

RESEARCH ON SPEECH PERCEPTION

Progress Report No. 15 (1989)

Indiana University

Stress Clash in Isolated Phrases and Sentence Contexts¹

Keith Johnson and Michael S. Cluff

*Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405*

¹An earlier version of this paper was presented at the 117th meeting of the Acoustical Society of America in Syracuse, New York. We appreciate the suggestions and comments of Mary Beckman and Stuart Davis. This research was supported by NIH Training Grant NS-07134-09 to Indiana University.

Abstract

The experiment reported in this paper investigates the possibility that stress clash should be called "accent clash". Stress clash in the phonological literature has been defined in terms of lexically specified stresses, while experimental investigation of stress clash candidates defined in this way has failed to find evidence for such rhythmic effects in speech production. This study tests the hypothesis that lexical stress defines the candidates for a clash and that the placement of accents during speech production determines whether a clash will actually occur. Subjects produced pairs of adjective noun phrases in which the presence of a stress clash was manipulated (loose cannon/loose canoe) both as isolated phrases and within sentence environments. It was found that subjects showed the segmental duration effects which are predicted by metrical phonology when the clashing stressed syllables were both accented (the phrase contexts) but not when only the noun was accented (the sentence contexts). It was also found that the temporal location of the peak of F1 was not affected by the clash/nonclash manipulation and that F1 trajectories for the clash and nonclash conditions diverged after the peak of F1.

Stress Clash in Isolated Phrases and Sentence Contexts

Patterns of segmental duration variation can be attributed to a number of segmental and structural factors. Among the factors commonly acknowledged are (1) inherent segment duration, (2) segmental context, (3) lexical stress, (4) location of a word within a syntactic structure, and (5) the presence or absence of emphasis. Klatt (1976) demonstrated that most of the observed phenomena in the literature on segmental duration in American English can be accounted for if these factors are included in a concatenative model of segment duration.

In contrast to a concatenative model of segmental duration, a hierarchical model has been proposed as appropriate for the description of rhythmic properties of speech production (Pike, 1943; Lehiste, 1977; Huggins, 1975, 1978; Fourakis and Monahan, 1988), and thus, indirectly also for some aspects of segmental duration. According to this hypothesis, in addition to the segmental and structural factors which play a role in concatenative models, rhythmic structure affects segmental durations in speech production.

Although this metrical approach is intuitively appealing, there is very little experimental evidence for rhythmic effects in speech production. It has been clearly demonstrated that, even in the most favorable situations, English speakers do not produce speech in which the intervals between stressed syllables are isochronous (see for example Nakatani et al., 1981). Rather, the strongest evidence that rhythm plays a role in speech comes from studies of speech perception (see Lehiste, 1977; Darwin and Donovan, 1980; Martin, 1970; and Allen, 1975). The evidence against isochrony in speech production taken together with the evidence that rhythm plays a role in speech perception indicates that rhythm (if it plays a role in speech production at all) is only one of a number of factors which determine segmental durations. This further suggests that evidence of rhythmic effects in speech production will be in the form of subtle (but predictable) changes in segmental durations within the constraints of other, concatenative factors affecting segmental duration.

Cooper and Eady (1986) took this attitude and measured utterances for which the metrical theory of phonology (Liberman and Prince, 1977; Selkirk, 1984) makes explicit predictions concerning rhythmic effects in speech production. They investigated two putative phenomena, (1) stress shift and (2) stress clash. In stress shift, it is predicted that the location of stress in a multi-syllabic word will be shifted to the left (earlier in the word) in order to preserve an alternating stress pattern. For example, "bamboo" is normally stressed on the last syllable (e.g. bamBOO), but in the phrase "bamboo tables" the stress seems to fall on the first syllable (BAMboo TABLES). This stress shift results in an alternating stress pattern which is the preferred state of affairs in English and perhaps universally (Selkirk, 1984, p. 12, *passim*). Stress clash has a similar description and motivation, except that in stress clash situations it is not actually possible to shift the location of stress within the word because there is no "stressable" syllable to the left of the lexically stressed syllable and thus, other rhythmically motivated effects occur. For example, the phrase "cement tables" involves the same type of situation which results in stress shift in "bamboo tables", but the reduced vowel

in the first syllable of "cement" (in some dialects of English) is unstressable. Of course, when there is only one syllable in a word, stress clash is the only possibility (e.g. "big tables"). In cases of this type, Selkirk (1984, p. 186ff) has suggested that the first stressed syllable (the "-ment" of "cement" or "big") is phonetically lengthened "as a manifestation of the tendency toward isochrony of beats." Cooper and Eady (1986) tested both the stress shift and stress clash predictions of (this version of) metrical phonology and found evidence for neither. So, even with rather more subtle expectations concerning the phonetic realization of speech rhythm, there is still very little evidence for rhythmic patterns in speech production.

Johnson and Evans (1987) replicated one of the experiments conducted by Cooper and Eady (1986) and added a manipulation of speaking style. Subjects read sentences in which the presence of a stress clash was manipulated in both a normal reading style and in a careful reading style (as if speaking over a bad telephone line). In neither reading style was there evidence for a durational adjustment resulting from a stress clash. Johnson and Evans suggested that rhythmic phenomena such as stress clash may depend less on lexically specified stressed syllables and more on accent placement within an utterance. This relates back to Bolinger's (1972) maxim, "Stress belongs to the lexicon. Accent belongs to the utterance." Lexically specified stressed syllables may or may not be accented when an utterance is pronounced, because the presence of a pitch prominence (intonational accent) on a particular word depends on semantic, pragmatic and perhaps syntactic factors (Bolinger, 1972; Bresnan, 1972; Schmerling, 1976; Selkirk, 1984). The hypothesis of Johnson and Evans (1987), which was tested in the experiment reported here, is that the "beats" in speech rhythm are the intonationally prominent syllables, not lexically stressed syllables. In order to test this hypothesis we had subjects read utterances which involved stress clash and nonclash environments (as defined by lexically specified stress) in two conditions. In one condition, only one of the lexically stressed syllables was given a pitch prominence; in the other condition, both lexically stressed syllables received a pitch prominence. We predicted that there would be a durational adjustment to stress clash only in those utterances which also involve an accent clash. The data of this experiment suggest that stress clash is best defined at the level of the utterance and in terms of pitch accents rather than at the level of lexical stress. "Stress clash" is therefore really "accent clash".

Method

Materials. In Webster's Pocket Dictionary, all nouns with primary lexical stress on the first or second syllable were identified through the use of a lexical search program. Sixty-five noun pairs which had the characteristics shown in Table 1(a) were selected. Examples of noun pairs that fit these criteria are: person-percent, broker-brochure, raven-ravine.

Insert Table 1 about here

Next, a semantically appropriate adjective was selected for each noun pair, and the nouns were paired with both the adjective and its comparative or superlative (two syllable) counterpart. Examples of the test noun phrases are shown in Table 1(b). The phrases were embedded as the first NP in simple NP-VP sentences. In addition, the number of syllables in the sentences were balanced, in order to control for possible effects of sentence length. Thus, 260 sentences were constructed, four for each noun pair. Examples of the sentences are shown in Table 1(c). Note that the first sentence of each pair with a monosyllabic adjective contains a lexical stress clash, while the second does not.

Procedure. Two native speakers of American English, one male and one female, who were unfamiliar with the purpose of the study, spoke each of the 260 items twice. In the first session, subjects read the test noun phrases in their sentence contexts. The subjects were instructed to emphasize the noun of the subject noun phrase in a normal declarative intonation. In the second session, the noun phrases were produced in isolation. The subjects were instructed to read the phrases as if someone had just made a statement involving the item in the phrase, and that they were repeating the phrase with some surprise. The subjects produced phrases which had two pitch accents as illustrated in Figure 1. In this figure, the average of the F0 contours of the items produced in *phrases* (by subject LM) have two pitch prominences, while the average F0 contour of the items produced in *sentences* is characterized by only one pitch accent.¹

Items to be read were presented to subjects on a video terminal located in a sound-attenuated booth and item presentation, randomization and digital sampling were all under computer control (Dedina, 1987).

Insert Figure 1 about here

Thus, there were three independent variables manipulated in this study; (1) clash versus nonclash environments (such as "loose cannon" versus "loose canoe"), (2) the number of syllables in the adjective ("loose canoe" versus "looser canoe"), and (3) the number of intonational accents in the subjects' productions of each item (items read in sentences were produced with one intonational accent, while items read as isolated phrases were produced

¹The contours can be transcribed in Pierrehumbert's (1980) system of intonational transcription as L+H* L* H H% (phrase condition) and H* L L% (sentence condition). Traces for Figure 1 are time normalized averages of voiced frames across all tokens.

Table 1

Criteria used in selecting materials, and examples of the test noun phrases and sentences used in the experiment.

(a) Criteria.

1. Members of the pair had the same number of syllables.
2. They had segmentally similar first syllables.
3. One member of the pair receives primary stress on the first syllable, while the other receives primary stress on the second syllable.

(b) Example Noun Phrases.

	First syllable stress	Second syllable stress
One syllable Adjective	"the loose cannon"	"the loose canoe"
Two syllable Adjective	"the loser cannon"	"the loser canoe"

(c) Example Sentences.

The loose/loser CANNON knows government secrets.
The loose/loser CANOE glides along the current.
The large/larger BUFFER fills the computer's memory.
The large/larger BUFFET fills a very long table.
The long/longer OVERTURE precedes a short opera.
The long/longer OVATION inspires the singer.

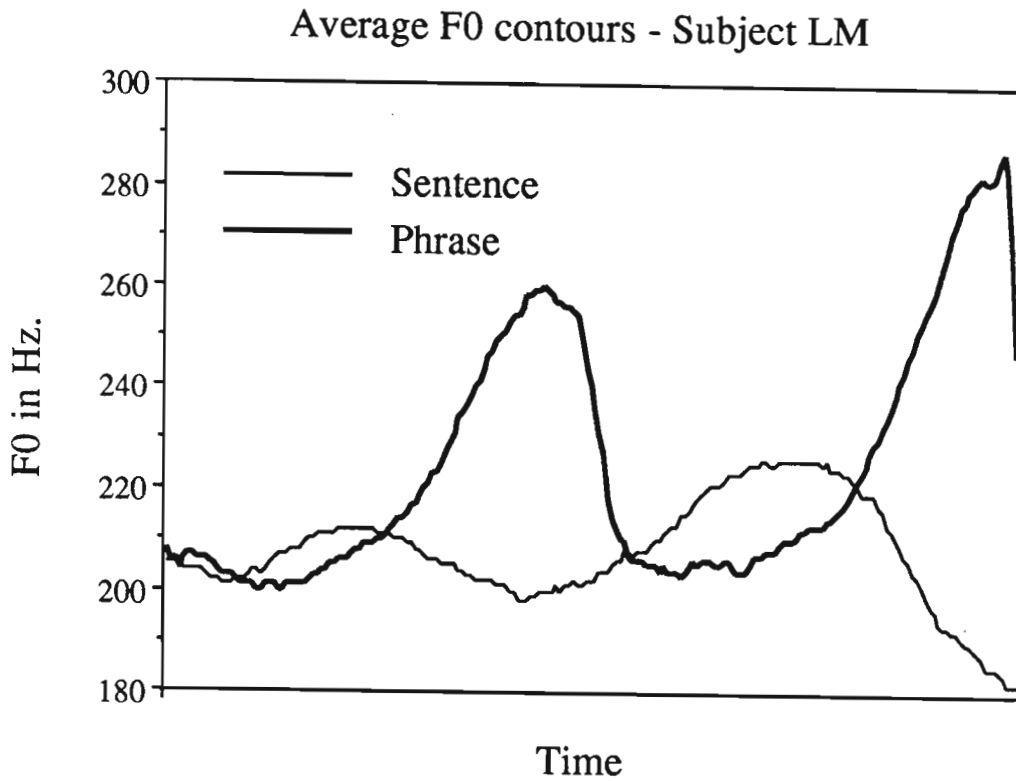


Figure 1. Average F0 contours of the test NP's for the items produced in sentence contexts and items produced in isolated phrases (subject LM). The phrase items show pitch prominence for both the adjective and the noun.

with two pitch accents). The dependent variables were the duration of the adjectives and the temporal locations of the peaks of F1 and RMS amplitude during the first syllable of the adjective.

Results

Durations of the adjectives were measured from waveform and pseudo-spectrogram displays of the digital signal.² F1 was estimated by LPC analysis at intervals of 12.8 ms throughout the adjective. The autocorrelation method of LPC was used to calculate LPC coefficients and a peak-picking algorithm was used to find F1. Gross errors and missing values were corrected by interpolation and the F1 contours were then smoothed by a three point smoothing filter.

Insert Figure 2 about here

The adjective duration results are shown in Figure 2. There was a main effect for sentence versus phrase contexts. Adjectives produced in phrases were longer than those read in sentences [$F(1, 128) = 2064.1, p < 0.001$] (phrase=469.6 ms, sentence=338.2 ms). We may note two possible sources of this effect. First, the adjectives produced in phrases carried an intonational accent and thus, may be considered more emphatic, or focused than those same adjectives in sentential context (see Klatt, 1976 on the effect of emphasis on segmental duration). Second, Huggins (1978) has suggested that, "The more words there are in a sentence, the shorter each word tends to become" (p. 287). Without addressing the issue of whether the unit of analysis for this generalization should be the sentence or the intonational phrase, we may point out that the extra length for adjectives produced in phrases fits Huggins' observation.

The statistical analysis also revealed (unsurprisingly) an effect of the number of syllables on overall word duration [$F(1, 128) = 383.3, p < 0.001$]. Two syllable adjectives were longer than one syllable adjectives (448.6 ms versus 359.3 ms).

There was a main effect for the stress clash factor [$F(1, 128) = 21.04, p < 0.001$] (non-clash=398.6 ms, clash=409.2 ms) as well as a significant interaction of stress clash and accentual structure (phrase context versus sentence context) [$F(1, 128) = 19.41, p < 0.001$].

²We also measured the interval between the onset of the adjective and the onset of the following noun (i.e. adjective plus pause). An analysis of the onset-to-onset data was also conducted, but because this analysis showed the same statistically reliable effects which were found in the analysis of the adjective duration data, only one set of data is presented here (the adjective duration data).

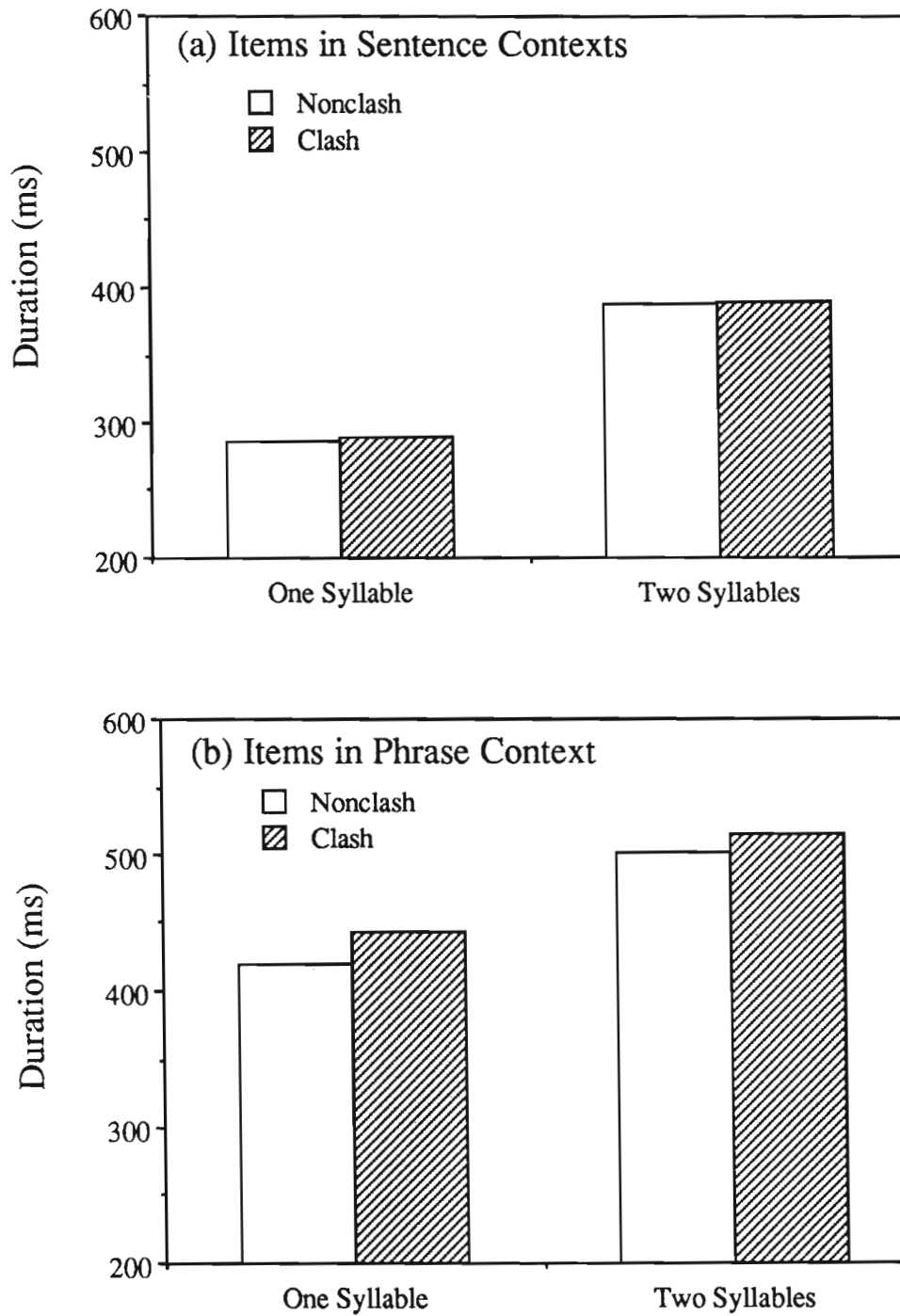


Figure 2. Adjective duration results averaged across subjects. (a) Items produced in sentence contexts. (b) Items produced in phrase contexts. The filled bars are the clash tokens and the open bars are the items produced in no-clash environments.

This interaction is shown in Figure 2. In a post-hoc comparison of means it was found that the difference between stress clash and no stress clash was reliable only for the items which had been produced in the phrase condition (for both monosyllabic and bisyllabic adjectives). This was true for both subjects. The fact that there was no effect of stress clash in the sentence condition replicates the findings of Cooper and Eady (1986), while the change in the duration of the adjective in the phrase condition supports our view of speech rhythm.

There was also an interaction between the accentual structure condition and the number of syllables [$F(1, 128) = 26.01, p < 0.0001$]. The difference between one and two syllable adjectives was smaller when the items were produced in a phrase context. Recall that items produced in phrasal context had a pitch accent on the adjective, while items produced in sentences did not. This interaction suggests that when accent is placed on a monosyllabic word, the relative overall increase in duration is greater than when accent is placed on a bisyllabic word.

An analysis of the temporal locations of the peak of F1 in these utterances revealed no differences in the location of F1 peaks as a function of stress clash. The only statistically reliable effect in the analysis of the F1 peak data was the accentual structure main effect [$F(1, 115) = 29.09, p < 0.0001$]. The peak of F1 occurred on average 178 ms into the word when items were produced in sentence contexts and 189 ms into the word when items were produced in phrase contexts. This effect is consistent with the large duration difference found for these contexts, although interestingly, the magnitude of peak shift is much smaller (11 ms) than the magnitude of duration difference (>130 ms).

Similarly, only the accentual structure effect was reliable in the analysis of the RMS peak data [$F(1, 128) = 110.12, p < 0.0001$]. The average location for the peak of RMS amplitude was 162.6 ms in tokens produced in sentences while the peak occurred later (210.3 ms) when the items were produced in phrases.

Figure 3 demonstrates the relevance of the peak data. This figure shows average F1 contours for subject LM's productions of monosyllabic adjectives in the phrase condition. The temporal locations of the peaks were not reliably different across the clash and no-clash environments while the clash items had longer duration. As illustrated in Figure 3 the durational difference between clash and nonclash items was realized on the end of the word.

Insert Figure 3 about here

Conclusions

Overall this pattern of results suggests: (1) that rhythmic effects in speech production are best described in terms of accents rather than lexically stressed syllables, and (2) that

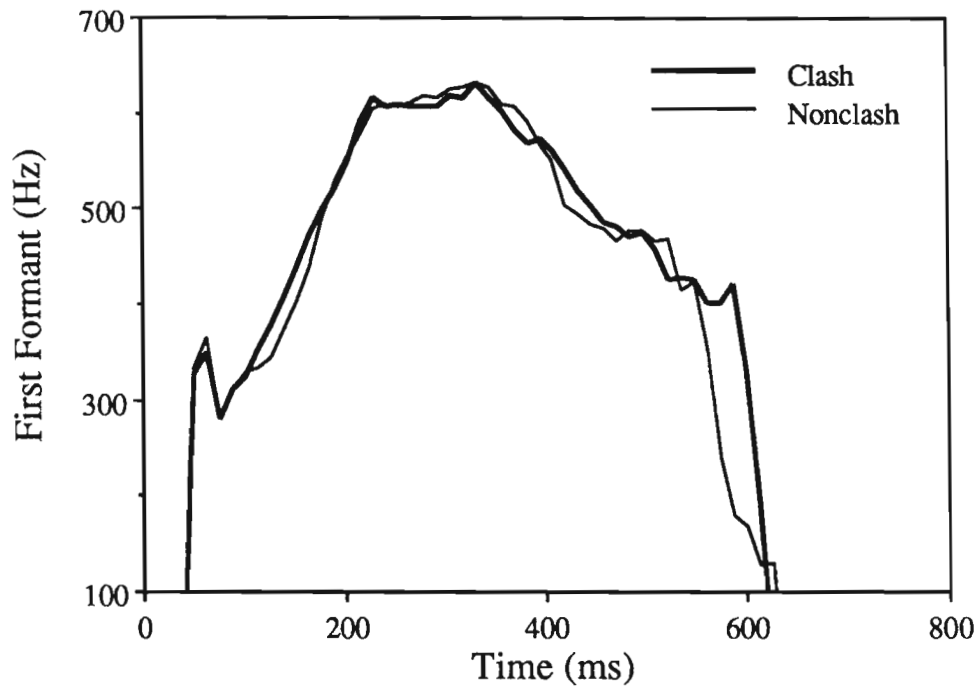


Figure 3. Average F1 contours for the monosyllabic adjectives produced in phrases by subject LM.

the rhythmic adjustment to stress clash involves a change in speech production localized on the final portion of the word and not a wholistic rescaling of articulation.

The first conclusion is motivated by the fact that we found the durational effect which is predicted by a metrical or hierarchical approach to the description of speech production, but that this effect was only found when subjects produced the utterances which involved a stress clash with accents on the "clashing" syllables. In other words, we found a durational adjustment to accent clash, but not to stress clash. Our null result in the case of stress clash is consistent with that reported by Cooper and Eady (1986) and so the two studies are in that sense mutually confirming. However, our results suggest a different conclusion. Cooper and Eady concluded, "At least some of the presumed 'facts' of rhythmic patterns presented in metrical phonology do not hold up under empirical testing" (p. 383). We would rather conclude that the rhythmic patterns described in metrical phonology have been wrongly attributed to lexically determined stresses when, in actuality, they are better described as properties of actual (rather than potential) pronunciations; the timing of pitch accents.

Insert Figure 4 about here

The second conclusion is motivated by the fact that the temporal locations of the peaks of F1 and RMS amplitude were not reliably affected by stress clash even in those cases in which overall duration was affected. This indicates that the durational difference between clash and no-clash environments occurs over the last part of the word. Note that we found the same pattern of results (durational difference coupled with no change in peak of F1 or RMS amplitude) in both monosyllabic and bisyllabic adjectives. This suggests that intervening syllables (whether one or two) impinge upon the closing gesture of an accented syllable and have almost no effect on the opening gesture. This is illustrated in Figure 4. The top panel of this figure shows hypothetical F1 contours of the four types of utterance used in this experiment. In these idealized schema the peaks of sonority for the clash and no-clash items are isochronous. In the nonclash case the intervening unstressed syllable(s) overlaps with the closing gesture of the first word and the opening gesture of the second word. The bottom panel in Figure 4 is data from subject EG. Each function is an average across the 65 tokens produced in each condition (items produced in phrases). The data diverge from the hypothetical situation in that the interval between the F1 peaks for the first and second words are clearly dependent upon the number of intervening syllables (anisochrony). They are, however, similar to the hypothetical data in that the closing portion of the first word in the clash condition is different from that in the no-clash condition. This data (for one syllable adjectives, at least) is comparable to that reported by Beckman (1989), "the jaw opening gesture was relatively shorter and the closing gesture relatively longer in the stress clash context."

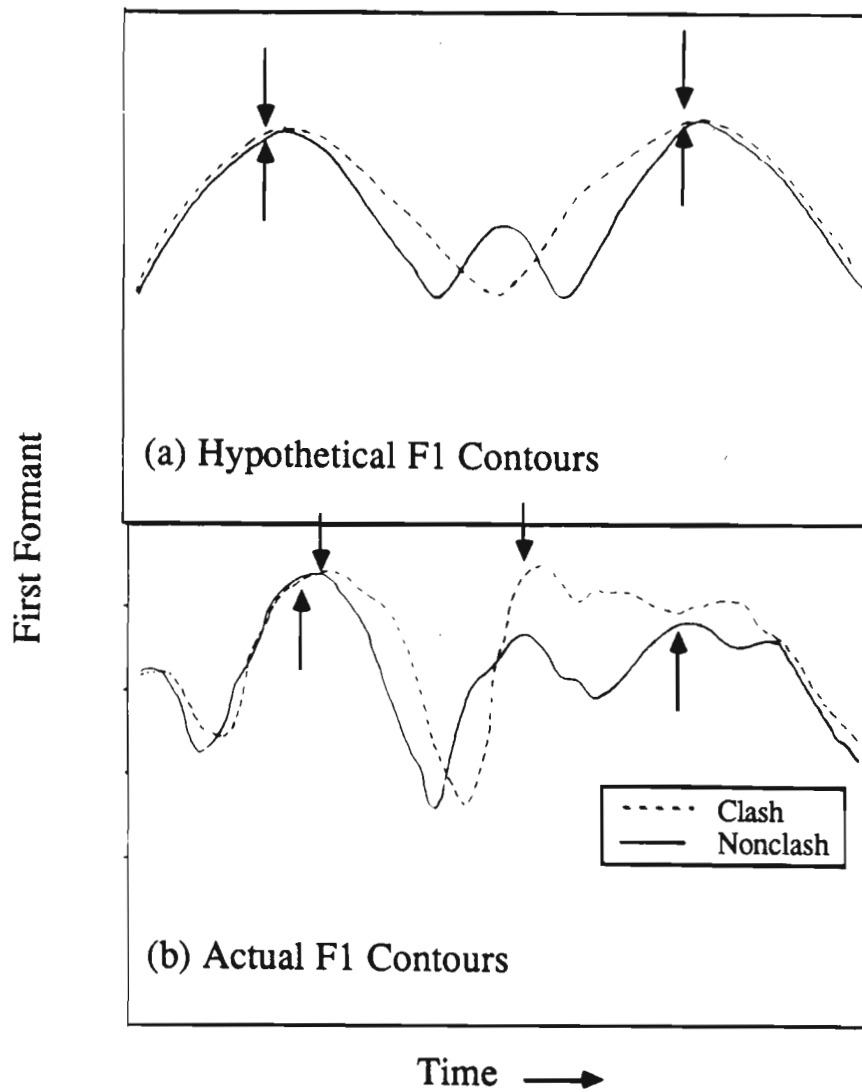


Figure 4. Hypothetical (a) and actual (b) F1 contours. The actual contours are averages ($n=65$) of subject EG's productions of the monosyllabic and bisyllabic adjectives (F1 traces of the entire noun phrase are shown) in clash and no-clash conditions in phrases. The contours were aligned at the onset of the phrase.

References

- Allen, G. D. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, **3**, 75-86.
- Beckman, M. E. (1989). Articulatory correlates of stress clash rhythms. *Journal of the Acoustical Society of America*, **85**, S96.
- Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, **48**, 633-644.
- Bresnan, J. (1972). Stress and syntax: A reply. *Language*, **48**, 326-342.
- Cooper, W. E. & Eady, S. J. (1986). Metrical phonology in speech production. *Journal of Memory and Language*, **25**, 369-384.
- Darwin, C. J. & Donovan, A. (1980). Perceptual studies of speech rhythm: Isochrony and intonation. In Simon, J., editor, *Spoken language generation and understanding*. D. Reidel, Dordrecht.
- Dedina, M. (1987). SAP: A speech acquisition program for the SRL-VAX. *Research on speech perception progress report no. 13*. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University.
- Fourakis, M. & Monahan, C. B. (1988). Effects of metrical foot structure on syllable timing. *Language and Speech*, **31**, 283-306.
- Huggins, A. (1975). On isochrony and syntax. In Fant, G. & Tatham, M., editors, *Auditory analysis and perception of speech*. Academic Press, New York.
- Huggins, A. W. F. (1978). Speech timing and intelligibility. In Requin, J., editor, *Attention and performance VII*. Lawrence Erlbaum Associates, Hillsdale, N.J..
- Johnson, K. & Evans, D. (1987). Phonetic reality and phonological prediction: Stress clash and rhythm in English. *Journal of the Acoustical Society of America*, **81**, S66.
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, 1208-1221.
- Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, **5**, 253-263.
- Lieberman, M. & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, **8**, 249-336.
- Martin, J. G. (1970). Rhythm-induced judgements of word stress in sentences. *Journal of Verbal Learning and Verbal Behavior*, **9**, 627-633.

Nakatani, L. H., O'Connor, K. D., & Aston, C. H. (1981). Prosodic aspects of american english speech rhythm. *Phonetica*, **38**, 84-106.

Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. *Unpublished Ph.D. dissertation, MIT*.

Pike, K. L. (1943). *Phonetics*. The University of Michigan Press, Ann Arbor, Michigan.

Schmerling, S. (1976). *Aspects of English sentence stress*. University of Texas Press, Austin, Texas.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. MIT Press, Cambridge, MA.