Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study

Jonathan Harrington,[a] Felicitas Kleber, and Ulrich Reubold
Institute of Phonetics and Speech Processing (IPS), University of Munich, Munich, Germany

(Received 17 September 2007; revised 10 February 2008; accepted 21 February 2008)

The aim of the study was to establish whether /u/-fronting, a sound change in progress in standard southern British, could be linked synchronically to the fronting effects of a preceding anterior consonant both in speech production and speech perception. For the production study, which consisted of acoustic analyses of isolated monosyllables produced by two different age groups, it was shown for younger speakers that /u/ was phonetically fronted and that the coarticulatory influence of consonants on /u/ was less than in older speakers. For the perception study, responses were elicited from the same subjects to two minimal word-pair continua that differed in the direction of the consonants' coarticulatory fronting effects on /u/. Consistent with their speech production, young listeners' /u/ category boundary was shifted toward /i/ and they compensated perceptually less for the fronting effects of the consonants on /u/ than older listeners. The findings support Ohala's model in which certain sound changes can be linked to the listener's failure to compensate for coarticulation. The results are also shown to be consistent with episodic models of speech perception in which phonological frequency effects bring about a realignment of the variants of a phonological category in speech production and perception.
[DOI: 10.1121/1.2897042]

PACS number(s): 43.71.Es, 43.71.An, 43.70.Mn [AJ] Pages: 2825–2835

I. INTRODUCTION

Many studies in recent years have shown that /u/ (lexical set GOOSE in Wells, 1982) of the standard accent of England, Received Pronunciation (RP) has become phonetically fronted. One of the earliest reports of this phenomenon was in Gimson (1966) and some 30 years later, Roach and Hartman (1997) described RP /u/-fronting as a radical shift that had taken place in the last 20–30 years. These and other similar auditory impressions (e.g., Wells, 1982, 1997) have been supported by various acoustic studies showing that the second formant frequency of /u/ both in RP and in standard southern British (SSB), a variety spoken by the majority of RP speakers, is raised for young compared with older speakers (e.g., Bauer, 1985; de Jong *et al.*, 2007; Hawkins and Midgley, 2005; Henton, 1983; McDougall and Nolan, 2007).

Diachronic /u/-fronting has also been found in longitudinal studies of Queen Elizabeth II (Harrington *et al.*, 2000a, b) who speaks a variety of the standard accent known as U-RP [see also Wells (1982, 1997) for the distinction between mainstream RP, or SSB, and U-RP and their relationship to Estuary English]. In these acoustic analyses of the Christmas broadcasts, /u/ was shown to have fronted between the 1950s and 1980s, but the Queen's 1980s /u/ was not as front as /u/ that was typical of SSB speakers of the 1980s. Harrington (2007) suggests that this sound change could be related to the prevalence with which RP /u/ follows consonants with a high F2 locus, both after alveolars in words like *noon*, *soon*, and after /j/, the latter context being somewhat more frequent in RP and SSB (and other related accents such as Australian and New Zealand English in which GOOSE is a high central vowel) because of the additional contexts with /coronal+j/ like *duty*, *news*, *tune* which are without /j/ in general American.

Thus, it is possible that the synchronic vowel fronting caused by consonant-on-vowel perseverative coarticulation in these contexts is related to diachronic /u/ -fronting in the last 50 years. Some evidence for this was provided in Harrington (2007), who showed that the distance between the F2 locus of the preceding consonant and the F2 target of /u/ had progressively diminished over a 50 year period in the Queen's Christmas broadcasts suggesting an increase in the influence of the preceding consonant on /u/ which, for various reasons that are also discussed in Harrington (2006), could not be explained away by a waning of the formality of speaking style leading to greater coarticulation and a shrinkage of the vowel space.

The above-provided explanation gives emphasis to speaker-hypoarticulation and speech production as a basis for diachronic change, but as argued by Ohala in a number of studies concerned with establishing a relationship between synchronic variability and diachronic change, there is also experimental evidence to show that the origin of sound changes is often as likely to be in the ear of the listener as in the mouth of the speaker (e.g., Ohala 1981, 1990, 1993, 2005; Ohala and Feder, 1994). For Ohala, the source of many hypoarticulation-induced sound changes—that is those that

[a]Author to whom correspondence should be addressed at Institut für Phonetik und Sprachverarbeitung, Ludwig-Maximilans Universität München, Schellingstraße 3, D 80799 München, Germany. Electronic mail: jmh@phonetik.uni-muenchen.de

can be related to contextual influences synchronically—is an unintentional error on the part of the listener that comes about because of a failure to compensate for the effects of coarticulation. Central to this idea is that listeners may incorrectly parse phonetic events from one phoneme that are temporally distributed and interwoven with those of another. For example, in a dissociation parsing error, two events in speech production that should be associated or parsed together are not (Ohala and Busà, 1995). Thus, as argued in Hombert *et al.* (1979), the development of phonological tone in certain daughter dialects of middle Chinese and some Southeast Asian languages can be attributed to listeners parsing the speaker's unintended f0 microperturbation not with the voicing status of the preceding consonant, but with tonal properties of the following vowel. The extension of Ohala's model to the analysis in the present study is that /u/-fronting may have come about because listeners fail to parse the fronting effect on /u/ either with a preceding palatal consonant or with a preceding coronal consonant that has a high F2 locus.

The evidence that listeners compensate for coarticulation at all has been demonstrated in various studies (see, e.g., Fujisaki and Kunisaki, 1976; Lindblom and Studdert-Kennedy, 1967; Mann and Repp, 1980 for some of the earliest of these) in which listeners are shown to categorize differently exactly the same acoustic token when it is embedded in two contexts that are known to exert different influences on its phonetic structure. Consider a hypothetical example which is entirely relevant to the present study from Lindblom *et al.* (1995), in which speakers are presumed to produce a back vowel as a fronted variant in an alveolar context because of consonant-on-vowel coarticulation: thus /tut/ which in the speaker's phonological plan is specified with a back vowel may be produced as [tʉt] with a central vowel due to the phonetic fronting effects of the flanking alveolars. However, assuming that listeners have an internalized knowledge of such rules of coarticulation, they can nevertheless recover the speaker's planned /tut/ by factoring out the proportion of /u/-fronting that they assume to be attributable to the alveolar context: that is they undo or compensate perceptually for coarticulation. However, if (for whatever reason) compensation for coarticulation should fail to apply, then the listener recovers not /tut/ but /tʉt/: that is, the listener assumes that a central /ʉ/ formed part of the speaker's phonological (rather than phonetic) speech production plan. A minisound change would then occur according to Ohala, if the listener turned speaker replaces /u/ with /ʉ/, especially if this is done in contexts like *move* in which a phonetically advanced /u/ has no coarticulatory *raison d'être*.

One of the main aims of the present study was to investigate whether there was any evidence for this relationship between the failure to compensate for coarticulation and /u/-fronting that has been taking place in SSB over the last 50 years. We reasoned that if young listeners are more likely to interpret a fronted production of /u/ as intended (i.e., as part of the speaker's phonological speech production plan), then they should show less evidence of compensation for the coarticulatory effects of fronting compared with older listeners from the same SSB-speaking community. A second and related aim was to assess whether, commensurate with the studies cited earlier showing /u/-fronting as a change in progress in SSB, the perceptual boundary along an /i-u/ continuum is shifted to the left for younger listeners—that is, whether young listeners are more likely to label a given token from an /i-u/ continuum as /u/ compared with older listeners from the same community.

In the present study, these perceptual effects were investigated by embedding /i-u/ continua in two sets of minimal pairs: *yeast-used* (*used*, past tense, as in "they used to study"), /jist-just/; and *sweep-swoop*, /swip-swup/.

Based on the foregoing discussion, we made the following predictions.

(1) The /i-u/ boundary should be left-shifted (greater proportion of /u/-responses) for younger listeners in both contexts. This is because, if young listeners' /u/ category boundary has fronted, then it is likely that they will interpret a greater number of stimuli with high F2 on an /i-u/ continuum as /u/.

(2) The /i-u/ boundary should be left-shifted for *yeast-used* compared with *sweep-swoop* for both young and old listeners, if a certain degree of variation in /u/ is interpreted as due to the different coarticulatory influences of the initial consonants. This follows entirely from the theory of, and experimental findings for, perceptual compensation for coarticulation (Beddor *et al.*, 2002; Beddor and Krakow, 1999; Fujisaki and Kunisaki, 1976; Kawasaki, 1986; Mann and Repp, 1980).

(3) The difference in the responses between these two sets of minimal pairs should be less for younger listeners. This is because, if young listeners compensate less for coarticulation—i.e., if they tend to ignore perceptually the influence of the consonant on /u/—then their responses to the same synthetic token whether embedded in a fronting or nonfronting context should be fairly similar, whereas they are predicted to diverge much more for older listeners, assuming that they compensate perceptually to a greater extent for the effects of coarticulation.

Another aim of this study was to relate these predictions from speech perception to speech production in two ways. First (and trivially), younger listeners should show evidence of a left-shifted category boundary in production, i.e., /u/ is expected to be phonetically fronted compared with that of an older group of speakers. Second, if perceptual compensation for coarticulation has weakened in younger listeners, then evidently the perceptual influence of the consonant on /u/ is reduced. We might then expect the coarticulatory influences of the consonant on /u/ also to be reduced in their speech production compared with that of older speakers. Thus we would expect diphones in which the consonant does (e.g., /ju/) and does not (e.g, /wu/) have a fronting effect on the vowel to be phonetically less divergent in the younger than in the older speakers.

In summary, then, there are five hypotheses to be tested, two in production and three in perception:

H1: /u/ is fronted in the production of young SSB speakers.

H2: The coarticulatory consonant-on-vowel influences in /Cu/ are less in younger SSB speakers.

H3: The /i-u/ perceptual boundary is left-shifted (greater proportion of /u/-responses) in younger listeners.

H4: Listeners (young and old) compensate perceptually for the expected coarticulatory influence of a consonant on /u/.

H5: Young listeners compensate less for the influence of a consonant on /u/ compared with older listeners.

II. SPEECH PRODUCTION

A. Method

1. Subjects

The subjects were recruited from the University of Cambridge and through the University College London and included 30 speakers of SSB. The subjects were recruited into a Young and an Old group. The Young group included 14 subjects (3 male, 11 female) aged between 18 and 20 and with an average age of 18.9 years; and the Old group 17 subjects (10 male, 7 female) over the age of 50 and with an average age of 69.2 years. Seven members of the Old group had been taken from the same subject pool as in Hawkins and Midgley (2005). Due to various scheduling difficulties, one subject from the Old group participated only in the production experiment (II) but not in the perception experiment (III); and four other subjects from the Old group participated only in the perception experiment (III) but not in the production experiment. Most of the subjects were tested and recorded in quiet rooms in the Phonetics Laboratory of the University of Cambridge, but some of the older subjects were tested and recorded in a quiet room at their homes. There were no apparent differences in the quality of the recorded speech signal and no differences in the subjects' performance related to the testing location. The equipment (for both the production and the perception experiment) was the same (Sennheiser stereo headset pc165 USB and a Toshiba Tecra notebook) for all recordings.

2. Materials, procedure, and parameters

Subjects produced a number of isolated monosyllabic words with different vowel nuclei to cover most of the RP vowel space. The words were displayed individually on a notebook computer screen in a quiet room and recordings were made with the "SpeechRecorder" software that is routinely used at the University of Munich for speech recording outside the laboratory (Draxler and Jänsch, 2004). A total of 540 words were produced in this way from a randomized list of 10 repetitions of 54 words. Any words that were mispronounced were excluded from the analysis.

The words that were analyzed in the present study included only a subset of these: /i/ (lexical set FLEECE), /u/ (lexical set GOOSE), and /ɑ/ (lexical set START) nuclei. The words with other types of nuclei, of which some will be analyzed in a future study, served mainly as filler and distractor items as far as the present study was concerned. The design of the words with /u/ and /i/ nuclei was shaped by five main criteria:

(1) The two sets of minimal word pairs used in the perception experiment *yeast*, *used*, *sweep*, *swoop* were included.
(2) There was a matched set of /Cu/ and /Cju/ words (e.g., *cooed* and *queued*)
(3) Words were included whose initial consonants had a minimal (*food*, *who'd*), a fronting (*soup*), and a backing (*cooed*) influence on /u/.
(4) There were matched minimal word-pairs between /Cid/ and /Cud/ (e.g. *keyed*, *cooed*)
(5) The final consonant was /d/ as far as possible.

A total of 4296 words were analyzed in this study including 2412 with /u/ nuclei, 1614 with /i/ nuclei, and 270 with /ɑ/ nuclei all taken from the word *hard*. The purpose of including /ɑ/, which in SSB is phonetically open, a monophthong (SSB is nonrhotic), and slightly fronter than cardinal vowel 5, was to provide a relative quantitative measure of /u/-fronting, as described in further detail in Sec. II A 2. The distribution of these 4296 words by age and gender is shown in Table I.

The words were digitized at 44.1 kHz and the first four formant frequencies were calculated using the EMU speech database analysis system (Cassidy and Harrington, 2001). The parameters for formant calculation were: LPC order of 10, a preemphasis of 0.95, and a 30 ms Blackman window with a frame shift of 5 ms. All of the data were segmented and labeled into phonetic segments.

The first two formant frequencies of all vowels in the words in Table I were checked manually both by inspecting the trajectories on the spectrogram and by identifying outliers from ellipse plots in the F1×F2 formant plane. All the formant frequencies were converted to bark using the formula in Traunmüller (1990).

The quantitative analysis of /u/-fronting was carried out in a three-dimensional space formed from the first three coefficients of the discrete cosine transformation (DCT) applied separately to the Bark-scaled F2 trajectory between the onset and offset of each vowel. The DCT breaks down any trajectory into $\frac{1}{2}$ cycle cosine waves which, if summed, reconstruct the original signal [Watson and Harrington (1999); see also Nossair and Zahorian (1991) and Milner and Shao (2006) for the relationship between DCT and cepstral coefficients]. For an N-point F2-trajectory, $x(n)$, extending in time from $n=0$ to $N-1$ points, the mth DCT coefficient, C_m $(m=0,1,2)$ was calculated with

$$C_m = \frac{2k_m}{N}\sum_{n=0}^{N-1} x(n)\cos\left(\frac{(2n+1)m\pi}{2N}\right)$$

$$k_m = \frac{1}{\sqrt{2}},\ m=0, \quad k_m = 1,\ m \neq 0 \qquad (1)$$

Thus the F2 trajectory between the acoustic onset and offset of each vowel was represented by a single point in a three-dimensional space whose axes were formed from the first three DCT coefficients or equivalently from the amplitudes of the cosine waves at frequencies $k=0$, 0.5, and 1 cycles. Since these first three DCT coefficients are proportional to

TABLE I. Number of tokens for each word analyzed in this experiment for young (Y), old (O), male (M), and female (F) speakers shown separately by vowel and with the left consonantal context shown in the left column.

		Y		O		
C(C)	Word	M	F	M	F	Σ
/u/						
j	**used**	30	109	70	60	269
fj	**feud**	29	109	70	58	266
hj	**hewed**	30	108	70	58	266
kj	**queued**	30	110	70	60	270
f	**food**	29	110	70	60	269
s	**soup**	30	110	70	60	270
k	**cooed**	30	110	69	55	264
h	**who'd**	30	110	70	60	270
sw	**swoop**	29	110	70	59	268
	Σ	267	986	629	530	2412
/i/						
j	**yeast**	30	109	70	60	269
f	**feed**	30	109	69	59	267
h	**heed**	30	110	70	60	270
k	**keyed**	30	110	70	60	270
s	**seep**	30	110	70	59	269
sw	**sweep**	30	109	70	60	269
	Σ	180	657	419	358	1614
/ɑ/						
h	**hard**	30	110	70	60	270

the trajectory's mean, linear slope, and curvature, respectively (Guzik and Harrington, 2007), this form of data reduction encodes a significant amount of dynamic information of the F2 trajectory's changing shape in time.

Using a methodology in Harrington (2006) and Guzik and Harrington (2007) we quantified the extent of /u/-fronting by calculating separately for each speaker the relative Euclidean distances of all the speaker's /u/ tokens to the same speaker's /ɑ/ (a back vowel) and /i/ (a front vowel) centroids in the three-dimensional DCT-space described earlier. This is illustrated in Fig. 1 in two dimensions, in which E_1 and E_2 are the Euclidean distances of a given /u/ token to the same speaker's /ɑ/ and /i/ centroids, respectively. The parameter d_u in the following:

$$d_u = \log(E_1/E_2) = \log(E_1) - \log(E_2), \tag{2}$$

is a quantification of the relative proximity of the /u/-token to these two centroids: When d_u is zero, then the token is equidistant between /i/ and /ɑ/; when it is positive, then it is closer to /i/ than to /ɑ/; and when it is negative, it is closer to /ɑ/ than to /i/. We calculated d_u over all /u/-tokens in this way

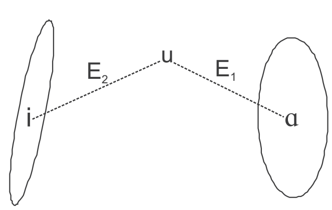

FIG. 1. E_1 and E_2 are the Euclidean distances in an arbitrary two-dimensional space from a single /u/ token to the same speaker's /ɑ/ and /i/ centroids, respectively.

separately for each speaker and using the speaker-specific /i/ and /u/ centroids as illustrated in Fig. 1 (but in a three-dimensional space). The hypothesis to be tested, then, is that if /u/ is phonetically fronted for younger speakers, then d_u should be smaller, reflecting its relatively closer Euclidean proximity to /i/ than to /ɑ/, compared with d_u for older speakers.

Finally, for the purposes of comparing the influence of C-on-/u/ coarticulation with perceptual responses, we measured the Euclidean distances again separately for each speaker but this time between *used* and *swoop* in the same three-parameter DCT space described earlier. The more that the consonants exert a coarticulatory influence on /u/, then the further apart *used* and *swoop* should be in the parametrized F2 space. Where m_{sw} is the (speaker specific) mean (centroid) of *swoop* in the DCT space, we calculated the Euclidean distances from all tokens of *used* to the same speaker's m_{sw}; and where m_{ju} is the mean of *used* in the same DCT space, we calculated the Euclidean distances from all tokens of *swoop* to the same speaker's m_{ju}. If, for the reasons described in Sec. I, the coarticulatory influences of the initial consonant on /u/ are greater for the older speakers, then they should show greater values on this Euclidean metric (commensurate with a greater F2-trajectory divergence between *used* and *swoop*) compared with younger speakers.

B. Results

We begin by presenting some graphical analyses which were then quantified using the metrics described in Sec. II A 2.

Figure 2 shows the average position in the Bark-scaled F1×F2 plane of formant values extracted at the temporal

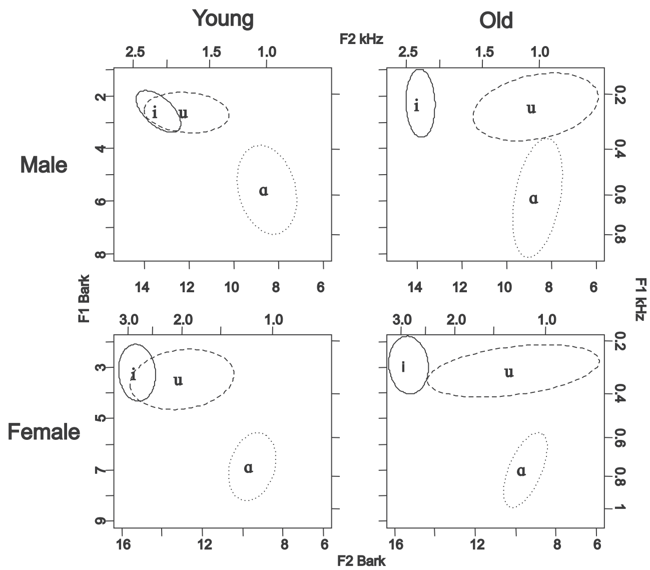

FIG. 2. 95% confidence ellipses for /iuɑ/ in a Bark-scaled formant plane shown separately for the age and gender groups in formant data extracted at the vowels' temporal midpoints.

midpoint of /i, u, ɑ/ separately for the gender and age groups. It is evident that F2 of /u/ is higher for young speakers, which is consistent with the various findings discussed in Sec. I that it is also phonetically fronter. The ellipses for /u/ are also somewhat larger for old speakers which would suggest that consonantal context had a more marked coarticulatory effect on /u/ than for young speakers.

Figure 3 shows linearly time-normalized trajectories for the second formant frequency over the extent of /ju/ in *used* and /wu/ in *swoop*, averaged separately for the four gender × age combinations. The formant trajectory shapes across gender (left and right panels) are fairly similar, but there are markedly different F2 patterns across the two age groups. In particular, three features characterize the old as opposed to the young group of speakers:

(1) F2 of *used* falls much more steeply and to a lower value (of roughly 1600 Hz in males and 1800 Hz in females) than in the young group for which F2 is more or less level (at around 2000 Hz for the male speakers and just over 2000 Hz for the female speakers).

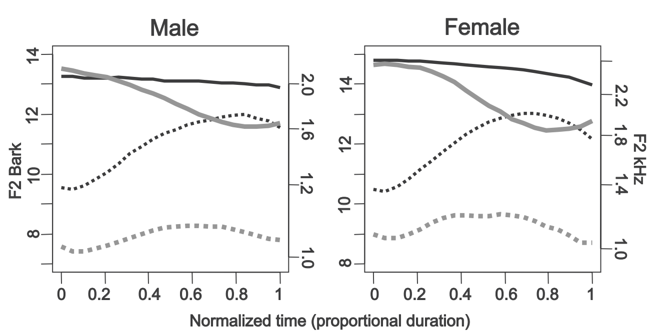

FIG. 3. Bark-scaled F2 trajectories between the onset and offset of /ju/ in *used* (solid) and /wu/ in *swoop* (dotted) that were linearly time-normalized and averaged across the young (black) and old (gray) groups shown separately for male (*left*) and female (*right*) speakers.

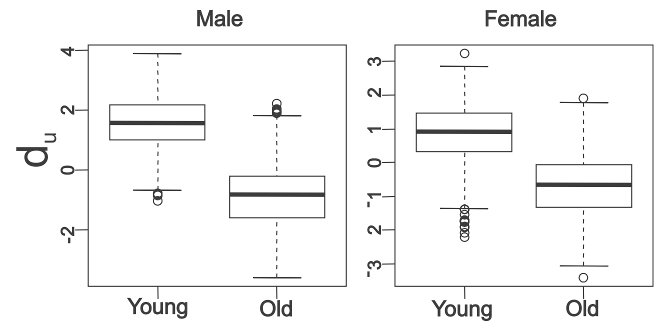

FIG. 4. Boxplots showing the median (thick horizontal bar), interquartile range and outliers (circles) on the parameter d_u, the log Euclidean distance ratio, across the young and old groups separately for male (*left*) and female (*right*) speakers.

(2) F2 of *swoop* rises but to a much lower frequency than for the young group.

(3) There is a considerably greater difference especially at F2 onset but also throughout the entire trajectories between *used* and *swoop* than for the young group. This suggests that the extent of C-on-V coarticulation is somewhat greater in the old group and for two reasons. First, because there is a more substantial deviation for the old group between the F2 onsets and by extension between the F2 loci of /w/ and /j/; and second because there is evidently a greater convergence in F2 toward a common vowel target across these two word contexts in the young group, as shown by the smaller separation between the *used* and *swoop* trajectories in both young male and female speakers.

We quantified the differences in the extent of /u/-fronting between the two age groups by calculating d_u, the log. Euclidean distance ratios in a three-dimensional DCT space, using the formula in Eq. (2) as described in Sec. II A 2. The results of these calculations are shown as boxplots for /u/ and /ju/ words in Fig. 4.

For both males and females, it is evident that the log Euclidean distance ratio, d_u, is greater for the younger speakers. Moreover, for the younger speakers, d_u is positive, which suggests that /u/ is closer to /i/ than to /ɑ/ on F2; by contrast, d_u is negative for the older speakers and so for them /u/ is closer on F2 to /ɑ/ than to /i/.

The results of a repeated measures analysis of variance (ANOVA) with dependent variable log Euclidean distance ratio, two between-subjects factors, Age (Young vs Old) and Gender, and a within-subjects factor for Word (9 levels corresponding to the 9 word types shown in Table I with a /u/ or /ju/ nucleus) showed a significant effect for Age $[F(1,13) = 117.1,\ p < 0.001]$. There was also an (unsurprising) significant effect for Word $[F(1,8) = 167.7,\ p < 0.001]$, which simply means that F2 in the different words in Table I with /u/ or /ju/ nuclei were not all equally close to /i/ as to /ɑ/. There was no significant effect for Gender which may have come about because of the small number of male speakers in the young group.

The only interaction that was significant was Word × Age $[F(1,8) = 10.1,\ p < 0.001]$. This means that the extent of

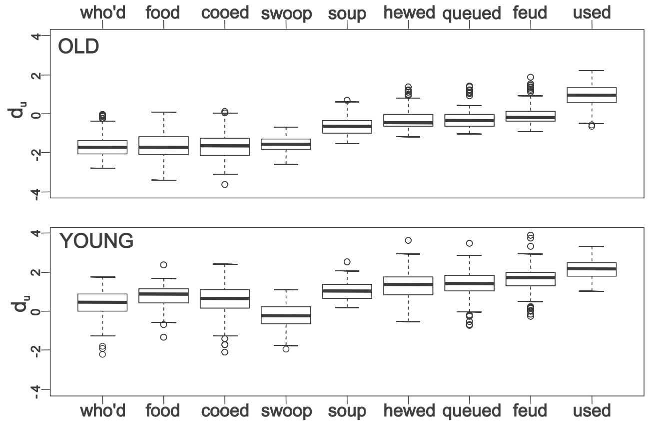

FIG. 5. Boxplots showing the median (thick horizontal bar), interquartile range, and outliers (circles) on the parameter d_u, the log Euclidean distance ratio, for the old (*above*) and young (*below*) groups separately for each word. The boxplots are rank ordered from lowest to highest median d_u in the old group in both cases.

/u/-fronting for the young relative to the old group was not the same in all words. Three findings emerged from a subsequent analysis of the word-specific distributions which are shown separately for the two age groups in Fig. 5. First, as also shown by *post-hoc* Tukey tests, there is a marked and significant ($p<0.001$) difference between the age groups for all words. As Fig. 5 shows, the median log Euclidean distance ratio is greater than zero for all words except *swoop* in the young speakers, which means they are all acoustically closer on (DCT and Bark-transformed) F2 to front /i/ than to back /ɑ/. There is exactly the opposite pattern in the older speakers for which the vowels of all words except *used* are closer to /ɑ/ (since median d_u is negative in these cases). Second, the relative positions and therefore rank order from lowest to highest d_u is very similar for both age groups: The main exception is *swoop* which has a roughly similar median value as the other back variants of /u/ in *cooed*, *who'd*, *food* for the older speakers, but which has a lower value in comparison with the three words in the younger speakers. The third finding falls out from the first two: The differences between the age groups were greatest in words with retracted /u/: that is, the difference between the age groups on the log Euclidean distance ratios was, as Fig. 5 shows, evidently greater for *cooed*, *who'd*, *food* than for *queued*, *hewed*, and *feud*.

We quantified these word-specific differences more precisely by measuring in the same Bark-transformed DCT space the Euclidean distances between the age groups separately for each word. More specifically, we calculated for *swoop* the Euclidean distances from all the young speakers' *swoop* tokens to the old group's *swoop* centroid; and the Euclidean distances from all the old speakers' *swoop* tokens to the young group's *swoop* centroid. The more these word-specific distributions of younger and older speakers overlap with each other, then the closer these distances should be to zero. We then carried out the same word-specific Euclidean measure for all word types separately. As Fig. 6 shows, the pattern of the resulting distribution on this measure is very similar to that obtained from the metric used for quantifying

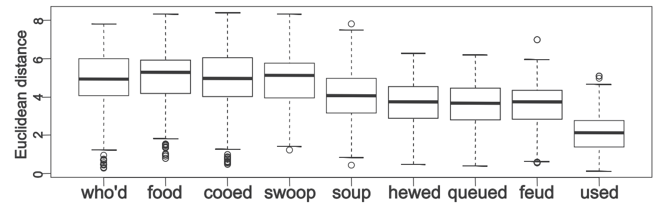

FIG. 6. Boxplots showing the median (thick horizontal bar), interquartile range and outliers (circles) on the inter-Euclidean distances (arbitrary units) between the young and old groups calculated in a Bark-scaled and DCT-transformed F2 space separately for each word. The words are arranged in the same order as in Fig. 5.

the relative position of the old speakers' vowels on a phonetic front–back dimension (Fig. 5). Thus, the greatest difference between the two age groups is for those vowels whose allophones are most retracted in the older speakers. This provides further evidence that the sound change involves a realignment in production of the phonetically back allophones of /u/ toward a phonetically front position (as a result of which, phonetically front variants of /u/ show the least displacement).

As far as the measures of coarticulation described in Sec. II A 2 are concerned, Fig. 7 shows the logarithm of the Euclidean distances from *used* tokens to the *swoop* centroid and from *swoop* tokens to the *used* centroid by age and gender: These distances are clearly greater for the old compared with the young group. A repeated measures ANOVA with dependent variable log Euclidean distances and two between-subjects factors, Age (Young vs Old) and Gender, and a within-subjects factor for Word (*used* and *swoop*) showed a significant effect for Age [$F(1,20)=36.5$, $p<0.001$] but no other significant effects.

III. SPEECH PERCEPTION

A. Method

For the perception experiment, HLsyn (High Level Parameter Speech Synthesis System, version 2.2) was used for creating the synthetic stimuli. With the exception of five subjects from the Old group, of which one participated only in the production experiment and of which four participated

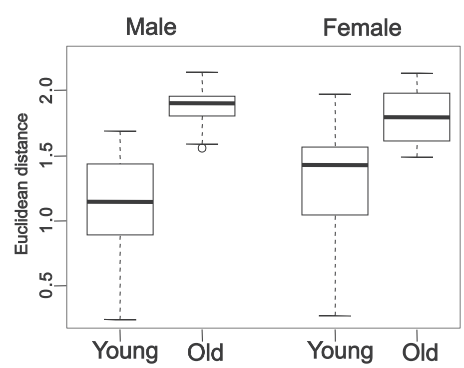

FIG. 7. Boxplots showing the median (thick horizontal bar), interquartile range and outliers (circles) on the logarithm of the inter-Euclidean distances (arbitrary units) calculated in a Bark-scaled and DCT-transformed F2 space between *used* and *sweep* shown separately for age and gender.

only in the perception experiment, listening responses were obtained from the same subjects who had participated in the production experiment (see Sec. II A 1 for further details).

We carried out various pretests with trained L1-English phoneticians to obtain naturalness judgments to a number of continua in the two word-contexts described more fully in the following, both in order to guarantee that our synthesized end points were intelligible as the intended words and to determine the perceptually most natural continuum as far as F2 changes were concerned.

We created two 13-step synthetic continua, one each between two sets of minimal-word pairs: *yeast-used* (/jist/-/just/ and henceforth /jVst/) and *sweep-swoop* (/swip/-/swup/ and henceforth /swVp/). These word pairs were chosen because the anterior /j_s/ context in the former is expected to have a phonetic fronting effect in speech production for which listeners might be expected to compensate perceptually relative to /sw_p/, as described in Sec. I.

The continua in both cases were created by varying F2 in the following ways. For the /jVst/ continuum, F2 was varied from 1278 Hz (stimulus 1 nearest /u/) to 2428 Hz (stimulus 13 nearest /i/) in equal 0.45 Bark-size steps. The following parameters were fixed for all stimuli in /jVst/: the F2-locus of the preceding /j/ at 2450 Hz; a transition phase of 90 ms in which F2 decreased linearly; a following steady-state vowel of 120 ms in duration; a level F1 and F3 from the periodic onset to the offset at 280 and 2700 Hz, respectively.

For /swVp/, F2 varied between 1014 Hz (stimulus 1) to 2320 Hz (stimulus 13) in 0.35 Bark steps. The /w/ locus was fixed at 600 Hz. There was a transition phase of 45 ms which was followed by a steady vowel of 140 ms with F1 and F3 fixed from the periodic onset to the offset at 280 and 2544 Hz, respectively. The reason why the two continua in /jVst/ and /swVp/ did not have the same formant end points was for reasons of perceived naturalness.

The 26 stimuli were randomized and individual synthetic word tokens were presented from both continua in one session 5 times (5 repetitions × 13 steps × 2 continua=130 randomized stimuli). For the perception experiment, subjects listened to each stimulus separately and carried out a forced-choice identification task in which they responded to each stimulus with one of *used*, *yeast*, *swoop*, or *sweep*.

Since, as described earlier, the two continua were synthesized with different end points, we had to map them onto the same F2-scale for the purposes of comparing the responses between them. The matching was done by aligning in frequency the responses to /jVst/ with those of /swVp/ at the closest F2 correspondences between the stimuli. For example, we compared responses to stimulus 4 from /swVp/ (F2=1260 Hz) with responses to stimulus 1 from /jVst/ with F2=1278 Hz because these were the two stimuli with the smallest F2 difference ($\Delta_{F2}=1278-1260=18$ Hz) between them. As Table II shows, we had to exclude from the analysis responses to stimuli 1–3 from /swVp/ and to stimulus 13 from /jVst/ for which no closest F2 matches from the other continuum were available. We also had to collapse the responses to stimuli 4 and 5, and to 9 and 10 in /jVst/ for the same reason (see Table II for further details). The responses

TABLE II. F2 values on the 13-point synthetic continuum for *sweep-swoop* (column 2), *yeast-used* (column 3), and the derived F2-aligned, 10-point continuum analyzed in this study (far right column).

Stimulus No.	/swVp/ F2 values (Hz)	/jVst/ F2 values (Hz)	Mean F2 value (Hz)
1	01. 1014		
2	02. 1092		
3	03. 1173		
4	04. 1260	01. 1278	1. 1269
5	05. 1351	02. 1350	2. 1351
6	06. 1447	03. 1426	3. 1437
7	07. 1549	04. 1505+05. 1588=1546.5	4. 1548
8	08. 1658	06. 1675	5. 1667
9	09. 1773	07. 1766	6. 1770
10	10. 1897	08. 1872	7. 1885
11	11. 2028	09. 1964+10. 2070=2017	8. 2023
12	12. 2169	11. 2183	9. 2176
13	13. 2320	12. 2302	10. 2311
		13. 2428	

to the continua in Sec. III B are therefore presented for 10 stimuli based on the matched F2 values shown in the right column of Table II.

Two types of analyses were carried out on the perceptual responses. First, the results were analyzed with a repeated measures ANOVA with the stimulus number from the synthetic continua as a repeated-measure dependent variable, and with Context (a within-subject factor with two levels, /jVst/ vs /swVp/) and Age (between-subject factor with two levels, Old and Young) as the independent variables. Second, we carried out another repeated measures ANOVA with the same independent variables but with the category boundary obtained separately for each listener (i.e., the F2 frequency for which responses with /i/ or /u/ were 50%) as the dependent variable (Fig. 8). The 50% category boundary was calculated using probit analysis following exactly the same procedure as in, e.g., Lotto *et al.* (1996).

B. Results

The results of the comparison of the entire continua showed significant effects for Age $[F(1,28)=6.163;\ p<0.05]$, and for Context $[F(1,28)=34.3,\ p<0.001]$, as well

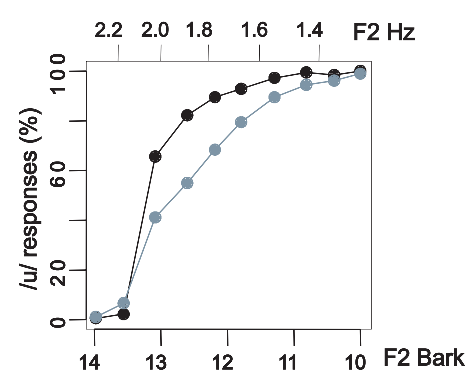

FIG. 8. (Color online) Percentage of /u/ responses as a function of decreasing F2 by young (*black*) and old (*gray*) listeners to synthetic continua pooled across *yeast-used* and *sweep-swoop*.

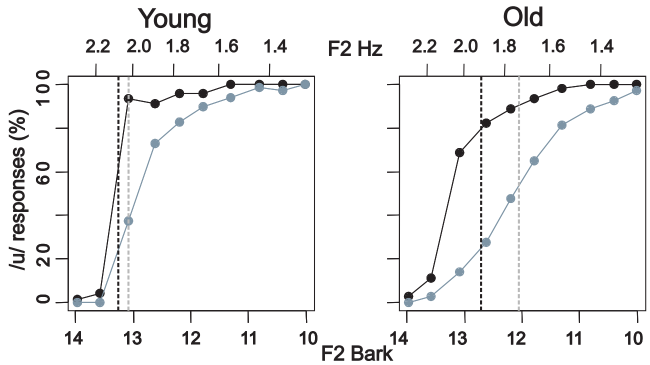

FIG. 9. (Color online) Percentage of /u/ responses as a function of decreasing F2 by young (*left panel*) and old (*right panel*) listeners to *used-yeast* (*black*) and *sweep-swoop* (*gray*). The vertical dotted lines show the mean category boundaries in *yeast-used* (*black*) and *sweep-swoop* (*gray*).

as a significant Context × Age interaction [$F(1,28)=4.3$, $p<0.05$]. The results of the comparison of the category boundaries, whose means for the four Age × Content conditions are superimposed on Fig. 9, were broadly consistent with those obtained from comparing the entire continua. There was a significant effect for Age [$F(1,28)=5.63$; $p<0.05$], for Context [$F(1,28)=39.7$, $p<0.001$], although a not quite significant Context × Age interaction [$F(1,28)=3.67$, $p=0.066$].

The effect across both sets of results for Age shows that, consistent with H3 (Sec. I), the /i-u/ boundary was left-shifted for the Young group: that is younger listeners were more likely to hear the same synthetic token as /u/ compared with older listeners, an effect which is clearly shown in Fig. 8. The effect across both sets of results for Context shows that the /i-u/ boundary was left-shifted in /jVst/: that is, listeners were more likely to hear the same synthetic token as /u/ in *yeast-used* compared with *sweep-swoop* (Fig. 9). This result is a replication of various experiments reviewed in Sec. I showing that listeners compensate perceptually for the effects of coarticulation.

The result of the Context × Age interaction which was significant in comparing the entire continua and not quite significant in comparing category boundaries, together with the data in Fig. 9 showing the responses separately to the two continua by age, is, in general, compatible with Hypothesis 5: Evidently, the young listeners did not compensate as much as the old listeners for coarticulation since the differences in their responses to *yeast-used* versus *sweep-swoop* were less than for older listeners. Thus, *post-hoc* Tukey tests showed that there were no significant differences between *yeast-used* versus *sweep-swoop* for the young listeners, neither in comparing the entire continua nor in comparing category boundaries; but that there were significant differences in older listeners' responses to *yeast-used* versus *sweep-swoop* both in comparing the entire continua ($p<0.001$) and in comparing category boundaries ($p<0.001$).

However, Fig. 9 also shows that the relatively diminished perceptual compensation for coarticulation in younger listeners came about because of differences in responses by the two age groups to /swVp/. More specifically, the results of *post-hoc* Tukey tests showed that there were no significant differences between the two age groups in their responses to /jVst/, either based on comparing the entire continua ($p=0.88$), or based on comparisons between the category boundaries ($p=0.86$). On the other hand, there were significant differences between the age groups in their responses to /swVp/, both in comparing continua ($p<0.01$) and in comparing category boundaries ($p<0.05$). This means that the main reason why there were overall different perceptual responses between the age groups (Fig. 8) was because the /i-u/ boundary on *sweep-swoop* was left-shifted in younger relative to that of the older listeners. We discuss the further implication of this result together with those from the acoustic analysis in Sec. IV.

IV. DISCUSSION

The hypothesis H4, that listeners compensate for the effects of coarticulation, is supported by the results from this experiment. Listeners attribute a certain proportion of vowel fronting on an /i-u/ continuum to the effects of consonantal context and factor this out: Accordingly, there is a greater probability that the same stimulus token will be perceived as /u/ in a fronting context such as /j_s/ than in a backing consonantal context such as /sw_p/. These findings are consistent with various studies showing perceptual compensation for coarticulation (Beddor *et al.*, 2002; Beddor and Krakow, 1999; Kawasaki, 1986; Fujisaki and Kunisaki, 1976; Mann and Repp, 1980), but also more generally with speech perception models (Fowler, 1984; Fowler and Smith, 1986; Ohala, 1993) in which listeners are sensitive to the nature and extent of coarticulatory overlap between segments.

The results of the production study showed unequivocally that younger speakers have a fronter realization of /u/ than older speakers (consistently with H1): this result lends support to other acoustic and auditory studies (Gimson, 1966; Hawkins and Midgley, 2005; Roach and Hartman, 1997) as well as to the longitudinal analysis in Harrington (2007) that /u/ in SSB has fronted in the last 40 years. The present investigation extends these analyses by showing that young listeners not only produce a fronter realization of /u/, but they also have a fronter category boundary in perceiving this vowel. Thus, H3, that young and old listeners respond differently to the same continuum that spans a sound-change in progress, was also supported by these data. In light of these data as well as other recent experimental studies on sound change in progress (e.g., Warren *et al.*, 2007), we have to abandon the notion that there is a uniform relationship between the phonological system and phonetic output for all members of the same speech community. Instead, phonological category boundaries are likely to be specific to different groups of speakers of the same speaking community and above all strongly related to the differences in their own speech production. The model that seems to be best able to account for these differences is an episodic or exemplar-based model of speech perception (Johnson, 1997; Pierrehumbert, 2002, 2003a, b, 2006). In this model, a phonological category is defined by a distribution in a perceptual space that depends on its remembered exemplars. Moreover, speech production involves the selection of one of the exemplars—typically one that is most probable and nearest the center of the distribution (Pierrehumbert, 2001). Two features from this model are consistent with the results from the present study. First, that the phonological category boundaries can be (and are likely to be) different from one individual to the next, given that the remembered exemplars from which a phonological category is constructed will certainly be different across different individuals, even of the same speaking community. Second that, since in speech production an exemplar is selected from the phonological category that is built out of perceived and remembered exemplars, there is a close link between the category boundaries in speech perception and production: Thus our interpretation of these data within this model is that younger members of the community who place the category boundary closer to /i/ on /i-u/ in perception are also more likely to produce a more fronted /u/ than older members of the same community (because if the category boundary shifts toward /i/, then so does the center of gravity from which tokens are most likely to be selected in speech production).

The most important hypothesis in this paper, H5, was that if, following Ohala (1981, 1993), sound change comes about because listeners fail to undo perceptually the effects of consonant on vowel coarticulation, then the differences in the responses to the two continua should be a good deal less for younger listeners. Before we assess in detail the extent to which this hypothesis was substantiated, we will briefly recap the three main findings from the perception experiments. First, the results showed that, compatible with H5, there was a significantly greater difference in responses to the two continua by the older listeners. Evidently this shows that the older listeners' responses were influenced by consonantal context to a far greater degree than those of the younger listeners. Here there is, once again, compatibility with their speech production because the acoustic influence of consonantal context on the vowel was also greater (H2), as shown by the older speakers' greater divergence between *used* and *swoop* in a (bark and DCT-parametrized) F2 space. Second, the two age groups differed minimally and not significantly in their perceptual responses to *yeast-used*. And third, they differed extensively and significantly in their perceptual responses to *sweep-swoop*.

We now consider in further detail whether these three findings can be reconciled with Ohala's (1981, 1993) model in which a minisound change occurs as a result of a failure to compensate for coarticulation which is then carried over into speech production. Our starting point in this interpretation is the finding that young and old listeners have more or less the same responses to the *yeast-used* continuum. Here we would argue that although both age groups cut up the *yeast-used* continuum in a similar way, the difference is in the classification: Older listeners classify the distinction as /i-u/ as opposed to the younger listeners who classify it as /i-ʉ/. Thus, suppose there is a token midway between the synthesis end points which we denote as [ʉ]. Older listeners compensate for coarticulation and attribute a certain amount of fronting to the consonantal context and classify it perceptually as a back vowel /u/. Younger listeners compensate far less, or minimally, for the effects of consonantal context and classify it perceptually as a central vowel /ʉ/. So, in both cases, the *position* of the category boundary on the *yeast-used* continuum may be roughly the same (left-*shifted* in old speakers because they compensate for coarticulation, but *positioned* to the left and near /i/ in young listeners because /ʉ/, being a central vowel, is characterized in speech production by a mid-high F2) but the *classification* is different. Now consider the results for *sweep-swoop*. Here there is of course no reason to compensate for the fronting effects of coarticulation and so the category boundary for older listeners shifts to the right commensurate with their classification of the low end of the F2 continuum as a back /u/, resulting in a relatively greater proportion of /i/ responses and a marked and significant difference compared with their responses to *yeast-used*. Since younger listeners compensate minimally for coarticulation, their category boundary is not significantly different from their *yeast-used* but it is significantly to the left of the older listeners' *sweep-swoop* boundary (who divide the continuum into /i-u/). Thus all of the findings from the perception experiment and their relationship to those from the production experiment are consistent with the following two ideas. First, older listeners compensate perceptually for the coarticulatory fronting effects of a consonantal context in *yeast-used* but younger listeners do not (or compensate much less). Second, there is a difference of *phonological* category between the two groups: Older listeners make a contrast between /i/ and /u/ whereas for younger listeners of the same speaking community, the contrast is between /i/ and a central vowel /ʉ/. It is because the two age groups cut up the *sweep-swoop* continuum differently and because the young listeners show neither significant evidence of coarticulatory compensation in perception nor of a back vowel allophone in contexts like *swoop* in production that we wish to argue for a difference in phonological category: this is, of course, very different from suggesting that the two age groups have the same /i-u/ contrast and differ because the young speaker-listeners realize /u/ phonetically as [ʉ].

The stages that give rise to the sound change within Ohala's model might then be as follows. The divergence between front and back variants of /u/ 40–50 years ago was very large (and is still considerable in older speakers) necessitating compensation for coarticulation in the front consonantal context to bring about a perceptual realignment of front realizations of /u/ with retracted variants, thereby maintaining the phonetic integrity of the phonological category /u/. With the assumed waning of perceptual compensation for coarticulation, the integrity of the phonological category was preserved by shifting diachronically *in production* the back toward the front variants. This suggests that the extent of this shift was greatest in the phonetically most retracted variants and some evidence to support this has been presented in Figs. 5 and 6 in which, for example, the difference between young and older speakers' production of *food* or *who'd* was greater than for *feud* or *hewed*.

In the above-presented analysis, the failure to compensate perceptually for coarticulation is the trigger for the sound change in which /u/ changes to /ʉ/. However, the data from the present study might be just as compatible with an episodic model of perception and production in which the frequency and probability with which /u/ occurs in a certain context is the trigger for sound change. An analysis of the CELEX database (Baayen *et al.*, 1997) in Harrington (2007) reported that just over 70% of SSB /u/ occur in words following a consonant with a high F2 locus, either after /j/ (*argue*, *cute*, *duty*, *few*) or after alveolars (*lose*, *noon*, *soon*). In an episodic model of speech perception, frequency of occurrence emerges as a direct consequence of remembered instances of words and it exerts an influence on speech perception by way of activation strength (Pierrehumbert, 2001). Essentially, more frequently occurring remembered exemplars have higher resting levels of activation and are therefore more likely to be selected than low frequency tokens in matching a remembered exemplar to an incoming speech signal in speech perception: This type of association between statistical frequency and activation strength can explain why in speech perception there is a bias toward perceiving more frequently occurring phonotactic consonant clusters, as shown by Hay *et al.* (2003). In speech production, an exemplar is, according to Pierrehumbert (2001), selected at random from a cloud of remembered tokens, but in such a way that the selection is weighted by the same frequency effects and activation strengths that bias speech perception. Since this model of speech production is probabilistic, then an exemplar may sometimes be selected from the exemplar cloud for /u/ that is not always quite appropriate for the consonantal context. Moreover, since fronted allophones of /u/ are statistically more frequent and therefore have a higher activation strength, the probability that the speaker inappropriately selects a front allophone of /u/ for a context like *swoop* is greater than the probability of inappropriately selecting a back allophone of /u/ for a fronting context like *yeast*. This greater probability of misselecting front allophones may be the mechanism by which the center of gravity of the entire /u/ cloud is incrementally shifted toward the front part of the vowel space [see Pierrehumbert (2001) for further details on the way in which this kind of bias and incremental shift can be introduced over time into speech production]. Thus the evidence that high frequency labels are advantaged perceptually because of their stronger perceptual representations (Pierrehumbert, 2003a) and that low frequency items are less robust than high frequency items (Hay *et al.*, 2003; Silverman, 2004) may then have been the trigger for back variants of /u/ to front. As the back variants were fronted, the center of gravity of the entire /u/ distribution would necessarily be fronted even further; and this may be the reason why there is some further minor phonetic advancement of vowels even in fronted /j/ contexts, as the difference between the age groups in the relative positions in production of words like *used* and *feud* shows (Fig. 5). With the progression of the sound change, listeners would need to compensate less for coarticulation in the front context because the divergence between the front and back variants of /u/ would be lessened as the sound change progressed (thus progressively reducing the need for perceptual realignment). So in this model, the waning of perceptual compensation for coarticulation would be an *effect*, not the *cause* or trigger for /u/-fronting.

In summary, we have found evidence that /u/-fronting is a sound change in progress in SSB both in production and perception. Since for the younger speaker-hearers, the phonological category is fronted relative to that of the older speaker-hearers, they produce phonetically fronted [ʉ] vowels and their category boundary is nearer /i/ than for older listeners. However, because older listeners compensate perceptually for the coarticulatory fronting influences of the consonant on a vowel, this age-group difference manifests itself perceptually only in contexts in which the consonant causes the vowel to be phonetically fronted. These coarticulatory-dependent perceptual differences between the age groups are commensurate with their speech production: The coarticulatory influences of the consonant in a fronting context are much more pronounced for the older speakers leading to consonant-dependent divergent vowel variants that are not in evidence for the young. Finally, the finding that there are perceptual differences between the age groups is consistent with Ohala's (1981, 1993) model in which a coarticulatory-dependent allophone may be phonologized (leading to sound change) if listeners give up on compensating perceptually for coarticulation in the same context. It may, however, also be compatible with an episodic model of speech perception and production in which phoneme frequency effects trigger a realignment in production between phonetically markedly divergent variants. Further research, possibly involving an analysis of a different hypo-articulation-induced sound change whose variants are less distinguished by statistical frequency effects, would be needed to resolve this issue further.

ACKNOWLEDGMENTS

The authors thank Ann-Kathrin Killguss for help with labeling, Sarah Hawkins of the Department of Linguistics, University of Cambridge, as well as Valerie Hazan and Moira Yip of the University College London for their help in recruiting subjects, and Allard Jongman and two reviewers for their comments. We are also very grateful to all the subjects at Cambridge and London for participating in this study. This research was supported by German Research Council Grant No. HA3512/3-3, "Sound change, lexical frequency, and variability: an experimental study of Southern British English, Received Pronunciation."

Baayen, R., Dijkstra, T., and Schreuder, R. (**1997**). The CELEX Lexical Database, Linguistic Data Consortium, University of Pennsylvania, Philadelphia (CD-ROM).

Bauer, L. (**1985**). "Tracing phonetic change in the received pronunciation of British English," J. Phonetics **13**, 61–81.

Beddor, P., Harnsberger, J., and Lindemann, S. (**2002**). "Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates," J. Phonetics **30**, 591–627.

Beddor, P., and Krakow, R. (**1999**). "Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation," J. Acoust. Soc. Am. **106**, 2868–2887.

Cassidy, S., and Harrington, J. (**2001**). "Multi-level annotation in the Emu speech database management system," Speech Commun. **33**, 61–77.

de Jong, G., McDougall, K., Hudson, T., and Nolan, F. (**2007**). "The speaker discriminating power of sounds undergoing historical change: A formant-based study," *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, pp. 1813–1816.

Draxler, C., and Jänsch, K. (**2004**). "SpeechRecorder—A universal platform independent multichannel audio recording software," *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, Lisbon, Portugal, pp. 559–562.

Fowler, C. A. (**1984**). "Segmentation of coarticulated speech in perception," Percept. Psychophys. **36**, 359–368.
Fowler, C. A., and Smith, M. R. (**1986**). "Speech perception as 'vector analysis:' An approach to the problems of segmentation and invariance," in *Invariance and Variability of Speech Processes*, edited by J. Perkell and D. Klatt (Erlbaum, Hillsdale, NJ), pp. 123–139.
Fujisaki, H., and Kunisaki, O. (**1976**). "Analysis, recognition and perception of voiceless fricative consonants in Japanese," Annual Bulletin RILP No. 10, 145–156.
Gimson, A. C. (**1966**). *An Introduction to the Pronunciation of English*, 1st ed. (Arnold, London).
Guzik, K., and Harrington, J. (**2007**). "The quantification of place of articulation assimilation in electropalatographic data using the similarity index (SI)," Advances in Speech–Language Pathology **9**, 109–119.
Harrington, J. (**2006**). "An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts," J. Phonetics **34**, 439–457.
Harrington, J. (**2007**). "Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts," in *Laboratory Phonology 9*, edited by J. Cole and J. Hualde (Mouton de Gruyter, Berlin), pp. 125–143.
Harrington, J., Palethorpe, S., and Watson, C. (**2000a**). "Does the Queen speak the Queen's English?" Nature (London) **408**, 927–928.
Harrington, J., Palethorpe, S., and Watson, C. (**2000b**). "Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts," J. Int. Phonetic Assoc. **30**, 63–78.
Hawkins, S., and Midgley, J. (**2005**). "Formant frequencies of RP monophthongs in four age groups of speakers," J. Int. Phonetic Assoc. **35**, 183–199.
Hay, J., Pierrehumbert, J., and Beckman, M. (**2003**). "Speech perception, wellformedness and the statistics of the lexicon," in *Papers in Laboratory Phonology VI: Phonetic Interpretation*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 58–74.
Henton, C. (**1983**). "Changes in the vowels of received pronunciation," J. Phonetics **11**, 353–371.
Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (**1979**). "Phonetic explanations for the development of tones," Language **55**, 37–58.
Johnson, K. (**1997**). "Speech perception without speaker normalization," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego), pp. 145–166.
Kawasaki, H. (**1986**). "Phonetic explanation for phonological universals: The case of distinctive vowel nasalization," in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic, New York), pp. 81–103.
Lindblom, B., Guion, S., Hura, S., Moon, S.-J., and Willerman, R. (**1995**). "Is sound change adaptive?" Rivista Di Linguistica **7**, 5–37.
Lindblom, B., and Studdert-Kennedy, M. (**1967**). "On the role of formant transitions in vowel recognition," J. Acoust. Soc. Am. **42**, 830–843.
Lotto, A. J., Kluender, K. R., and Green, K. P. (**1996**). "Spectral discontinuities and the vowel length effect," Percept. Psychophys. **58**, 1005–1014.
Mann, V. A., and Repp, B. H. (**1980**). "Influence of vocalic context on perception of the [ʃ]-[s] distinction," Percept. Psychophys. **28**, 213–228.
McDougall, K., and Nolan, F. (**2007**). "Discrimination of speakers using the formant dynamics of /u:/ in British English," *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, pp. 1825–1828.
Milner, B., and Shao, X. (**2006**). "Clean speech reconstruction from MFCC vectors and fundamental frequency using an integrated front-end," Speech Commun. **48**, 697–715.
Nossair, Z. B., and Zahorian, S. A. (**1991**). "Dynamical spectral features as acoustic correlates for initial stop consonants," J. Acoust. Soc. Am. **89**, 2978–2991.
Ohala, J. J. (**1981**). "The listener as a source of sound change," in *Papers from the Parasession on Language and Behavior*, edited by C. S. Masek, R. A. Hendrick, and M. F. Miller (Chicago Linguistic Society, Chicago), pp. 178–203.
Ohala, J. J. (**1990**). "The phonetics and phonology of aspects of assimilation," in *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, edited by J. Kingston and M. Beckman (Cambridge University Press, Cambridge), pp. 258–275.
Ohala, J. J. (**1993**). "The phonetics of sound change," in *Historical Linguistics: Problems and Perspectives*, edited by C. Jones (Longman, London), pp. 237–278.
Ohala, J. J. (**2005**). "Phonetic explanations for sound patterns. Implications for grammars of competence," in *A Figure of Speech. A Festschrift for John Laver*, edited by W. J. Hardcastle and J. M. Beck (Erlbaum, London), pp. 23–38.
Ohala, J. J., and Busà, M. G. (**1995**). "Nasal loss before voiceless fricatives: A perceptually-based sound change, " Rivista di Linguistica **7**, 125–144.
Ohala, J. J., and Feder, D. (**1994**). "Listeners' identification of speech sounds is influenced by adjacent 'restored' phonemes," Phonetica **51**, 111–118.
Pierrehumbert, J. (**2001**). "Exemplar dynamics: Word frequency, lenition, and contrast," in *Frequency Effects and the Emergence of Linguistic Structure*, edited by J. Bybee and P. Hopper (J Benjamins, Amsterdam), pp. 137–157.
Pierrehumbert, J. (**2002**). "Word-specific phonetics," in *Laboratory Phonology VII*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 101–140.
Pierrehumbert, J. (**2003a**). "Probabilistic phonology: Discrimination and robustness," in *Probabilistic Linguistics*, edited by R. Bod, J. Hay, and S. Jannedy (MIT Press, Cambridge), pp. 177–228.
Pierrehumbert, J. (**2003b**). "Phonetic diversity, statistical learning, and acquisition of phonology," Lang Speech **46**, 115–154.
Pierrehumbert, J. (**2006**). "The next toolkit," J. Phonetics **34**, 516–530.
Roach, P., and Hartman, J. (**1997**). *English Pronouncing Dictionary*, 15th ed. (Cambridge University Press, Cambridge).
Silverman, D. (**2004**). "The diachrony of labiality in Trique, and the functional relevance of gradience and variation," in *Papers in Laboratory Phonology 8*, edited by L. M. Goldstein, D. H. Whalen, and C. T. Best (Mouton de Gruyter, Berlin).
Traunmüller, H. (**1990**). "Analytical expressions for the tonotopic sensory scale," J. Acoust. Soc. Am. **88**, 97–100.
Warren, P., Hay, J., and Thomas, B. (**2007**). "The loci of sound change effects in recognition and perception," in *Laboratory Phonology 9*, edited by J. Cole and J. Hualde (Mouton de Gruyter, Berlin), pp. 87–112.
Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence for dynamic formant trajectories in Australian English Vowels," J. Acoust. Soc. Am. **106**, 458–468.
Wells, J. C. (**1982**). *Accents of English II: The British Isles* (Cambridge University Press, Cambridge).
Wells, J. C. (**1997**). "Whatever happened to Received Pronunciation?" in *II Jornadas de Estudios Ingleses*, edited by C. Medina and P. Soto (Universidad de Jaén, Spain), pp. 19–28.