A standard set of American-English voiced stop-consonant stimuli from morphed natural speech

Joseph D. W. Stephens and Lori L. Holt [a)]

Psychology Department and Center for the Neural Basis of Cognition,

Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Received

Running Title: Morphed speech stimuli

Abbreviated Title: A standard set of morphed speech stimuli

[a)] Electronic-mail: lholt@andrew.cmu.edu

ABSTRACT

Linear predictive coding (LPC) analysis was used to create morphed natural tokens of English voiced stop consonants ranging from /b/ to /d/ and /d/ to /g/ in four vowel contexts (/i/, /æ/, /a/, /u/). Both vowel-consonant-vowel (VCV) and consonant-vowel (CV) stimuli were created. A total of 320 natural-sounding acoustic speech stimuli were created, comprising 16 stimulus series. A behavioral experiment demonstrated that the stimuli varied perceptually from /b/ to /d/ to /g/, as expected. Acoustic analyses indicated that the stimuli compared favorably to standard characteristics of naturally-produced, voiced stop consonants. The entire set of stimuli is freely available on the Internet (http://www.psy.cmu.edu/~lholt/php/StephensHoltStimuli.php) for use in research applications.

PACS numbers: 43.71.Es, 43.72.Ja

INTRODUCTION

One of the mainstays of speech perception research is the use of series of speech stimuli that vary perceptually from one phonetic endpoint to another. Many experiments make use of perceptually ambiguous speech stimuli as a means of identifying factors that influence speech perception, including categorical perception (Liberman *et al.*, 1957), lexicality (Ganong, 1980; Elman and McClelland, 1988), transitional probability (Pitt and McQueen, 1998), lexical neighborhood size and density (Newman, Sawusch, and Luce, 1997), visual information (Massaro, 1987; 1998), native language experience (McCandliss *et al.*, 2002) and adjacent phonetic or acoustic context (Mann, 1980; Lotto and Kluender, 1998; Holt, 2005). Relatively straightforward sound-editing methods can be used to create ambiguous sounds from natural speech tokens of unvoiced fricatives (e.g., /s/ and /ʃ/; McQueen, 1991) or phonemes that differ in voice-onset time (e.g., /t/ and /d/; Ganong, 1980). However, ambiguous sounds between natural tokens of voiced stop consonants (e.g., /b/, /d/, /g/) are more difficult to create using simple waveform mixing or editing techniques. Stop place of articulation is signaled by several acoustic cues including formant transitions, which contain complex temporal and frequency information. Speech synthesizers like those of Klatt (1980; Klatt and Klatt, 1988) allow for explicit control of the acoustic characteristics of speech, but compared to natural speech such synthetic speech is acoustically impoverished and tends to sound relatively unnatural to listeners.

An alternative to the use of purely synthetic speech sounds is the application of digital signal processing techniques to “morph” between natural speech tokens and produce more realistic sounding stimuli. Linear predictive coding (LPC) analysis (Atal and Hanauer, 1971; Markel and Gray, 1976) can be used to model the vocal-tract filter properties of speech sounds and the resulting filters can be modified to produce acoustically and perceptually intermediate sounds. The interpolation of LPC-derived filters to create speech stimuli has been used successfully in speech perception experiments involving liquid consonants (e.g., McCandliss *et al.*, 2002), and recent efforts have been made to enhance the ease and automaticity of speech-morphing procedures (Slaney *et al.*, 1996; Pfitzinger, 2004). Nonetheless, the application of LPC methods can require significant time and effort, as well as trial and error, to successfully produce high-quality sounds that are free of noticeable acoustic artifacts. The purpose of the current project was to aid researchers by providing a useful set of LPC-morphed stimulus series, consisting of voiced stop-consonant utterances with high acoustic quality and natural characteristics. The series consist of vowel-consonant-vowel (VCV) and consonant-vowel (CV) utterances in which the consonants range from /b/ to /d/ and from /d/ to /g/, in four vowel contexts: /i/, /æ/, /a/, and /u/. The stimuli are freely available electronically (Appendix A, "Instructions for downloading and use of stimuli").

It is important to note that because the stimuli presented here were created by interpolating between natural tokens, each stimulus in the set may vary from others across multiple acoustic dimensions. As a result, the stimuli will be inappropriate for use in experiments that require explicit control of individual acoustic dimensions. Nonetheless, these naturalistic, ambiguous auditory speech stimuli should be of use in many research applications.

The following text describes the methods used to create the stimuli, a behavioral experiment that evaluated perception of the stimuli by native speakers of American English, and several acoustic analyses of the stimuli. The experiment and analyses, respectively, demonstrated that the LPC-morphing procedure succeeded in producing sounds that were reliably perceived as members of English /b/, /d/, and /g/ categories, and that the perceptually relevant features of the stimuli were consistent with previously established observations of naturally-produced, voiced stop consonants in American English.

I. MORPHING NATURAL TOKENS

Eight 20-member series of acoustic stimuli were created with the morphing procedure. Each series consisted of VCV tokens in which the consonant ranged from /b/ to /d/ or from /d/ to /g/, in one of four vowel contexts: /i/, /æ/, /a/ or /u/. The stimuli were created by adjusting filter parameters of natural utterances derived from LPC analysis and applying the adjusted filters to source waveforms extracted from the natural tokens.

An adult, Midwestern American male speaker (J.D.W.S.) produced three repetitions of each of the 12 VCV combinations used in the stimulus set. The tokens were recorded digitally on a personal computer using Computer Speech Laboratory (CSL; Kay Elemetrics Corp., Lincoln Park, NJ) with 16-bit precision at a sampling rate of 11.025 kHz. The tokens were isolated and saved separately as monaural PCM .wav files. The tokens were then matched in RMS power prior to further processing.

The alignment of temporal characteristics of endpoint sounds is one of the primary challenges of sound morphing (particularly for automatic methods; Slaney *et al.*, 1996; Pfitzinger, 2004). For the current stimuli, this problem was circumvented by maximizing the alignment of temporal characteristics between endpoints prior to morphing. Within each vowel context, the tokens for each consonant that were most compatible in pitch and temporal properties (i.e., speaking rate, burst length, and duration) were selected as series endpoints. The selected /b/, /d/ and /g/ tokens within each vowel context were then edited to produce further temporal alignment. The offsets and lengths of the initial and final vowels were aligned by deleting or duplicating pitch periods within the vocalic portions of the waveforms as necessary. The burst onsets of the consonants were aligned by shortening or lengthening the post-vocalic silences. In some cases it was necessary also to realign or shorten consonant bursts to maximize the compatibility of the endpoint tokens.[1]

An LPC analysis was performed on each of the edited natural endpoint tokens using the autocorrelation algorithm (Markel and Gray, 1976) implemented in the computer program Praat (version 4.3.19; Boersma, 2001) with the following parameters: prediction order 16; analysis width (Hamming window) 25 ms; time step 5 ms; pre-emphasis above 50 Hz. The resulting filter coefficients were saved in text files for editing. The /d/ tokens from each vowel context were inverse-filtered by their LPC coefficients to extract approximate voicing sources for each /d/ endpoint token. The resulting four source waves (one for each vowel context) were saved and used in the subsequent resynthesis of all stimuli within a corresponding vowel series.

To create series ranging perceptually between endpoint consonants, the LPC coefficients for each endpoint were read into a script that computed the differences between the endpoint tokens' filter coefficients at every 5-ms time step and recorded 20 new sets of coefficients (18 of which were intermediate sets that represented incremental adjustments of those differences). For any time step for which the endpoint filters possessed different numbers of coefficients, the lower number of coefficients was used in the output filter (as a result, the endpoint filters that were created as output were not necessarily exactly identical to the filters used as input). Thus, filters were created that ranged from /b/ to /d/ and from /d/ to /g/, in the /i/, /æ/, /a/, and /u/ VCV contexts.

After each series of LPC filters was created, the filters were applied to the source wave derived from the /d/ token with the corresponding vowel, so that all members of each VCV series were based on the same voicing source. Subsequent to this resynthesis all 160 VCV stimuli were

RMS-matched. A 100-ms silent interval was added to the beginning of each waveform file and the durations of all stimuli were standardized to 830 ms by adding silence to the end of each file as necessary. An additional set of 160 CV stimuli were created by excising the initial vowel from each of the VCV stimuli. Each VCV was cut at the time point corresponding to the last zero-crossing prior to the consonant release in the relevant source wave. After editing, all 160 CV stimuli were RMS-matched. A 100-ms silent interval was added to the beginning of each waveform file and the durations of all stimuli were standardized to 515 ms by adding silence to the end of each file as necessary. In all, 320 stimuli were created (20-step morphing × 2 perceptual continua × 4 vowel contexts × 2 syllable types).

II. BEHAVIORAL EXPERIMENT

The procedure described above defined sets of filters with parameters intermediate between two endpoint consonants. The usefulness of these sounds in speech research depends not on their filter parameters, but rather on their perceptual characteristics. Thus, to verify that the stimuli were perceived in the manner intended by the morphing process, a behavioral experiment was conducted in which participants identified the consonants in a 3-alternative, forced-choice task.

A. Method

1. Stimuli

The eight (two consonant series: /b/-/d/; /d/-/g/ × four vowels) 20-member series of VCV utterances and the eight 20-member series of CV utterances derived from the VCVs were used as stimuli in the experiment.

2. Participants

Thirteen participants were recruited from the Carnegie Mellon University community. One participant's data showed no discernible pattern in identification responses across consonant series; this participant's data were not included in group analyses. All participants were monolingual, native speakers of American English with no reported hearing impairment. Each participant received $7 after completing each session of the experiment (four sessions, $24 total).

3. Procedure

The two series ranging from /b/ to /d/ and /d/ to /g/ in each vowel context were combined to yield groups of 40 stimuli spanning the three response categories. Thus, each stimulus was presented within the context of the overall range of English voiced stop consonants. Each 40-member stimulus series was presented within a separate experimental trial block. Thus, there were eight blocks, one for each of the four vowels (/i/, /æ/, /a/, or /u/) within each utterance type (VCV or CV). These eight blocks were spread across four experimental sessions that were conducted on separate days. Each session separately tested perceptual identification for consonant stimuli within one of the four vowel contexts, with one block devoted to VCV stimuli and the other block devoted to the corresponding CV stimuli. The ordering of the four sessions for each participant was counterbalanced according to a Latin square design and the ordering of blocks within each session alternated from session to session. During each block, a set of 40 stimuli ranging from /b/ to /d/ to /g/ was repeated 10 times. For each repetition of the set, stimuli were presented in random order. Participants were instructed to listen to the stimuli and identify the consonants by pressing buttons on an electronic response box labeled "B," "D," and "G." Presentation of stimuli was controlled using Tucker-Davis Technologies System II hardware. After digital-to-analog conversion, stimuli were low-pass filtered at 4.8 kHz and output diotically to Beyer DT-150 headphones at approximately 65-70 dB. Participants were seated in sound-attenuated booths during the experiment.

During data collection, it was discovered that a faulty switch on the output line to the headphones caused some stimuli to be presented through only the left channel. Although this problem was corrected, it affected a total of eleven experimental sessions spread across five participants: three sessions for two participants; two sessions for two participants; and one session for one participant. The other seven participants were not affected. In order to ensure that this malfunction did not influence the interpretation of the experimental results, statistical analyses were performed both with and without the data from the affected sessions. The exclusion of the affected data did not substantially alter the results.

B. Results

Figure 1 displays identification data averaged across 12 participants for each of the /b/-/d/-/g/ consonant series, separately for VCV (solid lines) and CV (dashed lines) stimuli. For each stimulus series, participants sorted stimuli into three distinct categories, with clear boundaries between stimuli labeled “B,” “D,” and “G.” Thus, the LPC-morphing procedure succeeded in creating series that were perceived by listeners as shifting perceptually between good exemplars of the voiced stop-consonant categories.

One notable feature of the data is the apparent difference in responses due to presence or absence of an initial vowel. Across all four vowel contexts, participants responded with “D” more often to VCV stimuli than to CV stimuli. This pattern was examined using a 2 (initial vowel) × 40 (series member) repeated-measures, multivariate analysis of variance (MANOVA) on the three response types (“B,” “D,” and “G”), separately for each stimulus series. In each of the four vowel contexts, the multivariate effect of initial vowel was significant (for /i/, $F(2,10)$ =

11.8, $p = .002$; for /æ/, $F(2,10) = 18.9$, $p < .001$; for /a/, $F(2,10) = 4.21$, $p = .047$; for /u/, $F(2,10) = 9.32$, $p = .005$). Univariate tests of initial vowel were conducted for each response type; the results of these tests are summarized in Table I. Across all four vowel contexts, there was a significant increase in "D" responses for VCV stimuli relative to CV stimuli. Consistent with the clear shifts in identification across each series, a multivariate effect of series member was highly significant in all four vowel contexts (for /i/, $F(78,858) = 184.4$, $p < .001$; for /æ/, $F(78,858) = 361.2$, $p < .001$; for /a/, $F(78,858) = 174.7$, $p < .001$; for /u/, $F(78,858) = 162.7$, $p < .001$). A significant multivariate interaction of initial vowel and series member was also found in each of the four vowel contexts (for /i/, $F(78,858) = 6.13$, $p < .001$; for /æ/, $F(78,858) = 4.63$, $p < .001$; for /a/, $F(78,858) = 2.74$, $p < .001$; for /u/, $F(78,858) = 5.73$, $p < .001$). These interactions indicate that the effect of initial vowel was greatest near the perceptual boundaries between consonant categories.

As stated above, all statistical analyses were repeated, excluding data from sessions in which participants erroneously received headphone output in only one ear. Excluding these data, all of the multivariate effects found in the full data set remained statistically significant. Two of the univariate effects were no longer significant; these were the effect of initial vowel on "D" responses in the /æ/ series and the effect of initial vowel on "G" responses in the /u/ series (see Table I).

Inspection of data from individual participants indicated that, for some trial blocks, individuals' responses did not fit the overall pattern of categorization seen in the average data. That is, for the large majority of trial blocks run in the experiment (90 of 96) each of the three response options reached ceiling (i.e., 100%) in the participant's responses at some point along the stimulus series. However, in six of the trial blocks (three blocks for one participant and one block for each of three participants – three of these blocks were also affected by the output line problem described above), one of the response categories did not reach ceiling at any point along the stimulus series. In three of these blocks one response category did not exceed 50% for any stimulus. As a result, it could be the case that these few, anomalous blocks of data distorted the averages in a manner that caused spurious effects to be detected in the statistical analysis. The statistical analyses were therefore repeated excluding both data blocks (*i.e.*, CV and VCV) from any sessions that contained an anomalous data block. For the /a/ series, the multivariate effect of initial vowel was no longer significant, and the univariate effect of initial vowel on "G" responses was no longer significant. All other effects that were significant in the full data set were also significant when these anomalous data blocks were excluded (see Table I).

C. Discussion

The LPC interpolation procedure was successful in creating stimuli that transitioned between good perceptual examples of American-English /b/, /d/, and /g/ consonant categories. Although the averaged data show a few cases in which one of the responses did not reach ceiling levels, individual participants nearly always used each of the three response labels at the maximum rate somewhere along each of the series. Interestingly, the perception of stimuli near consonant category boundaries was affected by whether the stimuli included the initial vowels produced in the original VCV utterances. Specifically, participants identified consonants as "D" more often when the initial vowel was present than when the initial vowel was absent. This effect was observed across all four series, although it was more pronounced for the /i/ and /u/ series than for the /a/ and /æ/ series. The effect of initial vowel was also generally more reliable around the boundary between /d/ and /g/ than around the boundary between /b/ and /d/. Some possible sources for the effect of initial vowel will be considered below.

The results of the identification task demonstrate the efficacy of the LPC-interpolation technique for creating stimuli that morph perceptually between natural tokens. For these stimuli to be used effectively in speech perception research, it will also be desirable to have information regarding their acoustic properties and the extent to which these acoustic characteristics are consistent with prior observations of the nature of voiced stop consonants. The following section describes a set of systematic acoustic measurements that were made of the LPC-interpolated stimuli.

III. ACOUSTIC ANALYSES

LPC interpolation and resynthesis were used to create these stimuli because of the ability of such a procedure to simultaneously vary numerous acoustic properties of speech sounds while retaining the sounds' natural quality. Thus, in addition to demonstrating the perceptual reliability of the stimuli, it is important to document their specific acoustic characteristics. Several acoustic analyses were performed to verify that the morphing procedure produced gradual changes across several acoustic dimensions, and that the acoustic properties of the resynthesized stimuli were consistent with previous research on natural productions of voiced stop consonants. Some light was also shed on the perceptual differences caused by the deletion of the initial vowel from the VCV stimuli.

A. Formant frequencies and locus equations

1. Measurements

One of the primary correlates of stop-consonant place of articulation is formant frequency trajectory, particularly for the second ($F2$) and third ($F3$) formants, although the formant patterns corresponding to each consonant category are highly dependent on vowel context (Öhman, 1966; Lindblom, 1963). Despite this dependence of formant frequencies on vowel context, $F2$

frequency of consonant onsets relative to neighboring vowels often exhibits a reliable linear relationship within each consonant (i.e., place of articulation) category, which can be modeled using linear regression to derive “locus equations” for consonant place (Lindblom, 1963; Sussman, 1991). Therefore, the relation of the formant frequencies of the current stimuli to stop consonant perception was also examined by computing locus equations for the three consonant categories.

An automatic procedure was used to measure formant frequencies of the current stimuli at specific time points. The following time points were marked separately for each of the VCV series: the midpoint of the initial vowel; the offset of the initial vowel (defined here as the offset of harmonic structure, e.g., Sussman *et al.*, 1997); onset of the consonant (defined here as the onset of formants immediately after burst; e.g., Sussman *et al.*, 1991); early in the second vowel (after consonant transition); and the midpoint of the second vowel. Using Praat (version 4.3.19; Boersma, 2001), formant trajectories were computed for all 160 VCV stimuli using the Burg algorithm with the following parameters: 5 formants; maximum formant frequency 5 kHz; window length 25.4 ms; and time step 6.35 ms. The formant values for $F1$, $F2$, and $F3$ at the identified time points were then recorded for each of the 160 stimuli. For a few of the stimuli near /ugu/, the formant analysis yielded anomalous values that were obviously inconsistent with neighboring series members. In these cases, the algorithmically-generated values were corrected by visual inspection of spectral plots (FFT, in Praat) of the stimuli at the specified time points and measuring the formant center frequency.

Figure 2 displays formant frequencies measured at the midpoint of the initial vowel, at the offset of initial vowel, at the onset of the consonant, early in the final vowel, and at the midpoint of the final vowel; the formant frequencies are plotted separately for each /b/-/d/ and

/d/-/g/ series. The numerical values of the formant frequencies represented in Figure 2 are given in Appendix B, “Formant measurements,” and are also available in electronic format as described in Appendix A. The measurements indicate that the interpolation of LPC filter coefficients led to gradual steps in formant frequencies across each series, although not every step was the same size. The primary shifts in formant frequencies across consonants in each series are seen in the frequencies of $F2$ and $F3$ at initial vowel offset and consonant onset. Vowel formant frequency characteristics remained stable across most series, with the exception of the /u/ series, for which $F2$ of the final vowel showed an apparent influence of the consonant.

The stimuli representing the best perceptual instances of each consonant within each vowel context were used to compute locus equations. For each consonant category, the three series members that were most often identified by participants as instances of that that category in the perceptual experiment (averaged across VCV and CV blocks, see Fig. 1) were selected from each vowel context. Locus equations were computed by fitting regression lines to the points defined by $F2$ center frequency at consonant onset and at the midpoint of the final vowel (e.g., Sussman *et al.*, 1991). For stimuli representing /g/, separate lines were fit to approximate the locus equations for the palatal /g/ and velar /g/ allophones that occur in front vowel and back vowel contexts, respectively. Due to the small number of vowel contexts available here, an equation for palatal /g/ was estimated by excluding measurements from consonants in the /u/ vowel context and an equation for velar /g/ was estimated by excluding measurements from the /i/ vowel context.

Figure 3 displays locus equations for each consonant along with the measured points used to compute them. Also displayed are the locus equations for 15 male speakers of American English reported in two studies of naturally-produced voiced stop consonants (Sussman *et al.*, 1991; Sussman *et al.*, 1997[2]). It can be seen that the locus equations for stimuli identified strongly as belonging to one of the consonant categories were generally similar to the locus equations for the corresponding consonants found by previous studies. Differences between the current locus equations and those of previous studies are due mainly to the influence of a few stimuli for which the relation between $F2$ at consonant onset and at vowel midpoint were slightly outside the typical range. Specifically, for stimuli identified as /ubu/ and /idi/, $F2$ frequencies for the consonant were somewhat higher relative to the vowel than in previous studies. For stimuli identified as /ugu/, consonant $F2$ was somewhat lower relative to vowel $F2$ than in previous studies.

2. Influence of initial vowel

The above analysis indicates that the formant frequencies of consonant onsets in the current stimuli compare favorably to previous measurements of naturally-produced, voiced stop consonants. However, the observed differences in perception of VCV and CV stimuli indicate that consonant perception is also influenced by information present in the initial vowel. It is evident from the formant frequency measurements displayed in Figure 2 that many of the stimuli exhibit an asymmetry between formant frequencies at the offsets of the initial vowels and at the onsets of the consonants. These differences might account for the observed perceptual effect of the initial vowel.

One possible explanation for how the initial vowel could affect consonant identification is that the formant transitions at the offset of the vowel carried information that was more consistent with a given consonant category than the formant transitions after consonant release. In this case, perhaps the greater proportion of “D” responses seen across each VCV stimulus series was due to the formant transitions in the initial vowel of any given stimulus carrying more

/d/-like information than the formant transitions of the consonant alone. To evaluate the phonetic information present in the initial vowel, an additional set of locus equations was derived using data points defined by *F*2 frequency at the endpoint of the initial vowel and at the midpoint of the initial vowel. Table II compares these locus equations with the equations derived from measurements of *F*2 at consonant onset and at the midpoint of the final vowel. Table II also lists mean locus equation data from the male speakers in the study by Sussman *et al.* (1997), which compared syllable-initial and syllable-final consonants. Both sets of data are consistent in the differences observed between locus equations for consonants at syllable onset versus syllable offset. The slopes of locus equations for /b/ and velar /g/ were reduced for offsets, whereas the slopes for /d/ and palatal /g/ were increased. Overall, the differences in slope across all consonant categories were reduced for offsets compared to onsets; similar observations have been made in other studies (Lindblom, 1963; Krull, 1988). The general shift toward more intermediate slopes for offsets also resulted in slopes that were more similar to the slope for the /d/ category, particularly in the current stimulus set. Thus, the initial vowels could have shifted identification responses toward “D” by providing additional acoustic information that was relatively more consistent with the /d/ category than the information present from the onsets of the consonants onward.

Another explanation for how the presence of initial vowels led to differences in “D” identification is that the formant frequencies of the initial vowels may have shifted perception of subsequent formant frequencies in a direction favorable to perception of /d/. Such a shift would be similar to phonetic context effects originating from contrastive auditory mechanisms (e.g., Lotto and Kluender, 1998; Holt, 1999). For instance, manipulating the *F*2 frequency of an initial vowel can cause shifts in the consonant identification of subsequent CV syllables (Holt and Lotto, 2002). Similar effects have been found using a variety of speech and non-speech stimuli (for a summary, see Lotto and Holt, 2006) and have been characterized as resulting from a perceptual mechanism that exaggerates differences between formant frequencies of temporally adjacent sounds. In the current stimulus set, this type of contrastive mechanism could shift perception toward /d/ when formant frequencies at the offset of an initial vowel are further from the /d/-endpoint than the formant frequencies of the subsequent consonant. For example, in the /ubu/ - /udu/ series, $F2$ at consonant onset is higher for /d/ than for /b/ (see Figure 2). Additionally, for each individual stimulus along the series, $F2$ at initial vowel offset is lower than $F2$ at consonant onset. If a contrastive auditory process is at work, the presence of the relatively low $F2$ information at initial vowel offset in VCV stimuli should shift perception of $F2$ at consonant onset toward higher perceived frequencies, and thus toward more “D” identifications. Inspection of the formant data in Figure 2 indicates that the relations between vowel offset and consonant onset could allow for a contrastive mechanism to shift perception of formant frequencies toward those of /d/ via either $F2$ or $F3$ spectral contrast (or both).

B. Gross spectral shape

Another acoustic correlate of stop-consonant place of articulation is the shape of the “short-time” spectrum at consonant onset. Acoustic theory indicates that the gross spectral shape sampled over 10-20 ms following consonantal release should exhibit distinctive characteristics depending on place of articulation (Fant, 1960). These theoretically derived gross spectral shapes have been shown to be good approximations of natural consonant productions (Blumstein and Stevens, 1979), and to be relevant for perceptual identification of consonants (Stevens and Blumstein, 1978; Blumstein and Stevens, 1980). Therefore the extent to which the current stimuli agree with theoretical predictions concerning gross spectral shape at consonant onset was examined, particularly with regard to perceptual identification of the stimuli.

Following the method of Blumstein and Stevens (1979), the first difference of each stimulus waveform was computed (to pre-emphasize higher frequencies) and was multiplied by a modified raised cosine window over the first 26 ms after consonant release. A time point corresponding to consonant release was defined for the members of each stimulus series as the last zero-crossing prior to the onset of burst noise in the unfiltered source waveform used to generate that series. The shape of the window (see Figure 4) was such that the earliest portion of the burst contributed most to the spectrum. The power spectral density was then computed using a 14-pole Burg algorithm (in MatLab, version 6.0.0.88; The MathWorks, Inc., Natick, MA).

Figure 5 displays power spectral density plots for the consonant onsets of stimuli identified as one consonant on at least 90% of trials in the experiment. Blumstein and Stevens (1979) described the distinctive features of the onset spectra of consonants: “In the case of velar consonants, the theoretically predicted common attribute of the spectrum is a major spectral prominence in the midfrequency range; for alveolar consonants, the spectral energy is diffuse or distributed throughout the frequency range, but with greater spectral energy at higher frequencies; when the consonant constriction is made at the lips, the spectral energy is again diffuse, but the spectrum is weighted toward the lower frequencies” (p. 1002).

The overall shapes of the short-time spectra for each consonant in the current study are quite consistent with these expected shapes. For most of the consonant-vowel combinations, the spectra fit the templates defined by Blumstein and Stevens (1979) in their study of consonant productions: labials exhibited a diffuse-falling spectrum, alveolars exhibited a diffuse-rising spectrum, and velars exhibited a compact spectrum with one prominent peak.

The notable exception to the overall pattern was the group of /d/ consonants in the context of /u/. For these stimuli, the higher-frequency peaks are not greater in amplitude than the lower-frequency peaks, and these spectra would be rejected by Blumstein and Stevens' criteria for the diffuse-rising template. It is interesting to note that the effect of initial vowel on "D" responses was particularly strong for stimuli in the series with /u/ vowel context. For these stimuli, the dependence of perception on the initial vowel may have been partly due to the relative lack of gross spectral cues for /d/ at consonant onset.

C. Onset temporal characteristics

In addition to formant frequencies and onset spectra, place of articulation in stop consonants is related to certain temporal characteristics of consonant onsets. For instance, velar stops tend to have longer voice-onset times (VOTs) than labial stops (Lisker and Abramson, 1964; Kewley-Port, 1982; Stevens, 1998), and VOT can be used to predict differences in place of articulation in situations where formant frequencies are not very diagnostic (Engstrand, Krull, and Lindblom, 2000). In the current study, the endpoints' burst onsets were aligned prior to morphing, and the same voicing source (from the /d/ endpoint) was used for all stimuli within each vowel context. These procedures eliminated any VOT differences that may have existed between the consonants in the original utterances. However, other temporal onset correlates of consonant place of articulation exist. For example, velar stops tend to exhibit a more gradual release of constriction than labials (Stevens, 1998). Thus, the amplitude of the acoustic waveform for velars tends to increase more slowly at consonant onset. Although this property of consonants is produced by different articulatory mechanisms than VOT, it has some of the same consequences: for instance, both longer VOT and more gradual release attenuate formant amplitudes at consonant onset. Differences in abruptness of consonant onset are also an interesting correlate of place of articulation regardless of VOT, so an attempt was made here to quantify these changes in abruptness across the LPC-interpolated stimuli.

It was possible to achieve a rough measure of how gradual the releases of the stops were by tracking the amplitudes of the waveforms over a brief period after consonant onset. The consonant release point for each vowel series was defined in the same manner as for the gross spectral shape measurements described above. For each of the 160 stimuli, the absolute value of its waveform was measured over the 25 ms period following the release point identified for the corresponding vowel series, and a simple linear regression was fit to the sample points within that window. The slope of the resulting line was taken as a measure of “onset velocity.” Figure 6 illustrates the methods used to obtain these values.

Figure 7 displays the onset velocity measurements obtained from the consonants in each of the four vowel contexts. As expected, onset velocity showed an overall decrease from /b/ to /d/ to /g/. Each series fell within somewhat different ranges on the absolute scale of the onset velocity measure, but specific comparisons between the series are not warranted considering that different voicing sources were used in the synthesis of stimuli in the different vowel contexts. The measurement of onset velocity devised here was intended mainly as a relative measure to compare stimuli within the same vowel context.

D. Summary of acoustic analyses

Three general conclusions can be drawn from the acoustic analyses of the LPC-morphed stimuli described here:

1. The LPC-morphing procedure produced gradual acoustic shifts along each series. Formant frequencies, onset velocities, and onset spectra were all incrementally modified via the LPC-interpolation procedure. A gradual manipulation of multiple acoustic dimensions was the

objective of the current endeavor and accounts for the naturalistic quality of the stimuli. It should be noted that the stepwise changes in LPC filter parameters did not correspond to acoustic changes of equal magnitude. Thus, when using these stimuli, no assumptions should be made that steps along each series correspond to equal steps along any acoustic dimension. For reference, acoustic measurements of the formant frequencies of the stimuli are provided in Appendix B.

2. The stimuli exhibited acoustic properties typical of voiced stop consonants. For most of the stimuli, all acoustic measurements fell within expected ranges based on previous research. The exceptions to this general observation were the points representing /ubu/, /idi/, and /ugu/ in locus equation space, and the gross spectral shape at consonant onset for /udu/, which did not exhibit as much energy at higher frequencies as expected from theory (Blumstein and Stevens, 1979). In addition, the onset velocity measurements did not monotonically decrease from /b/ to /d/ to /g/ across all four vowel contexts. Nonetheless, no stimulus exhibited gross acoustic abnormalities, and for those stimuli with slight anomalies in one acoustic dimension, normal characteristics were observed in the other acoustic dimensions studied. Further, the normal identification functions found in the perceptual experiment suggest that any perceptual difficulty that might be introduced by a minor irregularity in one acoustic property is made up for by other characteristics, and supports the notion that consonant perception is accomplished through the recognition of patterns across more than just one acoustic dimension (e.g., Engstrand *et al.*, 2000).

3. The effect of initial vowel observed in the perceptual experiment was probably due to acoustic information in the initial vowels that differed slightly from information in the consonant onsets. The locus equations computed from measurements of initial vowels revealed that they

contained more /d/-like *F*2 information than the subsequent consonants. For many of the stimuli, formant frequencies in the initial vowels could also have shifted perception of consonant formants toward /d/ as a result of contrastive auditory effects.

IV. GENERAL DISCUSSION

Interpolation of LPC coefficients for naturally-produced stop consonant utterances was used to create series of stimuli ranging perceptually from /b/ to /d/ and /d/ to /g/ in four vowel contexts and in both VCV and CV form. The purpose this effort was to create natural-sounding, perceptually-ambiguous tokens that may be used by researchers for whom standard synthesis methods are either impractical or result in stimuli that are too acoustically impoverished for their purposes. The stimuli were shown to be reliably perceived by listeners as belonging to the three intended voiced stop-consonant categories in a 3-alternative forced-choice listening task. The VCV and CV stimuli were both found to be reliably identified by participants, although some differences in perception were observed as a function of the presence of the initial vowel. Acoustic analyses of the stimuli demonstrated that the LPC-interpolation procedure was successful in gradually shifting several acoustic properties of the original tokens and that the acoustic characteristics of the resulting stimuli were generally consistent with previous observations of naturally-produced utterances. These stimuli were created with the aim of providing researchers with a naturalistic consonant series varying along a place-of-articulation continuum. As such, the stimulus corpus is freely available for download, and readers interested in hearing or using the stimuli may see Appendix A for further information on accessing the materials via the Internet.

ACKNOWLEDGMENTS

The authors thank Christi Adams Gomez and Tony Kelly for help with data collection and preparation of online materials, respectively. This work was supported by a National Research Service Award (1 F31 DC007284-01) from the National Institute on Deafness and Other Communication Disorders to J.D.W.S., by a grant from the National Science Foundation (BCS-0345773) to L.L.H., and by the Center for the Neural Basis of Cognition.

APPENDIX A: INSTRUCTIONS FOR DOWNLOADING AND USE OF STIMULI

The full set of 320 stimuli is available on the Internet (with accompanying documentation) at http://www.psy.cmu.edu/~lholt/php/StephensHoltStimuli.php. The files named "VCV_stimuli.zip" and "CV_stimuli.zip" each contain eight 20-member morphed series, arranged hierarchically within directories. The directories corresponding to each vowel in "VCV_stimuli.zip" also include the original recorded utterances as well as the edited versions of the recordings and source waveforms used as input in the morphing procedure. The file named "all_stimuli.zip" contains all 320 stimuli and associated files within a single directory. Finally, the file named “formant_frequencies.pdf” contains the table of measured formant frequencies given in Appendix B.

Researchers are advised to refer to this report and the online documentation for important information regarding the acoustic and perceptual properties of the stimuli. The stimuli may also be obtained via electronic mail (lholt@andrew.cmu.edu) or in CD format by postal mail addressed to Lori Holt, Psychology Department, 5000 Forbes Ave., Pittsburgh, PA 15213.

APPENDIX B: TABLE OF ALL MEASURED FORMANT FREQUENCIES

Table B1 lists formant frequencies of the VCV stimuli (in Hz) for $F1$, $F2$, and $F3$ as measured by the automatic method described in the text. Formant frequency measurements that were corrected by hand are marked with asterisks (*). Listed alongside the formant measurements are the filenames for the digital waveform files included in the online archive. The measurements for the CV stimuli are identical except for the absence of the initial vowel.

FOOTNOTES

[1] The original recorded tokens as well as the edited tokens used for the morphing procedure are also available online.

[2] The locus equations from Sussman et al. (1997) represent the measurements reported from initial consonants in CVC utterances. Although Sussman et al. (1997) also studied VCV utterances, the initial and final vowels were independently varied in that study to evaluate coarticulatory influences of the initial vowel.

REFERENCES

Atal, B. S., and Hanauer, S. L. (**1971**). “Speech analysis and synthesis by linear prediction of the speech wave,” J. Acoust. Soc. Am. **50**, 637-655.

Blumstein, S. E., and Stevens, K. N. (**1979**). “Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants,” J. Acoust. Soc. Am. **66**, 1001-1017.

Blumstein, S. E., and Stevens, K. N. (**1980**). “Perceptual invariance and onset spectra for stop consonants in different vowel environments,” J. Acoust. Soc. Am. **67**, 648-662.

Boersma, P. (**2001**). “PRAAT, a system for doing phonetics by computer,” Glot International **5**, 341-345.

Elman, J. L., and McClelland, J. L. (**1988**). “Cognitive penetration of the mechanisms of perception: compensation for coarticulation of lexically restored phonemes,” J. Memory Lang. **27**, 143-165.

Engstrand, O., Krull, D. and Lindblom, B. (**2000**). “Sorting stops by place in acoustic space,” in Proceedings of the XIIIth Swedish Phonetics Conference (FONETIK 2000), Skövde, Sweden, pp. 53-56.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague, The Netherlands).

Ganong, W. F., III. (**1980**). “Phonetic categorization in auditory word perception,” J. Exp. Psychol. Hum. Percept. Perform. **6**, 110-125.

Holt, L. L. (**1999**). “Auditory constraints on speech perception: An examination of spectral contrast,” Unpublished doctoral dissertation, University of Wisconsin, Madison, WI.

Holt, L. L. (**2005**). “Temporally non-adjacent non-linguistic sounds affect speech categorization,” Psychol. Sci. **16**, 305-312.

Holt, L. L. and Lotto, A. J. (**2002**). “Behavioral examinations of the neural mechanisms of speech context effects,” Hear. Res. **167**, 156-169.

Kewley-Port, D. (**1982**). “Measurement of formant transitions in naturally produced stop consonant–vowel syllables,” J. Acoust. Soc. Am. **72**, 379-389.

Klatt, D. H. (**1980**). “Software for a cascade/parallel formant synthesizer.” J. Acoust. Soc. Am. **67**, 971-995.

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

Krull, D. (**1988**). “Acoustic properties as predictors of perceptual responses: A study of Swedish voiced stops,” Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm (PERILUS), Vol. VII, pp. 66-70.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). “The discrimination of speech sounds within and across phoneme boundaries,” J. Exp. Psychol. **54**, 358-368.

Lindblom, B. (**1963**). “On vowel reduction,” Report #29, The Royal Institute of Technology, Speech Transmission Laboratory, Stockholm, Sweden.

Lisker, L., and Abramson, A. S. (**1964**). “A cross-language study of voicing in initial stops: Acoustical measurements,” Word **20**, 384-422.

Lotto, A. J., and Kluender, K. R. (**1998**). “General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification,” Percept. Psychophys. **60**, 602-619.

Lotto, A. J., and Holt, L. L. (**2006**). “Putting phonetic context effects into context: A commentary on Fowler (2006),” Percept. Psychophys. **68**, XXX-XXX.

Mann, V. A. (**1980**). “Influence of preceding liquid on stop-consonant perception,” Percept. Psychophys. **28**, 407-412.

Markel, J. D., and Gray, A. H., Jr. (**1976**). *Linear Prediction of Speech* (Springer-Verlag, New York).

Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Lawrence Erlbaum Associates, Hillsdale, NJ).

Massaro, D. W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT, Cambridge, MA).

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., and McClelland, J. L. (**2002**). "Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception," Cogn. Affect. Behav. Neurosci. **2**, 89-108.

McQueen, J. M. (**1991**). "The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity," J. Exp. Psychol. Hum. Percept. Perform. **17**, 433-443.

Newman, R. S, Sawusch, J.R., and Luce, P. A. (**1997**). "Lexical neighborhood effects in phonetic processing," J. Exp. Psychol. Hum. Percept. Perform. **23**, 873-889.

Ohman, S. E. G. (**1966**). "Coarticulation in VCV utterances: Spectrographic measurements," J. Acoust. Soc. Am. **39**, 151-168.

Pfitzinger, H. R. (**2004**). "Unsupervised morphing between utterances of any speakers," in Proceedings of the 10th Australian International Conference on Speech Science and Technology, Sydney, Australia, pp. 545-550.

Pitt, M. A., and McQueen, J. M. (**1998**). "Is compensation for coarticulation mediated by the lexicon?" J. Memory Lang. **39**, 347-370.

Slaney, M., Covell, M., and Lassiter, B. (**1996**) “Automatic audio morphing,” in Proc. 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP96), Atlanta, GA, Vol. 2, pp. 1001-1004.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).

Stevens, K. N., and Blumstein, S. E. (**1978**). “Invariant cues for place of articulation in stop consonants,” J. Acoust. Soc. Am. **64**, 1358-1368.

Sussman, H. M., Bessell, N., Dalston, E., and Majors, T. (**1997**). “An investigation of stop place articulation as a function of syllable position: A locus equation perspective,” J. Acoust. Soc. Am. **101**, 2826-2838.

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (**1991**). “An investigation of locus equations as a source of relational invariance for stop place categorization,” J. Acoust. Soc. Am, **90**, 1309-1325.

TABLE I. Summary of effects of initial vowel on consonant identification

			Univariate effects by response type		
Vowel context	Data	Multivariate effect	“B”	“D”	“G”
/i/	All	*p* = .002	*NS*	*p* < .001	*p* = .003
	Limited 1[a]	*p* = .021	*NS*	*p* = .004	*p* = .011
	Limited 2[b]	*p* = .006	*p* = .021	*p* = .001	*p* = .007
/æ/	All	*p* < .001	*NS*	*p* = .019	*p* = .001
	Limited 1	*p* = .002	*NS*	*NS*	*p* = .004
	Limited 2	*p* < .001	*NS*	*p* = .019	*p* = .001
/a/	All	*p* = .047	*NS*	*p* = .011	*p* = .038
	Limited 1	*p* = .044	*NS*	*p* = .010	*p* = .025
	Limited 2	*NS*	*NS*	*p* = .021	*NS*
/u/	All	*p* = .005	*p* = .003	*p* = .001	*p* = .008
	Limited 1	*p* = .020	*p* = .005	*p* = .004	*NS*
	Limited 2	*p* = .003	*p* = .002	*p* = .001	*p* = .043

[a] Limited 1 excludes data that were affected by the output line malfunction described in the text.

[b] Limited 2 data from sessions in which participants did not use each of the three responses maximally for at least one stimulus in each series (see text).

TABLE II. Locus equation slopes and *y*-intercepts for consonants at syllable onset and syllable offset. The data from Sussman *et al.* (1997) represent means of 5 male speakers.

		/b/		/d/		/g/ palatal		/g/ velar	
	Syllable	slope	*y*-int.	slope	*y*-int.	slope	*y*-int.	slope	*y*-int.
Current stimuli	CV	.795	309	.553	933	.435	1298	1.27	-67
	VC	.713	216	.606	779	.603	990	1.26	-96
Sussman *et al.* (1997)	CV	.734	290	.330	1280	.290	1453	1.07	316
	VC	.621	216	.351	1135	.436	1338	.857	303

TABLE B1. Measured formant frequencies.

					Formant frequencies (Hz)				
					V1		C	V2	
Vowel	Consonant	Filename	Formant		Midpoint	Endpoint	Onset	Early	Midpoint
				t=	195 ms	307 ms	425 ms	457 ms	550 ms
/i/	/b/-/d/	i_bd01_VCV	*F*1		231	230	235	243	253
			*F*2		2422	1713	2262	2390	2388
			*F*3		3299	2448	2839	3188	3276
		i_bd02_VCV	*F*1		232	231	234	243	253
			*F*2		2420	1824	2261	2385	2387
			*F*3		3300	2471	2835	3179	3273
		i_bd03_VCV	*F*1		232	232	234	244	252
			*F*2		2419	1912	2260	2379	2389
			*F*3		3291	2509	2840	3170	3273
		i_bd04_VCV	*F*1		233	232	233	244	251
			*F*2		2418	1994	2258	2372	2388
			*F*3		3294	2525	2839	3160	3272
		i_bd05_VCV	*F*1		234	232	233	244	250
			*F*2		2417	2044	2257	2366	2387
			*F*3		3303	2556	2839	3153	3271
		i_bd06_VCV	*F*1		235	232	232	244	250
			*F*2		2415	2074	2256	2362	2385
			*F*3		3300	2606	2839	3146	3267
		i_bd07_VCV	*F*1		236	232	232	244	249
			*F*2		2414	2104	2255	2357	2384
			*F*3		3303	2638	2838	3138	3266
		i_bd08_VCV	*F*1		237	232	232	245	249
			*F*2		2413	2126	2254	2352	2384
			*F*3		3304	2668	2838	3134	3266
		i_bd09_VCV	*F*1		238	233	231	245	248
			*F*2		2412	2163	2254	2347	2382
			*F*3		3304	2687	2837	3128	3267
		i_bd10_VCV	*F*1		239	233	231	245	248
			*F*2		2410	2183	2252	2342	2382
			*F*3		3319	2713	2835	3121	3265
		i_bd11_VCV	*F*1		240	232	231	245	247
			*F*2		2408	2205	2252	2338	2381
			*F*3		3326	2741	2836	3115	3263

		i_bd12_VCV	*F*1	241	232	231	245	247
			*F*2	2406	2219	2251	2334	2381
			*F*3	3329	2764	2835	3110	3260
		i_bd13_VCV	*F*1	243	231	231	245	247
			*F*2	2405	2229	2253	2331	2380
			*F*3	3344	2784	2837	3101	3258
		i_bd14_VCV	*F*1	244	232	230	246	246
			*F*2	2404	2243	2252	2327	2379
			*F*3	3352	2798	2838	3090	3257
		i_bd15_VCV	*F*1	245	231	230	246	246
			*F*2	2402	2249	2249	2323	2378
			*F*3	3360	2814	2835	3078	3247
		i_bd16_VCV	*F*1	247	231	230	246	245
			*F*2	2400	2256	2250	2320	2378
			*F*3	3371	2839	2839	3065	3242
		i_bd17_VCV	*F*1	248	231	230	246	245
			*F*2	2398	2261	2256	2317	2378
			*F*3	3372	2861	2844	3051	3238
		i_bd18_VCV	*F*1	250	230	230	247	245
			*F*2	2397	2264	2257	2314	2377
			*F*3	3381	2884	2847	3039	3232
		i_bd19_VCV	*F*1	251	230	230	247	244
			*F*2	2396	2271	2260	2309	2376
			*F*3	3372	2913	2846	3025	3224
		i_bd20_VCV	*F*1	253	230	230	247	244
			*F*2	2393	2277	2252	2304	2375
			*F*3	3374	2935	2842	3006	3215
/i/	/d/-/g/	i_dg01_VCV	*F*1	253	230	230	247	244
			*F*2	2393	2277	2252	2304	2375
			*F*3	3374	2935	2842	3006	3215
		i_dg02_VCV	*F*1	254	229	230	245	245
			*F*2	2393	2290	2263	2308	2376
			*F*3	3366	2960	2867	3023	3218
		i_dg03_VCV	*F*1	255	228	230	243	246
			*F*2	2391	2304	2273	2314	2377
			*F*3	3358	2997	2889	3049	3221
		i_dg04_VCV	*F*1	256	228	230	241	247
			*F*2	2391	2322	2295	2320	2379

	*F*3	3360	3041	2938	3068	3223
i_dg05_VCV	*F*1	257	227	231	240	249
	*F*2	2390	2341	2303	2326	2381
	*F*3	3356	3092	2966	3083	3225
i_dg06_VCV	*F*1	258	226	231	238	250
	*F*2	2389	2356	2311	2333	2382
	*F*3	3349	3140	2993	3096	3227
i_dg07_VCV	*F*1	260	225	231	237	251
	*F*2	2387	2366	2304	2343	2385
	*F*3	3351	3178	3002	3111	3232
i_dg08_VCV	*F*1	261	224	231	236	253
	*F*2	2385	2375	2308	2352	2387
	*F*3	3340	3208	3023	3125	3235
i_dg09_VCV	*F*1	262	224	231	235	254
	*F*2	2384	2381	2313	2363	2390
	*F*3	3334	3230	3045	3142	3237
i_dg10_VCV	*F*1	264	223	231	234	256
	*F*2	2382	2386	2317	2374	2393
	*F*3	3327	3249	3068	3161	3240
i_dg11_VCV	*F*1	265	223	231	233	257
	*F*2	2380	2388	2309	2383	2395
	*F*3	3318	3263	3081	3180	3239
i_dg12_VCV	*F*1	266	222	231	233	259
	*F*2	2378	2390	2314	2389	2399
	*F*3	3333	3276	3110	3196	3241
i_dg13_VCV	*F*1	267	222	231	232	261
	*F*2	2376	2392	2318	2393	2403
	*F*3	3329	3290	3140	3211	3243
i_dg14_VCV	*F*1	269	221	231	231	262
	*F*2	2374	2393	2321	2396	2408
	*F*3	3323	3298	3173	3225	3246
i_dg15_VCV	*F*1	270	221	230	231	264
	*F*2	2373	2394	2334	2399	2408
	*F*3	3307	3315	3270	3239	3248
i_dg16_VCV	*F*1	272	221	229	231	266
	*F*2	2371	2395	2335	2401	2413
	*F*3	3303	3334	3329	3251	3250

		i_dg17_VCV	*F*1	273	220	226	230	267
			*F*2	2368	2396	2336	2402	2418
			*F*3	3301	3359	3375	3264	3254
		i_dg18_VCV	*F*1	274	220	220	230	269
			*F*2	2366	2397	2336	2404	2424
			*F*3	3303	3407	3399	3274	3255
		i_dg19_VCV	*F*1	276	219	207	229	271
			*F*2	2364	2400	2343	2405	2430
			*F*3	3276	3416	3332	3284	3253
		i_dg20_VCV	*F*1	277	219	175	229	272
			*F*2	2361	2402	2352	2406	2435
			*F*3	3278	3409	3164	3295	3256

			t=	213 ms	338 ms	420 ms	488 ms	605 ms
/ae/	/b/-/d/	ae_bd01_VCV	*F*1	754	541	380	703	760
			*F*2	1693	1378	1501	1706	1703
			*F*3	2593	2458	2465	2710	2608
		ae_bd02_VCV	*F*1	757	530	380	695	759
			*F*2	1697	1403	1514	1717	1703
			*F*3	2592	2480	2449	2725	2614
		ae_bd03_VCV	*F*1	760	519	384	689	760
			*F*2	1703	1438	1543	1729	1702
			*F*3	2594	2524	2442	2733	2646
		ae_bd04_VCV	*F*1	763	510	391	684	760
			*F*2	1711	1461	1590	1741	1706
			*F*3	2612	2540	2511	2732	2657
		ae_bd05_VCV	*F*1	765	507	396	679	760
			*F*2	1716	1470	1615	1751	1712
			*F*3	2623	2536	2563	2738	2604
		ae_bd06_VCV	*F*1	767	509	396	675	761
			*F*2	1720	1500	1610	1759	1720
			*F*3	2627	2625	2511	2747	2666
		ae_bd07_VCV	*F*1	769	508	401	672	763
			*F*2	1724	1531	1637	1766	1728
			*F*3	2633	2659	2562	2752	2769
		ae_bd08_VCV	*F*1	771	501	403	668	763
			*F*2	1727	1554	1662	1775	1727
			*F*3	2639	2650	2604	2756	2662
		ae_bd09_VCV	*F*1	772	497	402	664	762
			*F*2	1731	1572	1685	1783	1721

			*F*3	2648	2673	2610	2762	2650
		ae_bd10_VCV	*F*1	774	493	402	661	760
			*F*2	1734	1585	1714	1790	1720
			*F*3	2655	2700	2628	2762	2717
		ae_bd11_VCV	*F*1	775	491	398	657	759
			*F*2	1737	1604	1728	1799	1728
			*F*3	2658	2710	2637	2772	2805
		ae_bd12_VCV	*F*1	777	490	395	653	754
			*F*2	1740	1625	1752	1809	1735
			*F*3	2663	2721	2676	2783	2707
		ae_bd13_VCV	*F*1	778	489	390	650	748
			*F*2	1743	1639	1761	1821	1736
			*F*3	2669	2739	2670	2786	2625
		ae_bd14_VCV	*F*1	779	487	384	645	746
			*F*2	1747	1649	1780	1827	1740
			*F*3	2679	2755	2694	2788	2754
		ae_bd15_VCV	*F*1	781	485	381	642	746
			*F*2	1751	1665	1800	1832	1737
			*F*3	2691	2776	2738	2796	2806
		ae_bd16_VCV	*F*1	783	485	376	638	750
			*F*2	1755	1689	1809	1836	1739
			*F*3	2693	2794	2739	2806	2801
		ae_bd17_VCV	*F*1	785	484	370	635	756
			*F*2	1758	1703	1821	1840	1747
			*F*3	2683	2812	2767	2814	2813
		ae_bd18_VCV	*F*1	786	482	367	631	757
			*F*2	1759	1719	1833	1846	1746
			*F*3	2692	2828	2803	2817	2780
		ae_bd19_VCV	*F*1	787	480	364	626	757
			*F*2	1761	1731	1846	1851	1744
			*F*3	2727	2840	2812	2824	2779
		ae_bd20_VCV	*F*1	788	476	362	621	757
			*F*2	1765	1740	1859	1855	1747
			*F*3	2730	2859	2834	2835	2831
/ae/	/d/-/g/	ae_dg01_VCV	*F*1	788	476	362	621	757
			*F*2	1765	1740	1859	1855	1747
			*F*3	2730	2859	2834	2835	2831

ae_dg02_VCV	*F*1	787	464	362	626	758
	*F*2	1763	1755	1870	1851	1746
	*F*3	2721	2858	2832	2828	2830
ae_dg03_VCV	*F*1	785	451	362	630	758
	*F*2	1761	1772	1884	1847	1746
	*F*3	2712	2855	2829	2821	2828
ae_dg04_VCV	*F*1	784	437	362	634	758
	*F*2	1760	1787	1899	1843	1745
	*F*3	2705	2851	2828	2814	2827
ae_dg05_VCV	*F*1	782	424	362	638	759
	*F*2	1758	1800	1914	1839	1744
	*F*3	2699	2845	2826	2806	2826
ae_dg06_VCV	*F*1	781	411	362	641	760
	*F*2	1757	1813	1926	1835	1743
	*F*3	2696	2838	2824	2798	2825
ae_dg07_VCV	*F*1	780	399	362	644	760
	*F*2	1757	1828	1939	1832	1743
	*F*3	2698	2823	2821	2791	2825
ae_dg08_VCV	*F*1	779	385	362	647	761
	*F*2	1756	1842	1949	1828	1742
	*F*3	2699	2807	2817	2783	2824
ae_dg09_VCV	*F*1	776	372	362	649	762
	*F*2	1760	1858	1960	1825	1741
	*F*3	2661	2776	2812	2776	2824
ae_dg10_VCV	*F*1	775	360	362	652	761
	*F*2	1760	1877	1971	1822	1739
	*F*3	2668	2737	2807	2766	2826
ae_dg11_VCV	*F*1	775	348	363	654	762
	*F*2	1761	1897	1984	1819	1738
	*F*3	2678	2695	2801	2759	2825
ae_dg12_VCV	*F*1	774	338	363	657	763
	*F*2	1761	1918	1997	1816	1737
	*F*3	2687	2653	2795	2752	2825
ae_dg13_VCV	*F*1	774	328	363	659	764
	*F*2	1762	1938	2011	1813	1736
	*F*3	2695	2613	2789	2745	2825
ae_dg14_VCV	*F*1	774	319	363	661	765
	*F*2	1763	1958	2025	1810	1735

			*F*3		2701	2573	2783	2740	2825
		ae_dg15_VCV	*F*1		773	311	364	663	766
			*F*2		1763	1980	2042	1808	1734
			*F*3		2706	2533	2780	2734	2825
		ae_dg16_VCV	*F*1		774	304	365	666	768
			*F*2		1763	2005	2056	1805	1733
			*F*3		2713	2485	2772	2730	2825
		ae_dg17_VCV	*F*1		774	298	365	668	769
			*F*2		1764	2048	2074	1804	1732
			*F*3		2716	2426	2763	2727	2825
		ae_dg18_VCV	*F*1		769	292	366	671	770
			*F*2		1765	2108	2094	1801	1731
			*F*3		2717	2357	2754	2722	2825
		ae_dg19_VCV	*F*1		768	287	367	673	768
			*F*2		1766	2130	2118	1798	1736
			*F*3		2718	2320	2743	2720	2815
		ae_dg20_VCV	*F*1		768	282	367	675	769
			*F*2		1766	2176	2149	1796	1735
			*F*3		2719	2300	2732	2716	2815
				t=	176 ms	282 ms	363 ms	426 ms	503 ms
/a/	/b/-/d/	a_bd01_VCV	*F*1		805	590	540	671	749
			*F*2		1274	1050	1098	1240	1274
			*F*3		2876	2770	2331	2833	2863
		a_bd02_VCV	*F*1		805	552	504	673	750
			*F*2		1275	1150	1109	1247	1275
			*F*3		2874	2780	2344	2822	2860
		a_bd03_VCV	*F*1		798	512	514	675	752
			*F*2		1273	1177	1114	1254	1275
			*F*3		2875	2780	2329	2809	2853
		a_bd04_VCV	*F*1		792	483	551	677	753
			*F*2		1269	1181	1135	1261	1277
			*F*3		2877	2772	2345	2799	2849
		a_bd05_VCV	*F*1		784	471	580	679	755
			*F*2		1265	1210	1211	1270	1279
			*F*3		2877	2775	2375	2791	2848
		a_bd06_VCV	*F*1		783	466	580	680	758
			*F*2		1264	1261	1356	1292	1283
			*F*3		2880	2783	2454	2784	2844

a_bd07_VCV	*F*1	782	455	564	680	760
	*F*2	1263	1275	1472	1313	1288
	*F*3	2881	2789	2530	2769	2843
a_bd08_VCV	*F*1	787	446	548	680	762
	*F*2	1268	1282	1531	1331	1294
	*F*3	2884	2785	2619	2759	2840
a_bd09_VCV	*F*1	788	443	534	678	764
	*F*2	1269	1287	1565	1340	1301
	*F*3	2884	2776	2661	2752	2837
a_bd10_VCV	*F*1	785	448	519	677	765
	*F*2	1268	1357	1597	1344	1308
	*F*3	2890	2789	2671	2739	2836
a_bd11_VCV	*F*1	785	448	512	677	767
	*F*2	1266	1374	1617	1354	1318
	*F*3	2886	2803	2695	2724	2833
a_bd12_VCV	*F*1	785	449	505	677	768
	*F*2	1267	1395	1630	1379	1328
	*F*3	2887	2806	2700	2707	2832
a_bd13_VCV	*F*1	786	453	493	675	768
	*F*2	1268	1458	1636	1401	1335
	*F*3	2888	2804	2689	2702	2826
a_bd14_VCV	*F*1	788	449	488	674	769
	*F*2	1269	1482	1649	1417	1340
	*F*3	2887	2809	2714	2699	2820
a_bd15_VCV	*F*1	787	449	484	672	770
	*F*2	1268	1490	1657	1426	1345
	*F*3	2892	2815	2720	2679	2815
a_bd16_VCV	*F*1	787	449	476	670	771
	*F*2	1268	1504	1664	1426	1349
	*F*3	2894	2822	2735	2652	2809
a_bd17_VCV	*F*1	788	451	472	669	772
	*F*2	1268	1550	1672	1428	1352
	*F*3	2892	2824	2755	2642	2803
a_bd18_VCV	*F*1	790	450	465	669	774
	*F*2	1268	1570	1675	1447	1357
	*F*3	2893	2827	2745	2638	2796
a_bd19_VCV	*F*1	791	447	465	670	775

			*F*2	1267	1593	1683	1480	1361
			*F*3	2898	2835	2762	2634	2789
		a_bd20_VCV	*F*1	792	446	467	670	777
			*F*2	1267	1634	1693	1513	1368
			*F*3	2899	2842	2782	2630	2782
/a/	/d/-/g/	a_dg01_VCV	*F*1	792	446	467	670	777
			*F*2	1267	1633	1693	1513	1368
			*F*3	2899	2842	2782	2630	2782
		a_dg02_VCV	*F*1	788	444	468	668	777
			*F*2	1269	1639	1686	1512	1371
			*F*3	2893	2830	2759	2625	2779
		a_dg03_VCV	*F*1	785	440	471	668	779
			*F*2	1273	1649	1684	1512	1375
			*F*3	2881	2817	2748	2616	2776
		a_dg04_VCV	*F*1	783	436	480	669	780
			*F*2	1275	1663	1684	1511	1379
			*F*3	2877	2812	2746	2610	2771
		a_dg05_VCV	*F*1	780	431	488	670	781
			*F*2	1278	1668	1689	1511	1383
			*F*3	2872	2818	2743	2617	2768
		a_dg06_VCV	*F*1	780	427	499	672	782
			*F*2	1280	1668	1691	1509	1387
			*F*3	2858	2792	2725	2623	2766
		a_dg07_VCV	*F*1	778	422	504	673	782
			*F*2	1281	1666	1695	1508	1392
			*F*3	2847	2754	2711	2618	2762
		a_dg08_VCV	*F*1	776	416	511	674	783
			*F*2	1282	1668	1701	1507	1397
			*F*3	2838	2603	2703	2609	2760
		a_dg09_VCV	*F*1	773	412	516	674	785
			*F*2	1284	1682	1704	1506	1402
			*F*3	2827	2518	2690	2610	2757
		a_dg10_VCV	*F*1	771	406	522	675	784
			*F*2	1285	1696	1715	1505	1407
			*F*3	2815	2443	2678	2611	2754
		a_dg11_VCV	*F*1	769	400	528	675	784
			*F*2	1286	1700	1720	1504	1412
			*F*3	2804	2346	2654	2621	2751

		a_dg12_VCV	*F*1		766	394	533	675	783
			*F*2		1286	1703	1733	1502	1415
			*F*3		2793	2300	2624	2617	2745
		a_dg13_VCV	*F*1		764	388	539	675	784
			*F*2		1287	1704	1744	1501	1419
			*F*3		2781	2251	2582	2613	2743
		a_dg14_VCV	*F*1		762	382	543	674	785
			*F*2		1288	1708	1759	1500	1424
			*F*3		2770	2212	2535	2614	2739
		a_dg15_VCV	*F*1		759	375	548	674	785
			*F*2		1288	1714	1783	1498	1428
			*F*3		2757	2171	2493	2614	2736
		a_dg16_VCV	*F*1		757	368	555	674	785
			*F*2		1286	1725	1828	1497	1432
			*F*3		2746	2142	2435	2613	2738
		a_dg17_VCV	*F*1		756	360	565	674	784
			*F*2		1284	1733	1928	1496	1434
			*F*3		2740	2126	2338	2613	2738
		a_dg18_VCV	*F*1		756	352	566	674	783
			*F*2		1281	1736	1904	1494	1437
			*F*3		2736	2107	2249	2612	2735
		a_dg19_VCV	*F*1		756	343	550	674	781
			*F*2		1283	1739	1841	1492	1439
			*F*3		2726	2084	2228	2611	2732
		a_dg20_VCV	*F*1		756	336	543	674	777
			*F*2		1287	1755	1940	1490	1442
			*F*3		2717	2059	2200	2609	2736
				t=	150 ms	300 ms	381 ms	457 ms	525 ms
/u/	/b/-/d/	u_bd01_VCV	*F*1		321	303	326	379	368
			*F*2		1120	837	1352	1267	1170
			*F*3		2365	2382	2471	2417	2408
		u_bd02_VCV	*F*1		321	297	318	371	362
			*F*2		1133	875	1370	1285	1179
			*F*3		2367	2374	2482	2418	2408
		u_bd03_VCV	*F*1		321	293	312	364	355
			*F*2		1140	932	1450	1305	1194
			*F*3		2370	2374	2489	2418	2405

u_bd04_VCV	*F*1	321	287	307	356	349
	*F*2	1145	992	1510	1333	1213
	*F*3	2370	2378	2493	2421	2397
u_bd05_VCV	*F*1	321	281	299	349	344
	*F*2	1151	1030	1523	1355	1246
	*F*3	2371	2384	2490	2424	2399
u_bd06_VCV	*F*1	321	276	290	343	338
	*F*2	1157	1072	1535	1367	1272
	*F*3	2375	2383	2500	2427	2403
u_bd07_VCV	*F*1	321	272	285	337	333
	*F*2	1163	1085	1563	1373	1275
	*F*3	2377	2384	2508	2428	2400
u_bd08_VCV	*F*1	321	269	281	332	329
	*F*2	1168	1101	1588	1384	1280
	*F*3	2377	2395	2514	2428	2401
u_bd09_VCV	*F*1	321	266	275	328	325
	*F*2	1170	1137	1622	1401	1287
	*F*3	2378	2391	2520	2429	2398
u_bd10_VCV	*F*1	321	264	270	324	322
	*F*2	1171	1168	1636	1427	1300
	*F*3	2381	2399	2529	2430	2395
u_bd11_VCV	*F*1	321	261	265	321	319
	*F*2	1172	1182	1653	1453	1313
	*F*3	2382	2406	2545	2431	2395
u_bd12_VCV	*F*1	321	259	257	318	317
	*F*2	1172	1200	1661	1474	1343
	*F*3	2383	2402	2528	2434	2391
u_bd13_VCV	*F*1	321	257	258	315	314
	*F*2	1173	1237	1676	1493	1373
	*F*3	2382	2408	2529	2436	2391
u_bd14_VCV	*F*1	321	255	260	313	311
	*F*2	1175	1278	1701	1514	1375
	*F*3	2386	2407	2551	2434	2392
u_bd15_VCV	*F*1	321	253	259	311	309
	*F*2	1176	1305	1725	1535	1380
	*F*3	2386	2404	2559	2435	2390
u_bd16_VCV	*F*1	320	252	259	309	307
	*F*2	1177	1328	1731	1556	1389

			*F*3	2395	2394	2551	2436	2385
		u_bd17_VCV	*F*1	320	250	259	307	306
			*F*2	1178	1360	1739	1576	1400
			*F*3	2397	2390	2573	2437	2381
		u_bd18_VCV	*F*1	320	249	259	306	305
			*F*2	1179	1411	1751	1595	1414
			*F*3	2398	2393	2583	2436	2379
		u_bd19_VCV	*F*1	320	247	258	304	305
			*F*2	1181	1442	1761	1616	1440
			*F*3	2395	2390	2582	2434	2375
		u_bd20_VCV	*F*1	320	245	257	304	304
			*F*2	1182	1507	1771	1635	1467
			*F*3	2391	2393	2590	2438	2373
/u/	/d/-/g/	u_dg01_VCV	*F*1	320	245	257	304	304
			*F*2	1182	1507	1771	1635	1467
			*F*3	2391	2393	2590	2438	2373
		u_dg02_VCV	*F*1	319	244	255	303	305
			*F*2	1182	1463	1758	1619	1454
			*F*3	2393	2445	2581	2432	2376
		u_dg03_VCV	*F*1	318	243	254	304	305
			*F*2	1182	1439	1746	1602	1434
			*F*3	2399	2486	2572	2425	2378
		u_dg04_VCV	*F*1	317	241	252	304	306
			*F*2	1182	1416	1733	1587	1417
			*F*3	2401	2518	2563	2415	2382
		u_dg05_VCV	*F*1	316	239	251	304	306
			*F*2	1181	1379	1722	1571	1407
			*F*3	2405	2541	2556	2406	2390
		u_dg06_VCV	*F*1	315	236	249	304	306
			*F*2	1182	1358	1712	1556	1398
			*F*3	2408	2565	2549	2398	2394
		u_dg07_VCV	*F*1	314	234	246	305	306
			*F*2	1181	1343	1701	1539	1391
			*F*3	2410	2580	2539	2390	2392
		u_dg08_VCV	*F*1	314	232	237	305	307
			*F*2	1182	1329	1692	1519	1385
			*F*3	2411	2590	2540	2383	2394

u_dg09_VCV	*F*1	313	229	234	305	308
	*F*2	1182	1309	1678	1502	1381
	*F*3	2413	2594	2528	2377	2391
u_dg10_VCV	*F*1	313	227	230	306	309
	*F*2	1182	1280	1659	1487	1379
	*F*3	2415	2591	2512	2372	2392
u_dg11_VCV	*F*1	312	224	211	306	309
	*F*2	1183	1261	1634	1475	1378
	*F*3	2418	2586	2500	2366	2392
u_dg12_VCV	*F*1	312	221	181*	307	310
	*F*2	1183	1250	1608	1460	1373
	*F*3	2416	2582	2473	2361	2393
u_dg13_VCV	*F*1	312	219	184*	307	311
	*F*2	1184	1237	1586	1439	1354
	*F*3	2417	2582	2412	2357	2395
u_dg14_VCV	*F*1	312	216	181*	308	312
	*F*2	1184	1222	1564	1414	1336
	*F*3	2415	2583	2376	2354	2395
u_dg15_VCV	*F*1	312	214	187*	308	312
	*F*2	1184	1214	1540	1394	1319
	*F*3	2419	2617	2408	2352	2397
u_dg16_VCV	*F*1	311	212	184*	308	313
	*F*2	1185	1208	1512	1380	1305
	*F*3	2422	2644	2416*	2350	2402
u_dg17_VCV	*F*1	312	210	177*	309	313
	*F*2	1185	1222	1487	1372	1295
	*F*3	2436	2648	2416*	2349	2405
u_dg18_VCV	*F*1	311	207	181*	310	314
	*F*2	1186	1214	1465	1366	1288
	*F*3	2443	2653	2416*	2347	2406
u_dg19_VCV	*F*1	312	205	177*	311	316
	*F*2	1186	1197	1405	1359	1284
	*F*3	2458	2631	2416*	2345	2408
u_dg20_VCV	*F*1	312	203	173*	312	317
	*F*2	1187	1182	1300*	1345	1282
	*F*3	2465	2618	2419*	2343	2410

FIGURE CAPTIONS

FIG. 1. Proportion “B” (circles), “D” (squares), and “G” (triangles) responses across the four stimulus series. Solid lines represent responses to VCV stimuli; dashed lines represent responses to CV stimuli.

FIG. 2. Formant frequencies of F1, F2 and F3 measured at five time points in each VCV stimulus: the midpoint of the initial vowel; the endpoint of the initial vowel; the onset of the consonant; early in the final vowel; the midpoint of the final vowel. Solid lines indicate the /d/ endpoint of each series; dashed lines and dotted-dashed lines indicate the /b/ and /g/ endpoints of each series, respectively.

FIG. 3. Locus equations (solid lines) based on the best three instances of each consonant category from each stimulus series, as indicated by responses in the perceptual experiment (symbols). The y-axis represents F2 frequency at consonant onset and the x-axis represents F2 frequency at the midpoint of the final vowel. Dotted lines depict locus equations for ten male speakers reported by Sussman et al. (1991) and five male speakers reported by Sussman et al. (1997).

FIG. 4. Illustration of the window applied to each stimulus waveform in order to compute the short-time spectrum for consonant onset (after Blumstein & Stevens, 1979).

FIG. 5. Onset spectra for stimuli identified as one of the consonants on at least 90% of trials (averaged across VCV and CV blocks), for each of the four vowel contexts. Except for the /d/ stimuli with /u/ context, the spectra were consistent with the templates described by Blumstein and Stevens (1979).

FIG. 6. Illustration of the procedure used to compute “onset velocity.” A 25 ms segment of the waveform was sampled, starting at the time point corresponding to the last zero-crossing

before release in the original source wave used in the LPC resynthesis. A linear regression was fit to the rectified sample values from this time window; the slope of the line was taken as the onset velocity.

FIG. 7. Onset velocity measurements across each of the four stimulus series. Overall, onset velocity decreased from /b/ to /d/ to /g/.