Effects of obstruent consonants on fundamental frequency at vowel onset in English[a)]

Helen M. Hanson[b)]
Department of Electrical and Computer Engineering, Union College, 807 Union Street, Schenectady, New York 12308

(Received 9 February 2007; revised 6 October 2008; accepted 13 October 2008)

When a vowel follows an obstruent, the fundamental frequency in the first few tens of milliseconds of the vowel is known to be influenced by the voicing characteristics of the consonant. This influence was re-examined in the study reported here. Stops, fricatives, and the nasal /m/ were paired with the vowels /i,ɑ/ to form CVm syllables. Target syllables were embedded in carrier sentences, and intonation was varied to produce each syllable in either a high, low, or neutral pitch environment. In a high-pitch environment, *F*0 following voiceless obstruents is significantly increased relative to the baseline /m/, but following voiced obstruents it closely traces the baseline. In a low-pitch environment, *F*0 is very slightly increased following all obstruents, voiced and unvoiced. It is suggested that for certain pitch environments a conflict can occur between gestures corresponding to the segmental feature [stiff vocal folds] and intonational elements. The results are different acoustic manifestations of [stiff] in different pitch environments. The spreading of the vocal folds that occurs during unvoiced stops in certain contexts in English is an enhancing gesture, which aids the resolution of the gestural conflict by allowing the defining segmental gesture to be weakened without losing perceptual salience.
 [DOI: 10.1121/1.3021306]

PACS number(s): 43.70.Fq, 43.70.Bk [CHS] Pages: 425–441

I. INTRODUCTION

It is generally accepted that in American English (and other languages), the fundamental frequency (*F*0) in a vowel following an obstruent consonant is influenced by the voicing characteristics of the consonant. As we review in Sec. I A, methods of study and interpretation of the results have evolved somewhat over the past 50 years. However, it is agreed that there is a voiced/voiceless dichotomy, with *F*0 at vowel onset being significantly higher following voiceless obstruents than following voiced. This dichotomy is often called "pitch skip," and we sometimes refer to it as "obstruent intrinsic *F*0" (or obstruent I*F*0) in this paper.

The pitch-skip phenomenon is worth studying for several reasons. Studies have shown that the dichotomy in *F*0 at vowel onset cues listeners to voicing characteristics of the consonant (Sec. I A 2), suggesting that pitch skip should be included in speech synthesis systems to make synthesized speech both more natural and more intelligible. Likewise, speech-recognition systems could benefit by accounting for this dichotomy. In addition to these applications, studies of pitch skip are interesting because of what they can tell us about models of speech production and sound change. In some languages, lexical tone is believed to have derived from voicing contrasts of obstruents, and pitch skip is offered as evidence for a phonetic basis of such tonogenesis [e.g., see Hombert *et al.* (1979)]. Jun (1996) argued that pitch skip has affected the implementation of phrase-level tones in Korean. In Sec. IV we propose that the interaction of pitch skip and phrase-level tones in American English provides evidence supporting a model of speech production suggested by Keyser and Stevens (2006).

In the next section, we review both production and perception studies related to the pitch-skip phenomenon. We also review theories of the source of this effect.

A. Background

1. Production data

House and Fairbanks (1953) averaged *F*0 over the vowels in symmetrical CVC syllables and found that this average *F*0 was higher when the consonant was voiceless than when it was voiced. They observed the time course of a subset of the *F*0 contours and noted that the *F*0 difference occurred at the start of voicing, rather than occurring uniformly throughout the vowel. Lehiste and Peterson (1961) measured the peak *F*0 of CVC syllables and found that the peak *F*0 in vowels following voiceless consonants was higher than that in vowels following voiced. They also noted that the time course of the *F*0 contour varied depending on whether the initial consonant was voiced or unvoiced: following voiceless consonants, the peak *F*0 in the vowel occurred immediately after the consonant, while following voiced consonants the peak occurred at about the middle of the vowel. In addition, it was found that final consonants did not have any consistent effect on the *F*0 of the preceding vowel.

While those earlier studies were on data produced by American English speakers, Mohr (1971) purposely used speakers who had mixed language backgrounds (three native speakers of Chinese, Russian, and German, with English as a

[a)]Portions of this work were presented at meetings the Acoustical Society of America held in Chicago, IL (June 2001) and Austin, TX (Nov. 2003).
[b)]Electronic mail: hansonh@union.edu

second language) in order to test the universality of the pitch skip. Mohr (1971) measured *F*0 at vowel boundaries and at minima or maxima. In this way, he was able to confirm the earlier studies, quantify their observations that the influence of the preceding consonant is limited to the early part of the vowel, and show that the phenomenon occurs in languages other than English.

Lea (1973) recorded two types of utterances from two male speakers of American English. The first were bisyllabic nonsense words *hə′*CVC, in which C and V included nearly all the consonants and vowels of American English. The second were pairs of bisyllabic words such as the noun and verb forms of "compact." Both types of utterances were recorded in isolation. For the nonsense syllables, in which the medial consonant always followed an unstressed syllable and preceded a stressed syllable, Lea (1973) observed what may be called a "rise-fall dichotomy;" that is, the *F*0 contour following a voiced obstruent rises from a lower value at voice onset to a higher value at midvowel, while the *F*0 contour following a voiceless obstruent slopes down from voice onset to midvowel. However, for the second set of recordings, in which the medial consonant could precede or follow a stressed syllable, the results were more complicated, and Lea (1973) concluded that whether the *F*0 contour rises or falls at a CV transition is an interaction of both stress and segmental context.[1]

Hombert (1978) also focused on the time course of pitch skip rather than on peak or average vowel *F*0. He obtained *F*0 contours for several speakers of American English and, just as earlier studies had shown, found that vowel *F*0 following voiceless stops was higher than that following voiced stops, with the greatest difference occurring at vowel onset. When the *F*0 contours were averaged across all speakers, he observed the rise-fall dichotomy. However, Hombert (1978) noted that there were individual differences in the details of the dichotomy. For example, the *F*0 contour following voiceless consonants did not slope down for all speakers. Hombert (1977, 1978) also observed pitch skip for two speakers of the tone language Yoruba. While the dichotomy in *F*0 following voiced v. voiceless stops was observed for all three tones (high, mid, and low) of that language, it was greater for the high tone, both in magnitude and duration.

In a study on American English, Ohde (1984) recorded data both in isolation and in carrier phrases. The stimuli were the six voiceless aspirated and voiced stops paired with five vowels (/i,e,u,o,ɑ/) to form symmetrical CVC syllables. He observed that although the *F*0 contour following voiceless stop consonants is higher than that following voiced stop consonants, it cannot be described as a rise-fall dichotomy. Rather, the *F*0 contour falls after both voiced and unvoiced consonants for his three male subjects.

Most of the studies described thus far have observed *F*0 on CVC syllables in isolation or embedded in carrier sentences, for example, "Say CVC again." That is, the target syllable is in focus and most likely carries the phrase-level prominence and a high-pitch accent. Kohler (1982) was one of the first to consider that pitch skip, which he referred to as a type of "microprosody," might interact with the phrase-level tone variations, or "macroprosody." He recorded German words with medial stop consonants embedded in falling, rising, and monotone *F*0 contours. When the *F*0 contour was rising or monotone, the vowel *F*0 preceding the stop consonant was not affected, while the vowel *F*0 following the consonant showed the expected dichotomy (being higher after voiceless stops than after voiced). However, when the *F*0 contour was falling, the usual dichotomy was observed on the preceding vowel rather than on the following. Kohler (1982) proposed that utterance-level intonation can sometimes cancel the segmental-level pitch-skip effects.

Silverman (1984) observed the effects of voiced and voiceless obstruents on both preceding and following vowels in Southern British English. The obstruents were embedded in three-syllable nonsense words (e.g., /ə′pi:pi:p/), which were embedded in carrier phrases. The subjects were instructed to place lexical stress on the middle syllable of the nonsense word. For two male subjects, Silverman (1984) observed an effect of consonant voicing on the *F*0 of vowels that both precede and follow the consonant, with stressed syllables displaying a greater effect. He also found that *F*0 falls after both voiced and voiceless obstruents, in line with the Ohde (1984) data on stops.

A likely explanation for the observation of a rise-fall dichotomy by some researchers, versus a "fall-fall" dichotomy by others was proposed by Kohler (1985). He noted that the utterance types used by Hombert (1978), for example, favor an F0 contour that rises and falls on the target word, while the utterances used by Ohde (1984), for example, will most likely be produced with an *F*0 contour that falls throughout the target word. Therefore, it is not surprising if the perturbations introduced to these globally differing contours appear to differ.

Jun (1996) reported data for American English, Korean, and French. In her study, the intonation contour was varied such that the target consonants (both voiced and unvoiced) either carried a nuclear pitch accent or occurred before or after the nuclear pitch accent. For American English, she observed a dichotomy similar to that described by Ohde (1984), with the effect being somewhat less in certain intonational contexts, confirming the assertion of Kohler (1982) that the intonational contour interacts with segmental effects on *F*0. For the French data, a dichotomy also occurred, but *F*0 tended to rise following the voiced consonants, rather than fall. Korean has three stop series (lenis, tense, and aspirated), none of which tend to be voiced (although the lenis stops are voiced in certain environments). Jun (1996) found that the *F*0 contours following tense and aspirated stops were similar to each other, while the *F*0 contour following the lenis stops was more similar to the sonorants /m,l/. In addition, the difference in *F*0 following the lenis v. tense and aspirated stops persisted throughout the vowel in certain prosodic contexts, unlike what has generally been observed in other languages. Jun (1996) saw this effect as evidence that pitch skip is being phonologized at a prosodic level.

Similarly, Jessen and Roux (2002) found that pitch skip following voiced and unvoiced stops and clicks in Xhosa persists at least to midvowel. They theorized that this *F*0 dichotomy is phonologized to cue the voicing distinction because the voiced stops typically do not manifest much actual

voicing during the stop closure, and there are no other phonetic cues to voicing. Their target words carried low lexical tone; it is possible that the pitch skip is extended to enhance the weak acoustic cues to the voicing distinction [e.g., see Keyser and Stevens (2006) and Kingston and Diehl (1994)] and may lead to an eventual tone splitting (Matisoff, 1973).

Most of the work we have summarized compared *F*0 contours for vowels following, and sometimes preceding, voiced and voiceless obstruents. Some studies have also compared *F*0 behavior following different series of voiceless obstruents that can occur in languages. Danish, for example, has two stop series, both of which are voiceless; one series is aspirated and the other is not. Jeel (1975) and Reinholt Petersen (1983) found that in Danish, the *F*0 contour following both voiceless series is usually raised compared with the *F*0 contour following sonorants and /v/, although individual speakers differ in the degree of the dichotomy. *F*0 following the aspirated series tended to be somewhat higher than that following the unaspirated series, although again, speakers varied somewhat. Ohde (1984) had a similar result for *F*0 following aspirated stops compared to *F*0 following the unaspirated voiceless stops in /s/-stop clusters in American English.

In a study of Korean stops, Han and Weitzman (1970) observed that the *F*0 contour following so-called weak stops starts at a lower frequency than that following the strong stops. These results are in line with the Jun (1996) results for lenis stops v. tense and aspirated stops, described earlier. However, neither Jun (1996) nor Han and Weitzman (1970) specified whether a consistent difference occurred between the aspirated and tense/strong stops. On the other hand, Xu and Xu (2003), in a study of Mandarin, compared *F*0 following voiceless aspirated and unaspirated stops, and found that *F*0 following the aspirated stops was *lower* than that following the unaspirated stops. Francis *et al.* (2006) found the same result for Cantonese. Thus, the results of studies comparing aspirated voiceless stops to unaspirated tend to show a dichotomy, but the direction of the observed dichotomy is not consistent.

As our literature review of production data shows, methods of studying the pitch-skip phenomenon have evolved over the years, from studies that observe the average or maximum *F*0 over an entire vowel to studies that observe the time course, and from studies that give little consideration to higher-level pitch phenomenon to studies that do consider the interaction of segmental, lexical, and phrase-level *F*0. While some details of the results of the summarized studies are not consistent, it is clear that there is often a dichotomy in the *F*0 contour on vowels following contrasting series of obstruent consonants, and this dichotomy occurs in most, if not all, languages that have been studied. The dichotomy is not so much one of the slopes of the *F*0 contour at vowel onset as much as it is one of the absolute difference in *F*0 over the first few tens of milliseconds of the vowel. Although most studies indicate that the effect of obstruent consonants on vowel *F*0 occurs for vowels that follow these consonants, some studies have found evidence suggesting that, in the right context, the *F*0 of preceding vowels can also be affected.

2. Perception data

It has been shown through perception experiments that perturbations to *F*0 associated with voicing contrasts of obstruents provide cues to a listener concerning the consonantal voicing characteristics. For example, Haggard *et al.* (1969) and Fujimura (1971) observed that listeners were more likely to perceive a synthesized stop consonant as being voiceless when it was followed by a high *F*0 than a low *F*0. Massaro and Cohen (1976) observed a similar result for synthesized fricatives. Furthermore, Whalen *et al.* (1993) reported that listeners use the *F*0 cue even when voice-onset time is unambiguous. Therefore, pitch skip is not just an interesting artifact but a cue that speakers can extend or exaggerate to strengthen the perceptual saliency of a voicing contrast.

Through careful manipulation of both the global *F*0 contour and the portion of the *F*0 contour local to stop consonants, Kohler (1985) used synthesized German words to show that *F*0 *prior* to a stop consonant can also influence whether the stop is perceived as being voiced or unvoiced.

3. Source of pitch skip

Theories as to why this dichotomy occurs fall into two camps. Kingston and Diehl (1994) suggested that speakers intentionally lower *F*0 following a voiced obstruent relative to *F*0 following voiceless obstruents to signal the [+voice] feature to listeners. On the other hand, the dichotomy may simply fall out from the physiology or aerodynamics of obstruent production. For example, Halle and Stevens (1971) suggested that *F*0 is modified by obstruents because the vocal folds are stiffened to inhibit glottal vibration during a voiceless obstruent and are slackened to facilitate glottal vibration during a voiced obstruent. These stiffening or slackening gestures then carry over into the adjacent vowel. This view gained support from Löfqvist *et al.* (1989), who found that cricothyroid (CT) muscle activity increases for voiceless consonants relative to voiced consonants. An increase in CT activity increases the longitudinal tension of the folds, which not only increases the frequency of vibration, but also inhibits voicing in certain situations [see, for example, Löfqvist *et al.* (1989)].

Hombert *et al.* (1979) discussed the possibility of an aerodynamic basis of pitch skip. Reduced transglottal pressure at stop release would result in a lowered *F*0 for voiced stops, while the increased airflow due to the spread vocal folds of voiceless unaspirated stops would change the Bernoulli effect such that *F*0 is increased. However, for English at least, the *F*0 perturbation can last up to 100 ms into the following vowel, and therefore Hombert *et al.* (1979) ruled out the possibility of an aerodynamic basis.

Kohler (1985), on the other hand, disagreed with this conclusion; in experiments with monotone recordings, he found that there was no perturbation prior to a voiceless stop consonant, but that there was a perturbation following such a consonant. Because speakers were carefully controlling *F*0 to be constant, he claimed that the perturbation following the consonant could not be due to a change in vocal-fold tension but would have to be due to an intrinsic aerodynamic property of stop releases. Note, however, that Kohler (1985) did

not deny that vocal-fold tension can play a role in pitch skip. Rather, he proposed that an effect of vocal-fold tension is superimposed on an aerodynamic component, and because the aerodynamic component is intrinsic, it always occurs. He claimed, however, that the vocal-fold tension component will be overridden in certain intonation environments, such as monotone speech production.

Based on their study of Cantonese, Francis *et al.* (2006) have also suggested that pitch skip is composed of both intrinsic and controlled components. In particular, they proposed that in Cantonese, speakers sharply reduce the extent to which vocal-fold tension in obstruents is allowed to overlap with following vowels, in order to preserve the tone of a syllable, but suggested that in English, speakers extend the influence of obstruents on following vowels as a way of enhancing a voicing contrast.

B. Current work

The work presented in this paper was undertaken as part of a project to improve the generation of *F*0 contours in rule-based speech synthesis by including physiological constraints implemented in the HLsyn speech synthesizer (Stevens and Bickley, 1991; Hanson and Stevens, 2002). In HLsyn there is a parameter dc (delta compliance) that can be manipulated to produce the effect of stiffening or slackening the vocal folds. This parameter tends to either inhibit or facilitate glottal vibration during the obstruent consonants, when there is a reduced transglottal pressure difference. A change in dc for the consonant can carry over into the following vowel, and a perturbation due to dc is superimposed on parameter *f*0 in the vowel (along with other *f*0 perturbations due to intrinsic vowel effects and subglottal pressure).

One of our goals was to quantify the shifts in *F*0 due to obstruent effects and the time course of these shifts for different speakers. Previous studies have compared *F*0 contours following voiced and voiceless obstruents but tend not to compare these to contours following a neutral segment, that is, a segment not expected to perturb the contour [but see Hombert (1978) and Löfqvist *et al.* (1995)]. Therefore, it is difficult to use these data to model precisely how much *F*0 is raised or lowered relative to its unperturbed value. A second goal was to compare the effects of stop consonants to those of fricatives; most previous studies have focused on stop consonants only. A third goal was to observe the interaction of phrasal intonation and obstruent intrinsic *F*0. Previous studies have mainly looked at syllables carrying high-pitch accents [but see Kingston and Diehl (1994) and Silverman (1986)]. However, if intrinsic obstruent *F*0 is due to CT activity during obstruent production, in certain contexts such activity would conflict with changes in vocal-fold tension necessary to vary intonation. In this case, obstruent effects on *F*0 might be more or less obvious in different *F*0 environments. On the other hand, if speakers are intentionally lowering *F*0 following voiced consonants, the obstruent effects on *F*0 should be similar in all pitch environments.

In the next section we describe the corpus, subjects, recording procedure, and signal-processing methods for the study, and in Sec. III we present the results. Section IV presents a discussion of the results and interprets them within the framework of featural representation and enhancement theory. Finally, we summarize our findings and present some ideas for future work in Sec. V.

II. DATA COLLECTION AND PROCESSING

A. Corpus

Target syllables were formed by combining the consonants and consonant clusters /m,v,t,f,s,b,d,g,p^h,t^h,k^h, sp,st,sk/ with /im/ and /ɑm/ to form 28 syllables. Shadle (1985) reported that vowel intrinsic pitch was dependent on sentence position. To test whether sentence position would influence obstruent I*F*0, the target syllables in this study were inserted into the carrier phrase "My (s)CVm is called my (s)CVm again." The target syllables in each sentence were identical. In order to get the target syllables produced in different *F*0 environments, it was intended (by the experimenter) that the subjects produce the sentences with three intonation contours and two phrase-level prominence[2] patterns:

(1) Both target words are prominent. We intended speakers to produce pitch-cued prominence on the target words, with high *F*0 on the first target word, followed by a fall in *F*0 to the second target word, followed by a rise in *F*0 through the last word ("again").
(2) As with (1), it was intended (by the experimenter) that the target words would be produced with pitch-cued prominence; however in this case *F*0 would be higher on the first "my" than on the first target word, followed by a rise in *F*0 to the second target word, followed by a fall in *F*0 through the last word ("again").
(3) Ideally, neither target word would be prominent; rather the *F*0 contour would remain flat throughout the sentence until the last word, when *F*0 would be increased, reaching a peak in the second syllable of "again," to result in pitch-cued prominence on that word.

Schematics of the intended intonation contours are illustrated in Fig. 1.

We are purposely avoiding the use of terms associated with a particular intonational labeling system, such as pitch accent, under the belief that it is not relevant whether subjects produce prominences that are judged by listeners to be pitch accented or not; what is important is that the words were produced when the subject was also producing either high, low, or neutral *F*0, for whatever reason.

In sum, the goal was that there would be 84 sentences in the corpus; that each target word would occur in either high, low, or neutral *F*0 environment; and that each target word would occur in either early or late sentence position. We say more about how subjects were prompted to produce the intended *F*0 contours and prominence patterns, and about how successful our attempts were, in Secs. II C and II D 3.

B. Subjects

There were five female and five male subjects, all adults and all native speakers of American English. Nine of the

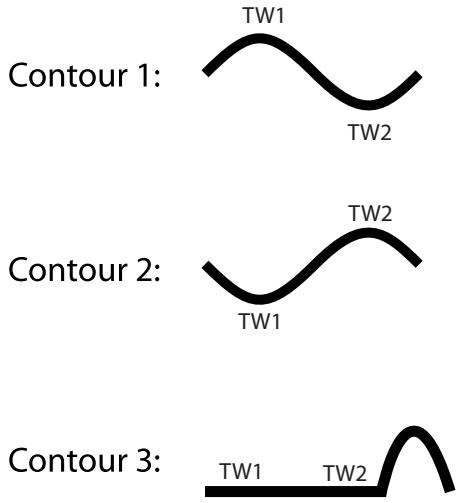

FIG. 1. Schematics of the *F*0 contours that the subjects were intended to produce. The abbreviations "TW1" and "TW2" refer to "target word 1" and "target word 2." See the text for more details.

subjects (including the author) were experienced as subjects in speech experiments (subject M5 was not experienced). Two of the subjects (F5 and M5) spent some of their early years in countries where English is not the native language but spoke English with no discernible accent. Except for the author, the subjects were paid for participating in the experiment.

C. Recording

Recordings were made in a sound-attenuated booth, either to digital audiotape (DAT) (48 000 samples/s) or directly to a hard drive (16 000 samples/s). Subjects were seated about 8 in. (20.3 cm) from an omnidirectional microphone. Prior to the recording, the experimenter demonstrated the intonation contours, and the subjects practiced producing a subset of the sentences until they were comfortable with the intonation contours and the target syllables. During the recording, each of the 84 sentences in the corpus was produced six times, yielding 504 utterances. To make the recording session easier for the subjects, blocks of 28 utterances were produced with a single intonation contour. The experimenter demonstrated the intonation contour at the beginning of a block, and the subject practiced it a few times before restarting the recording. Subjects could view a schematic of the desired contour throughout the block. Overall, subjects found contours 1 and 2 to be easy to produce, while contour 3 was somewhat unnatural. Subjects did not produce the same contour for consecutive blocks. Sentences were randomized within each block, and the randomizations were different for each subject. Care was taken so that a given target word occurred no more than once in the first or last position of a block. Subjects took a short break between each block and were offered the chance to take longer breaks as needed.

D. Processing

The data were downsampled to 10 000 samples/s using MATLAB software. Further processing was performed as follows.

1. Computation of *F*0 contours

The PRAAT software was used to label pitch periods. An autocorrelation-based algorithm included with Praat was used to estimate the pitch periods. These estimates were then corrected manually. The labels were converted to *F*0 contours and smoothed using software written in the C programming language (Xu, 1999). The smoothing algorithm compared each point in the *F*0 contour to its two neighbors; if the point was either higher than or lower than both of its neighbors (by 0.1 Hz or more), it was replaced by the average of the two neighbors. This method is effective for removing spikes that sometimes occur in *F*0 contours at the transition from one segment to the next. The smoothed F0 contours were then resampled at 5-ms intervals using linear interpolation.

2. Labeling of vowel onset

For each utterance, the starting time t_0 of the vowel in the target syllables was identified and tabulated. For the nasal and the voiced fricatives, t_0 was aligned with the release of the supraglottal constriction as identified by observations on the waveform and spectrogram. Specifically, for nasals we looked for changes in the waveform and spectrogram that indicated that formants higher than F1 were being more strongly excited; in the waveform, this stronger excitation manifests itself as higher frequency modulations on the F1 oscillations, and in the spectrogram it manifests as stronger energy at high frequencies. The CV transition for a nasal to a vowel was generally quite easy to label. For the voiced fricatives, we looked for points in the waveform where amplitude began to increase rapidly and the waveform became more periodic and less noisy. In the spectrogram these same properties were a rapid increase in energy and a switch from noise to harmonic excitation at high frequencies. In some cases the voicing ceased partway through the voiced fricatives. In these cases, t_0 was labeled at voice onset. For the stops and the unvoiced fricatives, t_0 was labeled at the onset of voicing. For voiced stops that were voiced throughout (meaning that voicing continued from the closure right through the release and into the vowel), t_0 was labeled at the point following the burst where the amplitude of voicing began to increase rapidly (similar to the way that voiced fricatives were labeled when they were voiced throughout).

3. Contour analysis

The *F*0 contours were evaluated to determine if subjects were successful in producing the target words in the desired *F*0 environments, and if they were consistent in doing so. This evaluation was purely visual and was performed by the author and some undergraduate students who assisted in the data processing.

The *F*0 contour for each utterance was examined and compared to the *F*0 contours for other tokens of the same sentence. Subjects were remarkably consistent with regard to utterance duration and shape of the *F*0 contour. Figure 2 shows the individual tokens for contours 2 and 3 of subject *F*1, target syllable "pom," limited to the target words in late sentence position. For contour 3, the target word was pro-

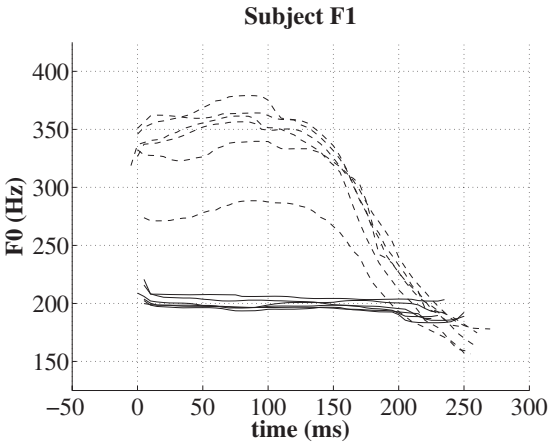

FIG. 2. *F*0 contours for the target syllable "pom" in late sentence position. Solid lines indicate contour 2 and dashed lines indicate contour 3. The speaker is subject F1.

duced in a neutral pitch environment, and it can be seen that there is little variation in average *F*0 among the tokens. For contour 2, the target word was produced in a high-pitch environment; here we observe a similar shape of the contours but great variability as to the maximum *F*0 at midvowel. This variability in maximum *F*0 was common for the words produced in high-pitch environment for both contours 1 and 2, and for all speakers. Note, however, that for this speaker, five of the contours cluster relatively closely together, while one of them deviates quite a bit from the others. Such outliers were not included in the analysis. Closer analysis revealed that token 1 was frequently an outlier, suggesting that the first round through each intonation contour was a kind of warmup for the speakers. In the later rounds the subjects were more comfortable with the target words and the intonation contours, and were therefore more consistent and natural in their productions. For this reason, the first repetition of each sentence (i.e., the first 84 sentences recorded) was not used for analysis, unless one of the later tokens (tokens 2–6) of the sentence was not usable. In the latter case, token 1 was used only if it did not deviate markedly from the other tokens. We note again that these judgments were purely subjective, based on visual inspection, but we are confident that we avoided the inclusion of outlier tokens.

Figure 3 compares intonation contour 1 for each speaker. Token 5 of the sentence "My mom is called my mom again" was used to generate the *F*0 contours. As intended, the first target word is produced with an *F*0 that is higher than that of the second target word. There is a great deal of variability among the subjects in the shape of the peak, the maximum *F*0, and the total *F*0 range. However, all the subjects seem to align their maximum *F*0 at or near the end of the vowel. Thus, all subjects produce a rising *F*0 on the vowel, but the degree of the rise varies greatly. Also, as intended, the second target word is produced with an *F*0 that is low in each speaker's pitch range, and that is lower than the word "again."

A similar plot for contour 2 is presented in Fig. 4. The *F*0 of the first target word is produced low in a speaker's pitch range, and *F*0 of the second target word is high in their range. Again, the maximum *F*0 is aligned at or near the end of the vowel, and there is great variety among the speakers as to the *F*0 range and shape of the peak. Finally, Fig. 5 shows examples of contour 3 for each speaker. The speakers all produced a level *F*0 contour up to the last word of the sentence, as hoped, but comparison with Figs. 3 and 4 show that *F*0 on the target words is about the same as the *F*0 produced in the low-*F*0 environment. Thus, we were not successful in getting an *F*0 that was intermediate between the high and low *F*0.

Subjects F5 and M5 were the least experienced at speech production experiments, but based on the contours in Figs. 3–5 their *F*0 contours are not grossly different from the more experienced subjects. In fact F5's contours are quite similar to those of F2, who is quite experienced. Similarly, M5's contours closely resemble those of M2, another well-experienced subject.

4. Averaging of F0 contours

For each subject, the ***F***0 contours for tokens 2–6 of each sentence were aligned at the sample occurring closest to t_0 (recall that the *F*0 contours were sampled at 5-ms intervals, while t_0 is essentially a continuous variable). The contours were averaged at points for which more than half of the contours had an *F*0 estimate (at other points the average *F*0 was discarded; these points were generally during obstruent closures when voicing ceased, but also sometimes at the ends of voiced segments when one token might be longer than others). Note that this alignment and averaging was done twice for each sentence: once for the first target syllable and once for the second target syllable. Informal observations of the averaged data for each sentence indicated that the data could be further reduced within a given intonation contour by averaging across place of constriction and across vowel. That is, while future work may show that there are significant differences in the magnitude of pitch skip due to vowel intrinsic pitch or to place of articulation, the general effect appears to be the same despite vowel or place, and we focus on that general effect in this paper. Averages across subject were also obtained, once for male subjects only and once for female subjects. For both of these additional types of average *F*0 contours—place/vowel and subject—averages were computed at points for which more than half of the contours had an *F*0 estimate. Thus, points in an average *F*0 contour across place and vowel include 5–10 data points for /m/, 10–20 for the fricatives, and 15–30 for the stops. Each point in an average *F*0 contour across subject includes 25–50 data points for /m/, 50–100 for the fricatives, and 75–150 for the stops.

5. Normalization of F0 contours

While *F*0 values at midvowel were not always observed to be the same across the voiced and unvoiced obstruents, they also did not seem to have systematic differences, as has been reported for some languages [e.g., see Jun (1996) and Cho *et al.* (2002)]. This observation leads us to believe that in general, English speakers have the same target *F*0 at midvowel for vowels following voiced and unvoiced obstruents. To prevent small differences in *F*0 at midvowel from obscuring or exaggerating obstruent effects at vowel onset, the *F*0

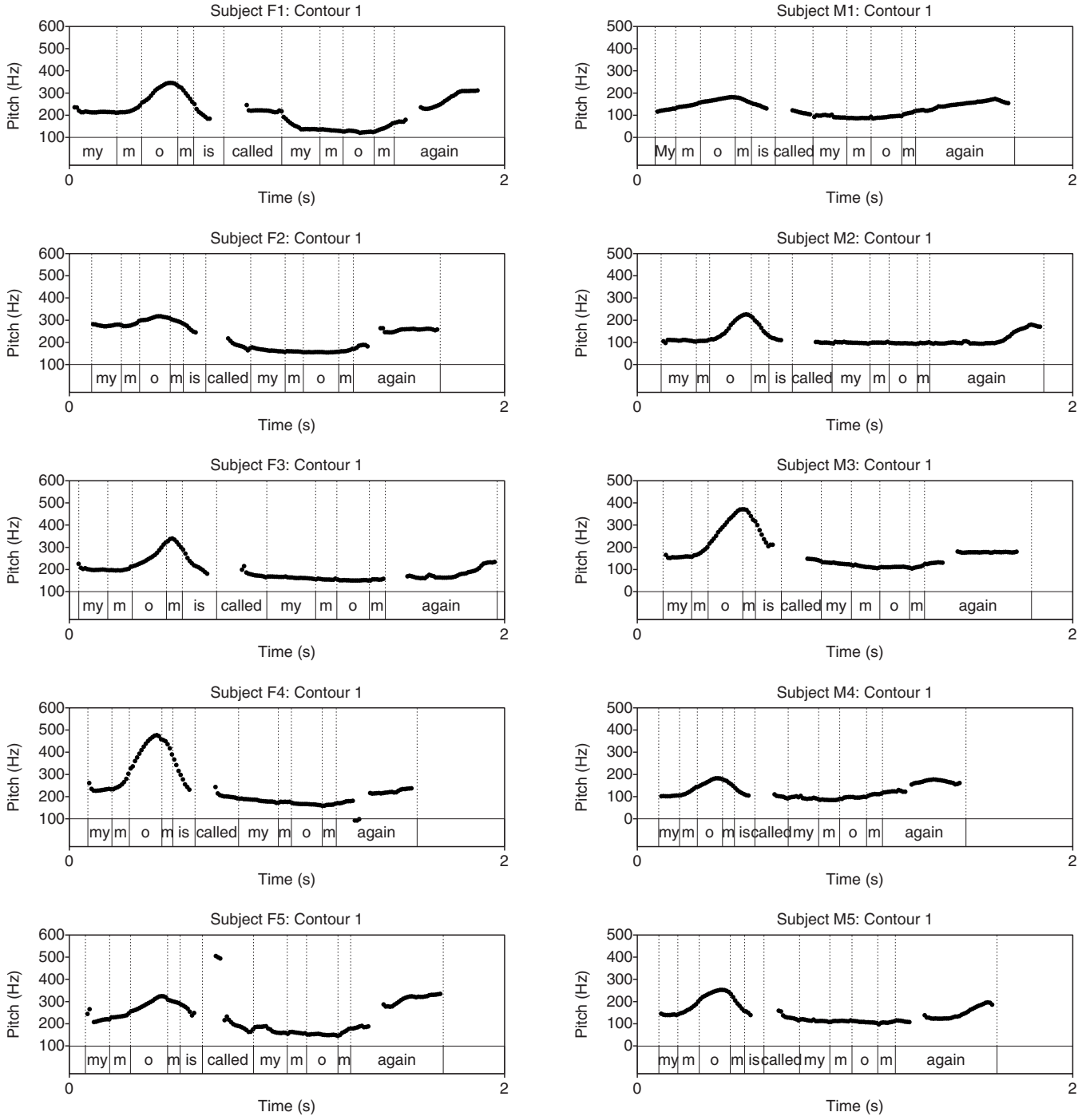

FIG. 3. Examples of the sentence "My mom is called my mom again" produced with intonation contour 1. The *F*0 contour for token 5 is shown for each subject. Female subjects are in the left column and male subjects are in the right column. (Note that these are "raw" *F*0 contours, which do not include the smoothing and manual corrections that were applied to the *F*0 contours used for the analysis.)

contours for the obstruent data were normalized in frequency to be similar to the *F*0 contours for the baseline /m/ data at midvowel. Normalized *F*0 at each point is

$$\widetilde{F0}(n) = F0(n) + N_{\text{nasal}} - N_{\text{obstruent}},$$

Where n=sample point, N_{nasal}=normalization factor for the baseline nasal, and $N_{\text{obstruent}}$=normalization factor for the obstruents.

The normalization factor N is obtained by averaging the *F*0 contour over a 50-ms window centered at either the peak *F*0 (most target syllables carrying a high *F*0) or 100 ms into the vowel (target syllables carrying low or neutral *F*0, for which the *F*0 contour tended to be flat, and target syllables carrying high pitch for which the *F*0 is flat or slopes down into the vowel). (The choice of 100 ms was based on our observations of where speakers tended to have their maximum *F*0.) Separate normalization factors were computed for contours averaged (1) across token, (2) across place, and (3) across subjects. Examples of this normalization are shown in Fig. 6.

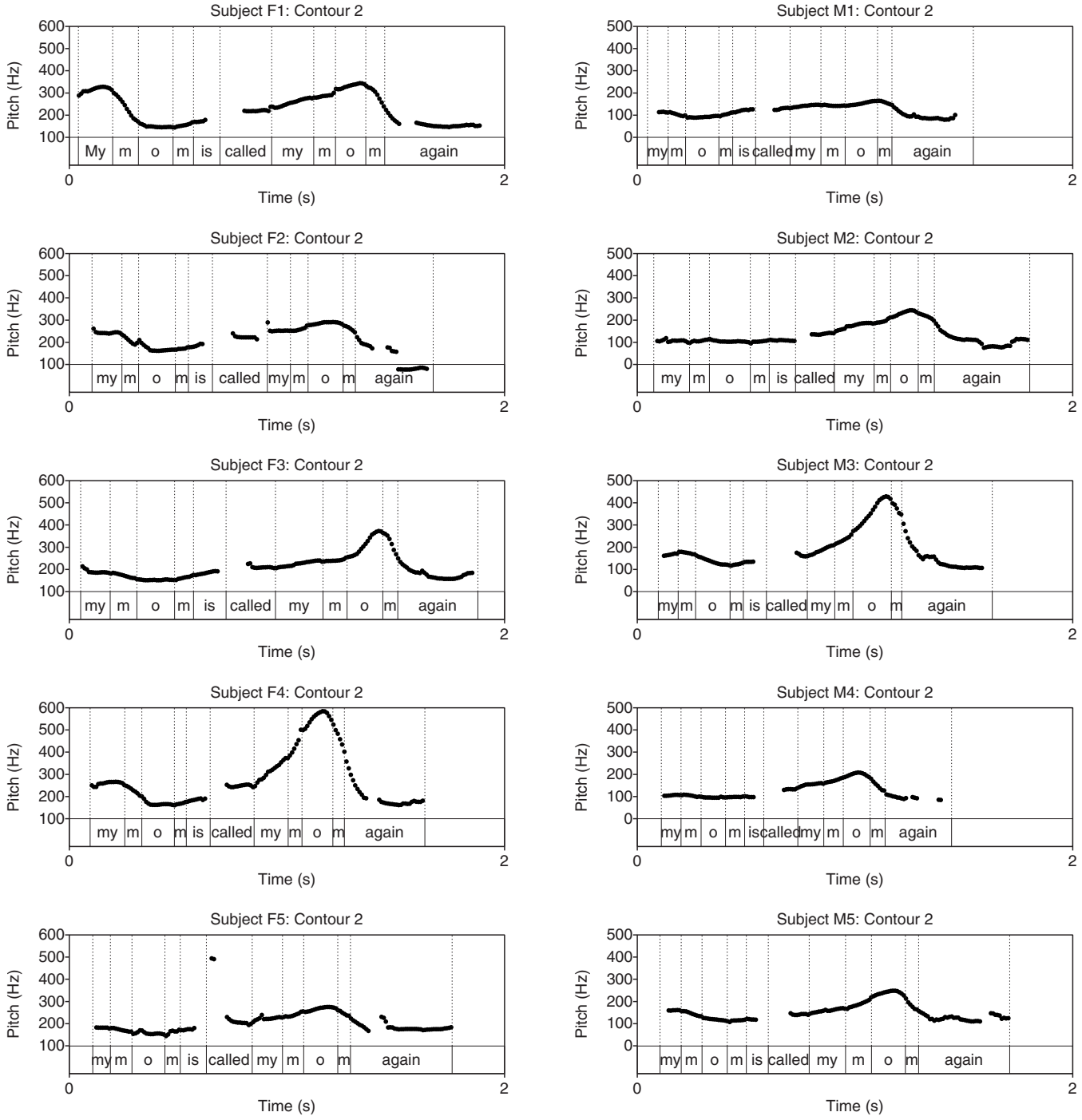

FIG. 4. Examples of the sentence "My mom is called my mom again" produced with intonation contour 2. The *F*0 contour for token 5 is shown for each subject. Female subjects are in the left column and male subjects are in the right column. (Note that these are raw *F*0 contours, which do not include the smoothing and manual corrections that were applied to the *F*0 contours used for the analysis.)

III. RESULTS

We first present results based on *F*0 contours averaged across subject (divided into male and female subjects). These results give us a general picture of the pitch-skip phenomenon. However, while all of the subjects exhibit the same general pattern, differences in the details of this pattern occur among them. Therefore, following the presentation of the general, across-subject results, we more closely examine the results for individual speakers.

A. Averages across subjects

Average *F*0 contours for the voiced and voiceless obstruents and the nasal consonant in the three pitch environments are summarized in Figs. 7–9.[3] Beginning with Fig. 7, the obstruents in high-pitch environment, we note that, as expected, *F*0 following the voiceless obstruents is raised relative to that of the voiced obstruents and the nasal consonant for about 100 ms into the vowel. However, *F*0 following the voiced obstruents tends to closely trace the baseline

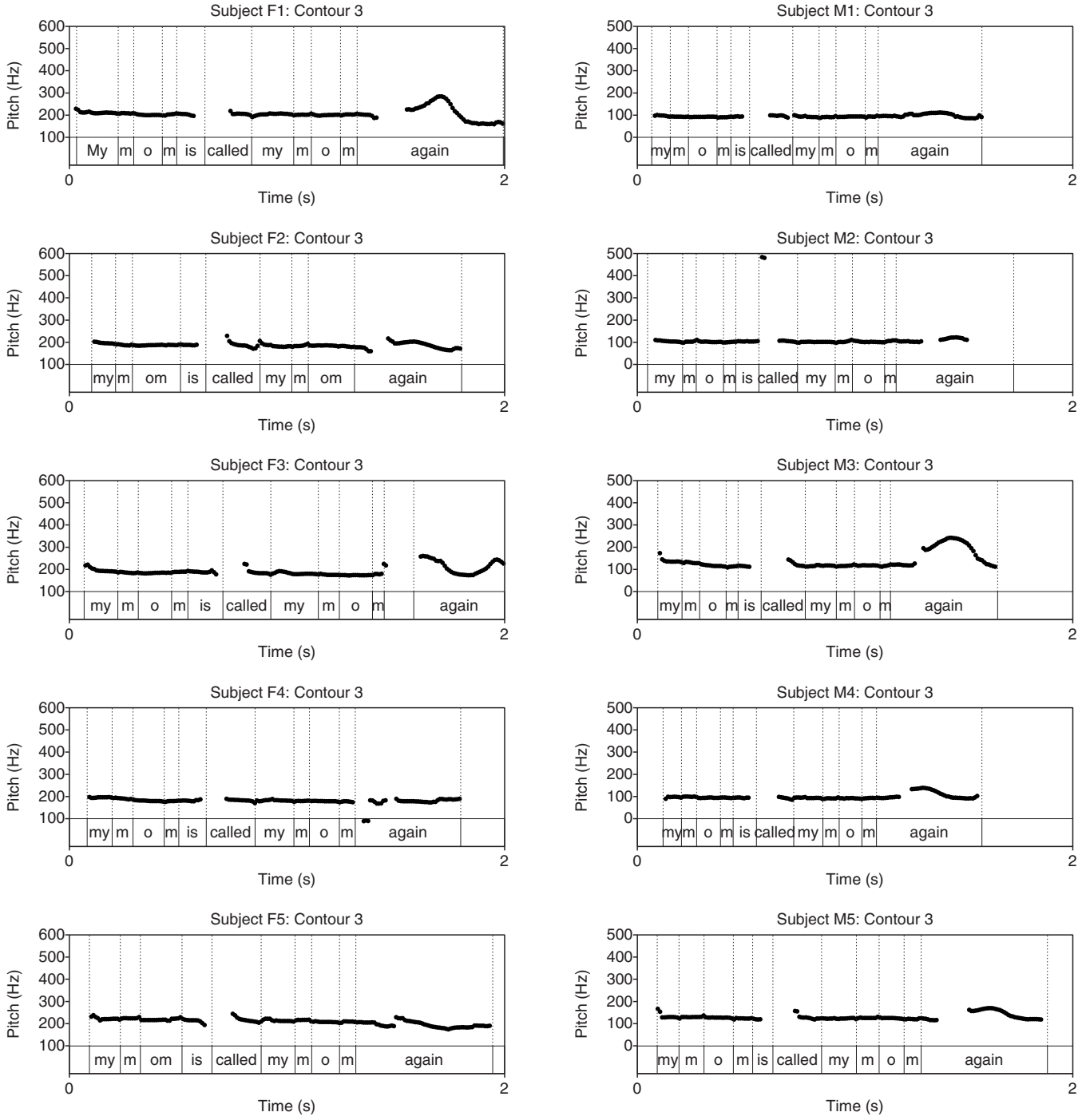

FIG. 5. Examples of the sentence "My mom is called my mom again" produced with intonation contour 3. The *F*0 contour for token 5 is shown for each subject. Female subjects are in the left column and male subjects are in the right column. (Note that these are raw *F*0 contours which do not include the smoothing and manual correction that were applied to the *F*0 contours used for the analysis.)

*F*0 following the nasal consonant, whereas one might have expected it to be lowered relative to the baseline (see Sec. I A 1). These results are consistent for both stop and fricative obstruents and for male and female speakers. Thus, assuming that the nasal consonant is truly neutral, it would seem that pitch skip is a phenomenon that applies only to voiceless obstruents and the vowels that follow them. We note also that there is little difference between the aspirated and unaspirated stop consonants.

We can also compare the results for the target syllables that occur early and late in the utterances. It appears that pitch skip occurs to a lesser degree for the target syllables late in the utterances. It is not clear if this effect is truly a function of syllable position because we note that the baseline *F*0 starts about 40 Hz higher for the late syllables than for the early syllables. For the female subjects, the maximum pitch in the target syllable is lower for the late syllable than for the early syllable. These two observations suggest that location in the pitch range or pitch excursion could also be a factor. We return to this point in the next section.

The *F*0 contours for the consonants in low-pitch environment (Fig. 8) seem to tell another story. There are few

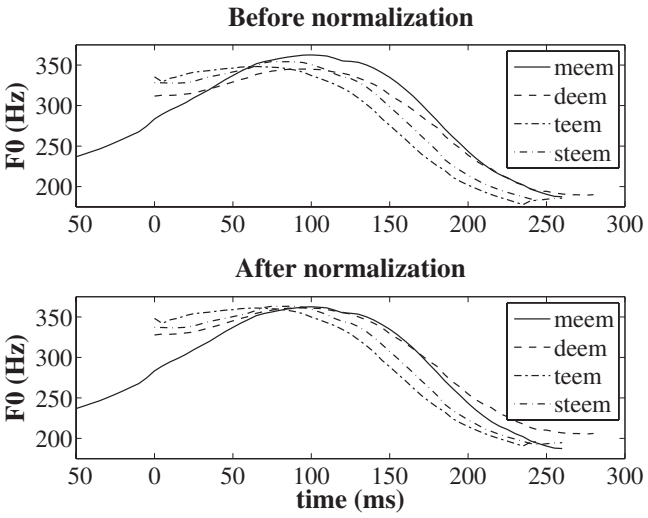

FIG. 6. An example of how averaged *F*0 contours for obstruents were normalized in frequency relative to averaged *F*0 contours for nasals (the baseline /m/ *F*0 contours). The example is for target syllables produced in a high-*F*0 environment by a female speaker. Top panel: averaged contours before normalization. Bottom panel: averaged contours following normalization. Note that $t=0$ in this and later graphs corresponds to the time t_0 of vowel onset for the CVm target syllables.

differences among the *F*0 contours for voiced and voiceless consonants, except in the first 10–15 ms, and those differences that occur are quite small. Again, this result is consistent for both male and female speakers and both stop and fricative consonants. The results for the consonants in the neutral pitch environment (Fig. 9) are nearly identical to those for the low-pitch environment. This similarity between the low and neutral *F*0 results is not surprising, given that the subjects produced similar *F*0 values for these target syllables. The contrast of the results for low and neutral *F*0 with those for the obstruents in high-pitch environment is quite striking. As Kohler (1982) suggested, there appears to be an interaction of segmental and phrase-level *F*0 effects; we discuss this result further in Sec. IV.

B. Comparisons among individual subjects

While the data for most of the subjects are qualitatively similar to the average *F*0 contours presented in Sec. III A, the details of the effects differ somewhat among the subjects. Although we do not think that these variations weaken the general result, it is worth acknowledging and describing them.

1. Degree of pitch skip in the three F0 environments

The general result is that pitch skip occurs in high-*F*0 environment for unvoiced obstruents only. However, some speakers deviate from this model in certain ways. For subject M1, the *F*0 contour following voiceless obstruents is only slightly higher than the baseline in all three *F*0 environments, as illustrated in Fig. 10 for fricatives early in the utterance. While it was generally true that *F*0 following voiced obstruents in a high-pitch environment closely followed the baseline *F*0 (following /m/), some subjects did exhibit an *F*0 that tended to be higher or lower than that following /m/ in certain environments. For example, subject F3's *F*0 following voiced obstruents early in an utterance tended to be lower than that following /m/, as illustrated in Fig. 11 for fricatives.

2. Aspirated stops v. unaspirated stops

Four of the subjects (F1, F2, F3, and M3) did show some difference between aspirated and unaspirated stops. For all four, *F*0 following aspirated stops tended to be somewhat higher than that following unaspirated stops when the target syllables occurred early in the utterance. Subjects F2 and M3 also exhibited this effect for the syllables occurring late in the utterances. Subject M3 displayed a particularly large effect, as seen in Fig. 12. Both the direction of the dichotomy and the variability among speakers are in line with previously reported data from Danish (Jeel, 1975; Reinholt Petersen, 1983) and American English (Ohde, 1984), described in Sec. I A 1.

3. Utterance position effect

In Sec. III A we mentioned that the degree of pitch skip in the high-*F*0 environment appeared to be stronger when the target syllables occurred early in the utterance than when they occurred later (see Fig. 7). Observation of the data for individual speakers suggests that most subjects do show some degree of change as a function of position (i.e., a smaller degree of pitch skip for late syllables), but the size of this change varies greatly across subjects. Figure 13 includes comparisons of *F*0 contours in early and late positions following fricatives for three subjects—M3 (no change with position), F1 (intermediate change with position, compared to other subjects), and M4 (extreme change with position, compared to other subjects). However, utterance position may not be the actual or only source of this apparent effect. We have observed for the baseline /mVm/ syllables that the maximum *F*0 in the vowel ($F0_{\max}$) tends to be lower in a subject's pitch range on the later syllables, while the *F*0 at vowel onset $[F0(t_0)]$ tends to be *higher* in a subject's pitch range. That is, for syllables starting with /m/ in a high-*F*0 environment, the pitch excursion $\Delta F0_m = F0_{\max} - F0(t_0)$ is more compressed for the target syllables late in the utterances than early in the utterances. There is, in effect, less "room" for pitch skip to occur.

This effect is illustrated in Fig. 14. Figure 14(a) is a bar chart comparing the pitch excursions $\Delta F0_m^{\text{Early}}$ and $\Delta F0_m^{\text{Late}}$ (averaged across vowel) for the /mVm/ syllables in the early and late positions, respectively, for each subject. Only one subject (M3) has a larger pitch excursion for the late syllable than for the early syllable. Figure 14(b) is a bar chart showing the difference in pitch excursion $\Delta F0_m^{\text{Late}} - \Delta F0_m^{\text{Early}}$ for each subject. As expected from part (a), the difference is negative for all subjects but M3. The two subjects showing the smallest differences (F2 and M3) are also the subjects least likely (as judged by the author) to show a change in degree of pitch skip for early v. late utterance position. What is most striking about part (b) is the degree of variation among the subjects, from subject F4 whose pitch excursion is compressed by almost 120 Hz, to subject M3, whose pitch excursion expands slightly. We note also that these changes

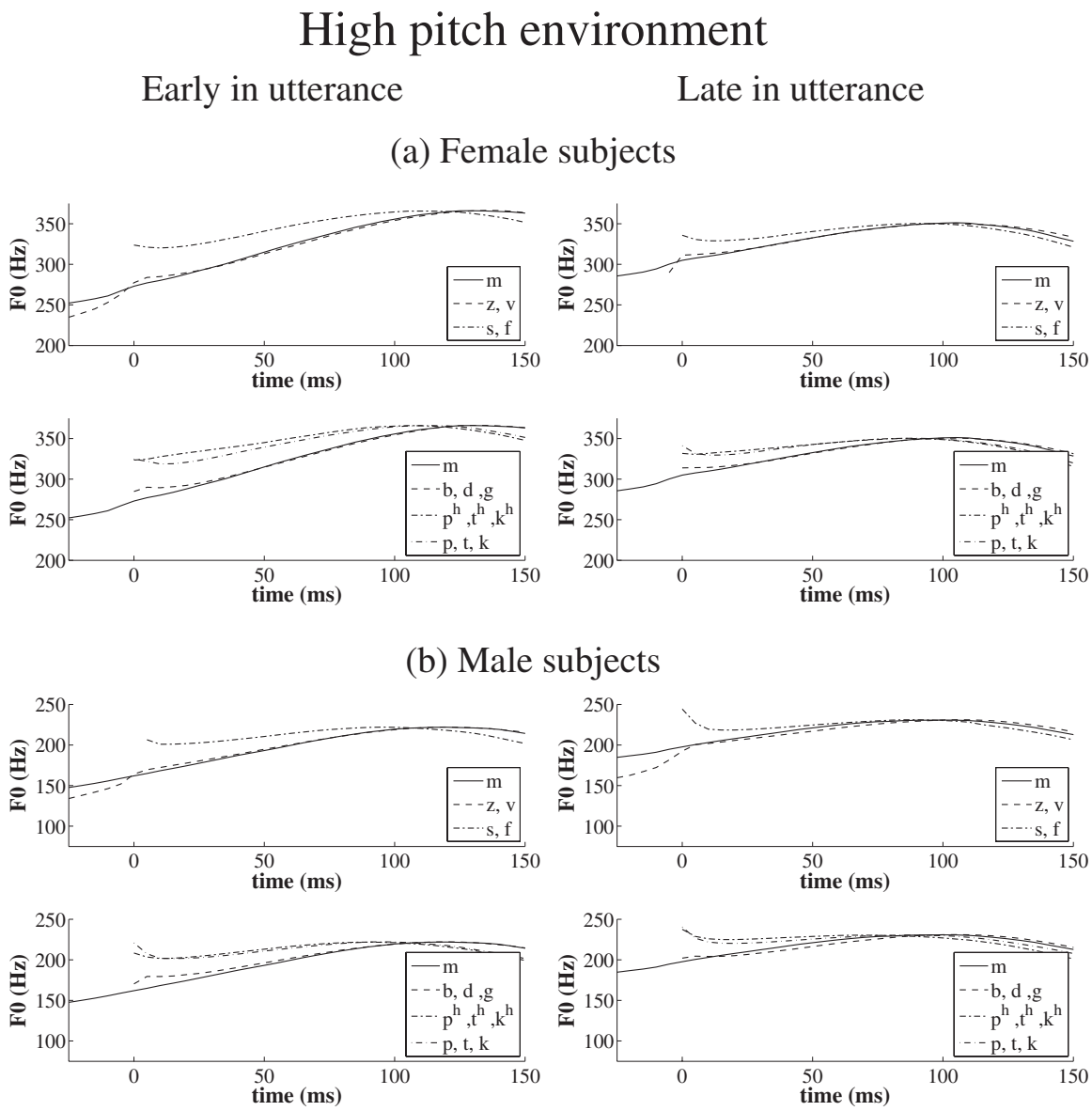

FIG. 7. *F*0 contours of target syllables in high-pitch environment, averaged across place of articulation and vowel for (a) female and (b) male subjects. The first column shows the contours for target syllables occurring early in the utterances, and the second column shows the contours for target syllables occurring late in the utterances. Each point in a contour includes 25–50 data points for /m/, 50–100 data points for the fricatives, and 75–150 for the stops.

in pitch excursion may be more perceptible for speakers with lower *F*0 ranges (100–150 Hz) than for speakers with higher-*F*0 ranges (200–400 Hz). Clearly there is a great deal more analysis that could be done to investigate the interaction of utterance position and pitch range in pitch skip.

IV. Summary and Discussion

The results can be broadly summarized as follows:

(1) When a (s)CVm syllable is in a high-pitch environment, *F*0 is greatly increased relative to a baseline *F*0 following voiceless obstruents, but *F*0 closely follows (or is only slightly higher than) the baseline following voiced obstruents.
(2) When a (s)CVm syllable is in a low-pitch environment, *F*0 is very slightly increased relative to a baseline following all obstruents.

These general results are consistent for both male and female subjects and both stop and fricative consonants. Details such as the degree of pitch skip (both in time and frequency) vary somewhat across subject, indicating that a model of this phenomenon should include speaker-specific parameters. In addition, an apparent effect of utterance position or pitch excursion was observed. Quantification of obstruent *F*0 then appears to be complicated, and must include considerations of speaker characteristics, prosodic stucture, and global as well as local *F*0 environments. Our original goal to quantify pitch shifts due to obstruent consonants turns out to have been somewhat naive. And yet, although we have not achieved that goal, our results have implications for models of speech production (Sec. IV A), which may be more relevant for improving speech synthesis than the original goal. Such an occurrence is common in research related to formant

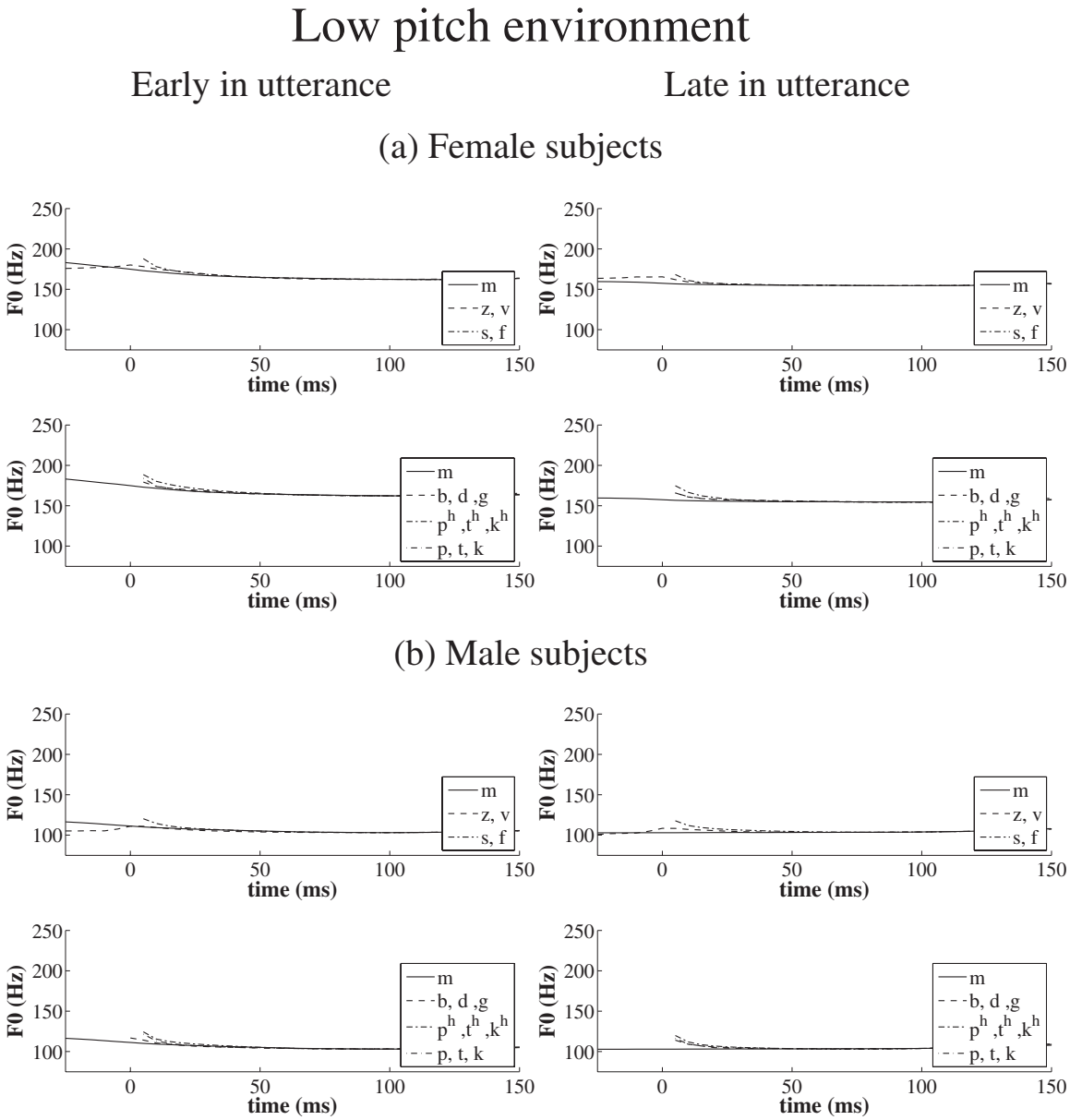

FIG. 8. *F*0 contours of target syllables in low-pitch environment, averaged across place of articulation and vowel for (a) female and (b) male subjects. The first column shows the contours for target syllables occurring early in the utterances, and the second column shows the contours for target syllables occurring late in the utterances. Each point in a contour includes 25–50 data points for /m/, 50–100 data points for the fricatives, and 75–150 for the stops.

or articulatory synthesis: insight is provided to issues of speech communication as synthesis is improved.

A. Interpretation of results

Because *F*0 following voiced stops is not lowered relative to that of /m/, it is probably not the result of a gesture intended to enhance the saliency of the [+voice] feature (Kingston and Diehl 1994). A more likely cause of the pitch skip observed in this data set is an increase in active vocal-fold stiffening during the voiceless obstruent consonants that carries over to the following vowel (Halle and Stevens, 1971; Löfqvist *et al.*, 1989), that is, an intrinsic effect that falls out from gestural overlap between a voiceless obstruent and a vowel that follows it. (We note, however, that an intrinsic effect can then be intentionally exaggerated by speakers to enhance the intended phonological contrast.) We propose that the observed difference between high- and low-pitch contexts occurs because of conflicts between the segmental and prosodic levels of speech production. The fact that we only observed a large effect in high-pitch environments is further support for the claim of Halle and Stevens (1971): high *F*0 in a vowel is in accord with vocal-fold stiffening in the preceding obstruent, so the gestural overlap in that case results in increased *F*0 at the vowel onset. However, when prosody demands low pitch in a vowel, a conflict with a stiffening gesture of a preceding voiceless obstruent arises; in such a case, the prosodic gesture trumps the segmental gesture, and the *F*0 contour is either not perturbed or perturbed to a degree that is probably not perceptible.

One might question the likelihood of a feature having a defining gesture that is so vulnerable to annihilation due to overlap with prosodic gestures. We suggest that this conun-

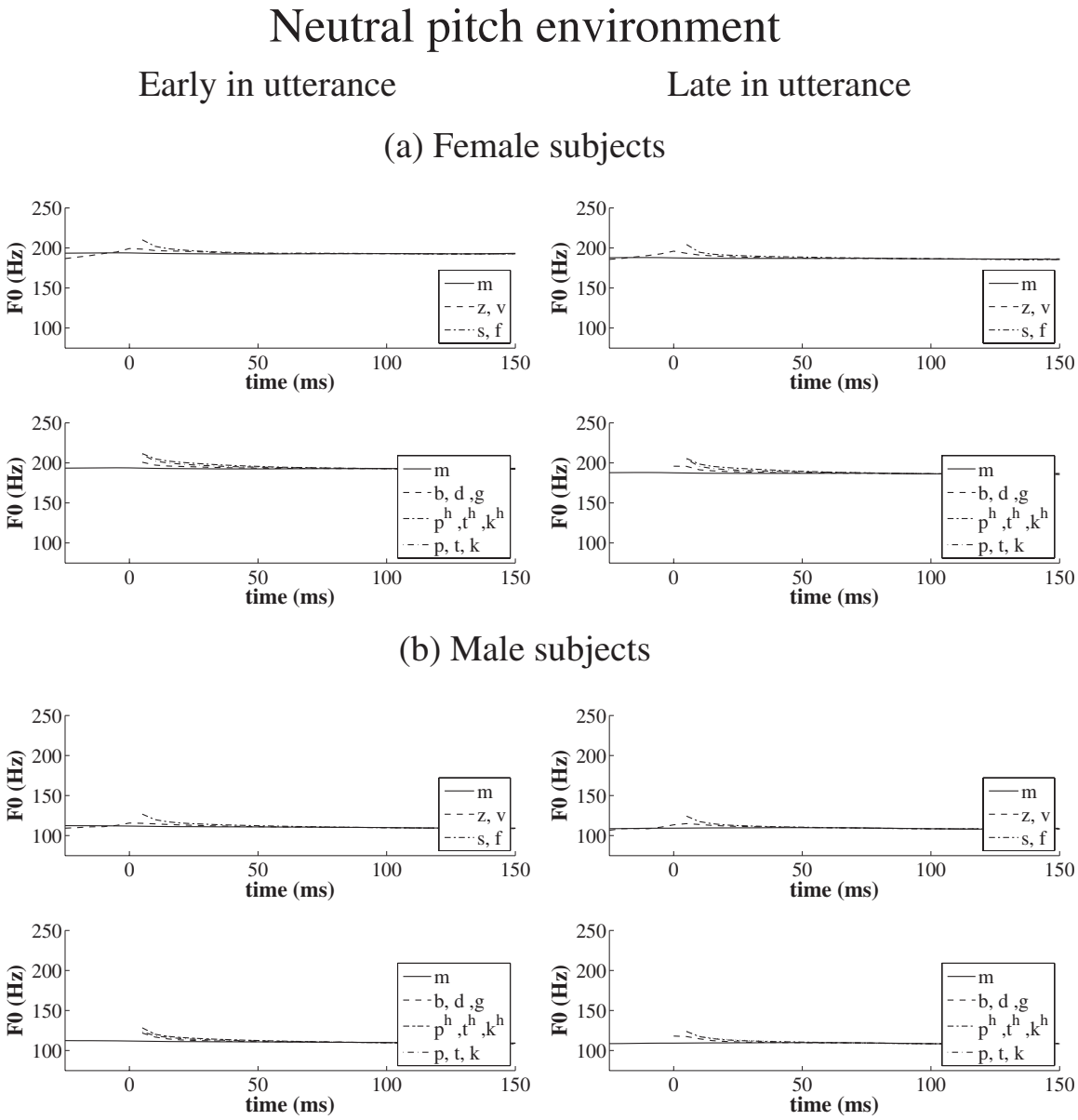

FIG. 9. *F*0 contours of target syllables in neutral pitch environment, averaged across place of articulation and vowel for (a) female and (b) male subjects. The first column shows the contours for target syllables occurring early in the utterances, and the second column shows the contours for target syllables occurring late in the utterances. Each point in a contour includes 25–50 data points for /m/, 50–100 data points for the fricatives, and 75–150 for the stops.

drum and our results can be explained within the framework of enhancement theory (Keyser and Stevens, 2006). According to this theory, a speech segment is defined as a bundle of binary features. A feature is defined as having a defining gesture and a corresponding defining acoustic characteristic. At the phonological planning stage, an utterance is comprised only of such bundles, but features are flagged if they are to occur in the context of other segmental features or prosodic elements that weaken their defining gestures and thus threaten the saliency of their defining acoustic characteristics. Features that are flagged as being vulnerable are reinforced with enhancing gestures. Enhancing gestures tend not to be subject to weakening or annihilation by overlap; they are specifically chosen to enhance the acoustic cues without fear of weakening. In such a scenario, our data can be explained as follows:

- In English, the voicing feature for obstruents is [stiff], and the defining gesture is stiffening of the vocal folds. This stiffening leads to inhibition of vocal-fold vibration during voiceless obstruents (Halle and Stevens, 1971). (Note that the same gesture during a vowel leads to an increase in *F*0.)
- The contrast between [+stiff] and [−stiff] is threatened when [+stiff] occurs in the context of a low *F*0 and when [−stiff] occurs in the context of a high *F*0. The contrast is thus made more salient through the use of enhancing gestures, specifically vocal fold spreading during voiceless obstruents (inhibiting vocal-fold vibration) and active expansion of the vocal tract during voiced obstruents (facilitating vocal-fold vibration) (Svirsky *et al.*, 1997). We note that spreading of the folds will be used as an enhancement

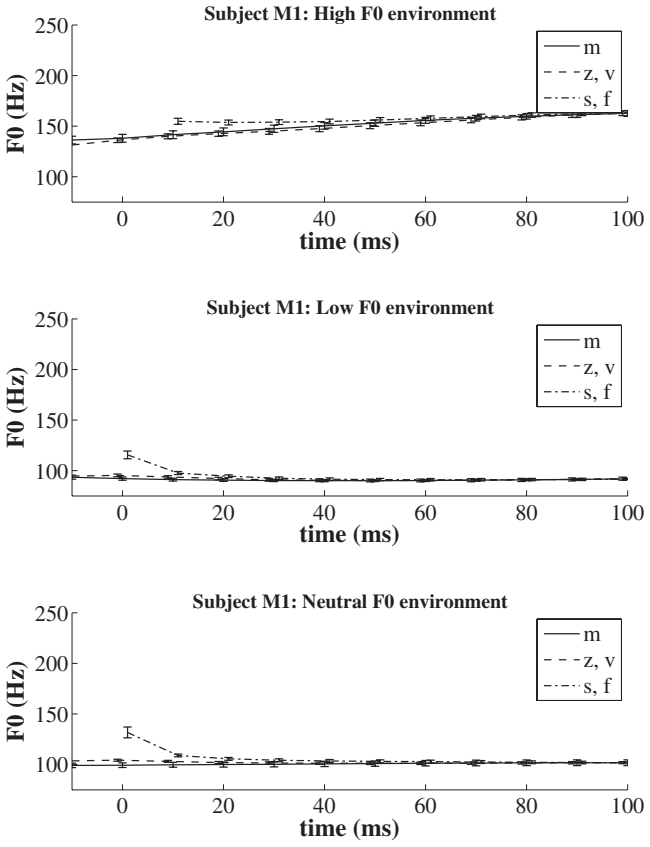

FIG. 10. *F*0 contours of target syllables occurring early in an utterance in high, low, and neutral *F*0 environment, for subject M1. Note that the voiceless obstruents exhibit a similar degree of pitch skip in all three *F*0 environments, contrary to the general results obtained by averaging across subject. Each point in a contour includes 5–10 data points for /m/ and 10–20 data points for the fricatives. Error bars indicate the standard error. (Contours for the obstruents are offset slightly along the time axis to improve clarity.)

for voiceless obstruents even in high-*F*0 environments. Likewise, vocal-tract expansion might occur even in low-*F*0 environments. While those claims may seem counterintuitive, the point of enhancement is to make *contrasts* more salient, and thus speakers may use enhancements for [+stiff] in an environment where [−stiff] is weakened, and vice versa.

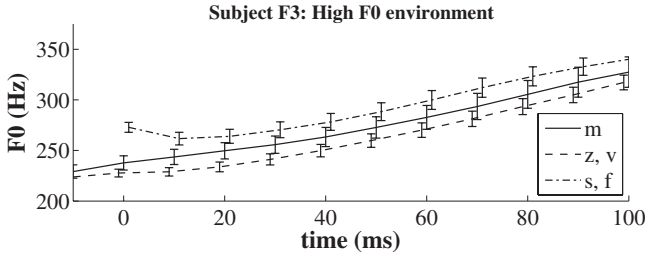

FIG. 11. *F*0 contours of target syllables occurring early in an utterance in high-*F*0 environment for subject F3. For this subject, *F*0 following the voiced obstruents in this environment tended to be low relative to the baseline contour following /m/. This result is contrary to the general result obtained by averaging across subjects. Each point in a contour includes 5–10 data points for /m/ and 10–20 data points for the fricatives. Error bars indicate the standard error. (Contours for the obstruents are offset slightly along the time axis to improve clarity.)

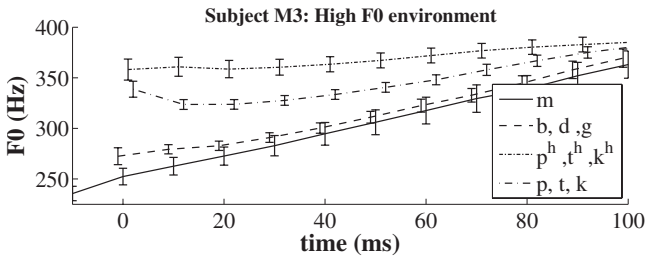

FIG. 12. *F*0 contours of target syllables occurring early in an utterance in high-*F*0 environment for subject M3. While most subjects did not show much difference between *F*0 following aspirated and unaspirated stop consonants, a few, such as this subject, exhibit a relatively strong difference. Each point in a contour includes 5–10 data points for /m/ and 15–30 data points for the stops. Error bars indicate the standard error. (Contours for the obstruents are offset slightly along the time axis to improve clarity.)

- Because of these enhancing gestures, gestures related to prosodic elements can override the defining segmental gestures in these cases, yet the saliency of the voicing contrast is preserved.

In contrast to features, which are believed to be universal in terms of both their defining gestures and acoustics, enhancing gestures are language specific (Keyser and Steven, 2006). Therefore, we emphasize that the proposed model of pitch skip is specific to English. Other languages may use the feature [stiff] contrastively; however, the acoustic manifestation could differ depending on how enhancing gestures are employed. For example, many tone languages include stop consonants, but the degree of observed pitch skip can vary greatly: Francis *et al.* (2006) reported that *F*0 perturbations are limited to the very early portion of the vowel in Cantonese (about 0–10 ms); Hombert (1977) reported perturbations that extend somewhat further into the vowel for Yoruba (about 50 ms for mid and low tones, and up to 100 ms for high tones); Kenstowicz and Park (2006) reported perturbations that extend at least until midvowel in Kyungsang Korean (specific times not provided). Francis *et al.* (2006) argued that Cantonese speakers intentionally limit the duration of pitch skip to protect the tone contrasts (see our interpretation below). On the other hand, the Kyungsang Korean speakers described by Kenstowicz and Park (2006) seem to have embraced the *F*0 perturbation as a means of shoring up the three-way unvoiced stop contrast, despite the competing demand of tonal contrasts on *F*0.

B. Implications for other studies

Our interpretation of the data within the theory of enhancement and overlap has implications for data reported in other studies. In Sec. I A 1 we described Korean data reported by Jun (1996). Although all three series of Korean stops are described as being unvoiced (underlyingly), an *F*0 dichotomy was observed at vowel onset following lenis stops, on the one hand, and tense or aspirated stops on the other. In our view, this finding suggests that the pitch-skip phenomenon is not due to voicing *per se* but to the manner in which voicing or its cessation is brought about. We would claim that a feature defining the lenis stops in Korean is [−stiff], while [+stiff] is a feature of the other two series of

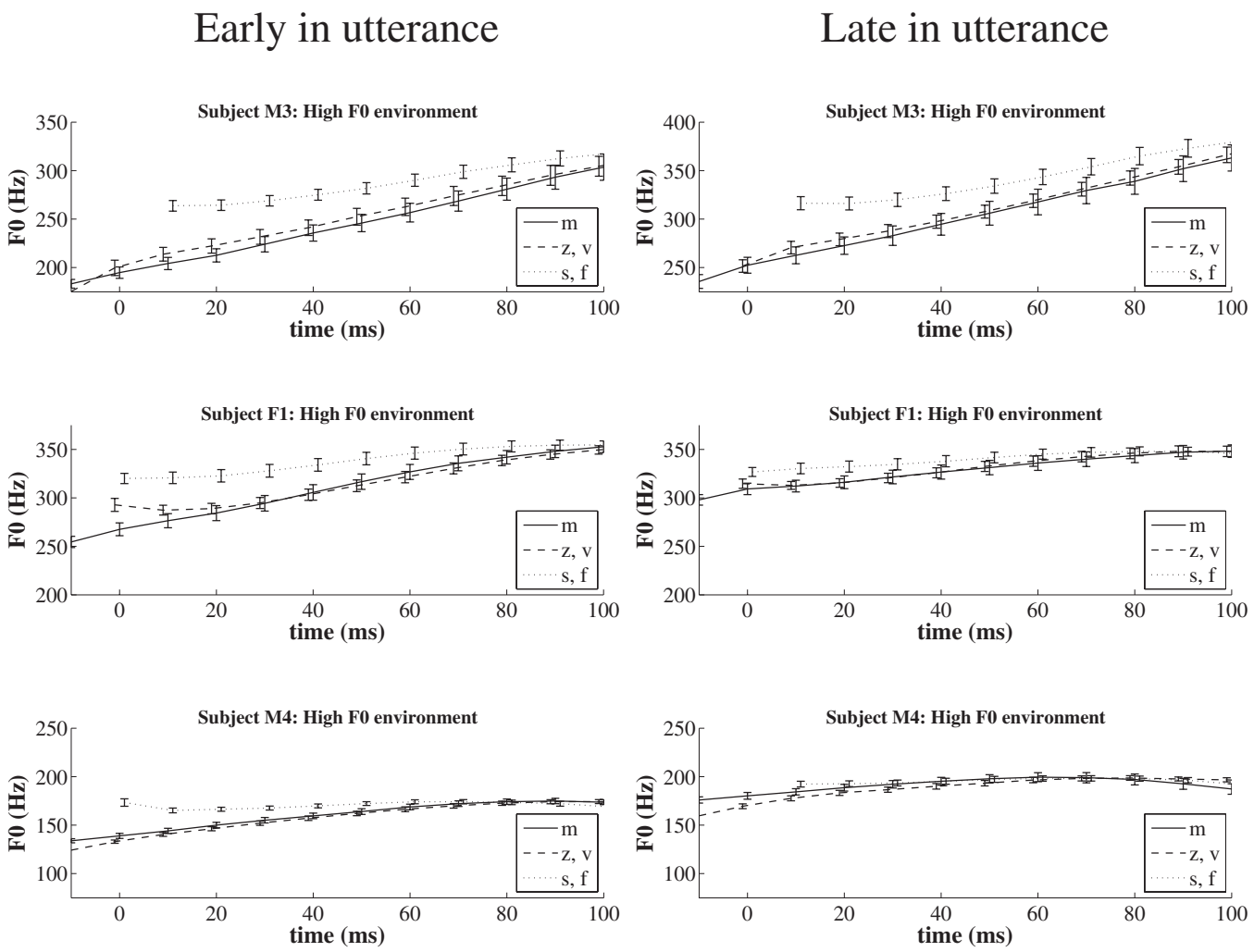

FIG. 13. *F*0 contours of target syllables occurring early and late in an utterance in high-*F*0 environment for subjects M3, F1, and M4. Subject M3 shows a similar degree of pitch skip across utterance position; subject F1 shows a reduced degree of pitch skip for syllables in late position, and subject M4 exhibits almost no pitch skip in late position. Each point in a contour includes 5–10 data points for /m/ and 10–20 data points for the fricatives. Error bars indicate the standard error.

stops. Further differentiation of the aspirated and tense stops would be due to the former having the feature [+spread], while the latter has the feature [−spread]. (The feature [spread] would not be defined for lenis stops.) Our interpretation of these Korean data is similar to the theory of Kim (1965) that proposes "tensity" as a feature independent of voicing and is similar in spirit to a proposal in Cho *et al.* (2002).

In their study of Cantonese, Francis *et al.* (2006) suggested that if pitch skip is intended to enhance a voicing contrast (Kingston and Diehl, 1994), it would be unlikely to be manifested in Cantonese because (1) there is no voice contrast and (2) it would interfere with lexical tone. While Francis *et al.* (2006) did observe a dichotomy between the aspirated and unaspirated stops in Cantonese, the effect was much smaller than in English, both in absolute difference and duration. This result was interpreted by them to indicate that while there was some intrinsic basis for pitch skip, the Cantonese speakers curtail this effect so as not to let it interfere with lexical tone. However, from our point of view, the feature that defines the contrast between aspirated and unaspirated stops in Cantonese would be [spread], not [stiff], and therefore, all else being the same, we would not expect to see a difference in the *F*0 of vowels following these obstruents. That Francis *et al.* (2006) did see a small effect may be due to other factors, such as aerodynamics, as suggested by Xu and Xu (2003) for Mandarin, or perhaps as a result of vocal-fold stiffening used to enhance voicelessness during the production of the unaspirated stops.

Similarly, our theory is in line with the observations of Kohler (1982) on medial stop consonants if one assumes that [stiff] is a feature for German stop consonants. When *F*0 is rising, the vowel *F*0 preceding the consonant is presumably low, and reduced vocal-fold stiffness during the vowel precludes the stiffening that might have occurred during the early phase of a voiceless consonant. Therefore, no effect of obstruent voicing is observed in the *F*0 of the preceding vowel. However, when *F*0 is falling, the relatively high *F*0 preceding the consonant does not conflict with stiffening of the folds early in the consonant and the pitch dichotomy is observed, while the drop in *F*0 for the following vowel masks the effects of stiffening later in the consonant.

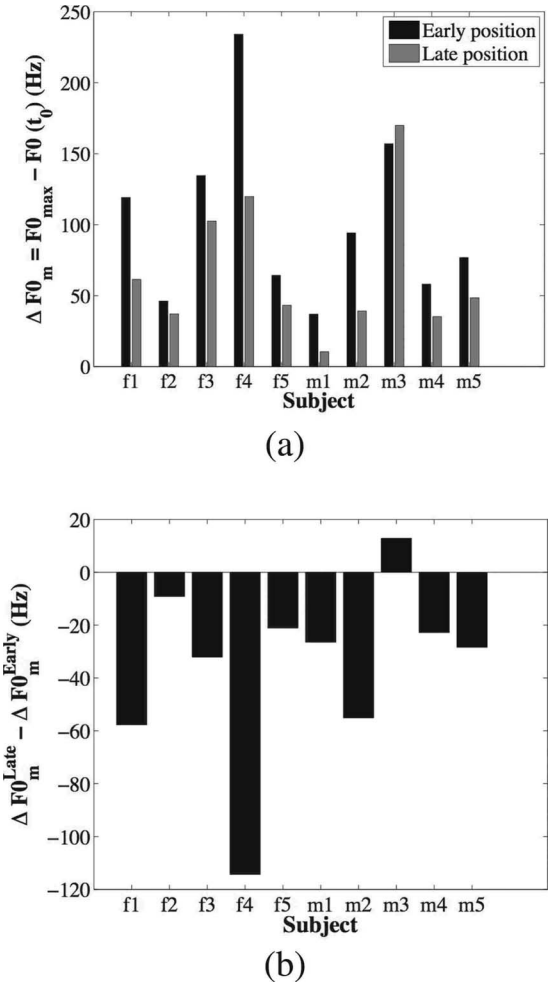

FIG. 14. Comparison of the pitch excursions $\Delta F0_m^{\text{early}}$ and $\Delta F0_m^{\text{late}}$ (averaged across vowel) for /mVm/ syllables in the early (dark bars) and late positions (gray bars). (b) Difference in pitch excursion $\Delta F0_m^{\text{late}} - \Delta F0_m^{\text{early}}$ for each subject. As expected from part (a), the difference is negative; i.e., pitch excursion is compressed in late position for all subjects but M3.

V. FUTURE DIRECTIONS

Certain predictions fall out from our model of pitch skip and are worth pursuing in future work. First, if implementation of a feature [stiff] is the source of pitch skip observed in most languages, the degree of pitch skip observed in vowels that either precede or follow unvoiced obstruents will depend on where *F*0 falls in a speaker's *F*0 range during those vowels. Therefore, a more complete test of this theory will observe obstruent effects on *F*0 contours that both rise and fall going into the obstruent. In addition, the effects should be observed on *F*0 contours that are neither rising nor falling, but rather constant at both low and high *F*0.

Likewise, our model predicts certain effects of *F*0 environment on the phonetic voicing of obstruents. For example, one might expect /b,d,g/ to show a lesser degree of vocal-fold vibration during the closure interval in a high-*F*0 environment than would be observed in a low-pitch environment. Preliminary data from two speakers support this prediction (Hanson, 2004).

As we have suggested, because pitch skip occurs consistently yet differently in the world's languages, studies of pitch skip can provide insight to models of speech production, speech perception, and sound change. Therefore, future studies of pitch skip will be most beneficial if they compare data across languages. Production and perception in bilingual speakers may be particularly insightful to studies of how different languages use enhancing gestures and acoustic cues to increase the saliency of universal feature contrasts.

ACKNOWLEDGMENTS

This work was carried out while the author was at the Speech Communication Group, MIT Research Laboratory of Electronics. The work was supported by NIH Grant Nos. DC04331 and DC00075. The programming assistance of Man Yin Yee is greatly appreciated. The PRAAT scripts for computing *F*0 contours and the C code for smoothing the contour were based on scripts and code provided by Dr. Yi Xu. The analysis and interpretation of the data have benefited from many conversations with Professor Kenneth N. Stevens. I thank Chris Shadle and three anonymous reviewers for their close reading of an earlier manuscript; their comments have greatly improved the paper.

[1]Note that lexical stress and utterance-level focus have been confounded in the Lea (1973) study, particularly for the second set of recordings. Therefore, it is not clear if segmental context interacts with stress or focus, or both. However, what is relevant for the discussion here is that segmental context appears to interact with suprasegmental elements that affect *F*0.

[2]By phrase-level prominence we mean what is often referred to as sentence stress or focus. Because these terms (especially the word *stress*) can mean different things to different readers, we use *prominence* to indicate that the speaker intended these words to be more prominent than the other words in a sentence, while not being specific about which acoustic cues were employed to signal that prominence.

[3]Error bars are omitted from these plots so as to avoid confusing variability among speakers for the variability within a given speaker.

Cho, T., Jun, S.-A., and Ladefoged, P. (**2002**). "Acoustic and aerodynamic correlates of Korean stops and fricatives," J. Phonetics **30**, 193–228.

Francis, A. L., Ciocca, V., Wong, V. K. M., and Chan, J. K. L. (**2006**). "Is fundamental frequency a cue to aspiration in initial stops?," J. Acoust. Soc. Am. **120**, 2884–2895.

Fujimura, O. (**1971**). "Remarks on stop consonants: Synthesis experiments and acoustic cues," in *Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen*, edited by L. L. Hammerich, R. Jakobson, and E. Zwirner (Akademisk Forlag, Copenhagen), pp. 221–232.

Haggard, M., Ambler, S., and Callow, M. (**1969**). "Pitch as a voicing cue," J. Acoust. Soc. Am. **47**, 613–617.

Halle, M., and Stevens, K. N. (**1971**). "A note on laryngeal features," Quarterly Progress Report No. 101 MIT Research Laboratory of Electronics, Cambridge, MA, pp. 198–213.

Han, M. S., and Weitzman, R. S. (**1970**). "Acoustic features of Korean /P,T,K/, /p,t,k/ and /p^h,t^h,k^h/," Phonetica **22**, 112–128.

Hanson, H. M. (**2004**). "The feature [stiff] interacts with intonation to affect vocal-fold vibration characteristics," J. Acoust. Soc. Am. **116**, 2546.

Hanson, H. M., and Stevens, K. N. (**2002**). "A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn," J. Acoust. Soc. Am. **112**, 1158–1182.

Hombert, J.-M. (**1977**). "Development of tones from vowel height?," J. Phonetics **5**, 9–16.

Hombert, J.-M. (**1978**). "Consonant types, vowel quality, and tone," in *Tone: A Linguistic Survey*, edited by V. A. Fromkin (Academic, New York), pp. 77–107.

Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (**1979**). "Phonetic explanations for the development of tones," Language **55**, 37–58.

House, A. S., and Fairbanks, G. (**1953**). "The influence of consonant environment upon the secondary acoustical characteristics of vowels," J. Acoust. Soc. Am. **25**, 105–113.

Jeel, V. (**1975**). "An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants," Technical

Report No. 9, Institute of Phonetics, University of Copenhagen, Copenhagen.
Jessen, M., and Roux, J. C. (**2002**). "Voice quality differences associated with stops and clicks in Xhosa," J. Phonetics **30**, 1–52.
Jun, S.-A. (**1996**). "Influence of microprosody on macroprosody: A case of phrase initial strengthening," Technical Report No. 92, University of California at Los Angeles, Los Angeles, CA.
Kenstowicz, M., and Park, C. (**2006**). "Laryngeal features and tone in Kyungsang Korean: A phonetic study," Studies in Phonetics, Phonology and Morphology **12**, 247–264.
Keyser, S. J., and Stevens, K. N. (**2006**). "Enhancement and overlap in the speech chain," Language **82**, 33–63.
Kim, C.-W. (**1965**). "On the autonomy of the tensity feature in stop classification (with special reference to Korean stops)," Word **21**, 339–359.
Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**, 419–454.
Kohler, K. J. (**1982**). "F_0 in the production of lenis and fortis plosives," Phonetica **39**, 199–218.
Kohler, K. J. (**1985**). "F_0 in the perception of lenis and fortis plosives," J. Acoust. Soc. Am. **78**, 21–32.
Lea, W. A. (**1973**). "Segmental and suprasegmental influences on fundamental frequency contours," in *Consonant Types and Tones*, Southern California Occasional Papers in Linguistics No. 1, edited by L. M. Hyman, (University of Southern California Press, Los Angeles), pp. 15–70.
Lehiste, I., and Peterson, G. E. (**1961**). "Some basic considerations in the analysis of intonation," J. Acoust. Soc. Am. **33**, 419–425.
Löfqvist, A., Baer, T., McGarr, N. S., and Story, R. S. (**1989**). "The cricothyroid muscle in voicing control," J. Acoust. Soc. Am. **85**, 1314–1321.
Löfqvist, A., Koenig, L. L., and McGowan, R. S. (**1995**). "Vocal tract aerodynamics in /aCa/ utterances: Measurements," Speech Commun. **16**, 49–66.
Massaro, D. W., and Cohen, M. M. (**1976**). "The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction," J. Acoust. Soc. Am. **60**, 704–717.
Matisoff, J. A. (**1973**). "Tonogenesis in Southeast Asia," in *Consonant Types and Tones*, edited by L. M. Hyman, Southern California Occasional Papers in Linguistics No. 1 (University of Southern California Press, Los Angeles), pp. 71–95.
Mohr, B. (**1971**). "Intrinsic variations in the speech signal," Phonetica **23**, 65–93.
Ohde, R. N. (**1984**). "Fundamental frequency as an acoustic correlate of stop consonant voicing," J. Acoust. Soc. Am. **75**, 224–230.
Reinholt Petersen, N. (**1983**). "The effect of consonant type on fundamental frequency and larynx height in Danish," Technical Report, Institute of Phonetics, University of Copenhagen, Copenhagen.
Shadle, C. H. (**1985**). "Intrinsic fundamental frequency of vowels in sentence context," J. Acoust. Soc. Am. **78**, 1562–1566.
Silverman, K. (**1984**). "*F*0 perturbations as a function of voicing of prevocalic and postvocalic stops and fricatives, and of syllable stress," in *Reproduced Sound: 1985 Autumn Conference, Windermere: Conference Handbook* (Institute of Acoustics, Windermere, Cumbria, Great Britain), Vol. **6**, pp. 445–452.
Silverman, K. (**1986**). "F_0 segmental cues depend on intonation: The case of the rise after voiced stops," Phonetica **43**, 76–91.
Stevens, K. N., and Bickley, C. A. (**1991**). "Constraints among parameters simplify control of Klatt formant synthesizer," J. Phonetics **19**, 161–174.
Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., and Wilhelms-Tricarico, R. (**1997**). "Tongue surface displacement during bilabial stops," J. Acoust. Soc. Am. **102**, 562–571.
Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1993**). "*F*0 gives voicing information even with unambiguous voice onset times," J. Acoust. Soc. Am. **93**, 2152–2159.
Xu, C. X., and Xu, Y. (**2003**). "Effects of consonant aspiration on Mandarin tones," J. Int. Phonetic Assoc. **33**, 165–181.
Xu, Y. (**1999**). "Effects of tone and focus on the formation and alignment of f_0 contours," J. Phonetics **27**, 55–105.