Vowel normalization for accent: A comparison of northern and southern British English speakers.

Bronwen G. Evans and Paul Iverson
Department of Phonetics and Linguistics, University College London, London, U.K.
E-mail: bron@phon.ucl.ac.uk, paul@phon.ucl.ac.uk

ABSTRACT

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 produced in either a northern or southern accent by a single male speaker. A computer minimization algorithm adjusted F1, F2, F3, and duration until the best exemplar of each vowel was found. The results demonstrated that some listeners normalize their vowel categorization decisions based on the accent of the carrier sentence, and that patterns of normalization are affected by individual differences in language background. The patterns of normalization corresponded with the changes in production that speakers typically make due to sociolinguistic factors, when living in multidialectal environments.

1. INTRODUCTION

In multidialectal environments (e.g., large cities such as London), native speakers of different accents regularly interact with one another, causing modifications to both production and perception ([4]). Speakers in these environments often avoid variants that are markedly regional or unusual, in order to facilitate communication and appear cosmopolitan, yet they also retain regional variants in order to show their allegiance to particular social or geographical groups (e.g. [4]). In order to understand speech, listeners must somehow tolerate or adjust to this phonetic variation, particularly when the accents that they hear strongly mismatch their own linguistic representations.

For British English, the focus of the present study, vowels are particularly important for distinguishing accents. Southern and northern accents of British English differ in the way that they use the vowels in *bath* and *luck*. Both accents use the vowels [a] and [ɑ:] but northerners produce *bath* with [a] and southerners use [ɑ:]. Southerners produce *luck* with [ʌ], but northerners do not have this vowel and produce *luck* with [ʊ], so that it becomes a homonym of *look*. When listening to a southerner, native speakers of northern English are thus required to map words that contain [ɑ:] and [ʌ] onto their lexical representations that may be based on [a] and [ʊ].

The present study investigated whether listeners adjust their vowel categorizations when listening to speech produced in different accents within the same language. The aim was to assess whether listeners change their best exemplar locations based on the accent of the carrier sentence.

2. EXPERIMENT 1

Experiment 1 investigated whether listeners from the north and south of England who were living in London, adjusted their vowel categorization decisions when listening to speech produced in SSBE (Standard Southern British English) and Sheffield English accents.

2.1 METHOD

2.1.1 Participants Twenty-three subjects, aged 20-45 years, 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, (minimum 1 year). Three subjects were dropped from the experiment because their best exemplar locations were not reliable. Of the remaining 20 subjects, 10 were from southern England and 10 were from northern England.

2.1.2 Stimuli and Apparatus The stimuli consisted of vowels in the phonetic environments /b/-V-/d/, /b/-V-/θ/, and /k/-V-/d/, embedded in recordings of the carrier sentence *I'm asking you to say the word ____ please*. The voiced portions of the vowel were synthesized ([3]) on-line to allow for fine-grained coverage of the entire vowel space. Each stimulus had a middle portion in which the formant frequencies were static, along with formant transitions appropriate for the consonants. The stimuli varied in terms of F1-F3 frequencies and duration of the middle portion. All other synthesis parameters were chosen to mimic the natural speech recordings (see [2] for more details). After synthesis, the stimuli were processed using a multi-band filter to fine-tune the match between the synthetic and natural speech. The carrier sentence was produced in both accents by the same male speaker, who was able to produce versions sounding like those of native speakers. In addition to the carrier sentences, the speaker recorded a 2 minute passage from a novel in both accents. The stimuli were played at a sampling rate of 11 kHz using a computer sound card and headphones.

2.1.3 Procedure There were two testing sessions, one for

each accent. Sessions were conducted on separate days to minimize the risk that listeners would be aware that the speaker was the same in both conditions. Each session was self-paced and lasted approximately 1 hour. 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.

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 on a computer screen. They responded by clicking with 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 designed to find the best exemplar location for that word in the 4-dimensional parameter space, and was derived from standard computer minimization algorithms (see [2]). The procedure had 5 stages, with 6 trials per stage, and was able to find the best exemplar after 30 trials. Subjects were allowed to repeat stages if they responded that the search algorithm had gone wrong at the end of the stage. The best exemplar found in Stage 5 was defined as the best exemplar of the word.

2.2. RESULTS

2.2.1 Bud and Cud As displayed in Figure 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, 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. There were no significant main effects or interactions for F2, $p>0.05$, but for F1 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. 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 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 Figure 2. In the Sheffield sentences, northerners chose a high vowel 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 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 for *bud* was relatively small for southerners, the direction of this shift was consistent. Separate repeated measures ANOVA analyses revealed that there were no significant main effects or interactions for F3 or duration, $p>0.05$, suggesting that vowel normalization for accent only took place in the F1 dimension.

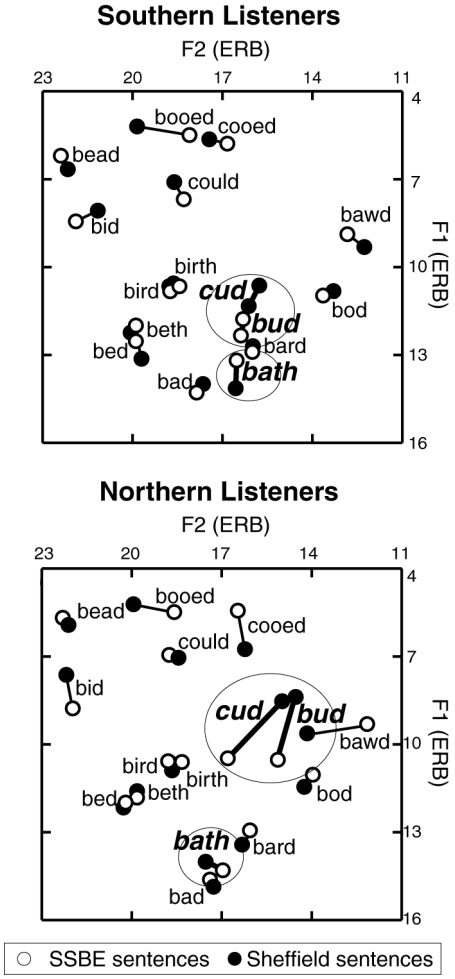

Figure 1: Average F1 and F2 formant frequencies of best exemplars for northern and southern listeners in SSBE and Sheffield carrier sentences.

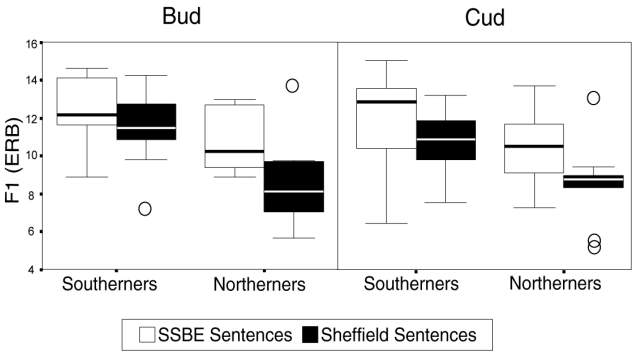

Figure 2: Boxplots of F1 formant frequency values for bud and cud in SSBE and Sheffield carrier sentences for northern and southern listeners.

2.2.2 Bath As displayed in Figure 1, listeners chose relatively similar formant frequencies for *bath* in both accents. Separate repeated measures ANOVA analyses for F1, F2, and F3 confirmed that there were no significant

main effects or interactions of sentence or subject background, $p>0.05$. There were no strong normalization effects for duration, 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 effects of duration can be seen clearly in Figure 3. Northerners preferred shorter vowels that corresponded to their production of [a] in *bath*, and southerners preferred longer vowels that corresponded to their production of [ɑ:]. Although the formant frequencies were not significantly different, the results trended in the same direction; the median values of F1 and F2 were more similar to *bad* than *bard* for northerners, and 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.

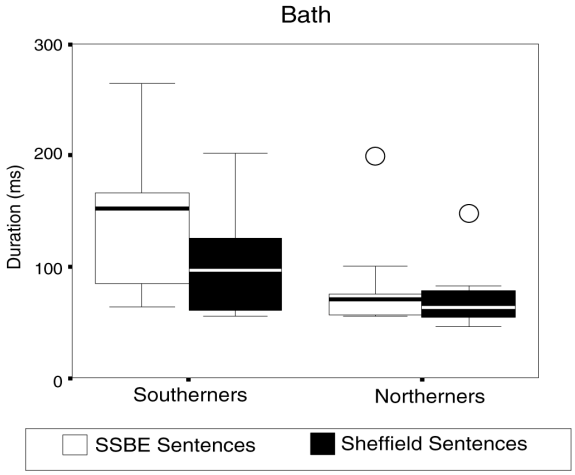

Figure 3: Boxplots of duration values for bath in SSBE and Sheffield carrier sentences for northern and southern listeners.

2.2.3 Other vowels For all other vowels, listeners chose similar formant frequencies and durations for each target word in SSBE and Sheffield English (Fig. 1), with small shifts in F1 and F2 for *bawd* and *booed*. These observations were tested in separate repeated measures ANOVA analyses[1]. There was a main effect of word for F1, $F(12,216)=203.98$, $p<0.01$, F2, $F(12,216)=113.76$, $p<0.01$, F3, $F(12,216)=7.78$, $p<0.01$, and duration, $F(12,216)=40.58$, $p<0.01$, demonstrating that different words had different formant frequency values and durations. However, there was no significant main effect of sentence context or subject background, $p>0.05$, and no significant interactions, $p>0.05$, suggesting that none of the words varied depending on accent. The small differences in *bawd* and *booed* shown in Figure 1 were thus not reliable.

[1] *Bud*, *cud* and *bath* were excluded from this analysis.

3. EXPERIMENT 2

Experiment 1 demonstrated that normalization depended on whether 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 aim was to determine whether the patterns of normalization for northerners in Experiment 1 were affected by the subjects' time living in London, or whether all northerners have the same patterns of normalization.

3.1 Method

3.1.1 Participants Twelve subjects were tested. All were native English speakers, born and raised in Ashby de la Zouch, where the dominant accent is similar to that spoken in Sheffield. Subjects were 16-17 years old and had not yet moved for employment or education. One subject was dropped from the experiment because her best exemplar locations were not reliable.

3.2.2 Stimuli and Apparatus The entire range of possible vowels was synthesized in advance with a resolution of 0.5 ERB in F1 and F2 so that the experiment could be run using a portable computer. Duration was quantized in 16 steps on a log scale from 20-403 ms. F3 was fixed to 2500 Hz for all stimuli as the results from Experiment 1 suggested that this parameter made only a modest contribution to perceived goodness.

3.3.3 Procedure There was a 4-stage search for best exemplars along the F1, F2, and duration dimensions, with 6 trials for each stage; the F3 adjustment stage was omitted. The procedure was the same as in Experiment 1 in all other respects.

3.2 Results

3.2.2 Bud and Cud As displayed in Figure 4, Ashby listeners chose similar formant frequencies for *cud* in both accents, although there was a possible difference in the F2 dimension for *bud* in the F2 dimension; listeners tended to choose a higher F2 in SSBE than in Sheffield 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. There was also little difference in terms of duration; a repeated measures ANOVA analysis revealed that there were no main effects of word, sentence context, or their interactions, $p>0.05$.

3.2.2 Bath Ashby listeners also chose similar formant frequencies for *bath* in each accent (Fig. 4). There was a trend for listeners to choose a longer vowel in the SSBE than in the Sheffield carrier sentences. However, separate repeated measures ANOVA analyses for F1, F2, and duration revealed that there was no main effect of sentence context, $p>0.05$.

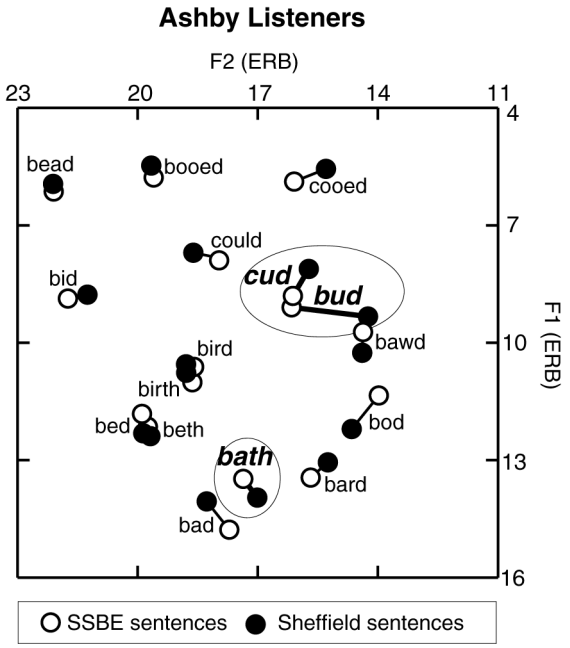

Figure 4: Average F1 and F2 formant frequencies of best exemplars for Ashby listeners in SSBE and Sheffield carrier sentences.

3.2.3 Other vowels For all other vowels, listeners chose similar formant frequencies and durations for each target word in SSBE and Sheffield English. 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 significant main effect of context and no significant interaction with word, p>0.05, suggesting that none of the other words varied depending on accent.

4. CONCLUSIONS

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 listeners' linguistic experience. Individuals living in London chose formant frequencies for *bud* and *cud* in their native accent that matched what speakers of that accent would produce, but chose centralized vowels for these words in their non-native accent. 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.

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 multidialectal environments ([4]). Northerners who live in the south of England typically modify some aspects of their accent in order to fit in with southerners; they tend to use a centralized vowel in words such as *bud* and *cud*, much like the centralized vowel that they chose as best exemplars for these words in the SSBE carrier sentence. Northerners also maintain some aspects of their regional identity; they retain their [a] when producing words like *bath*, 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; they likewise made 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. Although Ashby listeners are regularly exposed to southern English accents through the media, they had not had any experience of modifying their speech to fit into a new environment. It is plausible that these listeners chose northern vowels in the SSBE carrier sentences because they had not yet learned to change their speech when talking to southerners, even though they know how southerners talk.

This study is only a first attempt to investigate how vowel perception is modified to accommodate accent differences in the same language. The current results suggest that listeners can adjust their categorizations to accommodate to different accents within the same language, and that these adjustments are not simply determined by age and exposure; changes in best exemplar locations appear to follow sociolinguistic principles that help explain what happens when an individual chooses to fit in with a particular community or subculture.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Allen, J.S. & Miller, J.L., "Contextual influences on the internal structure of phonetic categories; A distinction between lexical status and speaking rate". *Perception and Psychophysics* 63, pp. 798-810, 2001.

[2] Evans B.G. & Iverson P., "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," *UCL Working Papers in Speech, Hearing and Language*, 2002.

[3] Klatt, D.H. & Klatt, L.C., "Analysis, synthesis, and perception of voice quality variations among female and male talkers". *Journal of Acoustic Society of America* 87, pp. 820-857, 1990.

[4] Trudgill, P., *Dialects in Contact*, Oxford: Basil Blackwell, 1986.