

# Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences

Bronwen G. Evans<sup>a)</sup> and Paul Iverson

*Department of Phonetics and Linguistics, University College London, 4 Stephenson Way, London NW1 2HE, United Kingdom*

(Received 26 December 2002; revised 17 October 2003; accepted 30 October 2003)

Two experiments investigated whether listeners change their vowel categorization decisions to adjust to different accents of British English. Listeners from different regions of England gave goodness ratings on synthesized vowels embedded in natural carrier sentences that were spoken with either a northern or southern English accent. A computer minimization algorithm adjusted F1, F2, F3, and duration on successive trials according to listeners' goodness ratings, until the best exemplar of each vowel was found. The results demonstrated that most listeners adjusted their vowel categorization decisions based on the accent of the carrier sentence. The patterns of perceptual normalization were affected by individual differences in language background (e.g., whether the individuals grew up in the north or south of England), and were linked to the changes in production that speakers typically make due to sociolinguistic factors when living in multidialectal environments. © 2004 Acoustical Society of America. [DOI: 10.1121/1.1635413]

PACS numbers: 43.71.Es, 43.71.Bp [PFA]

Pages: 352–361

## I. INTRODUCTION

In multidialectal environments (e.g., large cities such as London), native speakers of different accents regularly interact with one another. Speakers in these environments often avoid variants that are markedly regional or unusual, in order to facilitate communication (Trudgill, 1986) and to appear cosmopolitan (Foulkes and Docherty, 1999). However, they also retain some regional variants that show their allegiance to particular social or geographical groups (e.g., Foulkes and Docherty, 1999; Trudgill, 1986; Watt, 1998). In order to understand speech, listeners must somehow tolerate or adjust to this phonetic variation.

In British English, the focus of the present study, vowels are particularly important for distinguishing accents (e.g., Wells, 1982). Vowels can be used, for example, to broadly classify the many regional accents in England as northern or southern (Trudgill, 1986; Upton and Widdowson, 1996). Speakers of southern English accents use the vowel [ʌ] in words such as *buck*, but northern English speakers do not have this vowel; they say *buck* with a higher vowel, [ʊ], such that it becomes a homophone or near-homophone of *book*. Southerners and northerners both use the vowels [a] and [ɑ:], but with somewhat different lexical distributions; words such as *bath*, *dance*, and *ask* are produced with [ɑ:] by southerners and [a] by northerners, even though most words that have these vowels (e.g., *bad* using [a], and *bard* using [ɑ:]) do not differ between accents. When listening to a southerner, native speakers of northern English are thus required to map words that contain [ɑ:] and [ʌ] onto their own lexical representations that may be based on [a] and [ʊ].

Previous research on vowel normalization<sup>1</sup> has primarily examined how listeners adjust to the acoustic consequences

of anatomical variation, such as the length of the talker's vocal tract and characteristics of their glottis (e.g., Hillenbrand *et al.*, 1995; Ladefoged and Broadbent, 1957; Nearey, 1989). Listeners have been thought to normalize for anatomical variation by relying on gross acoustic—perhaps language-universal—characteristics of the speech signal such as the fundamental frequency of the vowel and the range of formant frequencies used in a carrier sentence [see Nearey (1989) for a review]. However, accent normalization is different because the realizations of phonemes in an accent cannot be predicted from such basic acoustic factors. That is, accent normalization must require more language-, accent-, and phoneme-specific processes.

How might listeners accomplish accent normalization? Exemplar models of speech perception (e.g., Goldinger, 1996, 1998) have theorized that listeners store phonetically detailed memory traces every time they listen to speech. Johnson (1997) has suggested that these exemplar representations can produce talker normalization effects, if listeners compare the words that they hear to stored exemplars of speech produced by similar talkers. Accent normalization can be viewed as an extreme example of this kind of talker normalization (see also Nygaard and Pisoni, 1998), in that the incoming speech could be compared to stored exemplars of speech produced by talkers with similar accents. For example, individuals who have experience with different British accents may be able to recognize northern-accented speech by mapping it onto similar stored exemplars produced by northern talkers, and recognize southern-accented speech by mapping it onto similar stored exemplars produced by southern talkers. Listeners may thus be able to fully adjust to vowel differences between accents, provided that they have had previous experience with similarly accented speech.

The evidence from cross-language speech research,

<sup>a)</sup>Electronic mail: bron@phon.ucl.ac.uk

however, suggests that individuals cannot easily adjust their phonemic categorizations to match the talker, at least when listening to foreign or foreign-accented speech. Novice listeners tend to assimilate foreign phonemes into the same categories that they use for native speech (Flege, 1992; Best, 1994; Best *et al.*, 1988, 2001). Although more experience with a foreign accent improves recognition abilities (e.g., Clarke, 2002), the category assimilation processes are difficult to modify; experienced listeners continue to assimilate most foreign phonemes into native categories, and create new categories for foreign speech primarily in cases where the foreign phonemes are too different from the native phonemes to be strongly assimilated (Flege, 1992, 1995).

Even bilingual listeners do not appear to strongly adjust their phonetic categorization processes when switching between different languages. One could expect, for example, that a Spanish–English bilingual would have a category boundary for /d/-/t/ at a shorter VOT for Spanish than for English, because Spanish speakers produce /t/ with a shorter VOT. However, late or weak bilingual listeners appear to use a single VOT boundary in both languages; they set their VOT boundary to a compromise location between English and Spanish (Flege, 1991, 1992). There is some evidence that early or strong bilinguals adjust their VOT boundaries for different languages (Elman *et al.*, 1977; Flege and Eefting 1987; Hazan and Boulakia, 1992), but the magnitudes of these boundary shifts are small and it is possible that these shifts may be caused by postperceptual processes (Bohn and Flege, 1993). The perception of different accents within the same language could operate similarly to these cross-language cases; listeners may assimilate the incoming speech to the phonetic categories of their own native accent without making specific adjustments for the accent of the speaker (see Flege, 1992) and create new categories only in cases where the non-native phonemes are too different from native phonemes to be strongly assimilated.

The present study investigated whether listeners adjust their vowel categorizations when listening to speech produced with different accents within the same language. The study contrasted two varieties of British English: Sheffield English, a northern variety, and Standard Southern British English (SSBE). Listeners with varying backgrounds were tested: northerners and southerners living in London (Experiment 1) and northerners living in the north of England (Experiment 2). Listeners heard synthesized vowels embedded in natural carrier sentences that were produced in either a Sheffield or SSBE accent. They gave goodness ratings on the vowels, and a computer program iteratively adjusted the F1, F2, F3, and duration values on successive trials until a best exemplar of each vowel was found. The aim was to assess whether listeners change their best exemplar locations based on the accent of the carrier sentence. Of particular interest were the vowels in *bath*, *bud*, and *cud*, which are produced very differently in northern and southern English accents (Wells, 1982).

## II. EXPERIMENT 1

Experiment 1 investigated whether listeners from the north and south of England who were living in London ad-

justed their vowel categorization decisions when listening to speech produced in SSBE and Sheffield English accents. London is a multidialectal community, and anyone living in the city for an extended period of time will have had experience with listening to and interacting with speakers of northern and southern English accents, as well as a wide variety of other accents. All listeners in experiment 1 thus had experience with different English accents, but their own native accents differed in terms of being northern or southern.

## A. Method

### 1. Participants

Twenty-three subjects were tested. All were native English speakers resident in London at the time of testing. They had lived in London for an average of 8.6 years, with a minimum of 1 year. The subjects were 20–45 years old, had no known hearing problems, and reported no speech or language difficulties. Three subjects were dropped from the experiment because their best exemplar locations were not reliable (i.e., their best exemplar locations for vowels that are produced the same in SSBE and Sheffield accents, such as in the words *bird* and *bed*, differed by more than 2 ERB in the two carrier sentences). Of the remaining 20 subjects, 10 were from southern England and 10 were from northern England. This classification of background was based on where they had lived between the ages of 5 and 18 years, an important period for the development of accent (Foulkes and Docherty, 1999).

### 2. Stimuli and apparatus

The stimuli consisted of synthesized vowels in the phonetic environments /b/-V-/d/, /b/-V-/θ/, and /k/-V-/d/, embedded in naturally spoken recordings of the carrier sentence *I'm asking you to say the word [ ] please*. The /b/-V-/θ/ words were included because northerners and southerners produce *bath* with different vowels. The /k/-V-/d/ words were included in case the potential shift in the *bud* and *cud* vowels with accent was affected by lexical influences (i.e., if these words were produced with the northern [ʊ] vowel, *cud* and *could* would become homophones, but *bud* would not become the same as any lexical competitor).

The carrier sentence was produced in both Sheffield and SSBE accents by the same male speaker ([əmaskɪnjə<sup>2</sup>seɪwɜːd---pliːs] in Sheffield and [ɑmɑːskɪnjuːtəseɪðwɜːd---pliːs] in SSBE). Multiple instances of the carrier sentences were recorded in each accent and one from each accent was selected for use in the experiment. The speaker had lived in Sheffield until the age of 19 and had then moved to the south of England, where he had lived for 7 years. This speaker was selected because he had an unusual ability to switch between accents at will, and was able to produce versions of both accents that sounded like those of native speakers.<sup>2</sup> In addition to the carrier sentences, the speaker was recorded reading a 2-min passage from a novel in both accents.

CVCs were embedded in the carrier sentences. The bursts, fricatives, and aspiration were spliced from the sentence recording, and the voiced portions were synthesized

on-line using the cascade branch of a Klatt synthesizer (Huckvale, 2003; Klatt and Klatt, 1990), to allow for fine-grained coverage of the entire vowel space. Each stimulus had a middle portion in which the formant frequencies were static, and had formant transitions appropriate for the consonants (see below). All transitions were linear. The stimuli varied in terms of F1–F3 frequencies and duration of the middle portion. F1 frequency was restricted so that it had a lower limit of 150 Hz and an upper limit of 950 Hz. F2 frequency was restricted to have a lower limit of  $F1+50$  Hz, and had an upper limit defined by the equation

$$F2_{\text{upper-limit}} = 3000 \text{ Hz} - 1.7 * F1. \quad (1)$$

F3 frequency was restricted to have a lower limit of 2000 Hz, an upper limit of 3150 Hz, and was always at least 100 Hz greater than F2. Duration of the middle portion was restricted to be greater than 20 ms and less than 403 ms.

All other synthesis parameters were chosen to mimic the natural speech recordings. For /b/-V-/d/, F1–F3 were 200, 1500, and 2400 Hz at the start of the formant transitions for /b/. The duration of the initial transition was 20 ms. F1–F3 were 200, 2300, and 3200 Hz at the end of the formant transitions for /d/. The final transition duration was 120 ms. F4 and F5 were fixed to 3200 and 4900 Hz throughout the stimulus. The bandwidths of F1–F5 were fixed to 100, 120, 150, 100, and 175 Hz. F0 (fundamental frequency) started at 116 Hz, rose to 126 Hz, and fell to 104 Hz. AV (amplitude of voicing) started at 45 dB, rose to 51 dB, and fell to 48 dB.

For /k/-V-/d/, F1–F3 began at the target formant frequencies of the vowel (i.e., there were no formant transitions at the onset of voicing, which is typical of voiceless stops). F1–F3 were 200, 1500, and 2600 Hz at the end of the formant transitions for /d/. The final transition duration was 40 ms. F4 and F5 were fixed to 3200 and 4450 Hz throughout the stimulus. The bandwidths of F1–F5 were fixed to 100, 120, 150, 150, and 175 Hz. F0 started at 125 Hz, rose to 128 Hz, and fell to 108 Hz. AV was ramped from 0 to 40 dB over the first 10 ms, rose to 50 dB, and fell to 45 dB.

For /b/-V-/θ/, F1–F3 were 200, 1300, and 2335 Hz at the start of the formant transitions for /b/. The duration of the initial transition was 20 ms. F2 and F3 were 1290 and 2400 Hz at the end of the formant transitions for [θ], and F1 ended on the target vowel frequency. The final transition duration was 20 ms. F4 and F5 were fixed to 3200 and 4900 Hz throughout the stimulus. The bandwidths of F1–F5 were fixed to 100, 160, 250, 150, and 175 Hz. F0 started at 115 Hz, rose to 125 Hz, and fell to 106 Hz. AV started at 40 dB, rose to 45 dB, and fell to 10 dB over the last 30 ms of the vowel.

After synthesis, the stimuli were processed using a multi-band filter to fine-tune the match between the long term average spectra of the synthetic and natural speech; frequencies between 0 and 1500 Hz were attenuated by 1.5 dB, frequencies between 1500 and 3500 Hz were amplified by 6 dB, and frequencies between 3500 and 5500 Hz were attenuated by 2 dB. This filter was necessary because adjustments of the Klatt synthesis parameters (e.g., formant bandwidths) were not entirely sufficient by themselves to match

the voice quality of the natural speech produced by this talker.

The stimuli were put together on-line during the experiment and played at a sampling rate of 11 kHz using a computer sound card and headphones (Sennheiser HD 414) in a sound attenuated booth.

### 3. Procedure

There were two testing sessions, one for each accent. The order of presentation was counterbalanced across subjects. Sessions were conducted on separate days to minimize the risk that subjects would be aware that the speaker was the same in both conditions (subjects were informally questioned after completing the experiment and no subject reported that the speaker was the same). Each session was self-paced and lasted approximately 1 h. At the start of each session, subjects listened to a short passage read by the speaker to familiarize them with the accent. They then found the best exemplar for one practice word (*kid*), and best exemplars for 16 experimental words: *bad, bard, bed, bird, bud, bod, bawd, bid, bead, booed, cud, could, cooed, Beth, birth, and bath*. To find the best exemplars, subjects heard a synthesized word embedded in a carrier sentence on each trial, and rated whether it was close to being a good exemplar of the target word that was displayed orthographically on a computer screen. They gave their response by positioning and clicking a computer mouse on a continuous scale from *close* to *far away*. The vowel parameters (F1, F2, F3, and duration) were adjusted after each trial using a customized procedure that was inspired by computer minimization algorithms (see Press *et al.*, 1992) and was designed to find the best exemplar location for that word in this four-dimensional parameter space.

The procedure had five stages, with six trials per stage, and was able to find the best exemplar locations within this large vowel space after 30 trials. In brief, the procedure adjusted F1 and F2 in Stages 1 and 2 (starting along a path in Stage 1 that would be likely to get close to best exemplars most quickly), adjusted the more secondary dimensions of F3 and duration in Stages 3 and 4, and then fine-tuned the best exemplar location in Stage 5.

The best exemplar was found in Stage 1 along a straight-line path through the F1/F2 plane that was defined by two points: the middle of the vowel space (F1=500 Hz and F2=1500 Hz) and the average F1 and F2 frequencies that the speaker of the carrier sentence had used for that word (averaged across the two accents). The path passed through these points and was terminated at the boundaries of the vowel space. For example, the Stage 1 search path for *bead* crossed diagonally through the vowel space, from the extreme high-front boundary of the space (i.e., low F1 and high F2), through the measured values for /i/ and the middle of the space, and through to the extreme low-back boundary of the space (i.e., high F1 and low F2). All other parameters were fixed to neutral values (F3=2500 Hz and duration=116 ms) in this stage.

On the first two trials of Stage 1, subjects heard the most extreme stimuli that it was possible to synthesize along the search path (e.g., in the case of *bead*, they heard extreme

high-front and low-back vowels, with the order of these two trials randomized). The selection of stimuli on the remaining trials was based on the subjects' judgments, using formulas that were designed to find stimuli along the path that would be perceived as better exemplars. On the third trial, subjects heard a stimulus that was selected by a weighted average of the first two stimuli, according to the equation

$$c = a * \frac{f(b)}{f(a) + f(b)} + b * \frac{f(a)}{f(a) + f(b)}, \quad (2)$$

where  $a$  and  $b$  are the positions on the search path for the first two trials,  $f(a)$  and  $f(b)$  are the goodness ratings for the stimuli on those trials (the goodness responses of *close* to *far away* were scaled from 0 to 1), and  $c$  is the new path position selected for the 3rd trial. On the 4th-6th trials, the stimuli were selected by finding the minimum of a parabola that was defined by the equation

$$\min = \frac{b - 0.5 * \{ [b - a]^2 * [f(b) - f(c)] - [b - c]^2 * [f(b) - f(a)] \}}{[b - a] * [f(b) - f(c)] - [b - c] * [f(b) - f(a)]}, \quad (3)$$

where  $b$  was the path position of the best stimulus found thus far;  $a$  and  $c$  were the most recently tested positions on either side of  $b$ ; and  $f(a)$ ,  $f(b)$ , and  $f(c)$  were the goodness ratings for those stimuli. In cases where Eq. (3) could not be calculated (i.e., if  $a$ ,  $b$ , and  $c$  were co-linear, or  $b$  was at an extreme position on the path), a weighted average [Eq. (2)] was calculated instead, based on the best exemplar found thus far and the last stimulus that had been played. At the completion of this stage, the parameters of the best stimulus found thus far were passed onto the next stage of the search algorithm.

The same six-trial search algorithm was used for the other stages, along different paths. Stage 2 found the best exemplar along a straight-line path that was orthogonal in the F1/F2 plane to the Stage 1 path, and included the best exemplar found in Stage 1. Stage 3 searched along the F3 dimension, keeping all other parameters fixed to the best exemplar values that had been found in Stage 2. Stage 4 searched along the duration dimension (log-scaled values), keeping all

other parameters fixed to the best exemplar values found in Stage 3. Stage 5 searched along a straight-line path through a three-dimensional F1, F2, and duration space (F3 did not vary), that began in the middle of the vowel space (F1=500 Hz, F2=1500 Hz, and duration=116 ms) and passed through the parameters of the best exemplar chosen thus far.

Subjects were allowed to repeat stages if they responded that the search algorithm had gone wrong (e.g., when the search was thrown off by an erroneous goodness rating). The best exemplar was defined as the stimulus given the highest goodness rating in Stage 5.

## B. Results

### 1. *Bud* and *cud*

As displayed in Fig. 1, listeners chose different formant frequencies for *bud* and *cud* in SSBE and Sheffield carrier sentences, indicating that they normalized these vowels for accent. The shift appeared to occur predominantly along the F1 dimension; both groups of listeners chose a higher F1 for *bud* and *cud* in SSBE (Standard Southern British English) sentences than in Sheffield sentences, although the size of the shift appeared to be larger for northerners. The differences in F1 and F2 were tested in separate repeated measures ANOVA analyses, with word (*bud* or *cud*) and sentence context (SSBE or Sheffield) coded as within-subject variables, and subject background (northern or southern) coded as a between-subject variable. For F2, there were no significant main effects or interactions,  $p > 0.05$ , suggesting that listeners were not normalizing for accent on this dimension. For F1, however, there was a main effect of sentence context,  $F(1,18) = 11.94$ ,  $p < 0.01$ , confirming that listeners overall chose higher F1 frequencies for *bud* and *cud* in the SSBE sentences.<sup>3</sup> There was also a main effect of subject background,  $F(1,18) = 12.08$ ,  $p < 0.01$ , demonstrating that northern listeners consistently chose higher F1 values for *bud* and *cud* than did southern listeners. There was no significant main effect of word and no significant interactions for F1,  $p > 0.05$ .

The effects of sentence and subject background on F1 can be seen clearly in Fig. 2. In the Sheffield sentences,

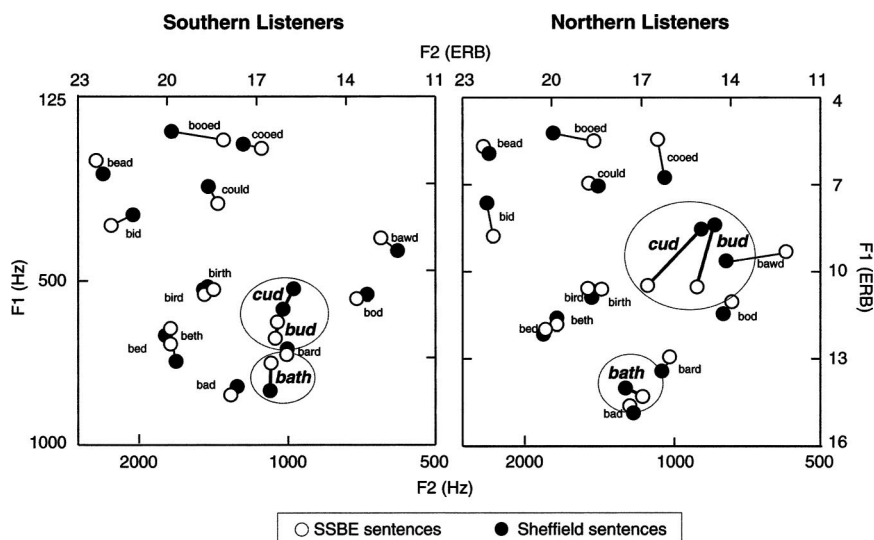


FIG. 1. Average F1 and F2 formant frequencies of best exemplars for northern and southern listeners in SSBE and Sheffield carrier sentences. The F1 frequencies of *bud* and *cud* were significantly different in the two carrier sentences for both groups of listeners, but no other words were reliably normalized for accent.

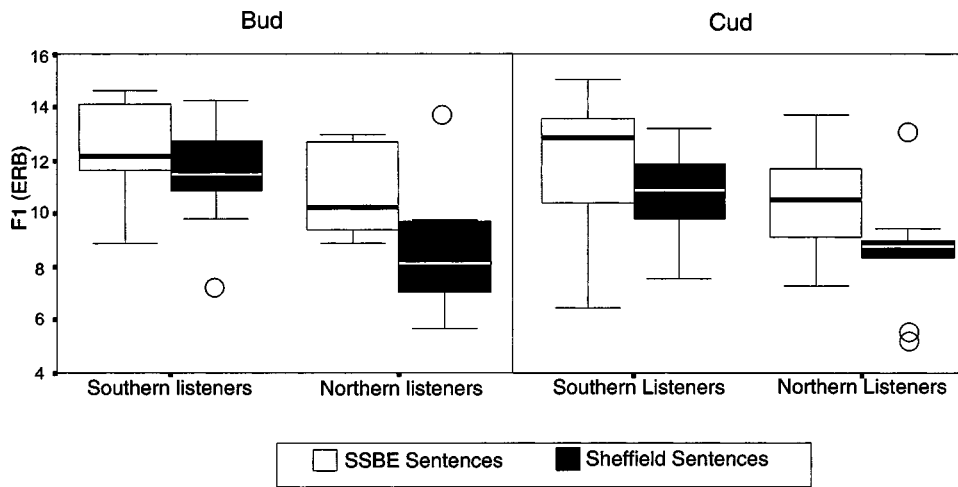


FIG. 2. Boxplots of F1 formant frequency values for *bud* and *cud* in SSBE and Sheffield carrier sentences for northern and southern listeners. Boxplots display the interquartile range of scores. The box shows the 25th to 75th percentiles, with a line at the median value. The lower and upper “whiskers” respectively show the first and last quartiles, with outliers represented by the unshaded circles. The best exemplar locations for northerners had higher F1 frequencies than those chosen by southerners, and both groups of listeners chose higher F1 frequencies in Sheffield than in SSBE carrier sentences.

northerners chose a high vowel (i.e., low F1 frequency) that was appropriate for that accent, but southerners chose a low-central vowel that was lower than Sheffield speakers actually produce. In the SSBE context, southerners chose a low vowel (i.e., high F1 frequency) that was appropriate for that accent, but northerners chose a central vowel that was higher than SSBE speakers produce. Although the size of the shift in *bud* was relatively small for southerners, the direction of this shift was consistent; nine of ten southerners chose higher F1 frequencies for *bud* in the Sheffield context.

As displayed in Tables I and II, there were few differences between *bud* and *cud* in terms of F3 or duration. Separate repeated measures ANOVA analyses for F3 and duration revealed that there were no significant main effects or interactions of sentence context, subject background, or word,  $p > 0.05$ , further suggesting that vowel normalization for accent only took place in the F1 dimension for *bud* and *cud*.

## 2. Bath

As displayed in Fig. 1, listeners chose relatively similar formant frequencies for *bath* in SSBE and Sheffield carrier

sentences, with perhaps a small shift in the F1 dimension for southern listeners. Separate repeated measures ANOVA analyses for F1, F2, and F3 revealed that there were no significant main effects or interactions of sentence or subject background,  $p > 0.05$ , suggesting that the formant frequencies of *bath* were not consistently normalized.

As displayed in Table II, there were no strong normalization effects for duration; listeners chose similar vowel durations in both sentence contexts, although there was a trend for southerners to choose shorter vowels in the Sheffield sentences. However, there was a consistent effect of subject background; southern listeners chose a longer vowel for *bath* in both sentence contexts than did northerners. A repeated measures ANOVA analysis verified that there was a main effect of subject background,  $F(1,18) = 8.09$ ,  $p < 0.01$ , but no significant main effect of sentence context or significant interactions,  $p > 0.05$ . The effect of subject background on duration can be seen clearly in Fig. 3. Northerners preferred shorter vowels that corresponded to their production of [a] in *bath*, and southerners preferred longer vowels that corre-

TABLE I. Average F3 frequencies (ERB) of best exemplars for northern and southern listeners in SSBE and Sheffield sentence contexts.

Word	Northern		Southern		Average
	SSBE	Sheffield	SSBE	Sheffield	
bud	23.4	23.0	22.8	23.5	23.2
cud	23.7	23.2	22.9	23.4	23.3
bath	22.7	22.1	23.1	22.1	22.5
bead	24.0	24.5	24.0	23.7	24.1
bid	24.0	24.2	23.8	23.2	23.8
bed	22.9	22.9	23.2	23.2	23.1
beth	23.4	22.4	23.2	23.4	23.1
bird	22.7	22.4	22.9	23.0	22.8
birth	22.3	22.4	23.5	23.1	22.8
bad	22.1	22.1	22.6	22.9	22.4
bard	22.8	22.9	23.6	23.2	23.1
bod	23.6	23.3	23.1	22.7	23.2
bawd	23.6	23.2	23.5	23.3	23.4
booed	22.2	22.4	22.4	22.4	22.4
cooed	22.5	22.8	22.2	23.4	22.7
could	23.3	23.0	23.4	22.8	23.1
Average	23.1	22.9	23.1	23.1	

TABLE II. Average durations (ms) of best exemplars for northern and southern listeners in SSBE and Sheffield sentence contexts.

Word	Northern		Southern		Average
	SSBE	Sheffield	SSBE	Sheffield	
bud	62.5	72.8	74.5	62.2	68.0
cud	63.7	62.0	81.8	69.7	69.3
bath	81.8	71.5	146.0	108.6	102.0
bead	138.9	129.4	120.9	130.7	130.0
bid	72.5	60.0	63.0	62.5	64.5
bed	78.1	66.8	78.1	66.8	72.5
beth	76.9	68.1	71.3	71.6	72.0
bird	182.6	144.2	119.5	130.6	144.2
birth	135.6	137.5	145.1	158.1	144.1
bad	101.0	70.6	104.6	84.7	90.2
bard	208.2	181.3	176.6	174.0	185.0
bod	68.5	79.5	77.5	74.4	75.0
bawd	187.8	170.1	176.3	151.7	171.5
booed	176.7	161.7	162.1	154.5	163.8
cooed	173.3	169.7	147.0	146.0	159.0
could	65.3	64.8	70.0	77.2	69.3
Average	117.1	106.9	113.4	107.7	

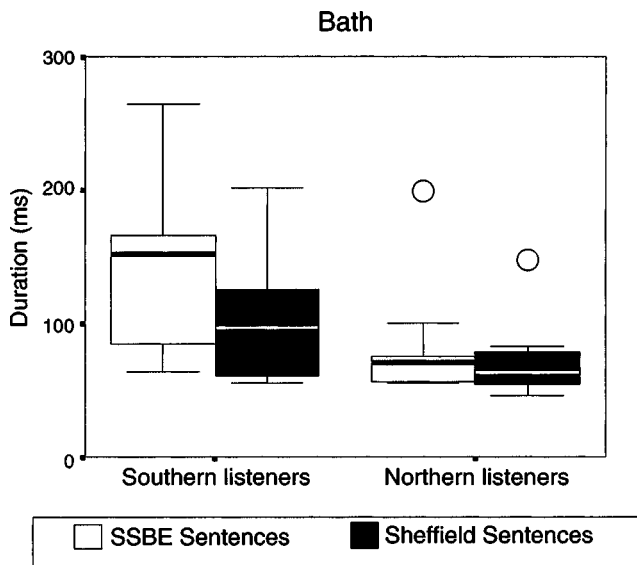


FIG. 3. Boxplots of duration values for *bath* in SSBE and Sheffield carrier sentences for northern and southern listeners. Northerners chose shorter vowels than southerners overall, but there was no normalization for the accent of the carrier sentences.

sponded to their production of [ɑ:] in *bath*. Although the formant frequencies were not significantly different, the results trended in the same direction (see Fig. 1); the median values of F1 and F2 for *bath* were more similar to *bad* ([bad]) than *bard* ([ba:d]) for northerners, and were more similar to *bard* than *bad* for southerners. This difference may have failed to reach significance because [a] and [ɑ:] have very similar formant frequencies overall; the vowels differ more markedly in duration.

### 3. Other words

As displayed in Fig. 1, listeners chose similar F1 and F2 frequencies in SSBE and Sheffield carrier sentences for most other words, with shifts for a few words such as *booed* and *bawd*. The potential differences in F1 and F2 were tested in separate repeated measures ANOVA analyses, with word (i.e., all words other than *bud*, *cud*, and *bath*) and sentence context coded as within-subject variables, and subject background coded as a between-subject variable. There was a main effect of word for F1,  $F(12,216)=203.98$ ,  $p<0.01$ , and F2,  $F(12,216)=113.76$ ,  $p<0.01$ , demonstrating that different words had different formant frequency values, but there were no main effects of sentence context or subject background, and no significant interactions,  $p>0.05$ . The differences in *bawd* and *booed* displayed in Fig. 1 were thus not reliable. As displayed in Tables I and II, listeners generally chose similar values for F3 and duration in SSBE and Sheffield sentence contexts. Separate repeated measures ANOVA analyses revealed that there was a main effect of word for F3,  $F(12,216)=7.78$ ,  $p<0.01$ , and for duration,  $F(12,216)=40.58$ ,  $p<0.01$ , demonstrating that different words had different F3 and duration values, but there were no main effects of sentence context or subject background, and no significant interactions,  $p>0.05$ .

It is notable that northern and southern listeners both chose a high-front vowel for *booed* and a high-central vowel

for *coed*, rather than high-back vowels with lower F2 frequencies (see Fig. 1). Although these preferences may seem unusual, they correspond to recent changes in the way that British English speakers produce these vowels; younger speakers in particular have begun to produce these traditionally high-back vowels with less lip rounding and with a more forward tongue position (Docherty and Foulkes, 1999; Torgersen, 1997; Williams and Kerswill, 1999).

## III. EXPERIMENT 2

Experiment 1 demonstrated that individuals living in London normalized *bud* and *cud* for accent, and that the patterns of normalization depended on whether the listeners were northern or southern. Experiment 2 further examined the role of language experience on vowel normalization by testing northerners who still live in the north of England. The subjects were born and raised in Ashby de la Zouch, a market town where the dominant accent is similar to that spoken in Sheffield. The subjects were 16–17 years old, and had not yet moved for employment or university education. The aim was to determine whether the patterns of normalization found for northerners in Experiment 1 were affected by the subjects' time living in London, or whether all northerners (i.e., even those who have not lived in the south) have the same patterns of normalization.

### A. Method

#### 1. Subject selection

Twelve subjects were tested. All were native English speakers, aged 16–17 years, born and raised in Ashby de la Zouch, and reported no hearing or language problems. One subject was dropped from the experiment because her best exemplar locations were not reliable (i.e., as in Experiment 1, subjects were dropped when the best exemplar locations for the vowels that were stable between accents differed by more than 2 ERB).

#### 2. Stimuli and apparatus

The stimuli were synthesized in advance so that the experiment could be run using a portable computer. The entire range of possible vowels was synthesized with a resolution of 0.5 ERB<sup>4</sup> (Moore *et al.*, 1997) in F1 and F2. Duration was quantized in 16 steps on a log scale, from 20 to 403 ms. F3 was fixed to 2500 Hz for all stimuli; although Experiment 1 had shown that F3 varied for different words, the results suggested that this parameter made only a modest contribution to perceived goodness. There were a total of 7616 stimuli synthesized for each of the CVC contexts. The stimuli and apparatus were the same as in Experiment 1 in all other respects.

#### 3. Procedure

There was a four-stage search for best exemplars along the F1, F2, and duration dimensions, with six trials for each stage; the F3 adjustment stage was omitted. The procedure was the same as in Experiment 1 in all other respects.

## Ashby Listeners

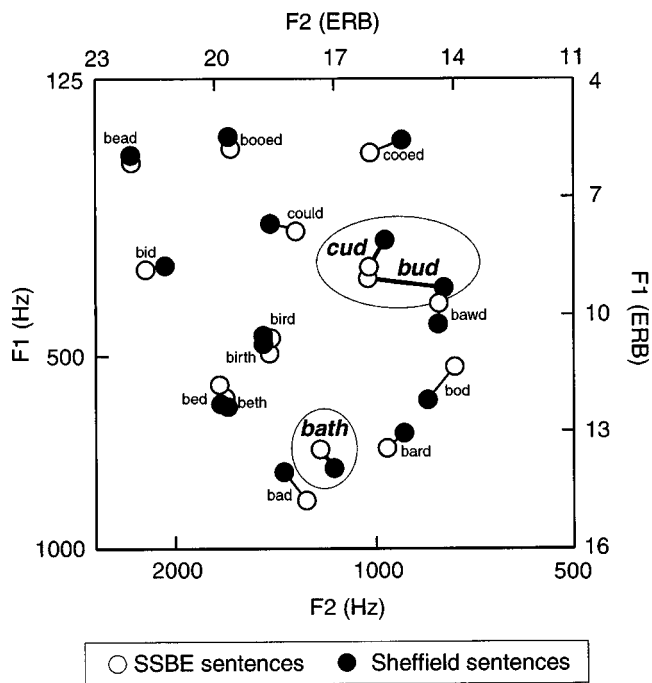


FIG. 4. Average F1 and F2 formant frequencies of best exemplar locations for Ashby listeners in SSBE and Sheffield carrier sentences. The F1 and F2 frequencies did not vary significantly between the two carrier sentences, suggesting that no vowels were normalized for accent.

## B. Results

### 1. Bud and cud

As displayed in Fig. 4, Ashby listeners chose similar formant frequencies for *cud* in SSBE and Sheffield carrier sentences, but for *bud* there was a possible difference in the F2 dimension; listeners tended to choose a higher F2 for *bud* in SSBE than in Sheffield carrier sentences. Separate repeated measures ANOVA analyses for F1 and F2 revealed that there were no main effects of word, sentence context, or their interactions,  $p > 0.05$ , demonstrating that the shift in the F2 dimension for *bud* was not reliable. As displayed in Table III, there was also little difference between *bud* and *cud* in terms of duration; a repeated measures ANOVA analysis revealed that there were no main effects of word, sentence context, or their interactions,  $p < 0.05$ . There was thus no consistent evidence that Ashby listeners normalized *bud* and *cud* for accent; they chose traditionally northern vowels in both carrier sentences.

### 2. Bath

Ashby listeners also chose similar formant frequencies for *bath* in SSBE and Sheffield carrier sentences (see Fig. 4), indicating that there was no normalization for accent. There was a trend for listeners to choose a longer vowel in the SSBE than in the Sheffield carrier sentences (Table III). However, separate repeated measures ANOVA analyses for F1, F2, and duration revealed that there was no main effect of sentence context,  $p < 0.05$ . There was thus no clear evidence that Ashby listeners normalized *bath* for accent.

TABLE III. Average durations (ms) of best exemplars for Ashby listeners in SSBE and Sheffield sentence contexts.

Word	SSBE	Sheffield	Average
bud	63.8	65.4	64.6
cud	76.9	55.8	66.4
bath	108.4	79.3	93.8
bead	99.0	115.7	107.4
bed	84.5	56.5	70.5
bid	71.9	68.5	70.2
beth	73.9	58.2	66.0
bird	139.9	117.3	128.6
birth	124.5	113.3	118.9
bad	85.4	77.8	81.6
bard	156.4	162.7	159.5
bod	72.4	72.0	72.2
bawd	140.1	142.5	132.3
booed	170.4	164.4	167.4
cooed	146.7	140.8	143.8
could	76.1	69.6	72.9
Average	105.6	96.4	

## 3. Other words

For all other words, listeners chose similar formant frequencies and durations for each target word in SSBE and Sheffield carrier sentences. Separate repeated measures ANOVA analyses for F1, F2, and duration revealed that there was a main effect of word for F1,  $F(12,120) = 76.08$ ,  $p < 0.01$ , F2,  $F(12,120) = 47.24$ ,  $p < 0.01$ , and duration,  $F(12,120) = 15.61$ ,  $p < 0.01$ . However, there was no main effect of sentence context and no significant interaction with word,  $p > 0.05$ , suggesting that none of the other words varied depending on accent.

## IV. DISCUSSION

The results demonstrated that individuals living in London normalized the vowels in *bud* and *cud*—but not *bath*—for southern and northern English accents, with the patterns of normalization reflecting each listener's linguistic experience. When individuals living in London heard sentences that were similar to their native accent, they chose formant frequencies for *bud* and *cud* that matched what speakers of that accent would produce; southerners living in London selected an [ʌ] vowel (i.e., high F1) when listening to SSBE sentences and northerners living in London selected an [ʊ] vowel (i.e., low F1) when listening to Sheffield sentences. When individuals living in London heard sentences that did not match their native accent (e.g., northerners listening to SSBE speech), they chose centralized vowels for *bud* and *cud* rather than the [ʌ] and [ʊ] vowels that would normally be produced in SSBE and Sheffield accents, respectively. Northerners who were less experienced with southern accents (i.e., Ashby listeners) did not normalize for accent at all, choosing vowels in Sheffield and SSBE sentences that would be appropriate for northern speakers.

Episodic memory research has shown that individuals store phonetically detailed representations of spoken words in long-term memory (e.g., Goldinger, 1998; Nygaard and Pisoni, 1998; Palmeri *et al.*, 1993), and our working hypoth-

esis based on this research was that listeners would choose best exemplars that matched their long-term memory representations for words spoken by speakers with similar accents. It was surprising then, that northerners living in London, for example, chose best exemplars for *bud* and *cud* in the SSBE sentences that do not match how southerners actually produce these vowels [see Wells (1982) for a phonetic description of SSBE vowel production]. Although this may suggest that listeners were not performing the task using stored exemplars, it is possible that listeners were using inaccurate exemplars that had been affected by perceptual magnet effects (Iverson and Kuhl, 1995, 1996, 2000; Iverson *et al.*, 2003) or category assimilation processes (Best *et al.*, 1988, 2001; Flege, 1992, 1995). That is, the northerners' perception of SSBE [ʌ] may have been distorted because they do not have a native /ʌ/ category; northerners may perceive the SSBE [ʌ] to be a member of their native /ə/ or /ɜ:/ categories [or a new category caused by merging /ʌ/-/ɜ:/; see MacKay *et al.*, (2001)], causing the SSBE [ʌ] to sound more centralized. It is plausible that such perceptual distortion caused northern listeners to remember mistakenly that southerners produce centralized vowels for *bud* and *cud*. In other words, the basic hypothesis that listeners choose best exemplars that match long-term memory representations may be correct, but the memories of listeners may be inaccurate.

There are two aspects of the present results that are inconsistent with this perceptually distorted exemplar account. First, the *bud* and *cud* vowels that southerners chose in the Sheffield sentences cannot be easily explained by perceptual magnet effects or category assimilation. Northerners produce *bud* and *cud* using [ʊ], and one would normally expect that southerners would assimilate this northern vowel into whatever native category is most similar perceptually (i.e., the southerners' own /ʊ/ category, that they use in words like *book* or *could*). Instead, southerners chose a low-central best exemplar for *bud* and *cud* that is on the other side of the vowel space from [ʊ]. It seems unlikely that southerners erroneously perceive the northern [ʊ] as a low-central vowel. Second, there was no normalization for *bath*. Southerners and northerners both use the vowels [a] and [ɑ:], and speakers of British English are very aware that the lexical distribution of these vowels is a clear marker of accent (Trudgill, 1986). Northerners in London thus know that southerners produce *bath* with [ɑ:], as do southerners in London know that northerners produce *bath* with [a]. Yet subjects in this experiment chose vowels for *bath* based on their own accent, rather than on their knowledge of what vowel would be expected based on the accent of the carrier sentence.

Although these patterns of normalization may seem idiosyncratic, they correspond closely with the changes in production that speakers tend to make when they live in multi-dialectal environments (Trudgill, 1986). Northerners who live in the south of England typically modify some aspects of their accent in order to fit in with southerners; they change their production of the vowel in *bud* and *cud* so that it becomes centralized (Trudgill, 1986), much like the centralized vowel that they chose as best exemplars for these words in the SSBE-accented carrier sentence. Northerners also maintain some aspects of their regional identity; they retain their

[a] when producing words like *bath* [and would rather “drop dead” than produce these words like a southerner (Trudgill, 1986)], much like they chose [a] for *bath* in both carrier sentences. Southerners living in London, however, speak the locally dominant dialect and are less apt to modify their productions when speaking to others; they likewise made relatively small adjustments to *bud* and *cud*, and preferred southern pronunciations of *bath* in both carrier sentences.

Production may also help explain why Ashby listeners did not perceptually normalize for accent. One could imagine that Ashby listeners did not normalize because they had not had enough perceptual experience with southern accents (e.g., see Labov and Ash, 1997). However, Ashby listeners are regularly exposed to southern English accents through the media (Foulkes and Docherty, 1999) and they are able to correctly identify the accents of southern speakers in perceptual tests (Evans, 2001). Moreover, all listeners heard a short passage read by the speaker in the relevant accent before starting the experiment, and such short-term familiarization can be enough to tune speech recognition processes to the characteristics of individual talkers (Nygaard and Pisoni, 1998). It may have been more important that these listeners (aged 16–17 years, born and raised in Ashby) had not had the experience of modifying their own speech in order to fit into a new environment (e.g., when attending a university). It is thus plausible that these Ashby listeners chose northern vowels in the SSBE carrier sentence because they had not yet learned to change their speech when talking to southerners, even though they know how southerners talk (see also Flege, 2003).

The mechanism responsible for this perception-production link is unclear. The results are consistent with motor theory's claim that listeners perceive speech in terms of their own articulatory gestures (Liberman *et al.*, 1967). That is, the acquisition of new articulatory targets to modify one's own accent may have directly changed how vowels in the southern and northern accents were perceived. However, it is possible too that the best exemplars found in perceptual experiments reflect auditory targets that a listener tries to achieve when speaking (Allen and Miller, 2001). That is, listeners may need to first change their notions of which phonemes sound good, in order to learn to modify the accent of their own speech.

The current study is only a first attempt to investigate how vowel perception is modified to accommodate accent differences in the same language, but the results thus far differ from descriptions of how listeners perceive foreign or foreign-accented speech. Cross-language research has suggested that foreign-accented phonemes are assimilated into the same categories that listeners use for native speech (e.g., Best, 1994; Flege, 1992), but the present results suggest that listeners can adjust their categorizations to accommodate different accents within the same language. Cross-language research has also emphasized the age and amount of exposure as determining factors in the ability to perceive and produce a foreign language (Flege *et al.*, 1999). However, the current results suggest that perceptual adjustments for accents within the same language are not simply determined by exposure; changes in best exemplar locations appear to follow sociol-

linguistic principles that help explain what happens to speech when an individual chooses to fit in with a particular community or subculture. It is likely that these sociolinguistic principles [e.g., a listener's motivation to learn; see Piske *et al.* (2001)] also have a role in cross-language cases. That is, although learning to perceive and produce the phonemes of a new language must depend on exposure to a great extent, losing one's accent may also be affected by one's willingness to be identified as a member of the same culture as a native speaker of that language.

## ACKNOWLEDGMENTS

This research was supported by an EPSRC Doctoral Training Award. We are grateful to Richard Dewire and Alexandra Evans for their assistance with the experiments.

<sup>1</sup>A broad definition of *normalization* is used here (i.e., the perceptual and cognitive adjustments that allow listeners to accommodate differences between speakers), rather than the narrow definition that is sometimes used [i.e., a hypothetical perceptual process in which speaker-specific information is discarded (Pisoni, 1997)].

<sup>2</sup>The speaker was coached to produce words like *bath* and *ask* with an [ɑ:] in SSBF sentence context, because he normally produced these words with [a].

<sup>3</sup>Note that there were no main effects of sentence context for the other vowels. Thus, the differences in the sentence contexts did not cause listeners to shift the overall vowel space due to some experimental artifact; the listeners specifically normalized *bud* and *cud*.

<sup>4</sup>Even though the quantization of the stimulus space used ERB values, the algorithm searched through the vowel space along linear Hz paths. The linear Hz values were rounded in the search algorithm to the values of the nearest available stimulus. This allowed the search algorithm to be the same as in experiment 1, and still have a stimulus quantization that was perceptually uniform.

Allen, J. S., and Miller, J. L. (2001). "Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate," *Percept. Psychophys.* **63**, 798–810.

Best, C. T. (1994). "The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model," in *The Development of Speech Perception*, edited by J. C. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 167–244.

Best, C. T., McRoberts, G. W., and Goodell, E. (2001). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," *J. Acoust. Soc. Am.* **109**, 775–794.

Best, C. T., McRoberts, G. W., and Sithole, N. N. (1988). "The phonological basis of perceptual loss for nonnative contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants," *Human Percept. Perform.* **14**, 345–360.

Bohn, O.-S., and Flege, J. E. (1993). "Perceptual switching in Spanish/English bilinguals," *J. Phonetics* **21**, 267–290.

Clarke, C. M. (2002). "Perceptual Adjustment to Foreign-Accented English With Short Term Exposure," paper presented at the 7th International Conference on Spoken Language Processing (ICSLP), Denver, CO.

Docherty, G. J., and Foulkes, P. (1999). "Derby and Newcastle: Instrumental phonetics and variationist studies," in *Urban Voices: Accent Studies in the British Isles*, edited by P. Foulkes and G. J. Docherty (Arnold, London), pp. 47–71.

Elman, J. L., Diehl, R. L., and Buchwald, S. E. (1977). "Perceptual switching in bilinguals," *J. Acoust. Soc. Am.* **62**, 971–974.

Evans, B. G. (2001). "An investigation into the identification of four varieties of English," unpublished master's thesis. University of Newcastle upon Tyne, Newcastle upon Tyne, U.K.

Flege, J. E. (1991). "Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language," *J. Acoust. Soc. Am.* **89**, 395–411.

Flege, J. E. (1992). "The intelligibility of English vowels spoken by British and Dutch talkers," in *Intelligibility in Speech Disorders: Theory, Measurement and Management*, edited by R. D. Kent (Benjamins, Amsterdam), pp. 157–232.

Flege, J. E. (1995). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 233–277.

Flege, J. E. (2003). "Assessing constraints on second-language segmental production and perception," in *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and Niels Schiller (Mouton de Gruyter, Berlin).

Flege, J. E., and Eefting, W. (1987). "Cross-language switching in stop consonant perception and production by Dutch speakers of English," *Speech Commun.* **6**, 185–202.

Flege, J. E., Mackay, I. R. A., and Meador, D. (1999). "Native Italian speakers' perception and production of English vowels," *J. Acoust. Soc. Am.* **106**, 2973–2987.

Foulkes, P., and Docherty, G. J. (1999). "Urban Voices—Overview," in *Urban Voices: Accent Studies in the British Isles*, edited by P. Foulkes and G. J. Docherty (Arnold, London), pp. 1–24.

Goldinger, S. D. (1996). "Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory," *Exp. Physiol.* **22**, 1166–1183.

Goldinger, S. D. (1998). "Echoes of Echoes? An Episodic Theory of Lexical Access," *Psychol. Rev.* **105**, 251–279.

Hazan, V. L., and Boulakia, G. (1992). "Perception and Production of a voicing contrast by French-English bilinguals," *Lang Speech* **36**, 17–38.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.

Huckvale, M. (2003). *Speech Filing System* (computer software) (University College, London).

Iverson, P., and Kuhl, P. K. (1995). "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling," *J. Acoust. Soc. Am.* **97**, 553–562.

Iverson, P., and Kuhl, P. K. (1996). "Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/," *J. Acoust. Soc. Am.* **99**, 1130–1140.

Iverson, P., and Kuhl, P. K. (2000). "Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?" *Percept. Psychophys.* **62**, 874–883.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). "A perceptual interference account of acquisition difficulties for non-native phonemes," *Cognition* **87**, B47–B57.

Johnson, K. (1997). "Speech Perception without Speaker Normalization: An Exemplar Model," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullenix (Academic, San Diego), pp. 145–165.

Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.* **87**, 820–857.

Labov, B., and Ash, S. (1997). "Understanding Birmingham," in *Language Variety in the South Revisited*, edited by C. Bernstein, T. Nunally, and R. Sabino (Alabama, Tuscaloosa).

Ladefoged, P., and Broadbent, D. (1957). "Information conveyed by vowels," *J. Acoust. Soc. Am.* **29**, 98–104.

Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). "Perception of the speech code," *Psychol. Rev.* **74**, 431–461.

MacKay, I. R. A., Flege, J. E., Piske, T., and Schirru, C. (2001). "Category restructuring during second language acquisition," *J. Acoust. Soc. Am.* **110**, 516–528.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (1997). "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness," *J. Audio Eng. Soc.* **45**, 224–240.

Nearey, T. M. (1989). "Static, dynamic, and relational properties in vowel perception," *J. Acoust. Soc. Am.* **85**, 2088–2113.

Nygaard, L. C., and Pisoni, D. B. (1998). "Talker-specific Learning in Speech Perception," *Percept. Psychophys.* **60**, 355–376.

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (1993). "Episodic encoding of voice attributes and recognition memory for spoken words," *J. Exp. Psychol. Learn. Mem. Cogn.* **19**, 309–328.

Piske, T., Mackay, I. R. A., and Flege, J. E. (2001). "Factors affecting degree of foreign accent in an L2: A review," *Phonetics* **29**, 191–215.

- Pisoni, D. B. (1997). "Some Thoughts on 'Normalization,' in Speech Perception," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullenix (Academic, San Diego), pp. 9–32.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. H. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge U.P., Cambridge).
- Torgersen, E. (1997). "Some phonological innovations in south-eastern British English," unpublished master's thesis, University of Bergen, Bergen, Norway.
- Trudgill, P. (1986). *Dialects in Contact* (Basil Blackwell, Oxford).
- Upton, C., and Widdowson, J. D. A. (1996). *An Atlas of English Dialects* (Oxford U.P., Oxford).
- Watt, D. J. L. (1998). "Variation and Change in the Vowel System of Tyne-side English," unpublished Ph.D. dissertation, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, U.K.
- Wells, J. C. (1982). *Accents of English* (Vol. 2). (Cambridge U.P., Cambridge).
- Williams, A., and Kerswill, P. (1999). "Dialect levelling: Change and continuity in Milton Keynes, Reading and Hull," in *Urban Voices: Accent Studies in the British Isles*, edited by P. Foulkes and G. J. Docherty (Arnold, London), pp. 141–162.