Perception and production of rise-fall intonation in American English

MARY L. STEPPLING

Columbia College, Columbia, South Carolina

and

ALLEN A. MONTGOMERY

University of South Carolina, Columbia, South Carolina

At the segmental level, the rate of speaking affects the degree of physical undershoot of articulatory targets and the resulting perception. Little is known regarding evidence of these effects at the suprasegmental level, particularly in intonation. In this study, the effect of rate of speaking on fundamental frequency and on perceptual judgments of peak pitch in a rise-fall intonation pattern was investigated. First, speakers produced rise-fall intonations in sentence contexts at slow, normal, and fast speaking rates. Peak fundamental frequencies (F_0) of the slow productions were significantly lower than those of the normal or fast productions. The mean normal rate production of the word *Miami* was used as a model for the target word in a series of subsequent perceptual experiments. Altering the duration of the target word to represent slow, normal, and fast rates of speaking did not affect listener judgment of peak pitch. Finally, the pitch of the target word was measured in a sentence context. No differences between peak pitch in isolation or in sentence context were found. It was concluded that the production and perception of this form of intonation was not subject to the effects of rate that are seen at the segmental level.

This study is concerned with the presence of acoustic and perceptual evidence of laryngeal coarticulation during intonation. Coarticulation is a general property of the control of a sequence of skilled movements. In speech, it may be loosely defined as the simultaneous presence of articulatory gestures associated with two different segments. The segments are typically adjoining but may be as far apart in the articulatory stream as four segments (Daniloff & Moll, 1968). Coarticulation among oral articulators is well established, as can be seen in the undershooting of articulatory targets owing to the surrounding phonetic context (Ohde & Sharf, 1975; Stevens & House, 1963). The extent to which coarticulation occurs between oral and laryngeal articulators, however, appears to be limited. Those effects that do appear, such as the lag in the onset of voicing in a sonorant following an aspirated, voiceless stop (the devoiced /r/ in *treat*), are due to the airflow requirements of the aspiration. The oral articulators influence the ongoing operation of the larynx, of course, as can be seen in the lowering of fundamental frequency (F_0) during stops (Ohde, 1984) and the elevation of F_0 during high vowels, the well-known *intrinsic pitch of vowels* (Peterson & Barney, 1952; Shadle, 1985). These effects are largely within a segment, however, and do not constitute coarticulation. Thus, there is little evidence of contextual influence of oral articula-

tors on the presence or absence of voicing at the segmental level.

There is more evidence of coarticulation within the larynx, however, as can be seen in the F_0 of vowels in adjoining syllables. That is, in addition to the expression of the voicing feature at the segmental level, the larynx is also used to control the pitch of voiced speech. There is evidence in tone languages in which pitch is used phonemically that some anticipatory adjustments in F_0 occur, depending on the tone requirements of the following segments. Xu (1994), for example, defined two kinds of phonetic environments in Mandarin. The first was a compatible sequence of tones in which a segment ended on a pitch that was the same (i.e., high or low) as the initiation of the following segment. The second was a conflicting circumstance in which a segment ending on a high tone was followed by a segment starting on a low tone or vice versa. This circumstance would allow coarticulation of the adjoining tones to occur. For example, a high tone followed by a low-to-high rising tone might lead to a lower pitch high tone in anticipation of the coming low pitch required for the rise. Xu (1994) found evidence of considerable coarticulation of tone in the conflicting context, but not in the compatible context. Abramson (1979), cited by Xu (1994), found very similar evidence of coarticulation of tone in Thai as well. Thus, there is some evidence of coarticulation in laryngeal activities in which variation in pitch conveys phonemic information.

One of the established methods for studying coarticulation is to vary the rate of speaking. The rationale is that

Correspondence concerning this article should be addressed to M. L. Stepling, Speech Language Pathology Program, Columbia College, Columbia, SC 29203 (e-mail: msteppling@collacol.edu).

rapid speech would provide less time for the articulators to reach or approach their targets and so would require more centrally planned coarticulation with surrounding gestures and, perhaps, more mechanical lag (Daniloff & Hammarberg, 1973; Gay, 1978; Ohde & Sharf, 1975; Stevens & House, 1963). Recently, several studies have used this method to investigate the use of pitch within and across syllabic boundaries.

Cooper and Sorensen (1981) found systematic changes in fundamental frequency as speaking rate varies. Participants were asked to read sentence lists at slow, normal, and fast rates. The peak F_0 in the slow sentences was consistently lower than the peak F_0 in fast sentences. Increased vocal fold tension at fast versus slow speaking rates was suggested as an explanation for the higher F_0 in the fast condition. Ladd, Faulkner, Faulkner, and Schepman (1999) studied the effect of changes in speech rate on the location and extent of pitch rise on stressed syllables in bi- and tri-syllabic words. They found only small and inconsistent effects of rate on the point of initiation of the rise and, more importantly, on the extent of the rise in hertz. That is, there was little evidence of rate-induced coarticulation in their materials. However, Ladd, Mennen, and Schepman (2000) found some evidence that the time pressure associated with producing an F_0 rise during a short vowel may influence the alignment of the F_0 within a syllable. Moreover, Fougeron and Jun (1998) found that their three speakers all showed smaller F_0 excursions at a faster rate. Xu (1998) found that rate of speaking, as seen in varying syllable duration in Mandarin, produced within-syllable adjustments of F_0 contours but gave little evidence of across-syllable influences. Finally, Caspers and van Heuven (1993) found that rising pitch accents are steeper at fast rates than at normal rates, whereas the slope of falling pitch accents is unaffected by rate. These studies, although somewhat inconsistent, appear to favor the interpretation that there are intralaryngeal influences between adjacent segments analogous to the articulatory undershoot seen in oral articulation. Even though the larynx is capable of rapid changes in F_0 (Sundberg, 1979), it may be helpful to the general goal of efficient articulation if anticipatory adjustments in F_0 are made.

It is generally agreed that linguistic and phonetic information contribute to the perception of intonation, and it is well established that F_0 influences the perception of rate of speaking (den Os, 1985; Feldsten & Bond, 1981; Kohler, 1986; Rietveld & Gussenhoven, 1987; Silverman, 1987). The effect of rate of speaking on the perception of F_0 , however, has received little study. Few studies of specific phonetic contributions to the perception of the pitch of intonation, regardless of rate variation, are found in the literature. Silverman, in his doctoral dissertation, completed a series of investigations to determine consonant and vowel effects on fundamental frequency variations in intonation contours. In exploring contributions to making synthetic speech sound more like natural speech, he made a case for intrinsic F_0 (IF_0) of vowels being perceptually

important for the listener to make correct judgments of intonation. The vowels /a/ and /i/ were contrasted in words that could be substituted in sentence pairs. For example, "we parked without peeking" could be altered to "we peeked without parking." The vowels were resynthesized with a range of differing fundamental frequency heights in the accent peaks while the rest of the sentence was unchanged. An attempt was also made to have each vowel appear equally loud. Listeners were asked to judge which word sounded more prominent in the sentence. Silverman found that F_0 had to be 2.5 Hz lower in the second accent in order for it to sound as prominent as the first. This suggests that listeners showed perceptual compensation for declination, which is consistent with Pierrehumbert's (1979) earlier findings. Although Silverman also found a difference in the perceptual prominence of high versus low vowels, with their associated high or low IF_0 , it was not of the magnitude corresponding to the actual range of IF_0 present in the vowels. The distance for which Silverman found that listeners compensated for intrinsic vowel F_0 had an extent of only 4.3 Hz, rather than the actual 10–14 Hz difference present between high and low vowels. That is, the listeners only partly compensated for the F_0 "expected" to accompany the various vowels. The experiment was repeated without adjusting loudness and without specific instructions to the listeners regarding pitch. The results proved to be somewhat inconsistent with the first study (Silverman, 1987).

An early work (Studdert-Kennedy & Hadding, 1973), which sought to illustrate the importance of incorporating intonation in perceptual judgments, used speech and non-speech comparisons. The word *November* was resynthesized to create 72 intonation contours that varied F_0 at peak stress and word end. In addition, analogous sine waves and frequency-modulated pulses of varying F_0 were constructed. Listeners were asked to judge whether the stimuli were representative of a statement or a question. It was concluded that peak F_0 and perceived terminal glide were most utilized by listeners in making judgments as to whether an utterance was a statement or a question.

The Present Experiments

The study investigated the relationship between the acoustic and the perceptual aspects of intonation in American English. In order to study these aspects, a database first was generated to provide descriptive information on the peak F_0 of this intonation pattern at slow, normal, and rapid rates of speech (Experiment 1). Rate of speaking is defined as words (or syllables) per unit of time. Obviously, as rate of speaking increases, overall duration of the entire utterance decreases. Thus, at higher rates of speaking, articulators must either move more rapidly, which may be difficult, or exhibit reduced amounts of movement in order to accommodate the greater number of phonemes per second that must be represented. At the segmental level, the demands that rapid speech place on the articulatory system often result in the latter "solution." That is, the artic-

ulators, especially the tongue, show undershoot or vowel reduction when required to reach extreme targets, such as the anterior placement of /θ/ or the high vowel locations associated with /i/ or /u/ (Lindblom, 1963). Perceptually, however, the listener is able to compensate for the effects of rate and shows little awareness of these complex articulatory accommodations.

The first experiment investigates whether or not such physical undershoot occurs at the suprasegmental level, specifically with regard to the use of fundamental frequency (F_0) to convey intonation. It is possible, of course, as was suggested by Cooper and Sorensen (1981), that as the rate of speaking increases, the extent and range of F_0 could increase as the speaker exerts additional effort. Although there have been many investigations of speaking and reading rate, only two have been found that address the specific issue of changes in F_0 at different rates of speaking (Cooper & Sorensen, 1981; Saxena, 1992). The findings of Saxena are difficult to interpret, since the study appears to be poorly controlled. Cooper and Sorensen, however, demonstrated general increases in mean F_0 of key words as rate of speaking increased in declarative sentences. Declarative sentences, however, do not provide an ideal test of the effects of rate on F_0 . The intonation patterns are generally rise, fall, or fall-rise, none of which requires rapid pitch changes that could produce undershoot. In addition, neither Cooper and Sorensen nor Saxena controlled for the effects of consonantal context on vowel F_0 . Thus, manipulation of rate allows comparison of known perceptual accommodations that occur in segmental processing with possible effects in suprasegmental processing.

In Experiments 2–4, the speaking rate of a single multisyllabic word produced as a rise-fall utterance was manipulated to assess its effects on perception of pitch at the peak of the rise-fall F_0 contour. There is evidence that people can perceive intonation in either a musical (pitch) mode or a linguistic (prominence) mode (Terken, 1991, 1993). Experiments 2–4 focused on the pitch of the peak of a rise-fall intonation. The pitch mode of perceiving was induced by four procedures: (1) use of the word *pitch* in the instructions for comparing the variable buzz and the intoned speech, (2) familiarization to the task of making judgments by giving feedback on the buzz-to-buzz comparison, (3) use of the buzz, which has a pitch contour but no prominence, as the variable standard, and (4) varying only the F_0 of the buzz, while allowing the speech to have the natural variations of F_0 , SPL, spectral tilt, and other acoustic qualities present during a stressed syllable.

The listeners compared speech and nonspeech rise-falls in order to assess judgments of this specific one-word contour without additional semantic influences (Experiment 2). Two control studies follow that allow interpretation of the effects of rate by holding first the duration of the nonspeech stimulus constant and then the word duration constant (Experiment 3). Finally, the word was placed in a sentence context to determine whether there are any

effects of sentence versus single word on peak pitch perception (Experiment 4). Specifically, the following questions are addressed. (1) Do talkers exhibit physiological undershoot or overshoot in the production of a rise-fall sentence-final intonation, as determined by detailed measurements of F_0 in speech produced at slow, normal, and fast rates (Experiment 1)? (2) Do listeners show evidence of undershoot or overshoot in the perception of peak F_0 in a rise-fall intonation under conditions of slow, normal, and fast speaking rate (Experiment 2)? (3) Do different durations of square waves and speech affect perceptual processing equally (Experiment 3)? (4) Does a sentence context affect the processing of pitch in the judgment of peak pitch in a rise-fall intonation (Experiment 4)?

EXPERIMENT 1

The words *Miami*, *Geneva*, and *Bahamas* produced with a rise-fall intonation in phrase-final position were studied. The utterances were generated to be in an “incredulous” form intended to yield a strong rise-fall pattern. Examples of the context used to induce this can be found in the Appendix. These all require stress on the middle syllable and provide a sonorant or fricative context for the stressed vowel. The rise-fall intonation pattern requires a rapid rise-fall F_0 pattern (coincident with the intensity peak), which must be produced solely by laryngeal articulation and breath stream dynamics. The stimuli provided a purer test of the undershoot-versus-effort explanation of laryngeal control in intonation than had been used previously.

Method

Participants. The participants were 10 adult (5 male and 5 female) speakers of English with no obvious speech, voice, or hearing disorders. The participants ranged in age from 22 to 47 years, with a mean age of 28 years. The participants did not evidence any regional dialects, as judged by the examiner.

Materials. The phonetic content of words contributes to F_0 patterns (Silverman, 1987). Furthermore, Ohde (1984) demonstrated that regardless of voicing status, F_0 consistently falls during stop-to-vowel transitions. In order to minimize possible consonantal influences, target words selected did not have stop consonants. A total of 3 three-syllable words (*Miami*, *Geneva*, and *Bahamas*), each with different vowels in the middle syllable, were selected and placed in a paragraph. The selected words were all representative of geographical locations, in order to maintain the same semantic category. The context of the paragraph, as well as the instructions given to the participant, required that additional stress be given the naturally stressed middle syllable. A list of the sentences used is found in the Appendix.

Procedure. Speakers read aloud the short paragraphs consisting of 3 sentences at three different rates designated as slow, normal, and fast. All the utterances were tape recorded. The normal rate was recorded first by asking speakers to read the passage as they typically would. In each sentence, the three-syllable words selected as targets were highlighted. The speakers were asked to emphasize these words with an incredulous rise-fall inflection, the peak of the rise being on the middle syllable. This type of rise-fall was demonstrated to the participant. Once the speaker had read each passage five times at the normal rate, it was explained that this rate should be consid-

Table 1
Means and Standard Deviations of Word Duration
(in Milliseconds) for Males and Females in Experiment 1

Word	Slow		Normal		Fast	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Males						
Miami	826	241	644	97	528	77
Geneva	817	167	617	50	510	81
Bahamas	946	215	737	79	580	85
Females						
Miami	866	159	661	62	565	48
Geneva	838	125	660	77	553	49
Bahamas	883	210	696	99	571	79

ered the standard, with an assigned rating of 100%. The participants were then asked to either slow the rate to 50% of the standard (half as fast) or speed the rate to 150% (one and a half times as fast). The rates were not demonstrated; however, the participants were asked to prolong speech to achieve the slow rate, rather than add more pauses to their readings, and this technique was demonstrated. Each participant read paragraphs containing 3 different sentences, five times each at normal, slow, and fast speaking rates, for a total of 45 sentences yielding 45 words with a rise-fall contour for analysis.

The speakers sat in a sound-attenuated room, using a microphone with mouth position a set distance of 10 cm from the microphone. Speech was recorded on a Yamaha MT100 tape recorder with a Sennheiser dynamic cardioid microphone. All the utterances were judged to be produced with acceptable rise-fall intonation.

Measurement. The targeted words were digitized at 22 kHz 16-bit samples/second on a Power 8100 AV Macintosh computer using Signalyze 3.0 software (Keller, 1994), which was also used to analyze the utterances. The Signalyze software enabled the examiner to mark and listen to any designated segments of the utterance, thereby allowing perceptual judgment to verify the beginning or the end, as well as other aspects, of the utterance. Measurements of the F_0 and duration of all the target words were made by one examiner. The duration of each target word was measured from the first visible waveform of the initial phonetic element to the last visible waveform. The duration of the rise and the duration of the fall of F_0 were also measured. The frequency at the beginning of the word, measured from the first visible waveform of the initial phonetic element, was termed base 1. The frequency measurement ended with the last sound audible that coincided with an obvious waveform. This endpoint was termed base 2. The F_0 of the peak of the stressed syllable was also measured by first performing an FFT pitch extraction on the entire word. This assisted the examiner in determining approximate location of the peak of the rise to be measured. The area of the waveform containing the peak F_0 was then enlarged, and the duration of 10 cycles was measured to determine the exact frequency. Several measurements were taken to ensure that the exact peak frequency had been located.

Reliability. Intraexaminer reliability and interjudge agreement were assessed in order to establish the accuracy of the measurements made in the study. The investigator first selected 50 of the stimulus items for reanalysis. These items were representative of the participant pool and included tokens of all the words and speaking rates. The target words were remeasured by the primary examiner and by a second examiner without knowledge of the initial F_0 and duration values. The mean values of 318, 323, and 319 Hz for Examiner 1, Examiner 1 remeasurement, and Examiner 2, respectively, indicated no significant difference in peak F_0 . All correlations among the three sets of measurements of peak F_0 were $>.99$. Corresponding values for duration were 665, 674, and 675 msec and again showed no sig-

nificant difference. Correlations for duration for both intra- and interjudge agreement ranged from .82 to .90. All correlations were significantly greater than .00 at $p < .01$.

In addition, the time scale of the Macintosh-Signalyze configuration was calibrated with a 999-Hz pure tone, using a Coulbourn signal generator and a Dynascan precision frequency counter. The same 10-cycle measurement process as that employed in the speech analysis was performed five times on the pure tone signal. The mean frequency was calculated to be 999 Hz, with a standard deviation of 1.8 Hz. It was concluded that the measurements of F_0 that were based on cycles/time and overall duration were reliable and accurate.

Results

A preliminary test of the effect of the five trials on mean peak F_0 and word duration found no significant differences owing to trials, so data pooled across trials were used in subsequent analysis. The data were then pooled across trials and used in subsequent analysis.

Duration. Durations of the target words (*Miami*, *Geneva*, and *Bahamas*), produced under three speaking rates of slow, normal, and fast, were obtained. Means and standard deviations of durations in milliseconds across five trials and 10 participants are displayed in Table 1. The standard deviations suggest that requiring people to slow their rate of speaking introduces increased variability in performance.

A three-way analysis of variance (ANOVA), with repeated measures on two factors, was conducted on the data to examine the variables of group (male or female), word, and rate. Mean duration for male speakers was 691 msec, and that for females was 705 msec. No sex differences or interactions were found. A significant main effect of word was observed [$F(1,32) = 9.86, p < .01$], indicating a significant difference in actual articulation time of the three different words. The mean durations of *Miami*, *Geneva*, and *Bahamas* were 684, 667, and 742 msec, respectively. Durations for *Miami* and *Geneva* did not differ significantly.

A main effect of rate on duration was found [$F(2,16) = 22.53, p < .01$], because duration of utterance decreases as speaking rate increases. Mean durations across words and groups were all reliably different. No sex differences or interactions were observed for duration.

Table 2
Means and Standard Deviations of Word Peak Fundamental Frequency (in Hertz) for Males and Females in Experiment 1

Word	Slow		Normal		Fast	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Males						
Miami	192	73	257	68	258	77
Geneva	224	62	245	83	265	73
Bahamas	205	46	252	84	241	83
Females						
Miami	345	75	400	52	391	65
Geneva	334	69	371	62	376	76
Bahamas	336	92	386	89	387	69

Figure 1. Mean fundamental frequency in hertz for three temporal locations across three words produced by five male and five female talkers at three speaking rates.

Fundamental frequency. Peak fundamental frequencies of the target words (*Miami*, *Geneva*, and *Bahamas*), produced under three speaking rates of slow, normal, and fast, were obtained. Means and standard deviations of frequencies across five trials and 10 participants are displayed in Table 2, and males and females show a similar pattern of lower values for the words produced at the slow rate and higher values at normal and fast rates.

A three-way ANOVA, with repeated measures on two factors, conducted on the data to examine the effect of the variables of group (male or female), word, and rate confirmed this finding [$F(2,32) = 33.72, p < .01$].

Discussion

The speakers did not undershoot F_0 targets during rapid rates of speaking. In fact, as rate of speaking increases from slow to normal or fast, the F_0 of the peak intonation rises. One obvious explanation is that increased speaking rates require increased effort, tensing the larynx, which would raise the F_0 . The three rates produced no laryngeal undershoot in peak frequency and thus provide little evidence for coarticulation. However, it is quite possible that the beginning frequency of the unaccented syllable prior to reaching peak frequency was elevated in anticipation of the need for a rapid shift to the targeted peak frequency. That is, the talker might anticipate the need for a rapid rise to a predetermined peak by increasing the frequency of the preceding syllable, thus requiring a less extreme jump in frequency to reach the peak, as was seen by Xu (1994). This possibility was considered by examining the base 1 frequency under the three rate conditions. These means were 167, 156, and 169 Hz. An ANOVA, with gender, rate, and word as independent variables, was performed. The results showed no significant interactions, a significant effect of gender, as was expected, and no effect of rate on base 1 frequency. Figure 1 contains the means for males and females separately. Note that no effect of rate on base 1 or base 2 is apparent even with genders considered

separately. This is consistent with Arvaniti, Ladd, and Mennen (1998), who found that duration had minimal effect on the F_0 excursions during rising intonations in Greek, and with Caspers and van Heuven (1993), who found no differences in extent of F_0 rise in normal and fast productions in Dutch.

The data are consistent with those of Cooper and Sorensen (1981). The investigations differed, however, in that Cooper and Sorensen measured the average frequency (not the peak) and the duration of four words each in four sentences as rates of speaking increased. The average frequencies of the means of the first word from their study are representative of all four words and were 193, 201, and 200 Hz, respectively, for slow, normal, and fast rates. It was demonstrated that the slow rate differed significantly from the other two. Both studies showed reduced F_0 at slower rates (or increased F_0 at normal and fast rates of speaking) despite the use of different measurements and very different intonation patterns. Undershoot, as evidenced at the segmental level of speech production, did not occur at the suprasegmental level under investigation here.

EXPERIMENT 2

Experiment 1 demonstrates that peak F_0 in a rise-fall intonation pattern is reduced at a slow rate of speaking. The purpose of Experiment 2 was to determine whether rate of speaking (with peak F_0 held constant) would influence participants' judgments of the pitch of that peak F_0 in the incredulous rise-fall intonation in the word *Miami*. If the participants were to hear a different (higher) pitch for a slow *Miami*, as compared with the pitch of the normal or fast production, it would indicate that their everyday listening experience had led them to "expect" a lower F_0 at slow rates of speaking and they had compensated accordingly. Alternatively, if pitch is unaffected by rate, it would suggest that intonation, at least of this type, is not normalized like those present in the perception of certain speech segments. Note that the choice of a simple rise-fall contour is consistent with several studies involving the perception of pitch in frequency-modulated pure tones and tone complexes that suggest that listeners are more sensitive to pitch peaks than to pitch minima (Demany & Clement, 1995a, 1995b; Demany & McAnally, 1994). The same is true for pitch changes in speech where the difference limen (DL) is less for rises than for falls (t'Hart, 1981).

Method

Participants. The participants were 10 adult speakers with English as their primary language and no obvious speech or hearing deficits. None had participated in Experiment 1. The participants were graduate students at the university, who voluntarily participated without compensation. Ages ranged from 22 to 35 years, with a mean age of 26.

Stimuli. The design of Experiment 2 called for the participants to provide data on the pitch of the peak of the rise-fall intonation in the word *Miami* in slow, normal, and fast forms. The three rates were produced by waveform editing of one normal rate production to

yield slow and fast versions with exactly the same peak F_0 . In addition, complex periodic waveforms with a range of peak F_0 were used as comparison stimuli or standards in a pitch judgment task, to provide measures of the pitch of the peak F_0 of three tokens. Nonspeech tokens that matched the speech tokens in F_0 contour and duration also had the peak pitch of their rise-fall F_0 contours measured in the same way. Measurements from speakers in Experiment 1 were used as guidelines to establish the duration of the F_0 changes present in the word *Miami*. One target production of the incredulous sentence *I said Miami* was then recorded by an adult male speaker (not used in Experiment 1) to approximate the calculated normal durations and produce a natural incredulous rise-fall. The sentence was then digitized as before on a Macintosh Power 8100 AV using Signalyze 3.0 software to edit the waveform. First, the utterance of *Miami* was separated from the sentence. Cycles were added to or deleted from the waveform in order to increase and decrease durations to represent slow and fast speaking rates of the word. A single cycle of the waveform was deleted every fourth cycle uniformly throughout the utterance to produce the appropriate duration replicating the mean fast rate of speaking. Extending the utterance proved less straightforward, because simply adding single cycles at a constant interval produced an utterance that was slightly distorted perceptually. A procedure was developed of locating every fifth cycle and duplicating those cycles twice, throughout the word. In order to further extend the utterance, additional cycles were added as needed to the steady F_0 portions of the vowels in the initial and final syllables. These procedures resulted in three durations of the word *Miami* spoken with an incredulous rise-fall intonation, categorized as slow, normal, and fast, respectively. The peak F_0 of each token was 220 Hz. Physical differences in rate prepared in this manner were verified as perceptually salient through an informal listening procedure requiring four listeners to judge the naturalness of the stimuli and which were slow, normal, and fast. The token durations measured 794, 627, and 500 msec for the slow, normal, and fast rates, respectively. All tokens were 220 Hz.

The participants judged the peak pitch of a rise-fall stimulus by comparing a target with fixed pitch with standards with one of seven discrete pitches in all. Utilizing a complex nonspeech stimulus allowed the listener to compare the intonation or melody with speech without involving additional semantic content. A square wave with duration and F_0 matching the target utterance provided a perceptually equivalent stimulus for use in such comparisons.

Standard stimuli of systematically varied rise-fall complex waveforms were created using a wave-generating program of the CSRE42 software and a Softek 286 computer. The temporal characteristics

and corresponding fundamental frequencies of the target utterances of *Miami* produced by the speaker were measured. Figure 2 shows the results for the normal rate *Miami*.

The peak frequency of the target utterance was found to be 220 Hz. This was designated as the peak frequency of the normal rate *Miami*. Square-wave waveforms were then generated with matching durations and frequencies differing only in the peak frequency in each. The peak frequency was varied systematically in steps of one semitone. For example, for the normal rate word with a peak frequency of 220 Hz, seven square-wave stimuli were created with varying peaks a distance of one semitone apart, with one stimulus being 220 Hz in peak F_0 and three being above and three below that frequency of 220 Hz. Figure 2 illustrates the set of nonspeech standards for judging pitch of the normal rate *Miami*. A similar set of square waves was also generated, matching the temporal values for the slow rate and fast rate *Miami* with different peak fundamental frequencies as before. Three sets of square waves with varying peak F_0 were generated, seven in each set.

The Macintosh 8100 AV was programmed with Hypercard 2.2 to create and present stimuli for the pitch measurement process. Each stimulus pair consisted of a target (whose pitch was to be judged) and a square-wave standard, separated by 1,030 msec. The standard was drawn randomly from the set of seven waveforms matched in duration to the target, but with variable peak fundamental frequencies. The participants performed two different tasks. In one task, the slow, normal, or fast production of *Miami* was the target. In the other task, the square wave (whose temporal characteristics matched the slow, normal, or fast *Miami*) was the target, and the seven square waves with variable peak F_0 were the standards. The selection of 7 stimulus pairs was repeated five times to create 1 Hypercard stack allowing 35 listener judgments. A run in each rate condition consisted of 10 such stacks, for a total of 350 stimulus pairs of normal rate/duration. Runs were then created in the same manner for slow and fast rate conditions. Stimuli to be judged consisted of sets of square-wave pairs and of speech and square-wave pairs in slow, normal, and fast rate conditions.

Familiarization to task. The experiment consisted of approximately six sessions 1 h in length. The first session included at least 1/2 h of instruction. The participant was exposed first to stimuli with square waves as both the target and the standard and then to stimuli with speech as the target and square waves as the standard. The participants judged whether the peak pitch of the standard was higher than, lower than, or equal to the peak pitch of the preceding target square wave. Feedback on the accuracy of the participants' pitch judgment was provided for the square wave to square wave stimuli

Figure 2. Plots of (F_0) across time showing results of extraction of F_0 from target variable and F_0 contours of seven square waves with peak F_0 one semitone apart used as a variable standard.

Table 3
Means and Standard Deviations of Frequency (in Hertz)
Judgments of the Participants in Experiment 2

	Mean Square-Wave Target			Speech Target		
	Slow	Normal	Fast	Slow	Normal	Fast
Grand mean	220	219	220	225	223	225
SD	2	1	1	6	5	7

during this session. If the participant erroneously judged the square-wave stimulus, the stimulus was replayed, and the actual pitch match (higher, lower, or equal) was identified for the participant. When the participants judged 80% or more of the stimuli correctly, they were considered ready to participate in the experiment. Following this, they were briefly exposed to the experimental task involving stimuli consisting of *Miami* followed by a square wave. Feedback was not given during this exposure.

Procedure. The participants were tested individually in a sound-attenuated test booth. The stimuli were presented in the sound field via a 6-in. loudspeaker. The stimuli were presented on line, using a Macintosh 8100 AV computer, and were output at a 20-kHz sampling rate, low pass filtered and amplified with a Crown D75 amplifier. The stimuli were adjusted so that the target and the standard were equally loud. The overall level was then set at 70 dB SPL on the C scale, at the participant's ear. The speaker was placed in front of the participant at a distance of 86 cm. Prior to presentation, sound levels had been measured in various locations around the room and indicated that no more than ± 1 dB variation in overall level occurred within 50 cm of the location of the participant's head.

The participants listened to each stimulus pair and judged whether the peak pitch of the standard was higher than, lower than, or equal to the preceding square wave or speech. The participant recorded responses on a sheet that was provided.

Peak fundamental frequencies of the standard in each stack were recorded on a key prior to initiation of the tasks. The standards judged as equal in pitch to the target by the participant were identified and associated with the actual peak F_0 . Each frequency was then multiplied by the number of times it was identified by the participant, and these were totaled and divided by the number of equal responses in order to derive a mean F_0 (in hertz) associated with the set of *equal* judgments for that task. This number is now referred to as the *pitch* of the target.

Results

The participants made pitch judgments of both square-wave targets and speech targets in Experiment 2. The mean pitch judgment in the task with the square wave as target (220 Hz) was significantly lower than the mean pitch judgment of the speech target [224 Hz; $F(1,18) = 18.05$, $p < .01$].

The primary focus of the experiment was to determine the effect of rate on judgments of the peak pitch of the rise-fall intonation in *Miami* at three different rates. The pitch of each of the three nonspeech tokens was also obtained. The means across rates did not vary significantly. They were 222, 221, and 223 Hz for slow, normal, and fast rates, respectively, when collapsed across speech and nonspeech targets. No significant interactions were noted. The mean frequency judgments are summarized in Table 3.

The data were further analyzed to determine whether rate had a differential effect on the pitch obtained from the speech targets, as opposed to the corresponding square-

wave targets. For each participant, the frequency difference was obtained by subtracting the pitch of the square-wave target in hertz from that of the corresponding speech target. The means were 5.05, 4.35, and 4.96 Hz for the slow, normal, and fast rates, respectively. The ANOVA indicated no significant differences.

A final analysis was performed on the data of the participants to examine intraparticipant reliability on speech to square-wave judgments. The mean participant judgments of the first five stacks (175 responses) were compared with same of the second five stacks. The means were 220.5 and 221.1 Hz. The difference was found to be nonsignificant, and the correlation was .70.

Discussion

The primary finding in Experiment 2 was the fact that the rate at which the target word *Miami* was produced had no significant effect on the pitch of the peak F_0 in the rise-fall intonation contour. No interactions between rate and the speech-versus-nonspeech factors were noted. There was no evidence for perceptual accommodation to the effects of rate of speaking on pitch. There was, however, a significantly higher pitch associated with the three speech targets than with the analogous nonspeech targets. This effect, in the range of 4 Hz (224 vs. 220 Hz, approximately 0.4 ST), suggests the existence of a modest effect of speech on pitch perception. Perhaps the rapid rise-fall intonation leads listeners to exhibit perceptual overshoot in general, not related to the rate of speaking at which the words were produced. On the other hand, the judgment of the pitch of the speech tokens was a more difficult judgment than that of the nonspeech, despite the acoustic similarity between them, and the higher pitch of the speech may reflect the uncertainty added by the linguistic content of the peak. In any event, the effect is small, much less than the standard deviation for the *equal* judgments.

The DL for pitch ranged from 0.5 ST to 2.2 ST. The DL for pitch in speech has been studied with a variety of stimuli, ranging from short vowel segments (Flanagan & Saslow, 1958; Klatt, 1973) to numbers, parts of sentences, and whole sentences (Harris & Umeda, 1987; Klatt, 1973; t'Hart, 1981). The imposition of sentence structure, with its linguistic content, added length, resynthesis uncertainty, and variety of possible contours and declinations, leads to much larger DLs than is found with short synthetic vowels. For vowels, the values are approximately 0.3 Hz (Flanagan & Saslow, 1958; Klatt, 1973). For phrases and sentences, however, at best, the DLs are in the range of 4–10 Hz (Harris & Umeda, 1987; Klatt, 1973; t'Hart, 1981). In addition, Rietveld and Gussenhoven (1985) found similar sensitivity to changes in prominence, approximately 1.5 ST. Interestingly, even the addition of a simple rising or falling contour on a vowel increases the DL to the 2–4 Hz range (Klatt, 1973), yet the shape of the contour (linear or parabolic) has minimal influence on perception (t'Hart, 1991). Thus, the DLs between 0.5 and 2.2 ST are slightly smaller than those for sentences, but not as

small as those for vowels, reflecting, perhaps, the multisyllabic nature of the stimulus *Miami*, which falls between an isolated vowel and a phrase in length and complexity.

In Experiment 2, stimuli of durations representing three rates of speaking were matched to three speech targets of corresponding durations. It is possible, however, that the duration of the standard square wave or the duration of the target word could affect perceptual judgments. Experiment 3 was designed to investigate such effects.

EXPERIMENT 3

In order to test for possible effects of the rate of the standard, an experiment was conducted in which the target was held constant, using only the normal rate target *Miami*, and the rate of the standard was varied (slow, normal, or fast). Recall that in Experiment 2, only targets and standards of identical duration were compared. It is possible that the duration of the speech targets interacted with the nonspeech standards of the same duration. Consequently, another set of experimental conditions was employed in which the standards were held constant at the normal rate and were paired with the *Miami* targets at three different rates. Alternatively, it is possible that the square-wave standard itself exerted an effect of rate, independent of the rate of the speech (or nonspeech) target. That is, for example, the fast square wave may simply be too brief to allow a good auditory image of its pitch to be available to the listener.

Method

Participants. The participants were 10 adult speakers with English as their primary language and no obvious speech or hearing deficits. None had participated in Experiments 1 or 2. The participants were graduate students at the university, who voluntarily participated without compensation.

Stimuli. The stimuli previously prepared for Experiment 2 were used. The stimulus pairs for each condition were prepared by matching the word *Miami*, altered to replicate slow, normal, and fast rates, to normal rate square waves.

Familiarization to task. Instruction was accomplished as previously described. The participant was first exposed to stimuli with feedback, where both the target and the standard were square waves. The participants were then exposed to stimuli where speech was the target and square waves were the standard, but without feedback.

Procedure. The listeners in this experiment also participated in approximately six 1-h sessions. The normal rate *Miami* square-wave condition with the normal rate standards was presented first, followed by the fast and slow rate *Miami* conditions, each with normal rate standards, alternated by participant. When this task was completed, the participants were presented once again with stimuli consisting of normal rate *Miami* and normal rate square-wave standards, followed by stimuli with fast rate *Miami* and normal rate standards and slow rate *Miami* and normal rate standards. The order of presentation of the slow and fast standard conditions was alternated by each participant. The same number of presentations (10 sets of 35 stimuli) as that used in Experiment 2 was used.

Results

The first part of the experiment required the speech rate to vary as the square-wave standards were held constant at

the normal rate. The mean pitch reported for slow, normal, and fast durations were 220, 219, and 221 Hz, respectively. Standard deviations were 7.2 Hz in the slow rate, 7.3 Hz in the normal rate, and 5.1 Hz in the fast rate conditions. No significant difference among mean frequencies was found.

The second part of the experiment required that speech rate be held constant as the square-wave standards varied in rate. Here again, the main effect of varied square-wave rate was not significant. The means were 220, 221, and 218 Hz for the square-wave durations of slow, normal, and fast. Standard deviations were 4.3 Hz in the slow rate, 6.2 Hz in the normal rate, and 7.4 Hz in the fast rate. Mean frequencies and standard deviations are listed in Table 4.

A final analysis was performed on the data from all 20 participants (Experiment 2 and Experiment 3 combined) to examine intraparticipant reliability. Participant performance on speech to square-wave judgments was examined for reliability. The mean participant judgments of the first five stacks (175 responses) were compared with the mean participant judgments of the second five stacks for each word to square-wave comparison in each duration condition of long, normal, and short. The mean for the judgments for the first half was 220.5 Hz, as opposed to 221.1 Hz for the second half. The difference was found to be nonsignificant. Correlations between the first half and the second half were found to be .70.

Discussion

The focus of Experiment 3 was to examine possible effects of varying the rates of either the square waves or the speech on perceptual judgment of pitch. The results show no effect of rate of target or standard, in that mean judgments of pitch were not significantly influenced by varying the rate of the speech target while holding the rate of the nonspeech standard constant or by holding rate of speech target constant and varying rate of nonspeech standard. The results of Experiment 2, then, would appear to be free of such contamination. Experiment 3 also provided a replication of one aspect of Experiment 2. In Experiment 2, a small but significant effect of using speech as the target was found in the pitch judgments. An elevation of approximately 5 Hz (0.4 ST) was found with the speech targets, but not with the nonspeech (square-wave) targets. In Experiment 3, the participants also judged the pitch of the normal rate *Miami* with normal rate (nonspeech) standards in both control conditions. The means for these were 219 and 221 Hz, which does not differ statistically from

Table 4
Means and Standard Deviations of Pitch (in Hertz)
of the Participants in Experiment 3

	Constant Square-Wave Target			Constant Speech Target		
	Slow	Normal	Fast	Slow	Normal	Fast
Grand mean	220	219	221	220	221	218
SD	7	7	5	4	6	7

the mean of pitch judgment of 220 Hz for nonspeech targets from Experiment 2. That is, there was no elevation in the pitch of the speech targets. The discrepancy between the results of Experiments 2 and 3 may be explained in several ways. First, the participants in Experiment 2 had continuous exposure to the square-wave targets, which may have provided an additional reminder of the existence of the speech–nonspeech contrast. The participants in Experiment 3 heard square-wave targets only during training, thus decreasing opportunities for the participants to reinforce the contrast in a similar manner. However, the effect is weak, if present at all, and the partial replication (Experiment 3 used only normal rate stimuli in a way comparable to Experiment 2) was unsuccessful.

EXPERIMENT 4

Words are most frequently spoken and heard within a sentence context. Obviously, perception of speech stimuli is not limited to short utterances, such as word units. Although the listener has the ability to determine an intonation pattern of an utterance as short as a monosyllabic word, connected speech processing takes place over longer utterances, such as those making up a phrase or a sentence.

In Experiment 4, the effect of sentence context on the perceptual processing ability of the participant was investigated. The sentence context was used to determine whether the results obtained in Experiment 2 could be replicated when the target was included in a meaningful sentence.

Method

Participants. The 20 participants who served as listeners in Experiment 2 and Experiment 3 participated in this experiment.

Stimuli. The sentence was created using the original digitized file of *I said Miami* from which the stimuli used in Experiments 2 and 3 were taken. The experimenters and several other listeners heard the stimulus and perceptually judged the stimulus to sound “natural” with no discernible perturbations of pitch. Essentially, the original sentence was restored. The sentence was again paired with seven normal duration square waves that varied by one semitone at the peak frequency only, and the resulting stimuli were arranged into random sets of five 35-stimulus runs.

Familiarization to task. All the participants utilized in this task had participated in either Experiment 2 or Experiment 3 and so were familiar with the task. No additional training took place.

Procedure. The procedure was as described for the previous experiments, with the participant seated in the sound-attenuated booth. The participant heard the sentence, followed by a square wave with frequency and duration characteristics to match the normal duration *Miami*. The participant was asked to judge whether the peak pitch of the square wave was higher than, lower than, or equal to the peak pitch of the preceding word *Miami*.

Results

The analysis required matching the participants’ performance in the normal duration word condition with their performance in the sentence condition. The participants were divided into two groups on the basis of their participation in the previous experiments. Group 1 had participated

in Experiment 2, and Group 2 had participated in Experiment 3.

Recall that the comparable mean judgment of Group 1 was 223 Hz (normal rate condition, speech stimulus), with a standard deviation of 5 Hz, whereas the comparable mean judgment of Group 2 was 221 Hz, with a standard deviation of 6 Hz. The means in the sentence task were 223 Hz ($SD = 6$) and 224 Hz ($SD = 9$), respectively. The ANOVA showed these differences to be nonsignificant

Discussion

The purpose of this experiment was to examine the possible effects of increased stimulus length and sentence context on perceptual judgments of peak pitch. The participants in Group 1 showed little effect of single-word versus sentence context, yielding a mean of 223 Hz in both conditions. Group 2 showed an increase of 3 Hz in the sentence task. This is interpreted as a modest increase, at best, and the overall mean for the two groups (223.5 Hz) reveals no dramatic effect of the sentence context. The analysis suggests that the participants were able to make pitch judgments consistent with their judgments of the single-word target when the speech stimulus was lengthened from three to five syllables and constituted a sentence.

GENERAL DISCUSSION

There is evidence to suggest that phoneme identification occurs at an early stage of speech processing. Marslen-Wilson (1984) has shown that word identification may be completed at 200 msec or earlier into the word, a location that may be only halfway through the word. Phoneme perception is seen as part of the early bottom-up perceptual process whereby speech provides a rich acoustic matrix from which the possible word candidates can be selected (Marslen-Wilson, 1987). Furthermore, the evidence suggests that phoneme identification requires, in many instances, knowledge of the talker’s rate of speaking; thus, Miller has proposed that rate of speaking is a general parameter extracted early, much like estimates of vocal tract length (Miller, 1987).

Intonation, on the other hand, because of the length of the patterns and the essentially relative nature of its contrasts, is of necessity derived only later in the comprehension of a sentence. It is reasonable, then, to find no clear association between rate of speaking and the extraction of a specific characteristic—in this case, peak pitch in a rise–fall intonation pattern.

Initially, of course, we believed that F_0 undershoot and its perceptual derivative might occur, although even articulatory undershoot is not universally found, depending on the consonantal context employed. Gay (1978), for example, found no lingual undershoot in bilabial stop contexts. Ainsworth and Lindsay (1986) speculated that intonation was categorical in its perception, and several studies had shown that higher pitch levels and greater pitch excursions yield judgments of faster rate of speaking in the absence

of any physical change in the rate of articulation (Ainsworth & Lindsay, 1986; den Os, 1985; Kohler, 1986). That is, pitch influences perception of rate. However, the reverse appears not to be true; rate of speaking showed no influence on pitch judgments. The association between pitch and rate of speaking appears to operate in only one direction. Furthermore, to the (unknown) extent that a sentence such as *I said Miami!* is representative of short sentences in general, the results of Experiment 2 are generalizable to larger contexts. Obviously, several sentence contexts would be necessary to allow more than tentative generalizations.

The absence of any such relationship implies that the complex accommodations necessary at the segmental level to process speech of varying rates and necessary to process rapid articulations such as stops, regardless of overall rate of production, are not required for processing the relatively slowly varying F_0 contours associated with intonation. If one accepts Flanagan's (1972) analogy of speech as a vowel carrier modulated by consonant constrictions, nonlinear accommodations are a property of the modulation, and intonation is a property of the carrier, with different mechanisms used in their decoding. Another way of saying this is that the articulations associated with producing even a rapid rise-fall intonation are not fast enough to stress or compromise the physiological capabilities of the larynx. Thus, no undershoot of the type seen in the production of individual phones is present, and no perceptual accommodations are necessary.

Experiment 1, in fact, indicates that the opposite effect is present. At normal and rapid rates, as compared with slow rates of speaking, the larynx, if anything, overshoots its F_0 target, presumably owing to overall energy or tension associated with normal or fast articulation. That is, assuming that articulatory tension associated with normal and faster rates is associated with increased vocal fold tension, the result would be higher fundamental frequencies (Zemlin, 1998, p. 153).

It remains to be seen whether both the physiological effects of rate on production and the lack of perceptual effects of rate of production in a carefully isolated rise-fall intonation pattern are generalizable to other more common, although "less clean," intonation contours. The rise-fall contour was selected for this very purity—that is, it is restricted to a single word, does not cross clause or phrase boundaries, and is subject to minimum consonantal influence. The generality of the findings in this study would be tested rigorously through similar research involving the pitch and F_0 of a terminal rise, for example, produced at varying rates of speaking. Such rises are longer, may span several words, and can certainly be subject to consonantal perturbations. As such, they pose measurement and analysis problems but might provide information on the robustness of the results of the present study.

To the extent that these findings may be generalized, it is possible to speculate on some aspects of the process of speech perception. There may indeed be different perceptual mechanisms applied to segmental and suprasegmen-

tal characteristics of speech. For example, it is thought that perception of melody is primarily a right-hemisphere phenomenon, although extensive musical training may alter this finding (Raffman, 1993). Furthermore, intonation, perhaps because of its melodic nature, appears to be processed in the right hemisphere. Thus, it may be possible that the brain directs intonation information to the right hemisphere and rapid segmental recognition to the left (Zatorre, Evans, Meyer, & Gjedde, 1992). It is also possible, as has been suggested by Mullennix, Johnson, Topcu-Durgun, and Farnsworth (1995), that F_0 of voice is processed primarily as an auditory event, as can be seen in the perception of gender. That is, the incredulous rise-fall used in the present investigation may be processed strictly as an analogue signal where the amount of pitch rise signals the amount of incredulity. Such processing would be in the right hemisphere and not subject to a strict linguistic/articulatory mode of perception.

The primary finding of the present study, the lack of influence of rate on the pitch on an intonation contour, is consistent with this conceptualization. However, this finding does not prove that suprasegmental characteristics of speech in general are perceived in a different nonsegmental manner. Theories and models of speech perception have been described as belonging to one of three approaches: invariant feature/cue-based, motor theory, or direct-realist (Nygaard & Pisoni, 1995). Inasmuch as all the models are primarily segmental in focus, the implication of the present findings would appear to apply equally to these theories, with the suggestion that mechanisms and processes underlying the perception of segmental properties are not directly applicable to suprasegmental events. It is hoped that one effect of these studies will be to turn research attention to suprasegmental effects in order to form more complete theories of speech perception that include both segmental and suprasegmental aspects of speech perception.

REFERENCES

- ABRAMSON, A. S. (1979). The coarticulation of tones: An acoustic study of Thai. In T. L. Thongkum, P. Kullavanijaya, V. Ponupong, & T. L. Tingsabath (Eds.), *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugenie J. A. Henderson* (pp. 1-9). Bangkok: Chulalongkorn University Press.
- AINSWORTH, W. A., & LINDSAY, D. (1986). Perception of pitch: Movement on tonic syllables in British English. *Journal of the Acoustical Society of America*, **79**, 472-480.
- ARVANITI, A., LADD, D. R., & MENNEN, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, **26**, 3-25.
- CASPERS, J., & VAN HEUVEN, V. J. (1993). Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica*, **50**, 161-171.
- COOPER, W., & SORESENSEN, J. (1981). *Fundamental frequency in sentence production*. New York: Springer-Verlag.
- DANILOFF, R., & HAMMARBERG, R. E. (1973). On defining coarticulation. *Journal of Phonetics*, **1**, 239-248.
- DANILOFF, R., & MOLL, K. (1968). Coarticulation of lip rounding. *Journal of Speech & Hearing Research*, **11**, 707-721.
- DEMANY, L., & CLEMENT, S. (1995a). The perception of frequency peaks and troughs in wide frequency modulations: II. Effects of fre-

- quency register, stimulus uncertainty and intensity. *Journal of the Acoustical Society of America*, **97**, 2454-2459.
- DEMANY, L., & CLEMENT, S. (1995b). The perception of frequency peaks and troughs in wide frequency modulations: III. Complex carriers. *Journal of the Acoustical Society of America*, **98**, 2515-2523.
- DEMANY, L., & MCANALLY, K. (1994). The perception of frequency peaks and troughs in wide frequency modulations. *Journal of the Acoustical Society of America*, **96**, 706-715.
- DEN OES, E. A. (1985). Perception of speech rate: A scaling experiment. *Progress Report Institute of Phonetics, University of Utrecht*, **10**, 35-43.
- FELDSTEN, S., & BOND, R. N. (1981). Perception of speech rate as a function of vocal intensity and frequency. *Language & Speech*, **24**, 387-395.
- FLANAGAN, J. L. (1972). *Speech analysis synthesis and perception*. New York: Springer-Verlag.
- FLANAGAN, J. L., & SASLOW, M. G. (1958). Pitch discrimination for synthetic vowels. *Journal of the Acoustical Society of America*, **30**, 435-442.
- FOUGERON, C., & JUN, S. A. (1998). Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics*, **26**, 45-69.
- GAY, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, **63**, 223-230.
- HARRIS, M., & UMEDA, N. (1987). Difference limes for fundamental frequency contours sentences. *Journal of the Acoustical Society of America*, **81**, 1139-1145.
- KELLER, E. (1994). *Signalize: Signal analysis for speech and sound*. Lausanne, Switzerland: Infosignal.
- KLATT, D. (1973). Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception. *Journal of the Acoustical Society of America*, **53**, 8-16.
- KOHLER, K. J. (1986). Parameters of speech rate perception in German words and sentences: Duration, F_0 movement and F_0 level. *Language & Speech*, **29**, 115-139.
- LADD, D. R., FAULKNER, D., FAULKNER, H., & SCHEPMAN, A. (1999). Constant "segmental anchoring" of F_0 movements under changes in speech rate. *Journal of the Acoustical Society of America*, **106**, 1543-1554.
- LADD, D. R., MENNEN, I., & SCHEPMAN, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America*, **107**, 2685-2696.
- LINDBLOM, B. E. F. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, **35**, 1773-1781.
- MARSLÉN-WILSON, W. D. (1984). Function and process in spoken word-recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 125-150). Hillsdale, NJ: Erlbaum.
- MARSLÉN-WILSON, W. D. (1987). Functional parallelism in spoken word-recognition. In U. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 71-102). Cambridge, MA: MIT Press.
- MILLER, J. L. (1987). Mandatory processing in speech perception. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 309-322). Cambridge, MA: MIT Press.
- MULLENNIX, J. W., JOHNSON, K. A., TOPCU-DURGUN, M. M., & FARNSWORTH, L. M. (1995). The perceptual representation of voice gender. *Journal of the Acoustical Society of America*, **98**, 3080-3095.
- NYGAARD, L., & PISONI, D. B. (1995). Speech perception: New directions in research and theory. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language and communication* (pp. 63-96). San Diego: Academic Press.
- OHDE, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America*, **75**, 224-230.
- OHDE, R. N., & SHARF, D. J. (1975). Coarticulatory effects of voiced stops on the reduction of acoustic vowel targets. *Journal of the Acoustical Society of America*, **58**, 923-927.
- PETERSON, G. E., & BARNEY, H. E. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, **24**, 693-703.
- PIERREHUMBERT, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, **66**, 363-369.
- RAFFMAN, D. (1993). *Language, music, and mind*. Cambridge, MA: MIT Press.
- RIETVELD, A. C. M., & GUSSENHOVEN, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, **13**, 299-308.
- RIETVELD, A. C. M., & GUSSENHOVEN, C. (1987). Perceived speech rate and intonation. *Journal of Phonetics*, **15**, 272-285.
- SAXENA, P. (1992). The effect of speaking rate on the vocal intensity and F_0 topline. *Psycho-Lingua*, **22**, 51-56.
- SHADLE, C. H. (1985). Intrinsic fundamental frequency of vowels in sentence context. *Journal of the Acoustical Society of America*, **78**, 1562-1567.
- SILVERMAN, K. E. A. (1987). *The structure and processing of fundamental frequency contours*. Unpublished doctoral dissertation, Cambridge University.
- STEVENS, K., & HOUSE, A. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech & Hearing Research*, **6**, 111-128.
- STUDDERT-KENNEDY, M., & HADDING, K. (1973). Auditory and linguistic processes in the perception of intonation contours. *Language & Speech*, **16**, 292-313.
- SUNDBERG, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics*, **7**, 71-79.
- TERKEN, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America*, **89**, 1768-1776.
- TERKEN, J. (1993). Baselines revisited: Reply to Ladd. *Language & Speech*, **36**, 453-459.
- T'HART, J. (1981). Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America*, **69**, 811-821.
- T'HART, J. (1991). F_0 stylization in speech: Straight lines versus parabolas. *Journal of the Acoustical Society of America*, **90**, 3368-3370.
- XU, Y. (1994). The production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, **95**, 2240-2253.
- XU, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, **55**, 179-203.
- ZATORRE, R. J., EVANS, A. C., MEYER, E., & GJEDDE, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, **256**, 846-849.
- ZEMLIN, W. R. (1998). *Speech and hearing science: Anatomy and physiology* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

APPENDIX

Sentences Used in Experiment 1

I saw my friend the other day. He said he had been to Miami. I said, "*Miami!* I thought you went to Missoula for your vacation!"

I saw my friend the other day. He said he had been to Geneva. I said, "*Geneva!* I thought you went to Savannah for your vacation!"

I saw my friend the other day. He said he had been to the Bahamas. I said, "*Bahamas!* I thought you went to Missoula for your vacation!"

(Manuscript received June 26, 2000;
revision accepted for publication July 4, 2001.)