This is a repository copy of *F2 slope as a perceptual cue for the front-back contrast in Standard Southern British English*.

Article:

F2 slope as a perceptual cue for the front-back contrast in Standard Southern British English

Kateřina Chládková[a], Silke Hamann[a], Daniel Williams[b], Sam Hellmuth[c]

[a]Amsterdam Center for Language and Communication, University of Amsterdam, Spuistraat 210, 1012VT Amsterdam, Netherlands

[b]Area of Excellence – Cognitive Sciences, Linguistics Department, University of Potsdam, Haus 14, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany

[c]Department of Language and Linguistic Science, University of York, Heslington, York, YO10 5DD, UK

Corresponding author's contact details:
Kateřina Chládková
Cognitive and Biological Psychology
Institute of Psychology, University of Leipzig
Neumarkt 9–19, 04109 Leipzig, Germany
email: katerina.chladkova [at] uni-leipzig.de; phone: +49 341 97-39574

Abstract
Acoustic studies of several languages indicate that second-formant (F2) slopes in high vowels have opposing directions (independent of consonantal context): front [iː]-like vowels are produced with a rising F2 slope while back [uː]-like vowels are produced with a falling F2 slope. The present study first reports acoustic measurements that confirm this pattern for the English variety of Standard Southern British English (SSBE), where /uː/ has shifted from the back to the front area of the vowel space and is now realized with higher midpoint F2 values than several decades ago. Subsequently, we test whether the direction of F2 slope also serves as a reliable cue to the /iː/-/uː/ contrast in perception. The findings show that F2 slope direction is used as a cue (additional to midpoint formant values) to distinguish /iː/ from /uː/ by both young and older SSBE listeners: an otherwise ambiguous token is identified as /iː/ if it has a rising F2 slope and as /uː/ if it has a falling F2 slope. Furthermore, our results indicate that listeners generalize their reliance on F2 slope to other contrasts, namely /ɛ/-/ɒ/ and /æ/-/ɒ/, even though F2 slope is not employed to differentiate these vowels in production. This suggests that in SSBE, a rising F2 seems to be perceptually associated with an abstract feature such as [+front] while a falling F2 with an abstract feature such as [-front].

Keywords: perceptual cues, front-back contrast, /uː/-fronting, Standard Southern British English, vowel perception, phonological feature.

1 Introduction

Vowels are acoustically differentiated in terms of their first and second formant (F1 and F2) values: for instance, the vowel [iː] has a low F1 and a high F2, while [uː] has a low F1 and a low F2. In addition to the values of formants measured at a stable portion of the vowel, for diphthongs the direction of formant trajectory also cues vowel identity: for instance, the diphthong [ɛi] has a falling F1 slope and a rising F2 slope, while [ɔu] has a falling F1 slope and a falling F2 slope. By formant slope (also called *inherent spectral change, formant trajectory* or *formant contour*) we refer in the present article exclusively to a vowel-inherent formant movement that is independent of the transitions to surrounding consonants; specifically we refer to the formant slope between 25% and the 75% point of the vowel.

Interestingly, cueing the identity of vowel diphthongs, formant trajectory seems to contribute to the identity of (some) nominally monophthongal vowels as well (for a review, see Hillenbrand, 2013). For instance, Nearey and Assmann (1986) tested whether Canadian English listeners attend to vowel formant trajectories when perceiving isolated monophthongal vowels. Nearey and Assmann extracted a short portion of the vowel nucleus (defined as the 30-ms portion centered at the vowel's 24%) and a short portion of vowel offglides (defined as the 30-ms portion centered at the vowel's 64%) and presented them to listeners in three conditions: the nucleus and the offglide portion in their natural order, in the reversed order, or the nucleus repeated twice (without the offglide). Compared to the results for non-manipulated full vowels, the manipulated stimuli yielded more misidentifications when the nucleus and offglide portion were reversed and when the nucleus was repeated twice than when the nucleus and offglide were presented in their natural order. Hillenbrand et al. (1995) analyzed the first three formants of American English vowels at 20%, 50% and 80% of the vowels' duration. In a discriminant analysis, classification accuracy was significantly better for a model that took into account formant values at 20% and 80% of the vowel than for a model that only considered the formant values at the vowel midpoint. The findings of Nearey and Assmann (1986) and Hillenbrand et al. (1995) thus suggest that vowel inherent spectral change in general may be an important cue to Canadian and American English vowels, respectively.

Relatedly, Watson and Harrington (1999) analyzed the values of formant targets and formant trajectories of Australian English monophthongs and diphthongs. Using a Gaussian classification technique, the authors showed that using formant trajectory data yielded significantly higher classification scores than using formant targets, only, for all diphthongs and for three monophthongs (namely, /iː/, /ɪ/, and /aː/). Watson and Harrington concluded that monophthongal vowels are sufficiently described by their formant target values (and duration) but that formant trajectory might help distinguishing members of some monophthongal lax-tense pairs (cf. a similar proposal by Hillenbrand et al., 1995: 3106).

Di Benedetto (1989a) reported a perception experiment that tested the effect of F1 slope on vowel identification in Italian, American English and Japanese, and found that stimuli with a falling F1 were more likely to be perceived as non-high vowels while

stimuli with a rising F1 were more likely to be perceived as high vowels (in line with Di Benedetto's [1989b] earlier production data that showed a pronounced rising F1 slope in a high-mid vowel /ɪ/ as opposed to a low-mid vowel /ɛ/). Several languages thus seem to have a relation between F1 slope direction and vowel height.

With respect to F2 slope direction, the literature lacks studies on its effect on the perceptual identification of monophthongal vowels. Despite that, a recurring pattern can be observed in acoustic vowel descriptions across languages: front vowels, i.e. vowels with a high midpoint F2, tend to have a rising F2 slope, while back vowels, i.e. those with a low midpoint F2, tend to have a falling F2 slope. For instance, Spanish /i/ and /e/ produced in isolation have a slightly rising F2, while /u/ and /o/ have a slightly falling F2 (Morrison and Escudero, 2007). A similar trend is seen in Dutch front /eː/ vs. back /oː/, and in British as well as American English front /iː/ vs. back /uː/, (see, respectively, Adank et al., 2007; Williams and Escudero, 2014; Stevens et al., 1966; Hillenbrand et al., 1995 and Munro, 1993); note also that for some languages or dialects this effect is more pronounced and the vowels are therefore sometimes referred to as diphthongal, as is the case for the Northern Standard Dutch /eː/ and /oː/. Figure 1 is a schematic illustration of the F1 and F2 slopes observed in non-low vowels that are phonetically front vs. back.

[Figure 1 here]

In the present study we test whether listeners are sensitive to the occurring correlation between midpoint F2 and the direction of F2 slope and whether they use it when identifying vowels. For this we focus on the long high vowels /uː/ (GOOSE lexical set) and /iː/ (FLEECE lexical set) of the English variety that is usually referred to as Standard Southern British English (SSBE). Impressionistic phonetic descriptions of several varieties of British English over the last 50 years have repeatedly mentioned a slight diphthongization of the two tense high vowels, with /iː/ sounding like [ɪi], [ɪiː], or [əi], and /uː/ like [ʊu], [ʊuː], [ʉ̈u], or [ʊʉ] (see Wells, 1962; Collins and Mees, 2008; and Roach, 2009: 20, on SSBE/RP; for other varieties of British English: see Stoddard et al., 1999, on the Sheffield dialect; Trudgill, 1999, on Norwich; and Docherty and Foulkes, 1999, on Derby and Newcastle). The first acoustic study supporting these observations was performed by Chládková and Hamann (2011), who analyzed young SSBE speakers' productions of /iː/ and /uː/. They found that speakers acoustically differentiate /iː/ and /uː/ not only by the vowels' midpoint F2 (and F3) values but also by the direction of F2 (and F3) slope: the formants had a rising slope in /iː/, but a falling slope in /uː/, irrespective of the consonantal context in which the vowels were embedded. This observation was confirmed by a recent acoustic description of southern and northern English vowels (Williams and Escudero, 2014). Williams and Escudero showed that in SSBE, as well as in the English variety spoken in Sheffield, /iː/ has a clearly rising F2 slope, while /uː/ has a falling F2 slope.

What makes SSBE particularly interesting with respect to F2 slopes and vowel backness is the fact that the vowel /uː/[1] has shifted from the back region of the vowel space towards the front (e.g., Henton, 1983; Bauer, 1985; Hawkins and Midgley, 2005; Harrington, Kleber and Reubold, 2008). That is, /uː/ – phonologically described as a back rounded vowel – that was originally produced with low midpoint values of F2, nowadays has higher midpoint F2 values. It should be noted here that acoustic changes in midpoint F2 can have various articulatory triggers, such as tongue fronting versus tongue backing, or lip rounding versus lip spreading, whose acoustic effects are difficult to tease apart (see e.g. Lindblom and Sundberg, 1971).[2] Figure 2 visualizes the acoustic fronting of /uː/ with data from old and young generations of speakers reported in the literature. As the process of /uː/-fronting is a comparatively recent change (and possibly still ongoing), Harrington et al. (2008) found a difference between young and older listeners in the use of the midpoint F2 as perceptual cue: the /iː/-/uː/ perceptual boundary along the F2 dimension was more fronted in young than in older listeners. On the basis of their findings one could expect a difference between young and older listeners for the use of F2 slope as well. That is, young listeners may rely on F2 slope as a cue to distinguish the two vowels more heavily than older listeners, because midpoint F2 is a less reliable cue to the /iː/-/uː/ contrast for them than it is for older listeners. Support for this hypothesis comes from reports that only older listeners seem to sometimes confuse /uː/ with /iː/ in the speech of younger speakers (Collins and Mees, 2008: 102).

[Figure 2 here]

Besides the primary goal, which is to assess the perceptual reliance on F2 slope in general, the present study follows previous work, such as Harrington et al. (2008), in that it also compares young and older speakers in their production and perception of the /iː/-/uː/ contrast. The present study adds to the existing literature in that it is the first to compare the two generations on the use of *F2 slope* (specifically, its falling vs. rising direction) as an acoustic and perceptual cue to the /iː/-/uː/ contrast. Speech production data will show whether older speakers *produce* similar F2 slopes as younger speakers (cf. Chládková and Hamann, 2011; Williams and Escudero, 2014), and a perception experiment will reveal whether young and older speakers differ in their use of F2 slope as a *perceptual* cue to the /iː/-/uː/ contrast.

If the /iː/-/uː/ contrast is, at least partially, cued by F2 slope direction, it is plausible that F2 slope is employed as a perceptual cue to other front-back contrasts as well. In that respect, results of various speech perception experiments suggest that

---

[1] Despite its considerable phonetic fronting by younger speakers, we transcribe the Goose vowel as /uː/ throughout this article.

[2] Since the acoustic effects of lip rounding and tongue backing are inseparable, we are not considering the various articulatory changes that may have lead to the acoustic change in midpoint F2. It is plausible that visual cues, such as the degree of lip rounding, are employed in the perception of the /iː/-/uː/ contrast, but these are not investigated in the present study, which focuses on *auditory* perceptual cues to the /iː/-/uː/ distinction.

listeners map the heard speech signal to phonological features rather than to single phonemes (e.g. Kraljic and Samuel, 2006; Scharinger et al., 2011). This indicates that the association between a specific direction of the F2 slope and a particular vowel, i.e. a falling F2 and the phonetically non-back high vowel /uː/, might not be phoneme specific but generalizable to other vowels, i.e. to a phonological feature. Such a generalization of a perceptual cue to front-back contrasts in general would be especially useful in a vowel system like that of English, where several vowels have changed their location in the F1 and F2 space, i.e. changed their identity in terms of midpoint F1 and/or F2 values (see e.g. the lowering and slight backing of /æ/, the lowering of /ɛ/, the fronting of /uː/ and /ʊ/, and the raising of /ɒ/ as illustrated by Hawkins and Midgley, 2005, and Wilkstrom, 2013). It would therefore be beneficial for English speakers and listeners to employ a cue additional to midpoint formant values to be able to reliably distinguish the vowels of their language, especially if this cue is already necessary to differentiate front rising from back rising diphthongs (/aɪ/ - /aʊ/). The present study therefore also tests the hypothesis that if SSBE listeners use F2 slope as a perceptual cue to the /iː/-/uː/ contrast, they might employ the same cue for other front-back contrasts, such as Kɪᴛ vs. Fooᴛ (i.e., /ɪ/-/ʊ/) or Dʀᴇss vs. Tʜᴏᴜɢʜᴛ (i.e., /ɛ/-/ɔː/).

The present study consists of three experiments. Experiment 1 is a speech production task and measures the F1, F2 and F3 slope in /iː/ and /uː/ produced by young and older speakers. Experiment 2 is a speech perception task and tests whether the direction of F2 slope affects the location of the perceptual /iː/-/uː/ boundary (i.e. whether F2 slope direction is used as a perceptual cue) and whether there is a difference between young and older listeners in their reliance on F2 slope. Experiment 3 examines whether F2 slope serves as a cue to front-back phoneme contrasts other than /iː/-/uː/.

2 Experiment 1

Experiment 1 assessed the production of /iː/ and /uː/ in the young and in the older generation of SSBE speakers. The aim of Experiment 1 was to find out whether older SSBE speakers produce /iː/ with a rising F2 slope and /uː/ with a falling F2 slope, which is the pattern reported previously for the young generation of SSBE speakers (Chládková and Hamann, 2011; Williams and Escudero, 2014).

2.1 Method
2.1.1 Participants

Four older (aged 66–69, 2 female) and four younger (aged 29–30, 2 female) speakers took part. They were all considered native speakers of SSBE because they spent most of their lives in a geographical area where SSBE is spoken and their accent was judged as SSBE by the experimenters (third and fourth author). The younger speakers were from Dorset (n=2) and London (n=2) and the older ones from London (n=2), East Sussex, and Hertfordshire. Before testing, the participants were not familiar with the purpose of the experiment. The experiment was approved by the ethical committee of the Faculty of Humanities, University of Amsterdam.

2.1.2 Materials and procedure

The speech material consisted of 24 CVC (consonant-vowel-consonant) English words, 8 of which were target $C_1[i:]C_2$ items, 8 were target $C_1[u:]C_2$ items, and 8 were $C[a:]C$ and $C[ɜ:]C$ fillers (see Table A1 in the Appendix for the target items). In half of the target words C_2 was a coronal consonant, and in the other half of the targets it was a labial consonant. Every $C_1[i:]C_2$ target (e.g. *team*) had a $C_1[u:]C_2$ counterpart with an identical $C_1\_C_2$ context (e.g. *tomb*). The words were embedded in the carrier phrase "I said CVC to you". The 24 phrases were pseudo-randomized to ensure that $C_1[i:]C_2$ targets and their $C_1[u:]C_2$ counterparts were not immediately following each other, and that there were at most 3 target items in series.

Participants read aloud the list of 24 phrases at a normal speaking rate three times. Older speakers' productions were recorded with a Marantz solid-state PMD661MkII recorder and external Shure SM10A head-mounted microphone (at a 44.1kHz sampling rate), and young speakers' productions with a Marantz solid-state recorder PMD620 with a built-in microphone (at a sampling rate of 48 kHz). The recording took place in a quiet room at the participants' or the experimenters' homes.

2.1.3 Acoustic analysis

The start and end points of vowel tokens were determined manually in the digitized waveform and were identified as the zero crossings of the first and last period that had considerable amplitude and a shape resembling the periods in the central part of the vowel. The first three formants were analyzed at 21 points equally spaced within the central 50% portion of the vowel, i.e. between 25% and 75% of the vowel's total duration. The initial and final 25% were not included in the analysis to discard the effects of the flanking consonants. Formants were measured in Praat (Boersma and Weenink, 1992–2016) by the Burg algorithm (Anderson, 1978) over a 25-ms window centered at each respective analysis point. The maximum number of formants that the algorithm searched for was 5, and the formant ceiling was fixed at 5000 Hz for male speakers and at 5500 Hz for female speakers. Tokens for which the algorithm failed to determine some of the formants at some of the analysis points were excluded from the data reported in Section 2.2 (this happened for 20 out of the total of 384 recorded tokens).

2.2 Results and Discussion

[Figure 3 here]

Figure 3 plots the average F1, F2 and F3 values measured at 25% and at 75% of the vowels' total duration. It can be seen that /i/ and /u/ are still distinguished by their average midpoint F1 and F2 values. Furthermore, it can be seen that the older speakers produce /u:/ with overall lower F2 values than the young speakers. A comparison of the formant trajectories demonstrates that both young and older speakers produce /i:/ with a rising F2 and /u:/ with a falling F2 (and a similar pattern is seen for F3 slope). The same vowel-specific direction of F2 trajectories is seen in the vowels produced by the

older speakers: they all produce /iː/ with a rising F2 and /uː/ with a falling F2. The data plotted in Figure 3 align well with the schematized observation shown in Figure 1.

Experiment 1 confirmed the pattern observed in previous acoustic descriptions of SSBE vowels produced by young speakers, namely, that /iː/ is produced with a rising F2 slope and /uː/ with a falling F2 slope despite the fact that the two vowels are still distinguished by their midpoint F2 values. This vowel-specific direction of F2 slope is also demonstrated in the speech of an older SSBE generation, whose /uː/s are not as fronted as those of young speakers.

3 Experiment 2

Experiment 2 tested whether SSBE speakers, who produce /iː/ with a rising F2 and /uː/ with a falling F2, also use F2 slope direction as a cue to these two vowels when perceiving speech. The goal of Experiment 2 was twofold. First, we aimed to show whether F2 slope serves as a perceptual cue to the /iː/-/uː/ contrast in SSBE at all. Second, we aimed to find out whether young and older listeners differ in their use of F2 slope as a perceptual cue.

3.1 Method
3.1.1 Stimuli

The stimuli were synthetic vowels made with a Klatt synthesizer (Klatt and Klatt, 1990) built into the program Praat (Boersma and Weenink, 1992–2016). A single F2 continuum ranging from 1800 Hz to 3200 Hz (measured at vowel midpoint) was divided into 12 values equidistant on an ERB scale (step size = 0.43 ERB). Each of the 12 F2 values was synthesized with two durations: 181 and 200 ms; including two different durations was to render the stimulus set more variable and thus more naturalistic. All stimuli had a midpoint F1 of 330 Hz and a midpoint F3 of 2700 Hz.[3] The stimuli were synthesized with three F2 slope types: rising, level, and falling. For 'level' stimuli, all formants were stable throughout the duration of the vowel. For 'rising' stimuli, F2 rose linearly by 0.5 ERB from the beginning to the end of the vowel, while for 'falling' stimuli, F2 fell linearly by 0.5 ERB.[4] The movement of the F3 mirrored the F2 movement. Both 'rising' and 'falling' stimuli contained a linear fall of 0.5 ERB in F1. The fundamental frequency (F0) rose linearly from 230 Hz at the beginning of the vowel up to 275 Hz at

[3] Note that the F2 ranged between 1800 and 3200 Hz, while the F3 was 2700 Hz. This means that for stimuli with F2 values higher than 2700 Hz, the two formants switched places and the value "defined" as F2 in fact became the physical F3 in the stimulus. Table A2 in the Appendix lists the F2 and F3 values of the stimuli. The scale of *defined* F2 values (i.e. 1800–3200 Hz) serves as a factor in the statistical analyses.

[4] The amount of 0.5 ERB for the formant trajectory change was used because it was large enough to be noticeable by a trained phonetician, but small enough not to be confused with a true diphthongal vowel. Also, 0.5 ERB is comparable to the average size of formant changes between vowels' 25% and 75% points that we measured in Experiment 1 (namely, 0.6 ERB for F1, 0.33 ERB for F2 of /iː/, and 0.85 ERB for F2 of /uː/).

15% of the vowel's duration and then decreased linearly to 175 Hz at the end of the vowel. The rather high F0 with this pronounced rise-fall contour imitated a young female voice. The movement in F0 was performed to acquire stimuli that sound more natural. There were in total 72 different stimuli: 12 F2 values × 2 durations × 3 slope types. Figure 4 illustrates the three F2 slope types as well as the pitch contour of the stimuli.

[Figure 4 here]

3.1.2 Participants

Forty-two young speakers and twelve older speakers of SSBE took part; they were different individuals than the participants in Experiment 1. The participants were considered native speakers of SSBE if they were born, and had been raised and educated in the south of England. All participants were paid for taking part in the experiment. The experiment was approved by the ethical committee of the Faculty of Humanities, University of Amsterdam.

The young speakers were university students between 18 and 33 years of age (mean age = 21.8; 16 male), they were recruited via posters and leaflets. The perception experiment involving these participants took place at the University of Sheffield, where the participants were tested in small groups. Before coming to study in Sheffield, the participants had lived all their lives in the south of England and themselves considered their dialect to be representative of that area. The young participants were randomly assigned to one of three groups that differed in the response labels that they were tested with, i.e. in the orthographically presented consonantal context of the answer categories: labial (n = 16, mean age = 21.1, 7 male), coronal (n = 14, mean age = 22.4, 6 male), and dorsal (n = 12, mean age = 22.2, 3 male).[5]

The older listeners were aged between 57 and 67 years (mean age = 63.2; 2 male). These participants were tested at their homes or work place: ten in London, and two in Royal Tunbridge Wells; they were recruited by the experimenters personally. All participants were healthy and reported normal hearing. Due to a limited number of recruited participants, older listeners were only tested with response labels with one of the orthographic contexts (coronal): this allowed comparable group sizes for the between-age comparison.

---

[5] Consonantal context was only present orthographically in the response labels. The reason for including CVC-labels instead of V-labels was to make the response options naturalistic because it is virtually impossible for English to unequivocally represent an isolated vowel in writing. Crucially, the *auditory* stimuli were *isolated* vowels synthesized without any consonant-specific formant transitions. Since auditorily present consonantal context has been found to affect the perceived /iː/-/uː/ boundary along the F2 dimension in young listeners (Harrington et al., 2008), we included the orthographic context as a factor in our statistical analyses to test whether orthographic-only context affects perception in a similar way as auditorily present context.

3.1.3 Procedure

The experiment was a two-alternative forced-choice identification task (implemented with the Praat software, Boersma and Weenink, 1992–2016). Participants were instructed that they would hear vowels cut from recordings of an English speaker, and they would have to identify which of two words the vowel came from. Answering categories were C_1VC_2-nonce words or rarely occurring words where C_1 was a voiceless obstruent and C_2 a voiced stop with the same place of articulation as C_1. Monosyllables that do not exist as meaningful words, or are rare words, in English were chosen in order to avoid response biases due to differences in word frequency or familiarity. To ensure that participants were familiar with how the orthographically-presented nonsense words would sound in English, they were given written instructions that the words rhymed with *leap* and *loop*, respectively. The place of articulation in the orthographically-presented answer categories varied between young listeners, and was coronal for all older listeners.

The vowel stimuli were presented over headphones. They were played in random order and there was no option of replaying the sound; if unsure, participants were asked to give their best guess. The experiment was preceded by a short practice round with 7 stimuli to ensure that participants understood the task. Each trial started with a 400-ms silent interval, after which the stimulus was played. Participants were asked to listen to the whole sound, and then indicate their response by clicking on one of the two buttons on the computer screen (labeled as e.g. *teed* and *tood*). After the participant's response, the following trial was presented. The whole randomized set of 72 stimuli was presented once to the older listeners, and twice to the young listeners. During the experiment, participants were prompted several times to take a short break and then resume the experiment (which they generally did within 2 minutes after pausing): two such breaks (i.e. after every 50th trial) were offered to the younger participants, and three (i.e. after every 20th trial) to the older participants. Older participants thus had fewer trials and more breaks than the young participants; this was because a pilot experiment showed that a task with 144 trials could be rather demanding for older listeners and that some of them had difficulties to complete it reliably and with full attention. To ensure that we collected data along the whole F2 range for older listeners, we presented them with only one instance of each stimulus (instead of two instances, as was the case for the young listeners). It took the participants between 15 and 30 minutes to complete the experiment.

3.2 Results and Discussion

In the identification task, participants classified each stimulus along the F2 range as either /iː/ or /uː/. The obtained binomial data were used to compute the location of the /iː/-/uː/ boundary along the F2 axis. Specifically, for each of the 42 young and 12 old listeners, we ran binomial logistic regression models with vowel midpoint F2 as the regression factor and proportion /iː/-responses as the dependent variable (the regression analysis was done with Praat; Boersma & Weenink, 1992–2016). The /iː/-

/uː/ boundary is located at such a midpoint F2 value *x* that would receive the label /iː/ with the probability of 0.5 (and, analogously, the label /uː/ with the probability 1 – 0.5):

$$ln\frac{0.5}{1-0.5} = \beta_0 + \beta_1 x$$

where β_0 and β_1 are the logistic regression coefficients. Since $ln\frac{0.5}{1-0.5} = 0$,

$$x = -\frac{\beta_0}{\beta_1}$$

The boundaries of the 42 young listeners were submitted to a repeated-measures analysis of variance (RM-ANOVA)[6] with slope type as the within-subjects factor (rising, level, falling) and orthographic context in the answer category as the between-subjects factor (labial, coronal, dorsal). The analysis revealed a main effect of slope type (F[2,78] = 37.847, p < .001). No significant main effects or interactions involving orthographic context were found. Pairwise comparisons (Fisher's LSD) of the mean boundary locations across the three slope types showed that the /iː/-/uː/ boundary for stimuli with rising F2 was at lower F2 values than the boundary for stimuli with level F2, which in turn was at lower F2 values than the boundary for stimuli with falling F2; see Table 1. Figure 5 (top graph) plots the logistic regression fit averaged across the 42 young listeners.

To test for the effect of age, the boundaries of the 12 older and the 14 young listeners who were tested with the same orthographic context were submitted to a second RM-ANOVA with slope type as the within-subjects factor and age group as the between-subjects factor. The analysis revealed a main effect of slope type (F[2ε,48ε, ε = .882] = 9.974, p < .001). There were no significant main effects or interactions involving age. Pairwise comparisons (Fisher's LSD) of the mean boundary locations across the three slope types showed that the /iː/-/uː/ boundary for stimuli with rising F2 was at lower F2 values than the boundary for stimuli with level F2, which in turn was at lower F2 values than the boundary for stimuli with falling F2; see Table 2. Figure 5 (middle and bottom graphs) plots the logistic regression fits of the 14 young and 12 old listeners who were tested with coronal context.

[Figure 5 here]
[Table 1 here]
[Table 2 here]

The results of Experiment 2 demonstrate that native speakers of SSBE use F2 (and/or F3) slope[7] as a perceptual cue to the /iː/-/uː/ contrast: the /iː/-/uː/ boundary

---

[6] In this and all subsequent analyses, if Mauchly's test of sphericity is not passed, we employ Huynh–Feldt's correction, which reduces the number of degrees of freedom by a factor ε.

[7] In our synthetic stimuli, F2 and F3 had identical slope directions (i.e. *both* formants were rising, falling, or level). We therefore cannot identify whether the perceptual effect was driven by slope direction in one of these formants or in both. However, since F2 is a lower frequency and thus perceptually more salient to the human ear than F3, we suspect the effect was caused by the slope of F2. Analogously, for the sake of simplicity, when referring to the present experiments

was at lower F2 values for stimuli with rising F2 slope than for stimuli with falling F2 slope. That is, listeners identified a stimulus with an ambiguous midpoint F2 more often as /uː/ when it had a falling F2 than when it had a rising F2. Besides midpoint F2 values that are used to distinguish /iː/ from /uː/, the direction of F2 slope functions as another (secondary) perceptual cue to the /iː/-/uː/ contrast. This finding on vowel perception is in line with the production data from Experiment 1 as well as recent acoustic studies on SSBE (Chládková and Hamann, 2011; Williams and Escudero, 2014), in which young SSBE speakers produced /iː/ with a rising F2 slope and /uː/ with a falling F2 slope. With respect to the age effects in the use of F2 slope, we did not find any difference between young and older listeners: both groups showed a similar influence of F2 slope on the perceptual boundary between /iː/-/uː/.

4 Experiment 3

To assess whether F2 slope direction is used as a cue to front-back contrasts in general we carried out Experiment 3. Additionally, the design of Experiment 3 improved several aspects of Experiment 2. It was a vowel identification task with a design that aimed at more closely replicating non-laboratory speech perception: stimuli were sampled from (1) a large F1-F2 vowel space (not just a single continuum), and the response labels consisted of (2) all the eleven British English monophthongal phonemes (not just two vowels). Experiment 3 was run with young SSBE speakers who have (3) always lived in the same single area of southern England (namely, Kent), and were slightly younger than the group of young participants in Experiment 2[8]. Experiment 3 thus investigated whether front-back contrasts other than /iː/-/uː/ are cued by F2 slope, and whether we can replicate the findings of Experiment 2 with a larger stimulus set, a larger number of response options, and a group of participants who are more homogenous with respect to linguistic experience and age.

4.1 Method
4.1.1 Stimuli

The stimuli were synthetic vowels sampled from a large F1-F2 vowel space spanning most of the possible vowel realizations of the modeled speaker, with relatively more stimuli from the upper region of the vowel space. Figure 6 shows the F1-F2 stimulus grid. F1 and F2 were both sampled into 11 values equidistant on an ERB scale. F1 ranged from 300 to 1000 Hz (7.28 to 15.29 ERB, step size was 0.80 ERB), F2 ranged from 800 to 3300 Hz (13.59 to 25.07 ERB, step size was 1.15 ERB). We excluded F1-F2 combinations that are by definition impossible (when F1 would be above F2, i.e. the lower right corner of the vowel grid) or highly unlikely, frog-like sounding, speech

---

we use the term "F2 slope" although it in fact represents the same physical change in both F2 and F3.

[8] Older participants were not included in Experiment 3 because the task of Experiment 3 was even more demanding than that of Experiment 2, and also because no differences were detected between younger and older participants' perception of F2 slope in Experiment 2.

sounds (high F1 values combined with high F2 values, i.e. the lower left corner of the vowel grid). This procedure yielded 93 unique F1-F2 pairs.

To test whether listeners rely on F2 slope as a cue to the front-back contrast among non-low vowels, we varied the direction of F2 slope of the 55 tokens in the upper part of the vowel grid (outlined by the grey rectangle in Figure 6). The upper 55 tokens were thus synthesized with three possible slope types: level, rising and falling (similarly to Experiment 2). Additionally, these 165 stimuli (i.e. 55 F1-F2 pairs x 3 slope types) from the upper part of the vowel grid were synthesized with two F3 values: 2200 Hz and 2800 Hz (21.72 and 23.72 ERB)[9], and two durations: 245 ms and 181 ms. The variation in duration and F3 values was included to achieve a more naturalistic stimulus set, but also to distract participants' attention from the systematic changes in F2/F3 slopes. The 38 tokens from the lower part of the vowel grid had level F2, an F3 of 2566 Hz (23 ERB) and a duration of 211 ms. These level-F2 low-vowel stimuli were included to render the stimulus set more variable and to make participants less aware of the fine acoustic detail in the upper part of the vowel space. All stimuli contained the same pattern of F0 contour as the stimuli in Experiment 2. Combining 55 F1-F2 values from the upper part of the vowel space with 2 F3 values, 2 durations, and 3 slope types, and adding the 38 tokens from the lower part of the vowel space yielded 698 stimuli in total.

[Figure 6 here]

4.1.2 Participants

The participants were 42 young monolingual native speakers of SSBE (38 female; different individuals from the subjects in Experiment 2). They were sixth-form high-school students between 17 and 19 years of age. They were first approached by their teachers, who gave them general information about the experiment. On the day of testing, interested students could ask the experimenters for more detail and/or could also express their interest in participating. At the time of testing, the participants had lived all their lives in Kent, UK. We tested seven additional participants but these were excluded because it turned out that they had been raised in a bilingual environment (5 participants) or they did not complete the perception task (2 participants). All participants were paid for taking part in the experiment. The experiment was approved by the ethical committee of the Faculty of Humanities, University of Amsterdam.

---

[9] Note that the F2 ranged up to 3300 Hz, while the F3 was either 2200 Hz or 2800 Hz. This means that for stimuli with defined F2 values higher than 2200 Hz and 2800 Hz, respectively, the two formants switched places and the value "defined" as F2 in fact became the physical F3 in the stimulus (as in Experiment 2; see footnote 3). Table A2 in the Appendix lists all the F2 and F3 combinations that were synthesized in our stimulus set. The scale of *defined* F2 values serves as a factor in the statistical analyses.

4.1.3 Procedure

The experiment was a multiple forced-choice identification task. Participants had to identify every vowel stimulus with one of 11 labels corresponding to nonce [10] monosyllabic words each containing one of the 11 SSBE monophthongal vowels /iː ɪ ɛ æ ɜː ʌ ɑː ɒ ɔː ʊ uː/. The words were presented orthographically on a computer screen as CeeC, CiC, CeC, CaC, CerC, CuC, CarC, CoC, CawC, CuCC, and CooC (the order corresponding to the 11 vowels listed above, C = consonant). The consonantal frames were fVb, tVd, and kVg (V = vowel) and participants were randomly assigned one of the three orthographic consonantal contexts for the whole experiment[11].

The 698 vowel stimuli were presented one at a time in random order over headphones. Each trial started with a 1000-ms silence, after which a stimulus was played. Participants were asked to wait until the entire stimulus was played and then give their answer by clicking on one of the 11 buttons on the computer screen containing the 11 English nonsense words. Participants were asked to give their best guess if unsure; there was no option to replay the sound. There was a 5-second break after every 88th stimulus; the fourth out of a total of 7 breaks was somewhat longer and participants could decide themselves when to resume the experiment. Participants were tested in small groups in a quiet computer room at the Charles Darwin School in Kent, UK. The experiment took between 45 and 60 minutes to complete.

Prior to the perception experiment, participants were presented with a printed list of their 11 answer categories together with a set of rhyming words embedded in a sentence. For instance, the text relevant for the /iː/-word in a labial frame were: "**Feeb** rhymes with feed and seek. In **feeb** we have an 'ee'. **Feeb**." The participants were asked to try to quietly learn the pronunciation of the 11 new words and were given approximately 5 minutes for this task. They were told that the purpose of the subsequent listening experiment was to test how well they had learnt the pronunciation of these eleven new words.

4.2 Results

Figure 7 shows the labeling results pooled across the 42 participants. For each stimulus, the figure plots the vowel category that was chosen by the majority of participants (in case of a tie, both response categories are plotted).

[Figure 7 here]

As can be seen from Figure 7, three response categories were hardly ever used: /ɔː/, /ʊ/, and /ʌ/. The labeling patterns also show that subjects used the labels *tudd* (the

---

[10] Some of the monosyllables do in fact represent words that exist in English. Since these are names, abbreviations, or rather infrequent words, we instructed the participants that the words they were going to learn do not exist in English. We supposed this would further draw participants' attention away from the possible existent meaning of these words. Therefore, we did not expect the possible-word status of some response labels to affect participants' identification of the stimuli.

[11] The design of response labels followed that of Experiment 2; see footnote 5 for details.

label for /ʊ/) and *tud* (the label for /ʌ/) interchangeably, most likely because /ʊ/-monosyllables spelled with a single vowel symbol <u> followed by a double consonant (e.g. *pull*), occur rarely as words in English. Due to the lack of reliable /ʊ/ responses, we could not include the /ɪ/-/ʊ/contrast in our analysis. Furthermore, participants were either not able to associate the label *tawd* with the vowel /ɔ:/, or did not consider the stimuli good renditions of this vowel, potentially because the duration of the stimuli was not long enough.

For stimuli from the upper vowel region (i.e., stimuli with an F1 between 300 and 515 Hz), we ran binomial logistic regression with mid-point F1 and F2 as the regression factors and proportion /i:/-responses as the dependent variable. The /i:/-/u:/ boundary in the two-dimensional F1-F2 space runs through such F1-F2 value pairs, i.e. y and x values, that would receive the label /i:/ with the probability of 0.5:

$$ln\frac{0.5}{1-0.5} = \beta_0 + \beta_1 y + \beta_2 x$$

where β_0, β_1, and β_2 are the logistic regression coefficients, y is the value of F1 and x is the value of F2. We are further interested in the boundary location on the F2 axis for an intermediate F1 value (i.e., for the value of y halfway between 300 and 515 Hz along an ERB scale). Therefore, since $ln\frac{0.5}{1-0.5} = 0$,

$$x = -\frac{\beta_0 + \beta_1 y}{\beta_2} = -\frac{\beta_0 + \beta_1 \cdot 8.88}{\beta_2}$$

The F2 locations of the boundaries were submitted to a RM-ANOVA with slope type as the within-subjects factor with three levels (rising, falling, level). Boundaries that were found to lie below 0 ERB or above 30 ERB were excluded from the statistical analysis: this happened for one participant's boundary for the level-F2 stimuli, thus leaving us with /i:/-/u:/ boundary data from 41 participants. The ANOVA yielded a significant main effect of F2 slope (F[2,80] = 3.800, p = .015). Pairwise comparisons showed that the F2 boundary was at significantly lower F2 values for stimuli with rising F2 than for stimuli with falling F2 (mean difference = 0.608 ERB, p = .013, 95% c.i. = 0.136…1.081).

Although we were not able to assess boundary locations for the /ɪ/-/ʊ/ contrast (most likely due to the confusion of the /ʊ/ and /ʌ/ labels), the data provide us with other front-back contrasts for which the boundary can be reliably determined. Figure 7 suggests that, apart from /i:/ and /u:/, stimuli from the upper region of the vowel space (i.e. the grey area of Figure 6) were often labeled as /æ/, /ɛ/, /ɜ:/, and /ɒ/. In SSBE, the vowels /æ/ and /ɛ/ are front, /ɒ/ back, and /ɜ:/ central (see e.g. Roach, 2009). Thus, to further examine whether F2 slope serves as a cue to a front-back contrast in general, we ran a binomial logistic regression for the two remaining front-back contrasts in our data: /æ/-/ɒ/ and /ɛ/-/ɒ/. Note that for /ɛ/-/ɒ/ in one subject and for /æ/-/ɒ/ in nine subjects there were not enough of the respective vowel responses to fit the logistic regression. From the regression coefficients we again computed, per participant, the location of the /æ/-/ɒ/ and /ɛ/-/ɒ/ boundaries for each slope type. As with /i:/-/u:/, boundaries below 0 ERB or above 30 ERB were excluded from further analyses. We thus had boundary data for all three contrasts from 32 subjects. We submitted the /æ/-/ɒ/

and /ɛ/-/ɒ/ boundaries together with the /iː/-/uː/ boundaries to a second RM-ANOVA with slope type and vowel contrast as the within-subjects factors with three levels each (i.e. slope: rising, falling, level; vowel contrast: /iː/-/uː/, /æ/-/ɒ/, /ɛ/-/ɒ/).

The ANOVA yielded a main effect of vowel contrast (F[2,62] = 16.051, p = .001) and a main effect of slope type (F[2,62] = 5.260, p = .008). The analysis did not detect a significant interaction between vowel contrast and slope type. The main effect of vowel contrast indicates that, unsurprisingly, the F2 boundary differed across the 3 vowel pairs. Pairwise comparisons of the means showed that the /ɛ/-/ɒ/ boundary was at lower F2 values than the /æ/-/ɒ/ boundary, which was in turn at lower F2 values than the /iː/-/uː/ boundary; see Table 3. As for the main effect of trajectory type, pairwise comparisons showed that the boundary for stimuli with rising F2 was at significantly lower F2 values than the boundary for stimuli with falling F2 (mean difference = -0.522 ERB, CI = -0.828…-0.216, p = .001). Figure 8 plots, for each slope type, the /iː/-/uː/, /æ/-/ɒ/, and /ɛ/-/ɒ/ boundaries in the two-dimensional F1-F2 space.

[Figure 8 here]
[Table 3 here]

4.3 Discussion
The findings of Experiment 3 replicated those of Experiment 2 in that the /iː/-/uː/ boundary was affected by the F2 slope of the stimuli: listeners identified a stimulus with ambiguous midpoint F2 values more often as /uː/ when it had a falling F2 than when it had a rising F2. The results of Experiment 3 further suggest that F2 slope affects boundary location in two other contrasts, namely /æ/-/ɒ/ and /ɛ/-/ɒ/, in a similar way as it does in the /iː/-/uː/ contrast. This indicates that (at least) in young SSBE listeners, front-back contrasts other than /iː/-/uː/ are also perceptually cued by the direction of F2 slope.

Though unrelated to our research question, we would like to report on the unexpected finding that stimuli with high F2 values and rather low F1 values (i.e. the space that is occupied by the vowel /ɛ/) were labeled /æ/, as can be seen in Figure 7. This effect is even stronger for the long stimuli. Given the fact that the vowel /æ/ has been reported to shift towards an [a]-like quality in the production of young SSBE speakers (Gimson, 2001: 83; de Jong et al., 2007; see also Harrington, 2007 for one older speaker), it is rather surprising that our listeners labeled stimuli with very low F1 values as /æ/. One could argue that this unexpected result is due to the nature of the stimuli: although they were carefully synthesized to model naturally produced vowels, the synthesis might not have captured all subtle cues that occur in natural speech and that may be important for identification of some vowels. To provide a more specific explanation for the unexpected labeling pattern, we propose the following: participants considered the front-vowel stimuli with rather low F1 values as being too long for an /ɛ/, and /æ/ is the only front vowel that is slightly longer in duration. This speculation seems to be supported by recent studies on SSBE vowel production and perception: /æ/ is produced with 1.1 times longer duration than /ɛ/ (Williams and Escudero, 2014:

Table IV), and listeners' perceptual judgments show that the best perceptual exemplar of /æ/ is 1.33 times longer than that of /ɛ/ (Evans and Iverson, 2004: Table II).

5 General discussion and Conclusions

The present study demonstrates that the direction of F2 (and F3) slope serves as a perceptual cue to the /iː/-/uː/ distinction in SSBE. When classifying vowels, adolescents, young and older adults use the slope of F2 in a similar way: a vowel with an ambiguous midpoint F2 is perceived as /iː/ if it has a rising F2 slope and as /uː/ if it has a falling F2 slope.

Our findings further indicate that F2 slope may not be specific to the /iː/-/uː/ contrast but seems to serve as a perceptual cue to a more general front-back contrast: F2 slope direction had the same effect on boundary location for the front-back contrasts /æ/-/ɒ/ and /ɛ/-/ɒ/ as it had for the /iː/-/uː/ contrast. For the latter, production data (from the present and previous studies) are in line with our perceptual findings, as /iː/ is produced with a rising F2 and /uː/ with a falling F2. For the other two pairs, however, production is not consistent with perception. Williams and Escudero (2014) show that SSBE speakers realize the back vowel /ɒ/ with a falling F2 slope, as expected on the basis of our perception data, but produce also the front vowels /æ/ and /ɛ/ with a falling or a level F2, which is not in line with our perception data. There is thus a mismatch between perception and production for some of these vowels. The homogeneous performance for front versus back vowels observed in the present perception experiment can therefore not be explained in terms of phoneme-specific learning of acoustic information. Instead, we propose that SSBE speakers may generalize a rising F2 slope from /iː/ to other front vowels by associating it with an abstract representation such as a feature [+front], and generalize a falling F2 slope from /uː/ to other back vowels via a feature such as [−front], regardless of their actual realization of these other front and back vowels in production. This proposed *generalization* of a perceptual cue across vowels sharing an abstract *feature* (such as [+/− front]) and the observed asymmetry between production and perception pose problems for exemplar-theoretic approaches (e.g. Johnson, 1997; Pierrehumbert, 2001), where language users are restricted in their storage and usage to phonetic properties that they have encountered in a given sound. Models of the phonology-phonetics interface, on the other hand, in which auditory cues can be mapped onto phonological feature representations (besides phoneme representations) can explain such generalization of a perceptual cue even if it is absent in production (e.g. Boersma and Hamann, 2009).

Our proposal that the perceptual reliance on F2 slope direction generalizes from the /iː/-/uː/ contrast to other front-back contrasts is based on the following: (1) the effect that F2 slope has on vowel categorization seems to be stronger for /iː/-/uː/ than for the other vowels (possibly because it is the /iː/-/uː/ contrast for which midpoint F2 – once a strong primary cue – seems to be becoming a less important cue than it was some 50 years ago), and (2) the perceptual effect of F2 slope that we found for /iː/-/uː/ aligns well with the *production* of these two vowels (while the perceptual effect found for the other vowels does not align well with the production of those vowels).

An objection that could be raised against our conclusion that SSBE listeners use F2 slope as a perceptual cue, is that our participants might not have been attending to the entire stimulus but instead listened only to its final part.[12] Such an assumption would imply that SSBE listeners in general ignore the first half of vowels, because it is not informative. This would then also hold for diphthongs, because listeners do not know in advance whether a vowel sound is a diphthong or a monophthong. However, SSBE has a contrast between the diphthongs /aɪ/ and /ɔɪ/ and the diphthongs /ɪə/ and /ʊə/, which requires that listeners pay attention to the first half of the vowel sound. A similar argument could be made against SSBE listeners focusing on the first part of the vowel only, namely the presence of diphthong pairs such as/aʊ/ and /aɪ/. In order to successfully perceive the vowels of their language, SSBE listeners thus need to attend to spectral information over the entire vowel: it is therefore likely that they did so also in the present experiments. Indeed, this interpretation is supported by Morrison's (2013) review of literature of evidence for theories of vowel inherent spectral change which finds that listeners' perceptual responses are best predicted with models that include acoustic information from both the onset and offset of vowels.

With respect to the role of F2 slope in the (still ongoing) process of /uː/-fronting in SSBE, we found that this cue is actively employed by both younger and older speakers and listeners. It is likely that F2 slope direction serves as a supplementing cue especially (or exclusively) for vowel tokens with ambiguous midpoint F2 values and less so for tokens with peripheral midpoint F2 values of typical [iː]- and [uː]-qualities. The relative perceptual weighting of midpoint F2 and of F2 slope direction (and any other cues such as a separate F3 slope direction) remains to be investigated in future work.

Note that previous phonetic literature on the emergence of SSBE /uː/-fronting has always focused on midpoint F2 and referred to its possible causes, such as articulatory ease (Harrington, Hoole, Kleber and Reubold, 2011; Harrington, Kleber and Reubold, 2011), a prevalence for /uː/ to occur post-coronally (Harrington 2007; Harrington et al., 2008), and a failure of the younger generation to perceptually compensate for coarticulation (Harrington et al., 2008, based on Ohala's 1981 hypocorrection account). Irrespective of what factors may have driven the fronting of /uː/, the present study shows that any potential decrease of the /iː/-/uː/ distinction along midpoint F2 values can well be accommodated for, because there is at least one additional auditory dimension, namely F2 slope, along which the two vowels are well differentiated. Our findings therefore provide novel insights supplementing the existing literature on /uː/-fronting in English.

---

[12] Note that a stimulus at a certain F2 midpoint value that had a falling F2 trajectory ended with a lower F2 than a stimulus at an identical midpoint F2 with a rising trajectory. That is, if listeners only attended to the end of stimuli, a stimulus with a falling F2 would be more likely perceived as /uː/ than the same stimulus with a rising F2, because the former ends at a lower F2 value than the latter stimulus.

References

Adank, P., van Hout, R., and van de Velde, H. (2007). An acoustic description of the vowels of northern and southern standard Dutch II: Regional varieties. *Journal of the Acoustical Society of America,* 121 (2), 1130–1141.

Anderson, N. (1978). On the calculation of filter coefficients for maximum entropy spectral analysis. In Childers, D. G. (ed.) *Modern Spectrum Analysis*, (pp. 252–255). IEEE Press: New York.

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

Boersma, P., and Hamann, S. (2009). Cue constraints and their interactions in phonological perception and production. In Boersma, P., Hamann, S. (Eds.), *Phonology in perception*, (pp. 55-110). Berlin: Mouton de Gruyter.

Boersma, P., and Weenink, D. (1992–2016). Praat: doing phonetics by computer (Version 5.2.26). [Computer program], Retrieved May 26, 2011, from http://www.praat.org/.

Chládková, K., and Hamann, S. (2011). High vowels in Standard British English: /u/-fronting does not result in merger. *Proceedings of XVII ICPhS 2011*, Hong Kong, 476-479.

Collins, B.S., and Mees, I.M. (2008). *Practical Phonetics and Phonology*. (2nd ed.). Abingdon: Routledge.

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 XVI ICPhS*, Saarbrücken, 1813–1816.

Di Benedetto, M.-G. (1989a). Frequency and time variations of the first formant: Properties relevant to the perception of vowel height. *Journal of the Acoustical Society of America*, 86 (1), 67–77.

Di Benedetto, M.-G. (1989b). Vowel representation: Some observations on temporal and spectral properties of the first formant frequency. *Journal of the Acoustical Society of America*, 86 (1), 55–66.

Docherty, G.J., and Foulkes, P. (1999). Derby and Newcastle: instrumental phonetics and variationist studies. In Foulkes, P., Docherty, G. (Eds.), *Urban Voices* (pp. 47–71). London: Arnold.

Evans, B. G., and Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America* 115 (1), 352–361.

Gimson, A. C. (2001). *Gimson's pronunciation of English* (6th ed.). London: Arnold.

Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In Cole, J., Hualde, J. (Eds.) *Laboratory Phonology 9* (pp. 125–143). Mouton: Berlin.

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

Harrington, J., Hoole, P., Kleber, F., and Reubold, U. (2011). The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels. *Journal of Phonetics*, *39,* 121–131.

Harrington, J., Kleber, F., and Reubold, U. (2011). The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English*. Journal of the International Phonetic Association*, *41,* 137–156.

Hawkins, S., and Midgley, J. (2005). Formant frequencies in RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association*, *35,* 183–195.

Henton, C. G. (1983). Changes in the vowels of Received Pronunciation. *Journal of Phonetics*, *123,* 353–371.

Hillenbrand, J. (2013). Static and dynamic approaches to vowel perception. In Morrison, G. S., Assmann, P. F. (Eds.) *Vowel Inherent Spectral Change* (pp. 9-30). Berlin: Springer.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97 (5), 3099–3111.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In Johnson, K., Mullennix, J.W. (Eds.) *Talker Variability in Speech Processing* (pp. 145–165). San Diego: Academic Press.

Klatt, D. H, and Klatt, L. C. (1990). Analysis, synthesis and perception of voice quality variations among male and female talkers. *Journal of the Acoustical Society of America*, 87 (2), 820-856.

Kraljic, T., and Samuel, A. G. (2006). Generalization in perceptual learning of speech. *Psychonomic Bulletin & Review*, 13 (2), 262–268.

Lindblom, B. E. F., and Sundberg, J. E. F. (1971). Acoustical consequences of lip, tongue, jaw and larynx movement. *Journal of the Acoustical Society of America*, 50 (4), 1166–1179.

Morrison, G. S. (2013). Theories of vowel inherent spectral change. In Morrison, G. S., Assmann, P. F. (Eds.) *Vowel Inherent Spectral Change* (pp. 31-48). Berlin: Springer.

Morrison, G. S., and Escudero, P. (2007). A cross-dialect comparison of Peninsular- and Peruvian-Spanish vowels. *Proceedings of XVI ICPhS*, Saarbrücken, 1505–1508.

Munro, M. J. (1993). English vowel production. *Language and Speech*, 36(1), 39–66.

Nearey. T. M., and Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, 80 (5), 1297–1308.

Ohala, J. J. (1981). The listener as a source of sound change. In C. Masek, R. A. Hendrik and M. F. Miller (Eds.), *Proceedings of the Chicago Linguistic Society* 17, 178–203. Chicago.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, J. L., Hopper, P. (Eds.) *Frequency and the Emergence of Linguistic Structure* (pp. 137–157). Amsterdam: Benjamins.

Roach, P. (2009). *English Phonetics and Phonology*. (4th ed.). Cambridge: Cambridge University Press.

Scharinger, M., Idsardi, W. J., and Poe, S. (2011). A comprehensive three-dimensional cortical map of vowel space. *Journal of Cognitive Neuroscience*, 23 (12), 3972–3982.

Stevens, K. N., House, A. S., and Paul, A. P. (1966). Acoustical description of syllabic nuclei: Interpretation in terms of a dynamic model of articulation. *Journal of the Acoustical Society of America*, 40 (1), 123–132.

Stoddart, J., Upton, C., and Widdowson, J.D.A. (1999). Sheffield dialect in the 1990s: revisiting the concept of NORMs. In Foulkes, P., Docherty, G. (Eds.), *Urban Voices* (pp. 72–89). London: Arnold.

Trudgill, P. (1999). Norwich: endogenous and exogenous linguistic change. In Foulkes, P., Docherty, G. (eds.), *Urban Voices* (pp. 124–140). London: Arnold.

Watson, C. I., and Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America*, 106 (1), 458–468.

Wells, J. (1962). *A study of the formants of the pure vowels of British English*. MA thesis University of London. [available at: http://www.phon.ucl.ac.uk/home/wells/formants/index.htm, accessed May 28, 2013]

Wilkstrom, J. (2013) An acoustic study of the RP English LOT and THOUGHT vowels. *Journal of the International Phonetic Association*, 43 (1), 37-47.

Williams, D. P., and Escudero, P. (2014) A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *Journal of the Acoustical Society of America,* 136 (5), 2751–2761.

Figures and Tables
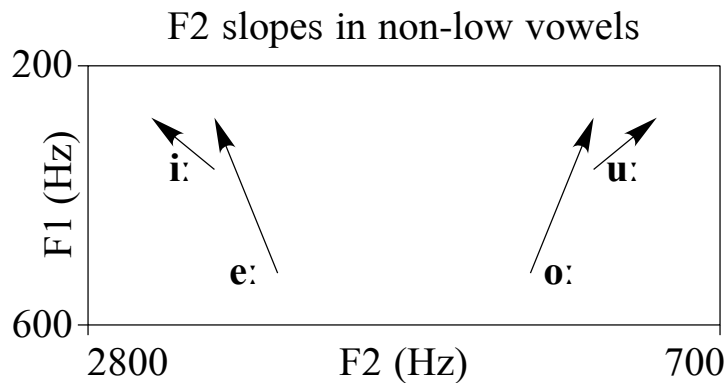

F2 slopes in non-low vowels

Figure 1: A visualization of the trend in formant slope directions for non-low vowels observed in several languages (see studies discussed in text): the F1 slope tends to be falling, while the F2 slope tends to be rising for the front vowels and falling for the back ones. This effect is seen for vowels produced in isolation as well as for vowels produced in various consonantal contexts (when consonant-specific formant transitions are removed, i.e. in the central 50% portion of vowels). Note that the exact F1 and F2 locations of the arrows are illustrative and depict the general tendencies observed across languages.
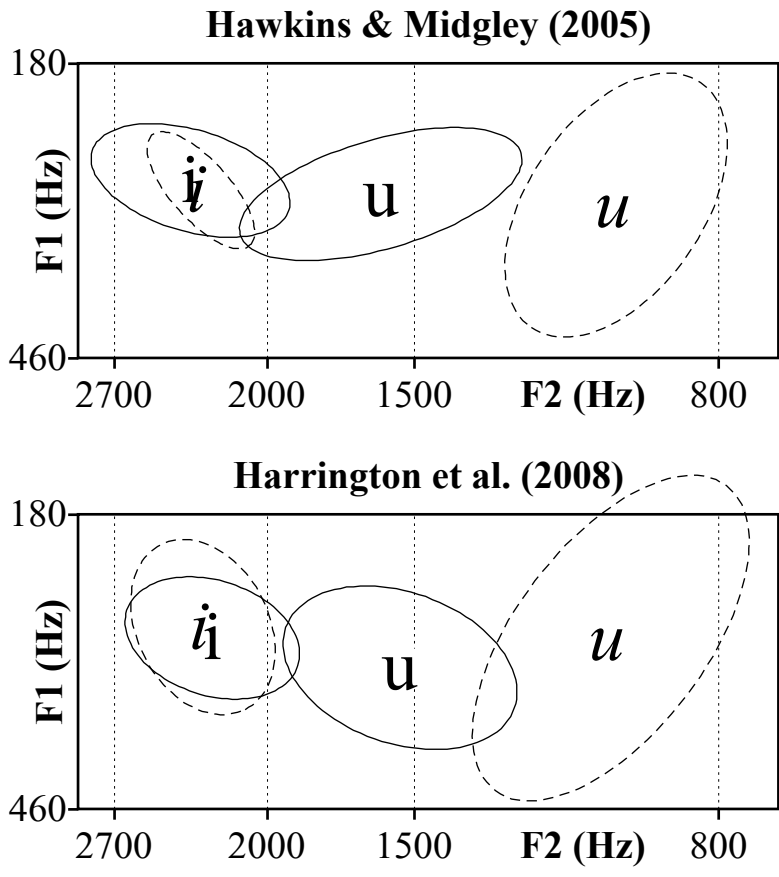
Figure 2: F1-F2 plot of /iː/ and /uː/ produced by male speakers of different ages from two previous studies. Symbols represent means and ellipses show 2 standard deviations. The figure shows data for the young(est) and the old(est) group of speakers in the respective studies. The data from young speakers is drawn with solid lines, the data from older speakers with dashed lines and italics.
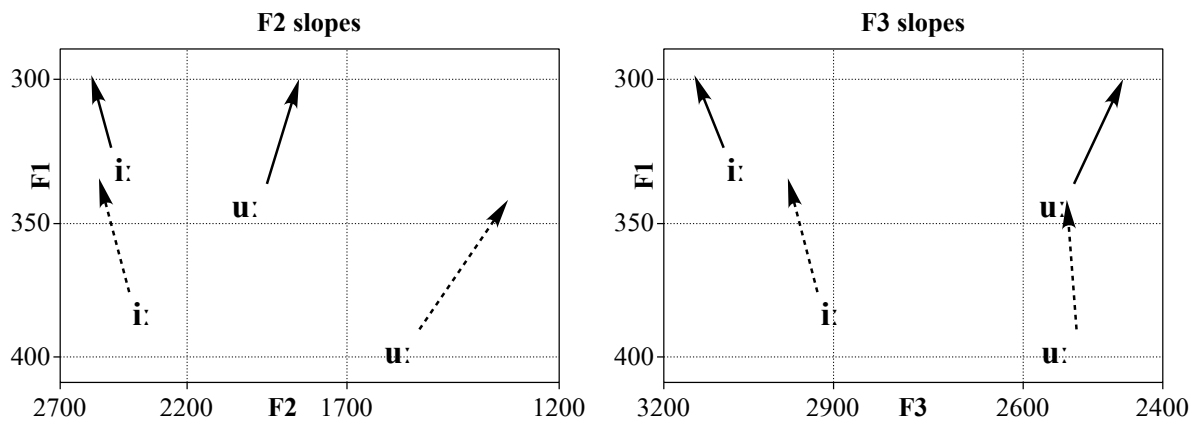
Figure 3: Formant values measured at 25% (start points of the arrows) and 75% (end points of the arrows) of the vowels' total duration, averaged across the 4 speakers per age group. Solid lines = younger speakers, dashed lines = older speakers. The average formant slopes were: 0.6-ERB fall for F1, 0.33-ERB rise for F2 of /iː/, 0.85-ERB fall for F2 of /uː/, 0.14-ERB rise for F3 of /iː/, and 0.09-ERB fall for F3 of /uː/. In order to show the magnitude of formant changes along a psychoacoustically plausible scale, the formant changes are given in ERB, and so is the scaling of the axes in the Figure.
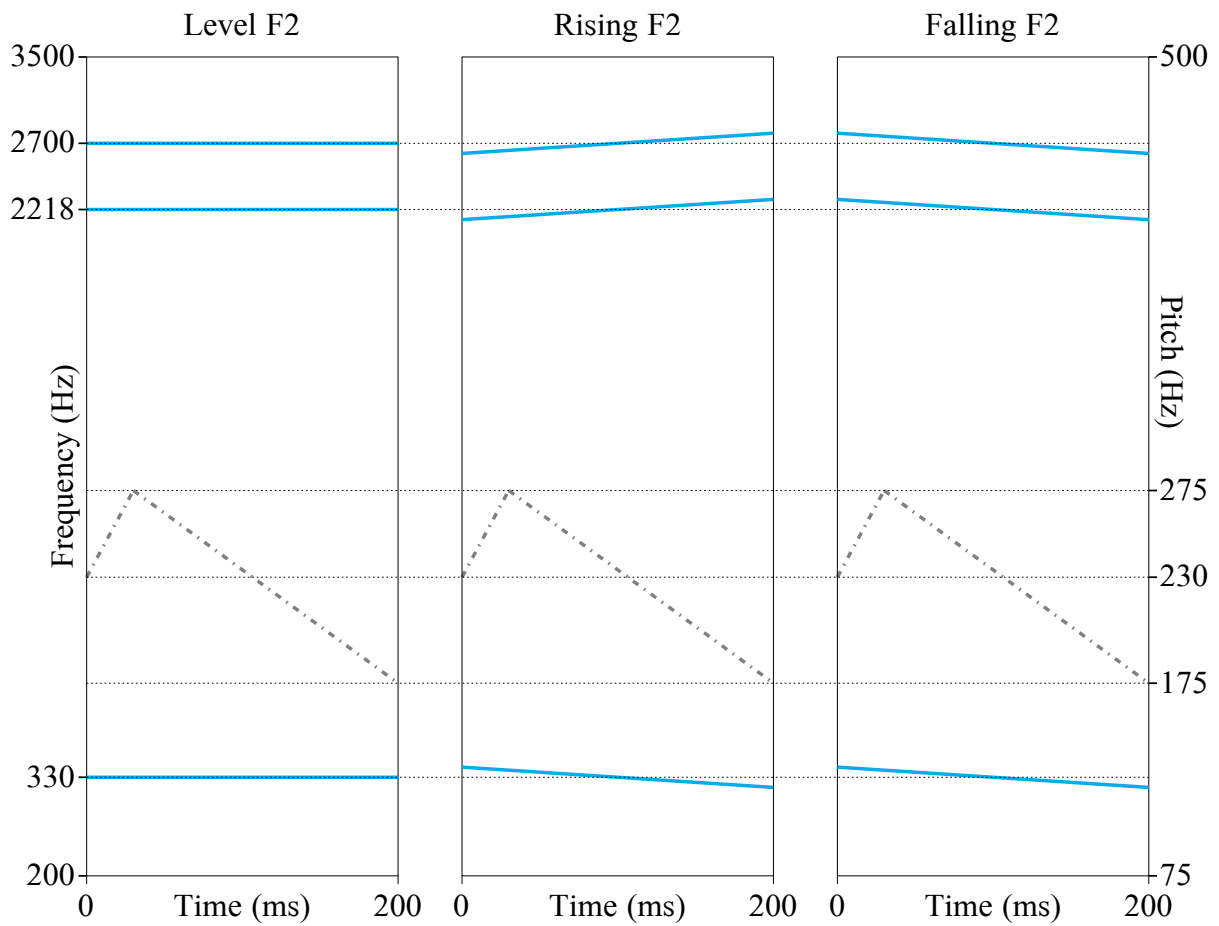
Figure 4: Illustration of stimuli from Experiment 2. The figure shows the three different slope types for a stimulus with mid-point F2 value of 2218 Hz and duration of 200 ms: the blue solid lines represent the first three formants (left axis), and the dotted-dashed line shows the pitch contour (right axis).
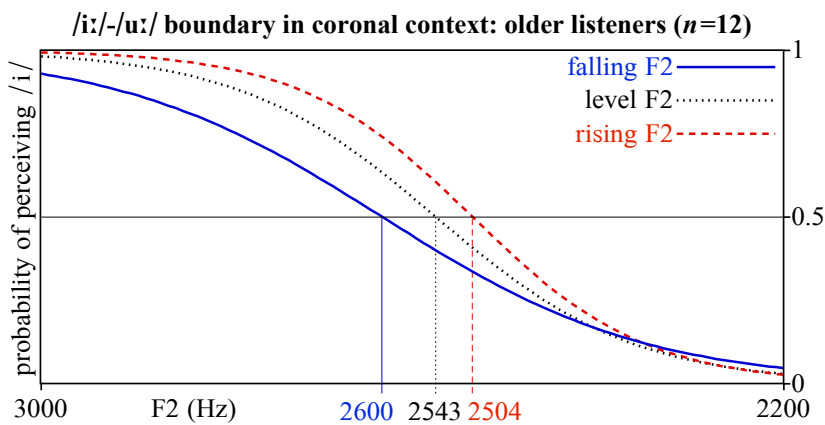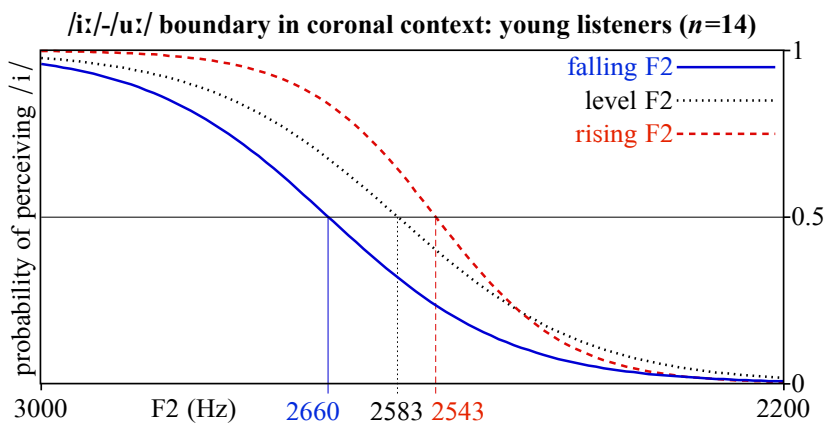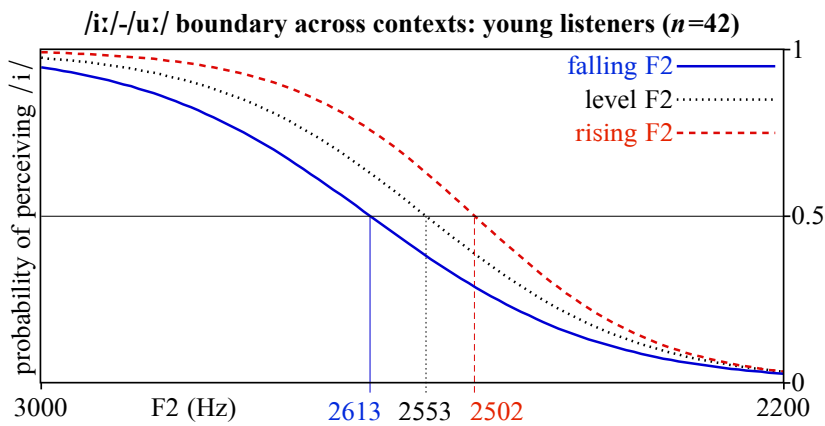
Figure 5: Experiment 2: perceptual /iː/-/uː/ boundaries on the F2 dimension, averaged across *n* listeners in each group. The top graph shows results for the 42 young listeners. The middle graph shows the subgroup of 14 young listeners who were directly compared to the 12 older listeners (shown in the bottom graph). Note that in order to zoom in on the boundary locations, the graphs show an F2 range between 2200 and 3000 Hz, however, the stimulus continuum in the experiment ranged from 1800 to 3200 Hz.
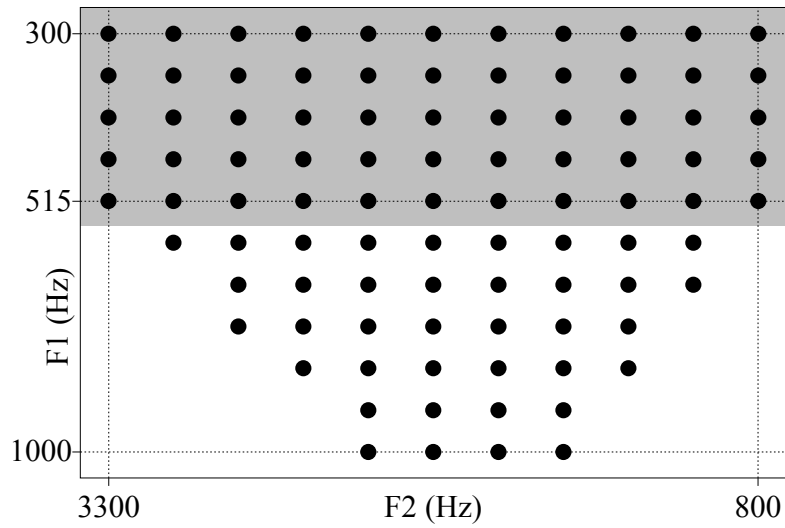
Figure 6: Experiment 3: the sampling of the F1-F2 stimulus space. The 55 F1-F2 pairs in the upper grey region were synthesized with two F3 values, two durations, and three trajectory types. The remaining F1-F2 pairs from the lower region were synthesized with one F3 value, one duration, and one trajectory type (level).
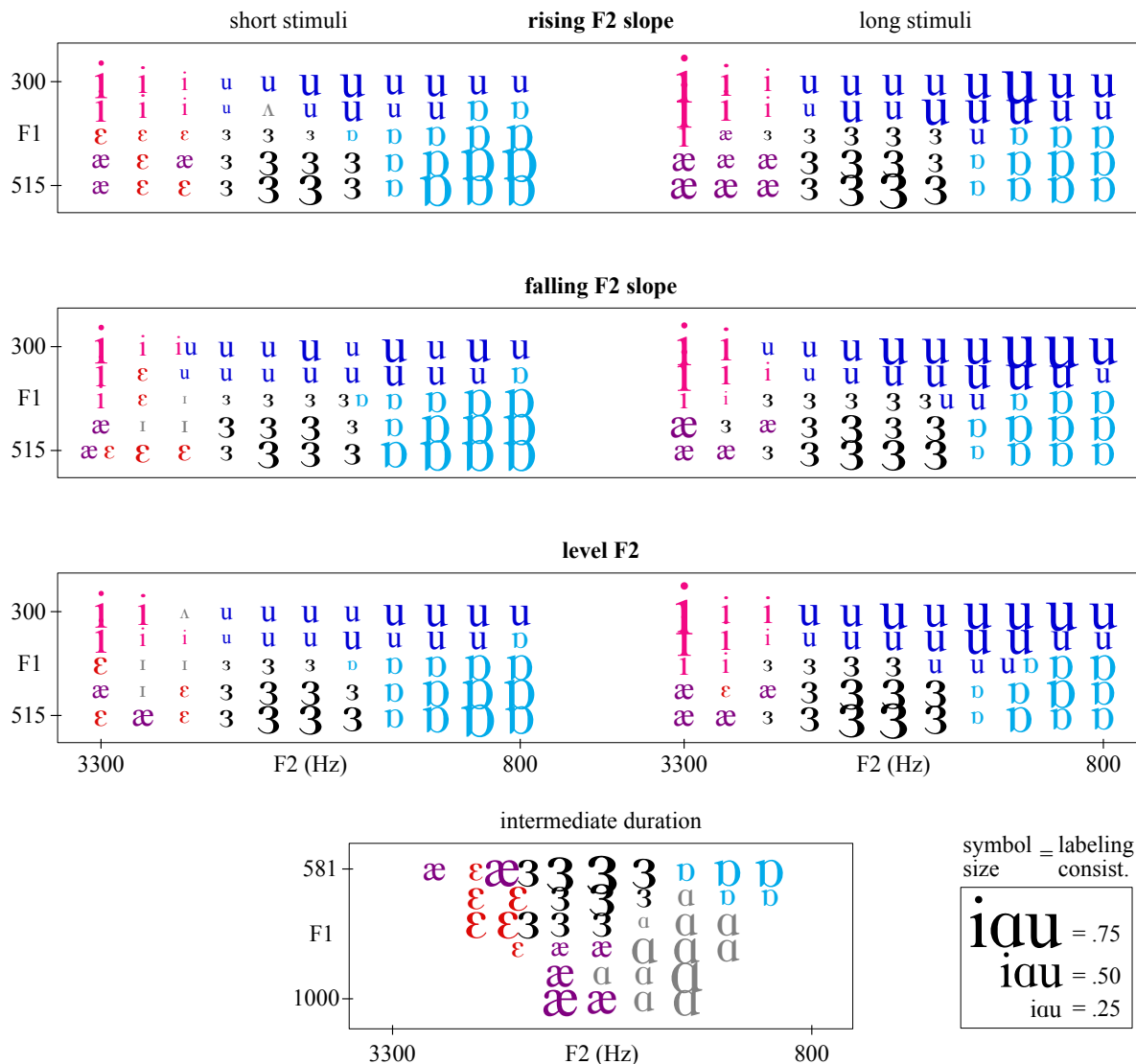
Figure 7: Experiment 3: response categories that were most often chosen for each stimulus (pooled across two different F3 values). For each stimulus, the label that was given by the majority of participants is plotted: the larger the symbol the more participants chose that label (in case of a tie both labels are plotted). The legend in the bottom right corner shows the correspondence between symbol size and the between-subjects labeling consistency. The F1 and F2 axes indicate formant values at the mid-point of the stimulus. Recall that for F1 greater than 515 Hz, i.e. the lower part of the vowel space, the stimuli all had level F2, all had one (intermediate) duration value, and one F3 value. Visual comparison of the two top graphs shows that for stimuli with falling F2 slope (second graph from top) there are more and larger back-vowel responses than for stimuli with rising F2 slope (top graph).
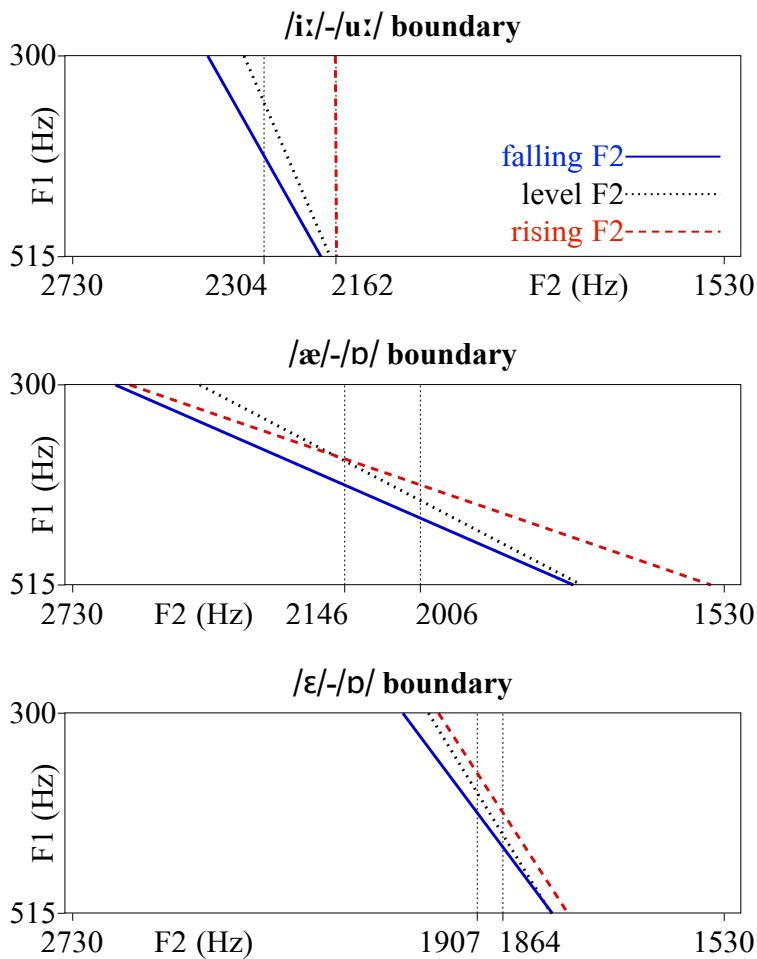
Figure 8: Experiment 3: perceptual front-back phoneme boundaries in the two-dimensional F1-F2 space, split into separate graphs for each vowel contrast. Boundaries are shown for every F2 slope type separately, coded by color and line-type (blue solid line = boundary for falling F2 slope; black dotted line = boundary for level F2; red dashed line = boundary for rising F2 slope). The boundaries were obtained from the logistic regression coefficients β_1 and β_2 for the regression factors midpoint F1 and midpoint F2, using the formula $x = -\frac{\beta_0 + \beta_1 y}{\beta_2}$, where x represents midpoint F1 and y represents midpoint F2 values. The x-axis also shows, for each vowel contrast, the numeric value in Hz (average across participants) for the boundaries found for falling F2 slope (value to the left) and those found for rising F2 slope (value to the right). Note that the boundaries shown in this Figure were computed from all the responses by individual participants, while Figure 7 plots only the winning labels, i.e. the response label that was given by the majority of participants for each stimulus.

Compared F2 slope types	Mean difference in boundary location along F2 (ERB)	95% CI of the difference	p
rising vs. level	-0.159	-0.228..-0.089	< .001
level vs. falling	-0.205	-0.292..-0.118	< .001
rising vs. falling	-0.364	-0.460..-0.268	< .001

Table 1: Pairwise comparisons of boundary locations across the three slope types; averaged over the 42 young listeners in Experiment 2.

Compared F2 slope types	Mean difference in boundary location along F2 (ERB)	95% CI of the difference	p
rising vs. level	-0.123	-0.230..-0.016	.026
level vs. falling	-0.203	-0.373..-0.033	.021
rising vs. falling	-0.326	-0.496..-0.155	.001

Table 2: Pairwise comparisons of boundary locations across the three slope types; averaged over 14 young and 12 older listeners in Experiment 2.

Compared vowel contrasts	Mean difference in boundary location along F2 (ERB)	95% CI of the difference	p
/ɛ/-/ɒ/ vs. /æ/-/ɒ/	-0.856	-1.331..-0.381,	.001
/æ/-/ɒ/ vs. /iː/-/uː/	-0.633	-1.213..-0.053	.033
/ɛ/-/ɒ/ vs. /iː/-/uː/	-1.489	-2.042..-0.936	< .001

Table 3: Pairwise comparisons of boundary locations across three vowel contrasts; averaged over the 42 young listeners in Experiment 3.

Appendix

Table A1. The 16 target words from Experiment 1. The voicing and place of articulation of the final consonant was systematically varied to cover both labial and coronal place of articulation as well as both values for consonant voicing.

	labial	*coronal*
voiceless	seep – soup	lease – loose
	leap – loop	heat – hoot
voiced	beam - boom	feed – food
	team – tomb	mean – moon

Table A2. F2-F3 combinations in the stimulus sets in Experiment 2 and Experiment 3.

	Defined F2 (Hz)	Defined F3 (Hz)	Actual F2 (Hz)	Actual F3 (Hz)
Experiment 2	1800	2700	1800	2700
	1896		1896	2700
	1998		1998	2700
	2105		2105	2700
	2217		2217	2700
	2336		2336	2700
	2461		2461	2700
	2593		2593	2700
	2732		2700	2732
	2879		2700	2879
	3035		2700	3035
	3200		2700	3200
Experiment 3	800	2200	800	2200
	931		931	2200
	1079		1079	2200
	1245		1245	2200
	1435		1435	2200
	1650		1650	2200
	1895		1895	2200
	2175		2175	2200
	2497		2200	2497
	2869		2200	2869
	3300		2200	3300
	800	2800	800	2800
	931		931	2800
	1079		1079	2800
	1245		1245	2800
	1435		1435	2800
	1650		1650	2800
	1895		1895	2800
	2175		2175	2800
	2497		2497	2800
	2869		2800	2869
	3300		2800	3300