less slope for the F1 of /æ/, /æ/, /oul/, and /aʊ/ due to a lower F1 at the first target; there is also a significantly reduced slope for the F2 of /æ/, /æ/, /oul/, with a similar trend for /æ/ and /aʊ/ \( p = 0.013 \), caused by a retraction of F2 at the second tar-
that there is a significantly greater F2 get. For the centering diphthong /aʊ/, post hoc analysis shows that there is a significantly greater F2 slope.

The above results are not unexpected. Since the monophthongal space of the DF speakers is reduced compared to the NH speakers, we would also expect the absolute formant trajectory difference between T1 and T2 to be reduced. The increased F2 slope in /aʊ/ is also compatible with this viewpoint in that the result is a more centralized second target. Other studies on normally hearing speakers have suggested the monophthong vowel space is strongly linked to the position of the first and second vowel targets for the diphthongs, a vowel shift in the monophthongs space results in a vowel shift in the diphthongs (Cox, 1996; Cox and Palethorpe, 2001).

FIG. 2. (a) Female F1–F2 vowel space in Hz for the individual deafened speakers (dashed line) superimposed on the average for the normally hearing speakers (solid line). The label “n” for /s/ indicates normally hearing speaker mean. (b) Male F1–F2 vowel space in Hz for the individual deafened speakers (dashed line) superimposed on the average for the normally hearing speakers (solid line). The label “n” for /s/ indicates normally hearing speaker mean.

C. Classification experiments

In the first part of this study we established ways in which the deafened speech differed acoustically from the speech of normally hearing. There were differences in duration, and formant values at the vowel target, and consequently in the curvature of the formant tracks. The next question is whether these differences are sufficient to cause problems in perceiving the correct vowels.

Zahorian and Jagharghi (1993) found there is a high correlation between vowel confusions using formant-based automatic classifiers and perceptual experiments. In addition, other acoustic classification studies have found that vowels of different quality are more effectively distinguished when the acoustic parameters are based on spectral information sampled at multiple time points rather than from just at the vowel target (Harrington and Cassidy, 1994; Hillenbrand et al., 1995; Watson and Harrington, 1999).

In this final experiment to investigate the types of vowel confusions that potentially may be made on the deafened data by a normally hearing speaker we performed an automatic Gaussian classification based on modeling the formant tracks of F1 and F2 with the coefficients of the discrete cosine transform (DCT), as described in the Methods section. Vowel duration was also used as a classification parameter because it is known to be an essential feature in the
separation of Australian English vowels (Cox, 1996; Watson and Harrington, 1999). In the Gaussian classification, the centroid and covariance matrix of the training set (the tokens of the normally hearing speakers) were estimated for each vowel class. Tokens from the test set (those of the deafened speakers) were then classified based on their Bayesian distances to each of the training class centroids. A “round-robin” procedure was used whereby the tokens for each deafened speaker were used in turn as a test set. This resulted in a classification score for each deafened speaker, and the overall classification score was a summation of the individual results.

Vowel classification was carried out on the monophthongs and diphthongs together. The purpose of using the dynamic information contained in the formant track for classification of monophthongs is that, among other studies, a previous study of Australian English (Watson and Harrington, 1999) showed a differential between lax and tense monophthongs in the timing from vowel onset to target. This suggests that the shape of the formant contour may be different in these pairs of vowels and that this difference can be encoded by the DCT information and used to separate the vowel pairs in classification studies. This is relevant in the case of the Australian high vowels, particularly /i/, which in many speakers is realized with a delayed target (see Fig. 1). In addition, for many speakers of Australian English the

FIG. 3. (a) Female F1 and F2 tracks time aligned to vowel target one for normally hearing (solid line) and deafened (dashed line) speakers. (b) Male F1 and F2 tracks time aligned to vowel target one for normally hearing (solid line) and deafened (dashed line) speakers.
diphthongs /aʊ/ and /eɪ/ are realized as monophthongized variants when preceding /d/ (Cox, 1999).

The confusion matrix from the female data is given in Table VI, and that for the male data is given in Table VII. For the deafened female data overall 70% were correctly classified. Watson and Harrington (1999) performed a similar vowel classification of monophthongs and diphthongs together using only normally hearing speakers from the ANDOSL database, speakers who were different from the normally hearing speakers in the present research. They found 90% of the female vowels and 94.3% of the male vowels were correctly identified. These values are clearly a lot higher than those found due to a combination of retracted vowel duration or a retracted vowel space. For the vowels spoken by a deafened speaker compared with those spoken by a normally hearing speaker.

The confusions occurred generally either as a result of increased vowel duration or a retracted vowel space. For the females there is some confusion between the high front vowels due to a combination of retracted F2 and slightly lowered F1. Also, both /aʊ/ and /eɪ/ were misclassified as /ou/; this result is not unexpected in that there would be a tendency in reduced vowel space in these female deafened speakers for these more peripheral diphthongs to be identified as the more central /ou/. In the male speakers, no /aɪ/ vowel was correctly classified; it was confused with /a/, /æ/, and /i/; in Australian English /aɪ/ and /ai/ are distinguished by length so an increased duration for /aɪ/ will produce more confusion with /aɪ/; the confusion with the other two vowels was due presumably to a lower F1 in /aɪ/. There was some confusion between /aʊ/ and some monophthongs. This is not an unexpected result given the changing nature of /aʊ/ in Australian English, where it can be heard in some speakers as a lengthened monophthong (Cox, 1999). However, most incorrectly classified deafened monophthongs were classified as another monophthong, and most incorrectly classified diphthongs were classified as another diphthong.

When we separated the data from the deafened speakers into either a diphthong or monophthong group and repeated the classification experiment, we got 93.6% accuracy in the separation of diphthongs and monophthongs for the deafened female speakers and 90% accuracy for the males. In their study using normally hearing speakers from the ANDOSL database, Watson and Harrington (1999) also separated the vowels into monophthong and diphthong groups, and carried out a vowel classification using these two parameters. They found a similar high accuracy of separation between monophthongs and diphthongs of 94.3% and 93.2% for female and male data, respectively.

In order to further clarify the effect of the retraction of F2 on the classification scores of the deafened female speakers, the vowel classification was repeated for both males and females using just the monophthong data, separated into two vowel sets: A front set containing the vowels /a/, /æ/, /i/, and /e/, and a central/back set containing the vowels /aʊ/, /aɪ/, /eɪ/, /ou/, and /ʌ/. The correct classification scores for the deafened male speakers were 90% for the front and 80% for the back sets. In contrast, the correct score for the deafened female speakers was 87% for the back set, and only 59% for the front set of vowels.
IV. DISCUSSION

The results of this study show that the deafened speakers generally preserve a differentiated vowel space for Australian English. However, there are some patterns in the vowel changes that do occur that are gender specific. The deafened females demonstrate a vowel space retraction in the $F_2$ plane for the front monophthongs and a reduction in $F_1$ for the open monophthongs. By contrast, the deafened males demonstrate a lowering of $F_1$ for the open monophthongs and a retraction of $i/e$, but no change for the other front monophthongs. The average values of $F_2$ for the front vowels ($i/ə$, $i/ʌ$, $i/ə$, and $i/e$) for the normally hearing female speakers were all much greater than 2 kHz, whereas for the normally hearing male only $i/ə$ had an $F_2$ greater than 2 kHz.

Since the deafened speakers have some hearing loss over 2 kHz, the differences in the vowel spaces between the female and male deafened speakers suggest that the pattern of front vowel retraction in the female speakers may be due to an attempt to bring these vowels into a more perceptible range. This change is less necessary for the deafened males, as the $F_2$ of front vowels for males is generally lower than that of female speakers. A similar pattern of reduction in vowel space can be seen in the data of Waldstein (1990). She comments that her deafened speakers had a reduced $F_1$ and $F_2$ range, and it can be seen in her data that the reduction for $F_2$ occurs mainly in the front vowels where the frequency value tends to be reduced in the male and female deafened speakers to approximately 2 kHz. The deafened female speaker in
a study by Plant and Hammarberg (1983) also had a reduced $F_2$ range where the reduction was mainly in the front vowels to about 2 kHz. It is also interesting to note that while changes in the point vowels /i/ and /a/ have been seen in our deafened data and that of Waldstein (1990), the high back vowel /u/ in our data, and the /u/ in Waldstein’s American English deafened data remain relatively unchanged. Therefore, it may not be all the point vowels that remain stable under changes auditory conditions, as suggested by Svirsky and Tobey (1991), but the high back vowels that are more stable due to their stronger orosensory attribution.

The deafened speakers’ production of the diphthongs showed the same relationship to monophthongs as for the normally hearing speakers. The position in $F_1/F_2$ space of the first target for all the diphthongs relative to the nearest monophthong is similar regardless of hearing type (cf. /a/
compared to the first target of /a/ in Figs. 4 and 5). The changes in the deafened monophthong space were also observed in the diphthong space. The most noticeable was a retracted F2 for both the front monophthongs and the diphthong targets which fell in the front region of the vowel space of the deafened females. These results suggest the deafened speakers have a built-in model of the vowel space, and that their ability to relate diphthongs to monophthongs is not lost despite a reduction in auditory feedback. Further, that for deafened speakers, a loss of formant movement in diphthongs is not necessarily a tendency towards steady state, as suggested by Plant and Hammarberg (1983), but rather it maybe a consequence of the smaller difference between the first and second target due to a reduced monophthong space.

We found no consistent relationship between the degree of retraction of the front vowels and degree or duration of deafness in the deafened female speakers, although this has been reported elsewhere (Cowie et al., 1982; Plant and Hammarberg, 1983; Waldstein, 1990). Much research in this area has shown idiosyncratic speaker responses to a reduction of auditory feedback and this may be the case here. For example, the female speaker who had been deafened for the longest time had a position involving some public speaking and this may have played a role in the maintenance of her speech intelligibility.

We found no reliable increases in the formant frequency variability around the vowel mean in the deafened speakers compared to normally hearing speakers, contrary to other studies (Langeres et al., 1997; Waldstein, 1990). This result may be due to more careful speech on the part of the deafened subjects, particularly as the total durations for the vowels of the deafened speakers showed no greater increase in variability compared with those of the normally hearing speakers. In addition, the consistency in the pattern of vowel space change in our deafened speakers is unusual when compared with other studies on deafened speech. However, most of these studies were concerned with the pre- and postcochlear implantation effects on the vowel formant space. Their results have shown that speakers with a disordered vowel space respond differently to the reintroduction of hearing (Economou et al., 1992; Kishon-Rabin et al., 1999; Perkell et al., 1992; Lane et al., 2001), although this variability may be, in part, a consequence of differing acoustic cues presented to the speakers by the particular speech processors used.

The variability in the mean formant frequencies of some vowels between the NH speakers merits further comment. Languages do undergo vowel change: in particular, Australian English has been known to have undergone a vowel change over the last 30 years (Cox, 1999). For example, note the large standard deviation in the F1 and F2 of /a/ for the normally hearing female speakers seen in Table III; this vowel is unstable in Australian English as a consequence both of accent difference along the broad-general-cultivated continuum and of an observed change in the /æ/ vowel, which has become a more open retracted vowel in the last 30 years (Cox, 1999). The /a/ vowel is also unstable in Australian English and the fronting of this vowel in recent years has been noted in Australian as well as other English dialects (e.g., Cox and Palethorpe, 2001; Harrington et al., 2000; Watson et al., 1998). Therefore, it could well be that the increased F2 range for this vowel in the deafened male speakers may be a consequence of vowel shift rather than of hearing impairment, with two speakers showing fronting and one a retraction compared with the normally hearing speakers. These results highlight the need to have an accurate picture of the speech of a normally hearing population that has been recorded at a similar time to that of the target population in order to remove, as far as possible, the confounding effects of accent change.

Two aspects of timing in vowel production for deafened speakers were investigated in this study; total duration and the timing of the vowel target. We found that durations for both monophthongs and diphthongs for the deafened speakers were significantly longer than for the normally hearing, with the exception of /ʊ/ in the male speakers. This increased duration is likely to reflect a more careful and precise vowel production in the deafened speakers. It may be an attempt to get more orosensory feedback or to enhance intelligibility, as increased duration has been shown by many studies to occur in clear speech that is more intelligible (see Hargus-Ferguson and Kewley-Port, 2002, for a summary). Also, the deafened speakers do preserve the duration distinctions between tense and lax vowels, as was also found by Economou et al. (1992) and Waldstein (1990). There were no significant differences in the time taken from the first to the second targets in diphthongs. However, since the formant space of the diphthong is reduced, the result may be in fact a slower transition between targets since the formant distance between the first and second targets will be less. In the present data, the gender differences in duration were compatible with gender differences found in normally hearing data (Simpson, 2001, 2002).

The classification experiment suggested that although the vowel quadrilateral remains intact, the reduced deafened vowel space would have an effect on vowel intelligibility in these speakers. The confusions mainly occur between neighboring vowels and can generally be predicted by the change in vowel space. However, unlike the intelligibility results in Plant (1984), the deafened speakers do preserve the monophthong/diphthong distinction. Further research is necessary to see if a listener panel can confirm these results.

Many of our monophthong findings are very similar to that of Waldstein (1990). In that study she argues that the errors produced by the deafened speakers are consistent with the claim that the speech production system is under some form of closed-loop feedback control. However, Perkell et al. (2000), among others, suggest that due to a long feedback delay, it is unlikely the auditory signal provides immediate feedback for the control of segmental features in speech.

Perkell et al. (2000) have developed a speech motor control theory that states that speech movements are programmed to achieve auditory/acoustic goals. The programming of articulatory movements to achieve auditory goals uses an internal mapping of the relationship between articulatory profiles and their acoustic consequences. The auditory/
acoustic targets of vowels, in particular, are described as depending on the consequences of economy of effort and sufficient perceptual contrast (cf. Lindblom and Engstrand, 1989). Once the internal model is acquired, it is maintained through auditory feedback. However, it is not closed-loop auditory feedback, for there is no need for continuous direct auditory feedback with an internal model. With the maturation of the speaker, the model becomes robust and peripheral feedback is only used intermittently. As a consequence, they propose, deafened speakers should have reasonably good intelligibility many years after the onset of profound deafness, in spite of a reduction in auditory feedback.

The findings in this study would also lend some support to this speech motor control model. Overall, the deafened speakers maintain relatively differentiated vowels in spite of some changes in the vowel space, together with a clear separation in the classification of monophthongs and diphthongs, and a tense/lax distinction between vowels. They are able to achieve this distinction after a long time with reduced and/or auditory feedback, suggesting they have a reasonably robust internal mapping between articulatory configurations and their acoustic consequences. Perkell et al. (2000) argue that auditory feedback is used to refine the internal model parameters. It is feasible then that a reduction in auditory feedback may result in some adaptation of the acoustic/auditory goals of the internal model to bolster the effectiveness of the remaining auditory feedback.

Given the interesting findings in the current study, we are continuing research in this area of deafened speech. The relatively small sample of deafened speakers needs to be expanded to include speakers with a wider degree and duration of deafness. We are interested in studying segmental changes in deafened speakers on a long-term basis, paying particular attention to inter- and intraspeaker variability and using both acoustic and articulatory measures. We also acknowledge the need to document more audiological information about the speakers. Information on hearing loss above 2 kHz would seem to be especially important, given the present results. This research topic is relevant as it not only can give descriptions of segmental distortions in postlingual deafness that can be related to possible audiological remediation, but also for the opportunity it provides for studying control strategies in speech production.

ACKNOWLEDGMENTS

This work was supported in part by a Macquarie University Research Grant. Our thanks to Felicity Cox for the use of some acoustic data from normal-hearing speakers collected by herself and Sallyanne Palethorpe in the course of another research project. We are grateful to Libby Harricks for her support of this project. We also thank Jonathan Harrington, Anders Löfqvist, and two anonymous reviewers for their helpful comments.


