2013

Asymmetries in English Vowel Perception Mirror Compression Effects

Jonah Katz
West Virginia University, katzlinguist@gmail.com

Asymmetries in English vowel perception mirror compression effects

Jonah Katz
University of California, Berkeley
Running head: Asymmetries in English vowel perception

Department of Linguistics
University of California
1223 Dwinelle Hall
Berkeley, CA 94720
United States
+1 (617) – 448 – 3598
katzlinguist@gmail.com

Abstract A series of vowel-identification experiments using gated consonant stimuli shows that English listeners are capable of recovering the vocalic context in which a consonant appears from information contained in the consonant alone. This is true for most consonants tested, including liquids, nasals, and stops in onset and coda position. Positional asymmetries in vowel sensitivity go in opposite directions for liquids (coda sensitivity > onset) and stops (onset > coda). Nasals pattern with liquids in terms of vowel sensitivity from consonant steady states alone, but pattern more closely with stops when portions outside the steady-state are taken into account. It is argued that these asymmetries are related to patterns of cluster-driven vowel compression (also called 'compensatory shortening') in speech production.

1 Introduction

This paper reports on an experiment concerning the identification of vowels from cues contained in adjacent consonants in English. The results have implications for theories of coarticulation, temporal coordination, and the influence of speech perception on speech production. The first finding is that subjects' ability to identify a vowel from hearing an adjacent consonant is not limited to obstruents, which have been the sole focus of earlier studies, but extends to nasal and sonorant consonants as well. The second finding is that contexts where an English vowel is independently known to be subject to a greater range and magnitude of compression ('compensatory shortening') effects (Lindblom & Rapp 1973, Munhall *et al*. 1992) also tend to contain more perceptual information about that vowel. The results suggest that properties of speech perception affect timing patterns in speech production in a fairly intricate way.

Most theories of language assume that at some level of mental representation, speech sounds are represented as temporally discrete, categorical entities such as phonemes. This foundational assumption is at odds with the fact that acoustic/perceptual information about speech sounds is distributed across the speech signal in ways that do not allow a straightforward mapping to discrete temporal segments (Liberman *et al.* 1967). This problem, referred to as coarticulation, has been noted since at least Menzerath & Lacerda's (1933) study, and has occupied a central place in the study of phonetics in the following decades (e.g. Joos 1948, Kozhevnikov & Chistovich 1965). Coproduction theories (e.g. Browman & Goldstein 1986) expand our understanding of coarticulation to incorporate the idea that adjacent articulatory gestures alter one another because they are coproduced with one another (that is, phonemes and gestures associated with them overlap). This paper is specifically concerned with coarticulation between vowels and consonants, where each of the segments exerts a coarticulatory effect on the other, changing the distribution of spectral energy in both segments (Delattre *et al.* 1955, Lindblom 1963, Zue 1976, Fowler 1994, Repp & Lin 1989).

Many studies have shown that listeners are capable of using this spectral variation alone to identify an adjacent segment at levels above chance, even when that segment is not present in the auditory stimulus. In English, subjects identify vowels at a level above chance from both preceding and following voiceless stops (Winitz *et al*. 1972). They also perform above chance with whispered transients, not including frication, from preceding voiced stops (Repp & Lin 1989). Subjects identify vowels at a level above chance from preceding (Yeni-Komshian & Soli 1981) and following (Whalen 1983) sibilant fricatives, both voiced and voiceless. Whalen reports that subjects are above chance at discriminating rounding contrasts and height contrasts. Nine of the ten subjects had higher percent correct for roundness than for height. Ohde & Sharf (1977) report that accuracy is greater for onset stops than coda stops. Some of these results have been reproduced in Swedish (Krull 1990), French (Bonneau 2000) and Dutch (Smits *et al*. 2003, Warner *et al*. 2005).

None of these studies examined the identification of vowels from adjacent non-obstruent consonants. The phenomenon of coarticulation is quite general; Ladefoged & Johnson (2010, p. 68), for instance, suggest that American English approximants are coarticulated with a following vowel. For this reason, liquids and nasals should contain perceptual information about an adjacent vowel; this basic prediction is unconfirmed to date.

Vowel identification from different manners of consonant is particularly interesting because asymmetries in the obscuring of vowels by adjacent consonants have been offered as an explanation for certain asymmetries in the timing of speech production related to vowel compression (Marin & Pouplier 2010, Katz 2012). Compression is the tendency for segments to be shorter when there are more of them in a string. Katz (2012) shows that English vowels followed by a liquid-voiced obstruent cluster (e.g. [dɪlb]) are shorter than those followed by a singleton liquid ([dɪl]); this *cluster-driven compression* does not obtain for similar pairs containing nasals ([dɪn]-[dɪnz]) or obstruents ([dɪs]-[dɪsp]) in place of the liquid. This pattern is different from what is observed with consonants preceding vowels, where both liquids ([lɪd]-[glɪd]) and nasals ([nɪd]-[snɪd]) induce cluster-driven vowel compression; obstruents ([p^hɪd]-[spɪd]) may as well, depending on how one treats the rather different phonetic intervals in an aspirated singleton vs. an unaspirated cluster. In addition, cluster-driven vowel compression is greater adjacent to coda than onset liquids, especially /r/.

Marin & Pouplier (2010) and Katz (2012) both propose that some asymmetries in cluster-driven compression relate to differences in the perceptual properties of various consonants in onset and coda position. In particular, the temporally- and gesturally-reduced tongue-tip component in English coda /l/ (Sproat & Fujimura 1993) may result in less obscuring of vowel contrasts than its onset counterpart. F1 and F2 frequencies during /l/ will be informative because /l/ includes a back-mid tongue-body constriction whose gestural dynamics will be affected by the adjacent vowel; the presence of a tongue-tip constriction overlapped with this gesture should have the effect of compressing the range of possible F1 and F2 frequencies. Katz (2012) also suggests that the greater temporal and gestural extent of the velic opening gesture in English coda nasals (Krakow 1999) may obscure the quality of a preceding vowel *more* than in onset position. The resistance of vowels to incremental compression before stops may be due to an asymmetry in gestural overlap between vowels and stops in onset and coda position. The principal acoustic signature of a stop capable of carrying vowel information is its burst; in VCV utterances, effects on formant trajectories associated with the second vowel are present well before stop closure is achieved (Öhman 1966). Browman & Goldstein (1988, 1990) show that in American English the release of an onset stop generally occurs at a point after the vowel gesture has begun, while in coda position stop closure is achieved near the offset of the preceding vowel gesture, meaning the release will not be as extensively colored by overlap with that vowel. The perceptual asymmetries brought about by these gestural asymmetries, by hypothesis, lead in turn to duration asymmetries: in this view, the availability of additional vowel-shortening driven by the presence of more consonants in a syllable is conditioned by how much vowel information is contained in the surrounding context. Vowels shorten more when there is more information about their features in the adjacent segment (coda liquids, onset nasals) than when there is less information (onset liquids, coda nasals and obstruents).

These explanations derive from theoretical considerations of how the different temporal and gestural properties of various consonants in onset and coda position ought to affect the acoustics and perception of an adjacent vowel. It remains to be seen whether those hypotheses are actually borne out by perceptual data. The most straightforward hypothesis is that increased 'vowel information' should correspond to increased sensitivity to vowel contrasts, i.e., ability to tell vowels apart and identify them. It is therefore predicted that listeners should show an asymmetry in sensitivity to vowel contrasts based on onset and coda liquids (sensitivity to vowel contrasts from coda liquids greater than that from onset liquids). They should show the opposite pattern conditioned by nasals and stops (onset > coda). The prediction regarding stops has already been confirmed by Ohde & Sharf (1977); the remainder are untested as yet.

One additional complication in studying vowel compression concerns the effect of consonant duration. Several previous models suggest that compression arises from the conflict between constraints on the duration of syllables or rimes, and constraints on the duration of segments (Fujimura 1987, Flemming 2001). This means that the way in which a consonant affects vowel compression is due to its inherent duration, not its internal spectral properties. Katz (2012) extends this general type of model, in view of compression asymmetries, to incorporate differences in the degree to which a consonant obscures the acoustic traces of a vowel in addition to differences in duration between consonants. To show that consonantal effects on vowel perceptibility brought about by such manner-related spectral properties are indeed important in characterizing compression phenomena, it must be shown that these differences make an independent contribution to vowel perceptibility, and are not merely the by-product of differences in duration between consonants. The analysis reported here makes use of variability in consonant duration to show that consonant manner differences make a contribution to explaining variance in vowel compression that goes above and beyond the contribution of consonant duration alone.

The current paper, then, has two main goals. First, to extend the finding that vowels may be identified from adjacent obstruents alone to nasals and liquids. Second, to explore the hypothesis that independently-attested asymmetries in cluster-driven vowel compression may be explained with regard to the perceptual effects of different kinds of consonants upon adjacent vowels. These goals were pursued using a forward- and backward-gating paradigm, which asks listeners to implicitly identify a vowel on the basis of fragments of the acoustic stream that do not contain the vowel 'proper'. This paradigm is described in the next section.

2 Methods

2.1 Materials

Materials consisted of word-pairs differing only in their vowels. The vowel pairs tested are /e/-/o/, /ɑ/-/u/, and /i/-/e/. The idea is to examine a small number of vowel contrasts that are representative of the types examined in prior studies: a pair differing along the backness/rounding dimension, which is generally found to be the most discriminable type of contrast; a pair differing in more than one step along the height dimension (high vs. low) as well as rounding, which should be roughly comparable to the backness/rounding contrast; and a pair differing in only one step (high vs. mid) on the height dimension, which is generally found to be the least discriminable type of contrast (Whalen 1983, Parker & Diehl 1984, Repp & Lin 1989).

Note that the vowels phonemically transcribed as /e/ and /o/ are actually realized as diphthongs in the dialect examined here. Some of the vowels transcribed as /ɑ/ may be realized as /ɔ/ by speakers who make this distinction; the two speakers in this study, however, did not consistently produce an /ɑ/-/ɔ/ distinction (see section 3.1 and figure 3).

Materials consist of all combinations of the relevant vowels with the consonants {r, l, n, p} in onset or coda position, except for /ur/, which was excluded due to its dubious phonotactic status in American English. The full set of materials is shown in Table 1.

[TABLE 1 HERE]

2.2 Speakers

Two native speakers of American English from eastern Massachusetts (1 female, 1 male) were recorded producing three repetitions of each lexical item in the carrier sentence ‘I bet ___ is the answer’. One token of each item from each speaker was selected for inclusion in the experiment. For each lexical item, the selected token was the one with consonant durations closest to each subject’s mean for the item. Almost all of the coda condition words were preceded by a weakly released /t/ in *bet*, followed by silence and a glottal stop; these were excluded from the excised stimuli. Although the speakers are from eastern New England, neither of them displays such regional characteristics as /r/-dropping (/r/ is clearly visible in the spectrogram of /er/ in figure 1 below) or intrusive /r/.

2.3 Segmentation

The selected tokens were segmented into two regions, referred to as ‘consonant’ and ‘gates’. Segmentation was done purely on the basis of particular landmarks within each token, which differed between consonants: F3 for /r/, presence of aspiration for [p^h], F1 and F2 for other consonants. It should be noted for future research that post-hoc inspection suggests post-vocalic /l/ may have been more reliably segmented at a small rise in F3 following the vowel. The general strategy for any acoustic parameter *P* was to mark the consonant boundary where *P* begins to slope noticeably from values associated with the consonant towards values associated with the vowel, except for onset [p^h], where the segmentation was done immediately following the burst and any frication; and coda [p], where the segmentation was done on the basis of cessation of acoustic energy. Figure 1 shows the segmentation strategy for creating stimuli from each of the consonants under investigation in onset and coda positions, illustrated here with the vowel /e/.

[FIGURE 1 HERE]

This procedure results in 'consonant' regions defined purely by acoustic targets, which may or may not correspond to mental units relevant to motor planning, speech perception and the phonological notion *segment*. The basic question addressed by this experiment, then, is whether the relatively static portion of a consonant, where acoustic parameters are not obviously moving to or from those of an adjacent vowel, still contains information about that vowel, and whether the amount of information differs between consonants and between positions. The segmentation was changed for /p/, because it does not have any portion that can be characterized as (relatively) acoustically static except for silence. For this reason, the burst and any following frication were included in the consonant region. Aspiration was not included in the consonant region for onset /p/, although this is not meant to suggest that aspiration is 'part of the consonant' or 'part of the vowel'. It was excluded in order to have comparable stimuli for onset and coda position, and to generate a more conservative test of the hypothesis that onset stops condition greater vowel sensitivity than coda stops.

Comparing the static portion as defined above across various consonants will not reveal much about how consonants obscure vowels if the resulting segments differ radically from those relevant to speech production and perception. Most notably, these criteria do not correspond straightforwardly to gestural landmarks; an alternative would have been to define a gestural notion of consonant and then attempt to find acoustic landmarks that correspond to the relevant gestural ones. For instance, consonant boundaries might be marked in a position corresponding to release in onset position and closure attainment in coda position. There are several reasons why this approach is not taken here.

First, it is not always straightforward to identify the relevant gestural landmarks. For instance, release in onset /l/ and apical closure in coda /l/ may be indicated by small amounts of noise in the spectrogram (this signature appears to be present in both tokens in Figure 1). In some tokens, however, this noise is not present; it would be exceedingly difficult to mark a release or closure point in these tokens. The acoustic criteria used here, shift in F1 and F2 frequencies, are present in all tokens. Second, theories of gestural phasing between consonants and vowels such as Browman & Goldstein's (1988, 1990) predict that some articulatory landmark in a singleton consonant gesture will bear a consistent temporal relationship to some point in the adjacent vowel gesture, but this consonantal landmark is either the onset or the target (closure attainment), *not* the release. So there is no reason to believe that segmenting the acoustic stream in this way would result in a more coherent or temporally-stable acoustic notion of 'consonant'. Finally, the gestural criteria would be less conservative than the acoustic criteria used here with regard to the experimental hypotheses. One of those hypotheses is that liquids and nasals contain information about an adjacent vowel. Using the gestural criteria mentioned above for /l/ and onset /n/ would result in more of the consonant being included in stimuli; if an effect is found with the current stimuli, therefore, it would hold *a fortiori* for stimuli demarcated according to the gestural criteria. The same logic applies to onset vs. coda distinctions. Coda /l/ is predicted to condition greater sensitivity to vowel contrasts than onset /l/; the gestural criteria would include more of coda /l/ in the stimuli (it would also include more of onset /l/, but to a lesser extent). Compared to the acoustic criteria, the gestural criteria would include a few more milliseconds of onset /n/ and any period of perseveratory closure voicing for coda /p/; neither of these small differences are likely to have a large effect on vowel perception, although a further effect stemming from this initial boundary difference is discussed in section 4.

The experiments included four versions of each selected token, referred to as *gates*, containing successively more of the acoustic material from the original tokens. The shortest fragments used in the experiment, referred to as gate 0, contain only the marked consonant portion as described above. Three succeeding gates incrementally added 20-27 ms. of the original token; for at least gates one and two, these portions generally consisted of the acoustic transition between consonants and preceding or following vowels, depending on whether the consonants were in coda or onset position.

All stimuli (across consonants) within each group consisting of a combination of vowel-pair and syllable position have the same gate duration, but the gate duration varies slightly between these groupings. Additionally, the stimuli were truncated at the zero-crossing closest to the chosen gate duration; this resulted in differences of up to 2 ms in gate duration between stimuli in the same condition. Some of the stimuli that included stops were segmented, and their closure portions run through a high-pass filter, in order to remove a noticeable electrical buzz from the recording.

Impressionistically, the sounds were rather easy to identify by the second gate. Short pilot studies were conducted for each vowel pair using gates 0, 1, and 2. The results indicated that most subjects obtained 80-90% accuracy by the second gate. At gate 0, accuracy ranged from slightly below chance to around 70%, depending on subject and stimulus. Subjects performed around chance (50% correct) at all gates for the coda consonant /i/-/e/ condition; this is presumably because /e/ includes an offglide that is nearly identical to [i]. This condition was dropped from the final study.

2.4 Design

There were five groups, each comprising a single Vowel Pair with either onsets or codas: /a/-/u/ onsets, /ɑ/-/u/ codas, /e/-/o/ onsets, /e/-/o/ codas, and /i/-/e/ onsets. The experimental stimuli within a Vowel Pair group were all the selected utterances for the particular vowel pair, from each of the two speakers, at each of the first three gates, for a total of 48 experimental stimuli per group, except for the /ɑ/-/u/ coda group, which had 36 due to the exclusion of /r/. Column 2 of Table 2 shows the number of experimental stimuli for each Vowel Pair group. The total number of stimuli in each Vowel Pair group (column 4 of Table 2) further differed because each included several additional stimuli that form part of a larger project examining other issues in consonantal timing. These additional stimuli were also monosyllables with the same vowel as the Group they were part of, but differed in consonants; they were also presented in three different gates. In the present experiment, the additional stimuli can thus be regarded as filler items.

[TABLE 2 HERE]

Each subject was assigned to one Vowel Pair group, thus each listener heard all and only the stimuli for one vowel pair in onset or coda. For each Vowel Pair group, 15 repetitions of each stimulus were obtained by randomising all stimuli 15 times in blocks. Thus the total number of trials for a given Vowel Pair group was 15 x the total number of stimuli for that group, that is, 15 x 48 = 720 for the two coda groups, and 15 x 72 = 1080 for the three onset groups.

A training session was also prepared, comprising 1 repetition of each lexical item for the particular Vowel Pair group from each of the two speakers at gate 3, randomized separately for each subject; the number of experimental trials in the training session is thus 12 for Group a-u coda, 16 for the other Groups.

2.5 Procedure

Subjects were tested in the Behavioral Research Lab at MIT, with up to 10 subjects simultaneously participating at workstations separated by dividers. Printed instructions informed them that on each trial they would hear a word with the beginning or end removed, and would have to choose which of two words they had heard part of. They were asked to respond as fast as possible and told that they would have a chance to take breaks and that they could stop for any reason if they wished to. The training session was presented first, without feedback, followed immediately by the experimental session.

The experiments were implemented using the Psyscope software, version B53. Audio files were played to the subjects over Koss UR50 headphones. As each file was played, a choice of two words appeared, one on each side of the screen; subjects used a left and right button to identify the corresponding word as the one they had heard part of. Pairs of words were not counterbalanced on the screen, as any advantage from subjects preferring the left or right button (or visual field) will only show up in the results as a shift in bias, not sensitivity. Subjects were given 1 second to respond; after this, the message *Timeout!* appeared at the center of the screen for 300 ms. The 1-second response window was used both to make the task more difficult and to limit the duration of the experiment. Subjects were fully capable of responding within one second; they timed out on about 4% of all trials. Subjects were given the option of taking a break after each block except the training session.

All word choices were existing lexical items of English; this sometimes required an orthographic consonant that was not present at all in the auditory stimulus. For instance, subjects were played a fragment of /op/ and asked whether it was *cape* or *cope*, despite the fact that there was no hint of a /k/ in the recording. Wherever possible, the choice of this ‘imaginary consonant’ was held constant across target consonants within each vowel pair (e.g. *care-core, kale-coal, cane-cone, cape-cope*); in a few cases this was not possible. The choice to use these consonants was made in part because using the orthography of (these particular) existing English words unambiguously encodes the intended phonetic string; orthographic representations of nonsense words (e.g. ‘ole’, ‘ain’) could be ambiguous in this regard. Because most of the coda-condition word pairs required such a consonant, the choice was made to simply include it in the orthography for all pairs, to make the task more uniform.

Although the use of these consonants solves the problem of phonetic ambiguity, it could potentially create another problem: perceptual compensation for coarticulation. That is, subjects may expect the vowels under examination to be coarticulated with the preceding /k/, and may respond in an aberrant manner when they find this not to be the case. Such an effect would only affect bias, however, not sensitivity: compensation for coarticulation shifts the boundary criterion between two phonemes, which in the current study would only affect the probability of one response relative to the other. The statistical models presented in section 3.2 find no evidence for such a bias effect.

Word pairs were not balanced for frequency; this is likely impossible given the nature of the task, and the statistical model of the results can correct for frequency effects by separating the effects of bias from the effects of similarity. Lexical bias, for instance, might lead subjects to respond with *knee* more often than *neigh*, but this would show up in the statistical model only as increased bias to respond /i/ in the context of /n/, not as increased sensitivity to the /i/-/e/ contrast.

2.6 Listeners

For the /e/-/o/ onset Vowel Pair group, 15 subjects were tested. For /ɑ/-/u/ coda, 10 were tested. For the other three groups, 11 subjects were tested. The total number of subjects was 58 (34 female, 24 male), with a mean age of 30 years. All reported being native speakers of American English who had never been diagnosed with any speech, hearing, or reading disorders. All subjects were compensated for their time.

2.7 Data analysis

The results were analyzed with a logit mixed effects model, implemented in the lme4 package (Bates 2007) for R. A logit model expresses how the likelihood of some binary response, e.g. ‘right button’, varies depending on properties of the stimulus. A mixed model allows us to analyze data with more than one *random variable*, variables whose levels are sampled from a larger population of possible levels, such as ‘word identity’ or ‘subject identity’. Excellent tutorials by Jaeger (2008), Quené & van den Bergh (2008), and Janda *et al*. (2010) describe and illustrate these models, and explain why modeling random effects is important.

The models described here attempt to distinguish between bias and sensitivity in a binary choice task. For instance, subjects may generally respond /i/ more often than /e/ regardless of what type of stimulus they are played, but this difference may be larger when the stimulus is extracted from a word with /i/ than a word with /e/. The logit mixed effects model checks whether the likelihood (in log odds, or *logits*) of responding /i/ is significantly higher when the stimulus originally contained /i/ than when it did not. In this model, the likelihood of responding /i/ to an /e/ stimulus (a *false alarm*) is related (though not equivalent) to response bias, while the difference between the likelihood of responding /i/ to an /e/ stimulus and the likelihood of responding /i/ to an /i/ stimulus (a *hit*) measures sensitivity. For instance, if the likelihood of responding /i/ to an /e/ stimulus is closer to 50% than the probability of responding /i/ to an /i/ stimulus is, there is evidence of a bias towards /i/. If the likelihood of responding /i/ to an /i/ stimulus is significantly higher than that of responding /i/ to an /e/ stimulus, then there is evidence of sensitivity to the /i/-/e/ contrast. If in contrast the likelihoods of responding /i/ to /i/ and to /e/ stimuli do not differ, then there is no evidence of sensitivity to the contrast, i.e. listeners cannot detect the vowel from the available information. Furthermore, if the difference between the likelihood of responding /i/ to /i/ and /i/ to /e/ is greater in, e.g., /pV/ than /Vp/ contexts, then there is evidence of greater sensitivity to the /i/-/e/ contrast conditioned by onset than coda consonants. In the presentation of results, the bias-related terms are listed in the description of the model for the sake of completeness, but are otherwise ignored. This is because the hypotheses being tested pertain to sensitivity to vowel contrasts, not to bias. The current experiment is in any case not a suitable design for a systematic study of factors affecting bias; such a study would require careful control of lexical and phonotactic frequencies, handedness, orthography, and possibly other factors.

The dependent variable here is one of two possible vowel responses, which differ by condition. This variable was coded as 1 if the subject pushed the button on the right, 0 otherwise. Random effects are speaker, lexical item, and listener. The model includes a fixed effect for each pair of vowels, each consonant in each syllabic position, and the interactions between them. These effects, which track false alarms, correspond to 'baseline' bias-related parameters for each contrast examined in the experiment. Separate fixed effects assess how the likelihood of the response variable changes depending on whether the presented stimulus was originally recorded with the vowel represented by the response choice on the right side of the screen, on which consonants are present, and on the total duration of those consonants. These effects, encoding differences between false alarms and hits, are sensitivity parameters, and they are the primary results of interest. Further fixed effects included whether or not each trial followed an error on the immediately preceding trial, whether it followed a timeout, log duration of the consonant stimulus, and the effect of these three parameters on sensitivity. Finally, the fixed effects of interest were tested to see if they vary significantly across levels of the random variables, through the use of random slopes. This step is crucial, because it allows us to express the main effects of bias and sensitivity in various contexts while taking into account variability between listeners, speakers, and individual words.

3 Results

3.1 Acoustic properties of the materials

The materials are described above as placing various consonants adjacent to the *same* vowel. This is obviously an idealization; the pronunciation of a vowel is likely to differ on the basis of which consonant is adjacent to it and whether that consonant precedes or follows. These differences may themselves be relevant to explaining any perceptual asymmetries that arise in the experiment. As a preliminary to the perception experiment, then, it is desirable to describe some acoustic asymmetries present in the stimuli. This section focuses on the offglide for /e/ and /o/ stimuli, which is clearly different before liquids than it is before other segments; and the /ɑ/ stimuli, whose vowels are phonemically distinct for some speakers.

Figure 2 shows F1 and F2 frequencies at the temporal midpoint of the marked vowel region and at the midpoint of the third gate (in the vicinity of the offglide) for /e/ and /o/ preceding various consonants. Values were extracted by script using the Praat formant tracker with the following settings: 5 kHz maximum formant (for male; 5.5 for female), 5 formants, 5 ms. window, 30 dB dynamic range. A few measurements from each vowel were checked by hand and it was confirmed that the script was extracting accurate frequency values.

[FIGURE 2 HERE]

The most obvious pattern here is that the temporal middle and end of the vowel characterized as /e/ have substantially higher F1 and lower F2 before liquids than before /n/ and /p/. These differences place pre-liquid /e/ and /o/ tokens closer to each other. Table 3 shows the Euclidean distance between /e/ and /o/ in F1 and F2 space measured at the vowel midpoint and third gate for each speaker. Although these measurements ignore F3, which plays some role in vowel contrasts, it should not strongly affect the particular contrasts examined here. While differences between /e/ and /o/ are comparable across consonants at the vowel midpoint, they are decidedly smaller for /l/ and /r/ than the other consonants towards the end of the vowel.

[TABLE 3 HERE]

Figure 3 shows both speakers' formant frequencies for /ɑ/ and and /u/. Several of the onset-condition words here (and possibly *call* as well) are expected to be produced with /ɔ/ in dialects that feature a contrast between /ɑ/ and /ɔ/. sign of producing an /ɑ/-/ɔ/ distinction (at least with regard to F1 and F2 frequencies). However, despite extensive spread in Speaker 2's productions of these vowels, neither speaker's data suggest they distinguish /ɑ/ from /ɔ/, at least with regard to F1 and F2 frequencies. These results accord with the author's judgment: the putative /ɔ/ vowels do not sound sufficiently distinct from their putative /ɑ/ counterparts to be assigned a different phonemic symbol.

[FIGURE 3 HERE]

There is no clear pattern in the mid-vowel Euclidian /ɑ/-/u/ distances, shown below in table 4. A possible overall trend for distances to be greater in the coda than onset condition is marred by the /n/ context, where, for speaker 1, the distance is less in the coda than in the onset.

[TABLE 4 HERE]

If the perception study does uncover differences between consonants and/or syllable positions, such differences may be due to spectral prominences internal to consonants or due to some other properties of consonants that make these prominences difficult to recover. As a preliminary to such issues, figure 4 presents the frequencies of the first two spectral prominences internal to the consonants. For sonorants (including /n/), frequencies were extracted by script from the temporal midpoint of the marked consonant region, using the Praat formant tracker with the settings listed above. For /p/, frequencies were extracted from the earliest point following the burst where the formant tracker identified two prominences in the region of 0-2500 Hz; this point was generally just about at the end of the burst itself, near the onset of frication. In about half of the /p/ tokens, the formant tracker could not reliably identify spectral prominences. For these cases, the entire burst and frication portion (not including aspiration) was analyzed by hand using an FFT spectrum with a Gaussian window of a duration determined by the length of the noise portion: the first two identifiable spectral prominences in the signal that corresponded to plausible formant frequencies (based on other tokens) were recorded for these tokens.

[FIGURE 4 HERE]

There are no striking asymmetries between consonants visible in these formant spaces. /n/ seems to have slightly less distinct formant frequencies adjacent to different vowels than the other consonants do, but the effect is not entirely consistent. Other differences between consonants in the magnitude of formant differences between vowel contexts vary across vowel and speaker.

In terms of asymmetries between various consonants in onset and coda positions, there are again very few systematic patterns in these materials. The formant frequencies indicate that, if anything, there is a small tendency for /p/ to be more distinct (in terms of changing with the adjacent vowel) in coda than onset position. Speaker 2 shows a similar pattern for /n/, but speaker 1 displays slightly more distinct /n/ in onset than coda position. Speaker 2 displays slightly more distinct coda than onset /l/, but speaker 1 shows the opposite pattern if anything. For /r/, neither subject shows a clear asymmetry between onset and coda position.

In addition to spectral properties, the duration of the consonants was also analyzed. This is because the statistical model of the perceptual data incorporates differences in consonant duration in order to isolate the duration-independent effect of consonant manner. Table 5 shows the measured consonant duration for each token present in the experiment, as well as the gate durations for each speaker used in the non zero-gate conditions.

[TABLE 5 HERE]

3.2 Identification experiment

Figure 5 shows two measures of sensitivity to the speaker's intended vowel contrasts in the context of various consonants in onset and coda position at the zero gate. Figure 5a uses the signal detection theoretic measure d' to summarize the observed distribution of sensitivity values for various consonants in onset and coda position. Figure 5b shows how the statistical model fits sensitivity parameters to this data, generalizing across subjects and lexical items and factoring out covariates.

[FIGURE 5 HERE]

Several patterns are noticeable here. First, subjects answer correctly more than half of the time (values above 0) for all consonants. This suggests that subjects can identify vowels from adjacent non-obstruent consonants alone. Previous work has shown that subjects generally perform above chance with obstruents; here, sensitivity seems to be even higher for many of the non-obstruent consonants than for the obstruents.

The statistical model of the zero-gate data examines differences between consonants in vowel sensitivity while factoring out effects due to consonant duration. Recall that our hypothesis predicts that consonant duration (by way of duration-trading) and manner (by way of differential vowel masking) should have independent effects on vowel compression; it is thus crucial to ensure that differences in vowel sensitivity across manners are not simply due to durational differences. The data indicate that this is not a concern: the effect of (natural logarithm of) consonant duration on vowel sensitivity is not even significant when manner differences are taken into account. The full model is shown in table 6.

[TABLE 6 HERE]

For onset consonants, vowel sensitivity in the context of /p/ is significantly greater than zero (chance). Because it is not possible to tell from this model whether sensitivity in the context of onset /n/ is significantly different from chance, the model was reparameterized post-hoc to test this comparison. The effect is significant: $\beta = 1.17$, $Z = 2.10$, $p = 0.036$. The significance of this contrast in conjunction with the significant effects in rows 2-4 of table 6 gives us the following partial ordering for onset-conditioned sensitivity: chance < /n/ < {/p/, /r/} < /l/.

Vowel sensitivity is significantly lower in the context of coda /p/ than onset /p/. This pattern is reversed for the other consonants (higher in coda than onset), resulting in significant interactions between consonant, syllable position, and sensitivity for /l/, /r/, and /n/.

The final model includes by-subject random slopes for most of the bias and sensitivity terms. This indicates that there is substantial variability between subjects in the magnitude of bias and sensitivity differences between manners of consonant. The model also includes a by-speaker random slope for general sensitivity: subjects are more sensitive to the vowel contrasts produced by the female speaker than the male one.

One further fixed sensitivity effect was significant: subjects performed significantly worse on trials immediately following an incorrect answer ($\beta = -0.23$, $Z = -2.63$, $p < 0.01$). This may be because subjects were sometimes aware that they had made an error, which distracted them on following trials.

Recall that /n/ was predicted to show the same kind of onset-coda asymmetry as /p/, as distinct from the asymmetries of /l/ and /r/. The zero-gate data contradict this hypothesis. Post-hoc analyses were conducted on the second gate, which included 40-54 ms. in addition to the consonant. The idea was to see whether there might be some property of CV and VC transitions, aspiration, or the vowels adjacent to /n/ and /p/ that would explain why the relation between their zero gate data is the opposite of what the production pattern from previous experiments would predict. This model included zero and second gate data for /p/ and /n/, summarized in figure 6.

[FIGURE 6 HERE]

Recall the unexpected effects with zero gate stimuli, shown in the left panel: /n/ conditioned greater sensitivity in coda than onset position, showing the opposite pattern from /p/. The second-gate data, shown in the right panel, suggest an explanation of why this might not matter for the purposes of compression, which affects the acoustic steady state of a vowel: although the 'consonant' portions of /p/ and /n/ as marked here show opposite patterns by syllable position, the transitions and aspiration associated with those consonants, included in gates 1 and 2, display an opposite interaction between consonant and syllable position. With 40-54 ms of this material included, as shown in the right panel, the difference between coda /n/ and /p/ shrinks to almost nothing. Statistical results are shown in table 7.

[TABLE 7 HERE]

Vowel sensitivity at gate zero in the context of coda /p/ is not significantly different from zero. Zero-gate sensitivity is significantly higher in the context of coda /n/ than coda /p/. This is all familiar from the previous model. What the second model shows is that this difference disappears by the second gate, resulting in a significant interaction between consonant and gate. This whole picture is reversed in onset position. Sensitivity to /n/ is lower at the zero gate, resulting in a significant interaction between consonant and syllable position. And this pattern in turn reverses by the second gate, resulting in a significant interaction between consonant, gate, and syllable position.

4 Discusssion

One finding from the current experiment is that subjects' ability to identify a vowel from an adjacent consonant alone extends quite generally across different manners of consonant, and is not limited to obstruents. This general finding is consistent with the idea that coarticulation affects all segments (though not equally) and that listeners are capable of using that coarticulation to extract cues to segmental identity from portions of the auditory stream that would not traditionally be considered part of the relevant segments themselves. This pattern does not extend to every single consonant and contrast examined here: vowel sensitivity in the context of coda /p/ is not significantly different from chance.

A second finding is that patterns of perceptual sensitivity to vowel contrasts broadly mirror production asymmetries attested in previous studies: the current study suggests that there is more information about vowels in contexts where previous studies find those vowels to be shorter. Stops condition greater vowel sensitivity in onset than in coda position, while this asymmetry is absent or reversed for liquids. This mirrors the fact that cluster-driven vowel compression is blocked in the context of coda obstruents (Marin & Pouplier 2010, Katz 2012); by hypothesis, this is due to coda stops' greater tendency to obscure vowel contrasts, while no such syllable-position effect is present in the context of liquids. The increase in sensitivity from coda consonants is clearly larger for /r/ than for /l/, and this also mirrors the reported magnitude of the production asymmetry for these two segments (Katz 2012). Importantly, all of these effects are significant even when differences in consonant duration are taken into account in the statistical model; they are thus truly manner-dependent, and not simply due to differences in inherent duration.

Patterns for /n/ are less clear: it is predicted on the basis of production data to pattern with stops in conditioning greater sensitivity in onset than coda. The results indicated that it instead conditions greater sensitivity as a coda; closer inspection of perceptual data suggests that the prediction of more vowel obscuring by coda nasals is borne out by gates 1 and 2 rather than steady states. Although this was characterized as a post-hoc analysis, one could argue that it ought to have been the initial hypothesis: one of the fundamental differences between English onset and coda nasals is that anticipatory nasalization has a greater physical and temporal magnitude than carryover nasalization (Krakow 1999). Nasalization is known to result in decreased sensitivity to vowel contrasts, at least along the height dimension (Wright 1975); response bias for such vowels is affected by perceptual compensation for coarticulation (Krakow *et al*. 1988). If the segmentation criteria used here (F1 and F2 'elbows') track oral constrictions more closely than they do velic aperture, the primary acoustic effect of the temporal asymmetry in nasalization should occur outside of the marked consonant steady state. The data are thus fully consistent with the idea that vowel compression depends in part on the characteristics of an adjacent consonant.

It was noted in section 2.3 that acoustic landmarks corresponding to gestural ones would also have been feasible segmentation criteria for this study, and it was argued that the particular gestural criteria discussed would result in less conservative tests of the experimental hypotheses regarding vowel perception from /n/ and /l/ and positional asymmetries for /l/. It also appears that these criteria would have generated a less conservative test of the positional asymmetries between /n/ and /p/ discussed here. In particular, those gestural criteria would have resulted in the consonant boundary of coda /p/ being marked earlier, to include closure voicing; and onset /n/ being marked later, to include release. While these differences in themselves are unlikely to have a large effect on vowel perception, the consequent shifting of gates 1 and 2 may well have a larger effect, especially in the case of coda /p/, where the temporal shift would be greater. This means that, if gestural landmarks are a better basis for consonant comparison, the current procedure systematically underestimates the amount of vowel information contained in the gated regions for onset /n/ and coda /p/. The fact that the relevant interaction between consonant, position, and gate was still significant in the current study suggests that it would hold *a fortiori* with stimuli created using gestural criteria.

One possible problem with the current results is that asymmetries in vowel sensitivity may not be due to the ways in which consonants and vowels overlap, but rather due to differences in the quality of vowels themselves, i.e. if vowel targets are less distinct adjacent to some consonants than others. For instance, the offglide and possibly the nucleus in sequences like /er/ are clearly different from those in sequences like /ep/. Subjects may discriminate between /er/ and /or/ better than they do /ep/ and /op/ because /r/ carries more information about an adjacent vowel than /p/ does, or it may be because the vowels notated as /e/ and /o/ are simply more distinct preceding an /r/ than a /p/; the symbolic transcription used here ignores systematic phonetic variation. This type of confound would also be a concern when comparing onset and coda liquids.

The acoustic analysis in section 3.1, however, strongly suggests that this is not the source of the effects found here. If anything, differences in vowel quality work *against* those effects. /e/ and /o/ are generally have less distinct F1 and F2 frequencies adjacent to coda liquids than adjacent to onset liquids, for instance, but vowel sensitivity is greater with coda liquids. For the vowels notated as /ɑ/, which are quite variable in these speakers' productions and may sometimes correspond to phonological /ɔ/, no systematic differences in F1 and F2 were found that could explain the perceptual results.

All of the hypothesized explanations for asymmetries in the extent to which consonants obscure vowels given in section 1 rely on the notion of overlap. The idea is that when the number, nature, or temporal extent of consonantal gestures overlaid on vocalic gestures differs across contexts or consonants, there will be consequent differences in gestural blending or obscuring, and hence differences in the acoustic reflexes of underlying vowel gestures. These differences should be reflected in the frequency of spectral prominences as measured internal to consonants; yet the acoustic analysis found very few systematic differences in this regard.

This suggests that spectral-prominence frequencies cannot offer a systematic explanation for the perceptual differences between consonants and contexts uncovered here. Other possible explanations include the influence of higher spectral prominences, differences in resonance bandwidth or intensity, differences in the ease of extracting frequency information, and any other factor that might differ between consonants. Loudness seems to offer a promising approach to some of the asymmetries discussed here: for instance, American English /p/ has far higher rms amplitude in word-initial position than word-final position (Redford & Diehl 1999), and the tokens shown in figure 1 suggest that acoustic energy in the region of F1-3 for /l/ and /r/ is higher in coda position than it is in onset (at least relative to the vowel).

The findings here have implications more generally for phonetic theory. They suggest that patterns of fine-grained timing in speech production are sensitive to perceptual properties of the sounds being produced. In other words, correctly characterizing the temporal coordination of articulatory gestures in speech production will require us to make reference to the perceptual consequences of those gestures, in addition to their inherent articulatory properties. Although this claim has been made frequently by researchers investigating the relationship between phonology and speech perception (e.g. Silverman 1995, Wright 1996, Steriade 1997, Gordon 2001, Jun 2002), it is invoked only occasionally in studies of articulatory phenomena (e.g. Browman & Goldstein 2000, Chitoran *et al*. 2002). This paper is a converging source of evidence that the grammar of timing cannot be computed solely over articulatory representations. More generally, it offers support for the claim that phonetic knowledge is organized in terms of perceptual goals or representations (Kingston & Diehl 1994).

References

Bates, D.: lme4: An R package for fitting and analyzing linear, nonlinear and generalized linear mixed models. Software application, 2007.

Bonneau, A.: Letter to the Editor: Identification of vocalic features from French stop bursts. J. Phonetics 28: 495-502 (2000).

Browman, C.; Goldstein, L.: Towards an Articulatory Phonology. Phonology 3: 219-252 (1986).

Browman, C.; Goldstein, L.: Some Notes on Syllable Structure in Articulatory Phonology. Phonetica 45: 140-155 (1988).

Browman, C.; Goldstein, L.: Tiers in Articulatory Phonology, with some Implications for Casual Speech. In J. Kingston & M. Beckman (eds.), Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech, 341-397 (Cambridge University Press, Cambridge, UK, 1990).

Browman, C.; Goldstein, L.: Competing constraints on intergestural coordination and self-organization of phonological structures. Bulletin de la Communication Parlée, 5: 25-34 (2000).

Chitoran, I.; Goldstein, L.; Byrd, D.: Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven & N. Warner (Eds.), Laboratory Phonology 7, 419-448 (Mouton de Gruyter, Berlin, 2002).

Delattre, P.; Liberman, A.; Cooper, F.: Acoustic loci and transitional cues for consonants. J. Acoust. Soc. Am. 27: 769-773 (1955).

Flemming, E.: Scalar and categorical phenomena in a unified model of phonetics and phonology. Phonology 18: 7-44 (2001).

Fowler, C.: Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. Perception and Psychophysics 55: 597–610 (1994).

Fujimura, O.: A Linear Model of Speech Timing. In Channon & Shockey (Eds.), In Honor of Ilse Lehiste, 109-124 (Foris Publications, Dordrecht, 1987).

Gordon, M.: Laryngeal timing and correspondence in Hupa. UCLA Working Papers in Phonology 5: 1-70 (2001).

Jaeger, T.: Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. J. Mem. Lang. 59: 434-456 (2008).

Janda, L.; Nesset, T.; Baayen, R.: Capturing Correlational Structure in Russian Paradigms: a Case Study in Logistic Mixed-Effects Modeling. *Corpus Linguistics and Linguistic Theory* 6: 29-48 (2010).

Joos, M.: Acoustic phonetics. Supplement to Language 24: (1948).

Jun, J.: Positional faithfulness, sympathy, and inferred input. Ms., Yeungnam University (2002).

Katz, J. Compression effects in English. J. Phonetics 40: 390-402 (2012).

Kingston, J.; Diehl, R.: Phonetic knowledge. Language 70: 419-454 (1994).

Kozhevnikov, V.; Chistovich, L.: Speech: Articulation and Perception (US Dept. of Commerce, Washington, DC, 1965).

Krakow, R.: Physiological organization of syllables: a review. J. Phonetics 27: 23-54 (1999).

Krakow, R.; Beddor, P.; Goldstein, L.; Fowler, C.: Coarticulatory influences on the perceived height of nasal vowels. J. Acoust. Soc. Am. 83: 1146–1158 (1988).

Krull, D.: Relating acoustic properties to perceptual responses: A study of Swedish voiced stops. J. Acoust. Soc. Am. 88: 2557-2570 (1990).

Ladefoged, P.; Johnson, K.: *A Course in Phonetics* (Wadsworth, Boston, 2010).

Liberman, A.; Cooper, F.; Shankweiler, D.; Studdert-Kennedy, M.: Perception of the speech code. Psychological Review 74: 431–461 (1967).

Lindblom, B.: Spectrographic study of vowel reduction. J. Acoust. Soc. Am. 35: 1773-1781 (1963).

Lindblom, B.; Rapp, K.: Some Temporal Regularities of Spoken Swedish. Papers in Linguistics from the University of Stockholm 21 (1973).

Marin, S.; Pouplier, M.: Temporal Organization of complex onsets and codas in American English: Testing the Predictions of a Gestural Coupling Model. Motor Control 14: 380-407 (2010).

Menzerath, P.; de Lacerda, A.: Koartikulation, Steuerung und Lautabgrenzung (Dümmler, Berlin, Bonn, 1933).

Munhall, K.; Fowler, C.; Hawkins, S.; Saltzman, E.: "Compensatory Shortening" in monosyllables of spoken English. J. Phonetics 20: 225-239 (1992).

Ohde, R.; Sharf, D.: Order effect of acoustic segments of VC and CV syllables on stop and vowel identification. J. Speech and Hearing Research 20: 543-554 (1977).

Öhman, S.: Coarticulation in VCV utterances: Spectrographic measurements. J. Acoust. Soc. Am. 39: 151-168 (1966).

Parker, E.; Diehl, R.: Identifying vowels in CVC syllables: Effects of inserting silence and noise. Perception & Psychophysics 36: 369-380 (1984).

Quené, H.; van den Bergh, H.: Examples of mixed-effects modeling with crossed random effects and with binomial data. J. Mem. Lang. 59: 413-425 (2008).

Redford, M.; Diehl, R.: The relative perceptual distinctiveness of initial and final consonants in CVC syllables. J. Acoust. Soc. Am. 106: 1555-1565 (1999).

Repp, B.; Lin, H.: Acoustic properties and perception of stop consonant release transients. J. Acoust. Soc. Am. 83: 379-396 (1989).

Silverman, D. Phasing and Recoverability. UCLA PhD Dissertation (1995).

Smits, R.; Warner, N.; McQueen, J.; Cutler, A.: Unfolding of phonetic information over time: A database of Dutch diphone perception. J. Acoust. Soc. Am. 113: 563-574 (2003).

Sproat, R.; Fujimura, O.: Allophonic variation in English /l/ and its implications for phonetic implementation. J. Phonetics 21: 291-311 (1993).

Steriade, D.: Phonetics in phonology: the case of laryngeal neutralization (Ms., UCLA, 1997).

Warner, N.; Smits, R.; McQueen, J.; Cutler, A.: Phonological and statistical effects on timing of speech perception: Insights from a database of Dutch diphone perception. Speech Communication 46: 53-72 (2005).

Whalen, D.: Vowel information in postvocalic fricative noises. Language and Speech 26: 91-100 (1983).

Winitz, H.; Scheib, M.; Reeds, J.: Identification of Stops and Vowels for the Burst Portion of /p, t, k/ Isolated from Conversational Speech. J. Acoust. Soc. Am. 51: 1309-1317 (1972).

Wright, J.: Effect of vowel nasalization on the perception of vowel height. In Ferguson, Hyman, & Ohala (eds.), Nasalfest: Papers from a Symposium on Nasals and Nasalization, 373-388 (Language Universals Project, Stanford, 1975).

Wright, R.: *Consonant clusters and cue preservation in Tsou*. UCLA PhD dissertation (1996).

Yeni-Komshian, G.; Soli, S.: Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. J. Acoust. Soc. Am. 70: 966-975 (1981).

Zue, V. W.: Acoustic characteristics of stop consonants: A controlled study (Lincoln Laboratory, Lexington Mass., 1976).

1

	/r/	/l/	/n/	/p/
[ɑ]-[u] + Coda	--	(c)all-(c)ool	(c)on-(c)oon	(c)op-(c)oop
[e]-[o] + Coda	(c)are-(c)ore	(k)ale-(c)oal	(c)ane-(c)one	(c)ape-(c)ope
[i]-[e] + Coda	(p)ier-(p)air	(p)eel-(p)ale	(p)een-(p)ain	(sh)eep-(sh)ape
[ɑ]-[u] + Onset	raw-rue	law-lou	gnaw-new	paw-pooh
[e]-[o] + Onset	ray-row	lay-low	neigh-no	pay-poe
[i]-[e] + Onset	ree(d)-rai(d)	lee-lay	knee-neigh	pea-pay

2

Vowel Pair Group	Stimuli		Total
	Experimental	Additional	
[ɑ]-[u] Coda	36	12	48
[e]-[o] Coda	48	0	48
[ɑ]-[u] Onset	48	24	72
[e]-[o] Onset	48	24	72
[i]-[e] Onset	48	24	72

3

Speaker 1	Vowel midpoint	Gate 3
er-or	6.44	4.81
el-ol	6.54	5.61
en-on	6.81	6.86
ep-op	6.00	7.07
Speaker 2		
er-or	7.12	3.43
el-ol	7.46	4.90
en-on	7.38	9.44
ep-op	7.14	8.58

4

Speaker 1	Onset	Coda
r	3.05	n/a
l	2.90	3.86
n	2.90	2.32
p	3.37	3.56
Speaker 2		
r	2.20	n/a
l	2.22	2.66
n	1.84	2.89
p	2.72	3.53

5

Condition	Segment	Item	Speaker 1	Speaker 2	Sp. 1 Gate	Sp. 2 Gate
a-u onset	r	ra	131	80	20	27
		ru	144	93		
	l	la	110	115		
		lu	147	116		
	n	na	108	90		
		nu	126	92		
	p	pa	88	105		
		pu	128	127		
a-u coda	l	al	47	88	27	25
		ul	117	83		
	n	an	68	108		
		un	70	131		
	p	ap	81	92		
		up	105	136		
e-o onset	r	re	113	104	22	21
		ro	152	79		
	l	le	124	125		
		lo	128	93		
	n	ne	131	104		
		no	128	96		
	p	pe	112	122		
		po	111	122		
e-o coda	r	er	62	104	24	20
		or	51	121		
	l	el	67	106		
		ol	71	76		
	n	en	54	94		
		on	47	84		
	p	ep	97	114		
		op	87	119		
i-e onset	r	ri	142	115	22	21
		re	113	104		
	l	li	130	97		
		le	124	125		
	n	ni	148	100		
		ne	131	104		
	p	pi	165	135		
		pe	112	122		

6

Sensitivity terms

Sensitivity to	compared to	β	Std. Error	Z	p	Sig.
pV	chance	1.72	0.54	3.15	0.002	*
nV	pV	-0.54	0.17	-3.16	0.002	*
lV	pV	0.67	0.26	2.53	0.011	*
rV	pV	0.35	0.25	1.40	0.162	
Vp	pV	-0.55	0.24	-2.25	0.025	*
(Vn vs. nV)	(Vp vs. pV)	1.50	0.30	5.06	0.000	*
(Vl vs. lV)	(Vp vs. pV)	0.97	0.47	2.07	0.039	*
(Vr vs. rV)	(Vp vs. pV)	2.57	0.48	5.39	0.000	*

Baseline (≈bias) terms

Likelihood of FA for response	compared to	β	Std. Error	Z	p	Sig.
poe	chance	-0.36	0.42	-0.86	0.388	
no	poe	0.70	0.22	3.15	0.002	*
low	poe	0.92	0.27	3.44	0.001	*
row	poe	0.95	0.23	4.13	0.000	*
pooh	poe	0.25	0.26	0.95	0.341	
new	pooh	0.65	0.31	2.14	0.032	*
lou	pooh	-1.44	0.38	-3.77	0.000	*
rue	pooh	-2.00	0.33	-5.98	0.000	*
pay	poe	0.34	0.26	1.29	0.196	
neigh	pay	-0.89	0.30	-2.98	0.003	*
lay	pay	-1.28	0.38	-3.36	0.001	*
raid	pay	-0.75	0.32	-2.31	0.021	*
cope	poe	0.33	0.29	1.15	0.250	
cone	cope	-0.93	0.35	-2.63	0.008	*
coal	cope	1.37	0.44	3.13	0.002	*
core	cope	-2.11	0.37	-5.70	0.000	*
coop	cope	0.07	0.39	0.17	0.864	
coon	coop	0.41	0.46	0.89	0.374	
cool	coop	-0.84	0.59	-1.44	0.150	

Other fixed effects

Change in	β	Std. Error	Z	p	Sig.
likelihood of FA for every log ms. of consonant duration	0.07	0.17	0.43	0.669	

likelihood of FA following a timeout on the previous trial	0.01	0.15	0.09	0.929	
likelihood of FA following an incorrect answer on the previous trial	0.13	0.06	2.12	0.034	*
sensitivity for every log ms. of consonant duration	0.38	0.23	1.60	0.109	
sensitivity following a timeout on the previous trial	-0.30	0.21	-1.41	0.157	
sensitivity following an incorrect answer on the previous trial	-0.23	0.09	-2.63	0.009	*

Random Slopes

Random slope for	By	D.F.	χ^2	p
FAs for /l/	Subject	3	105.4	0
FAs for /r/	Subject	5	92.7	0
FAs for /n/	Subject	7	137.4	0
Sensitivity to /l/	Subject	4	11.2	0.024
Sensitivity to /r/	Subject	6	42.3	0
Sensitivity to /n/	Subject	8	26.1	0.001
Overall sensitivity	Speaker	2	28.5	0

7

Sensitivity terms

Sensitivity to	compared to	β	Std. Error	Z	p	Sig.
Vp	chance	0.30	0.25	1.22	0.22	
Vn	Vp	0.86	0.30	2.88	0.00	*
pV	Vp	0.55	0.27	2.05	0.04	*
nV	pV	-1.41	0.36	-3.94	0.00	*
Change in sensitivity at gate 2 for						
Vp	chance	4.17	0.44	9.53	0.00	*
Vn	Vp	-0.85	0.32	-2.63	0.01	*
pV	Vp	-1.76	0.54	-3.26	0.00	*
nV	pV	1.77	0.39	4.53	0.00	*

Baseline (≈bias) terms

Likelihood of FA for response	compared to	β	Std. Error	Z	p	Sig.
op	chance	-0.18	0.26	-0.71	0.48	
on	op	-0.30	0.30	-0.99	0.32	
up	op	0.30	0.33	0.91	0.37	
un	op	1.14	0.39	2.95	0.00	*
po	op	-0.25	0.39	-0.66	0.51	
no	po	0.99	0.38	2.58	0.01	*
pu	po	-0.19	0.44	-0.42	0.68	
nu	pu	-0.44	0.52	-0.84	0.40	
pe	po	0.41	0.35	1.17	0.24	
ne	pe	-1.10	0.31	-3.55	0.00	*
Change in likelihood of FA at gate 2 for						
op	chance	-2.13	0.30	-7.11	0.00	*
on	op	0.67	0.27	2.52	0.01	*
up	op	0.39	0.27	1.41	0.16	
un	op	-1.21	0.33	-3.68	0.00	*
po	op	0.52	0.37	1.39	0.16	
no	po	-1.38	0.33	-4.19	0.00	*
pu	po	-0.22	0.36	-0.62	0.53	
nu	pu	0.84	0.43	1.94	0.05	*
pe	po	0.19	0.22	0.86	0.39	
ne	pe	0.64	0.26	2.51	0.01	*

Other terms

Effect	β	Std. Error	Z	p	Sig.
Change in likelihood of FA following a timeout	-0.09	0.14	-0.65	0.52	
Change in likelihood of FA following an incorrect answer	0.14	0.06	2.33	0.02	*
Change in sensitivity following a timeout	0.03	0.20	0.14	0.89	
Change in sensitivity following an incorrect answer	-0.15	0.09	-1.72	0.09	

Random slopes

Random slope for	By	D.F.	χ^2	p
likelihood of FA for /n/	subject	3	126.1	0
likelihood of FA at gate 2	subject	5	42.5	0
gate-0 sensitivity	subject	2	391.4	0
gate-2 sensitivity	subject	7	353.4	0
likelihood of FA for /Co/, /Cu/ responses	speaker	3	21.3	0
likelihood of FA for /Ce/ responses	speaker	4	35.6	0
overall sensitivity	speaker	2	9.1	0.01

Table 1. Materials used in the experiment. Parentheses indicate consonants that were not present in the recording but were presented as part of the response choice.

Table 2. Number of stimuli in each Vowel-Pair group.

Table 3. Euclidean distance between /e/ and /o/ in bark formant space in the context of various consonants, measured at the midpoint of the vowel and the midpoint of gate 3.

Table 4. Euclidean distance between /a/ and /u/ in bark space of the two lowest spectral prominences for various consonants, in onset (left) and coda (right) conditions.

Table 5. Measured consonant duration, in ms., for each lexical item in the experiment, with gate duration for each Vowel Pair group and each speaker.

Table 6. Statistical model for zero-gate data. Fixed effects are listed in terms of which stimulus parameter they quantify against which baseline level. Other columns show the estimated regression coefficient β, the standard error associated with that estimate, the Z statistic and p-value from a Wald test, and the significance of the effect at $\alpha = 0.05$. FA = false alarm. Random slopes are listed in terms of which fixed effects vary by which levels of random effects; other columns show the degrees of freedom, χ^2 statistic, and p-value associated with a likelihood ratio test.

Table 7. Statistical model for /n/ and /p/ data at gates 0 and 2. Fixed effects are listed in terms of which stimulus parameter they quantify against which baseline level. Other columns show the estimated regression coefficient β, the standard error associated with that estimate, the Z statistic and p-value from a Wald test, and the significance of the effect at $\alpha = 0.05$. FA = false alarm. Random slopes are listed in terms of which fixed effects vary by which levels of random effects; other columns show the degrees of freedom, χ^2 statistic, and p-value associated with a likelihood ratio test.

Figure 1. Tokens of each consonant used in the experiment, in coda (left) and onset (right) positions. Text grid shows three gates. Zero gate stimuli consist of only the portions marked with consonants here; successive gates add the portions labeled 'g' to that original stimulus, one 'g' section defining each gate.

Figure 2. F1 and F2 frequencies (bark) at the temporal midpoint of the measured vowel region (left panels) and the third gate (right panels) for materials in the /e/-/o/ coda condition. One speaker's data is shown in each row.

Figure 3. F1 and F2 frequencies (bark) at the temporal midpoint of the measured vowel region for materials in the /a/-/u/ conditions for each speaker.

Figure 4. Frequencies of two lowest spectral prominences (bark) at the temporal midpoint of the measured consonant region (or during the burst and frication of /p/) for both speakers in all conditions.

Figure 5. (a) Sensitivity to vowel contrasts from gate 0 of coda (left) and onset (right) consonants, pooled across vowels. Vertical axis in d' units. Plots represent the distribution of 58 subjects' data. The box indicates the inter-quartile range. The solid line indicates the median. The whiskers indicate the range. (b) Sensitivity to vowel contrasts from the steady states of coda (left) and onset (right) consonants, pooled across vowels and subjects, as fit by the statistical model with all other factors regressed out. Vertical axis in logit (log odds) units. Whiskers indicate the standard error estimated by the model.

Figure 6. (a) Sensitivity to vowel contrasts from onset and coda /p/ and /n/, at gates 0 (left) and 2 (right), pooled across vowels. Vertical axis in d' units. Plots represent the distribution of 58 subjects' data. The box indicates the inter-quartile range. The solid line indicates the median. The whiskers indicate the range. Open circle indicates a data point more than 1.5 times the inter-quartile range from the median. One negative outlier for /Vp/ not pictured here. (b) Sensitivity to vowel contrasts from onset and coda /p/ and /n/, at gates 0 (left) and 2 (right), pooled across vowels and subjects, as fit by the statistical model with all other factors regressed out. Vertical axis in logit (log odds) units. Whiskers indicate the standard error estimated by the model.

Acknowledgements

The author wishes to thank Adam Albright, Edward Flemming, Donca Steriade, Associate Editor Sarah Hawkins, and 2 anonymous reviewers for extensive commentary and discussion, and Sarah Wang at the MIT Behavioral Research Laboratory for practical assistance. This project has also benefited from feedback from audiences at the LSA Annual Meeting, the Ludwig-Maximilians-Universität Institute of Phonetics and Speech Processing, the University of Southern California, and the Université Paris Descartes Psychology of Perception Laboratory. This research was made possible in part by an NSF Graduate Research Fellowship Award, an ESF European Young Investigator Award, and a Mellon Postdoctoral Fellowship in the Humanities.