

Lund University, Dept. of Linguistics
Working Papers 35 (1989), 113-120

Perceptual Constraints and Tonal Features

David House

Abstract

This paper examines the tonal features High, Low, Falling and Rising from a perceptual point of view. Two questions are addressed, the first concerning the possibility of perceptual primacy of tonal levels over tonal movement and the second concerning the relationship between perception and the timing of tonal movement vis-à-vis segment boundaries. Results obtained from listener tests involving perception of tonal patterns in nonsense syllables provide an experimental basis for dealing with these questions. Results from these experiments conform to results obtained from perception experiments in the literature where utterances from Swedish, Chinese and German were used. Finally, these results are examined in terms of explanatory value concerning certain aspects of universals of tone. (A version of this paper was presented at the Sixth International Phonology Meeting, Krems, Austria, 1-4 July, 1988.)

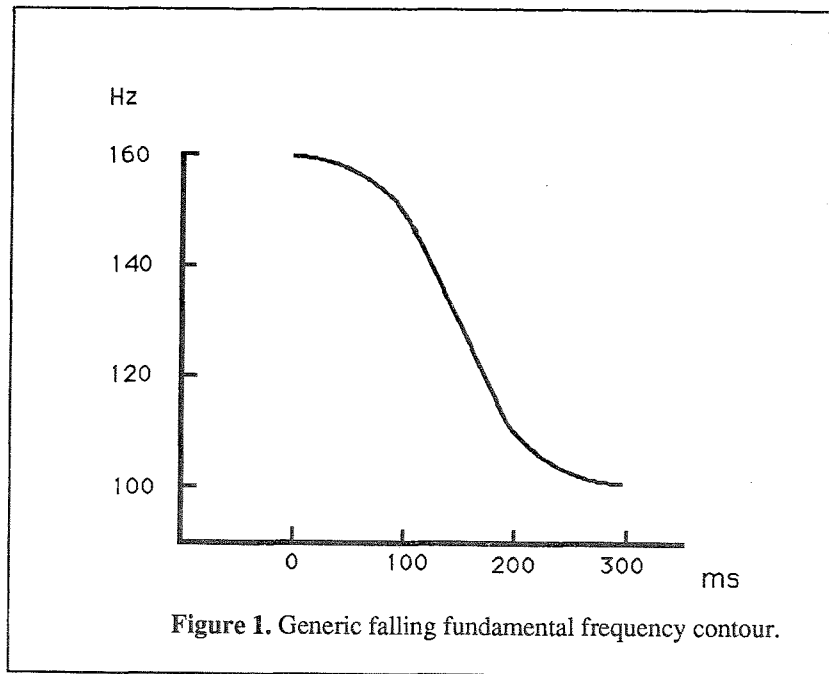
INTRODUCTION

The object of this paper is to present a model of optimal perceptual categorization of tonal movement based on the interplay between fundamental frequency movement and spectral change. Although language specific tonal contrast requirements may sometimes override an optimal perceptual structuring of tonal movement, a perceptual framework for the assignment of features should have explanatory value on a universal level.

Figure 1 represents an example of a generic falling fundamental frequency contour. What tonal feature or features should be assigned to this contour? A choice between movement or level features presents us with two ready possibilities: Falling or High + Low. (For a review of possible tonal features see Anderson 1978.)

In an experiment designed to compare the perception of tonal levels and tonal movement, Swedish listeners were given the task of categorizing rising-falling and falling-rising tonal patterns in a synthetic /a/ vowel (House 1987). Listeners were easily able to perform the task on the basis of tonal patterns. When a /b/ gap was introduced in different places in the vowel, listeners had difficulty categorizing the same tonal patterns. Categorization strategies shifted from what could be interpreted as tonal patterns or tonal movement to tonal levels or average tone frequency. These results seem to indicate that

actual pitch movement is optimally perceived during the steady-state portion of a vowel segment. When confronted with a more complex array of spectral and intensity shifts, the perceptual mechanism can recode the pitch movement in terms of levels. Pitch movement from one vowel to the next would then be interpolated as a movement from one level to another.

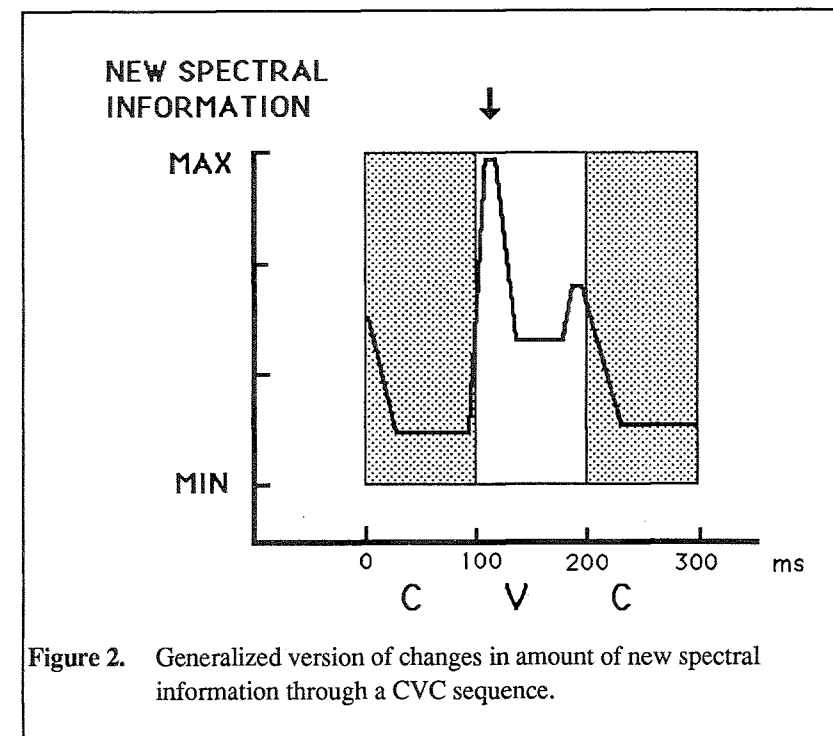


The assignment of tonal features to the contour presented in figure 1 would therefore depend upon the spectral configuration overlying the contour. If the spectrum were stable, as for example during a steady-state vowel, we would opt for the movement feature Falling. Spectral stability can even be seen as a prerequisite for the perception of a movement feature (Falling or Rising). A rapidly changing spectrum such as that caused by the insertion of a consonant would, on the other hand, optimize perception of levels (e.g. High + Low). It follows, then, that spectral changes during pitch movement introduce perceptual constraints, the first one being the perceptual primacy of tonal levels over movement. The second constraint involves the extraction of level features from the tonal movement contour. What portion of the contour is perceptually most salient for conversion into level features? To answer this

question we must look at the timing of tonal movement in terms of changing spectral information, i.e. segment boundaries.

THE INFLUENCE OF CHANGING SPECTRAL INFORMATION

Figure 2 presents a stylized prediction of how the amount of new spectral information can change through time in a CVC configuration. Of course, the degree of these changes varies greatly depending upon the individual vowel and consonants comprising the configuration. However, one element common to most CVC units is that a maximum of new spectral information coincides with rapidly rising intensity at vowel onset (cf. Stevens & Blumstein 1981).



We can speculate that because at vowel onset the perceptual mechanism is maximally loaded with the task of resolving spectral information its capacity to resolve fundamental frequency movement is decreased. The point at which new spectral information is at a maximum (marked by the arrow in figure 2) could be a significant point in time for the perception of tonal levels,

perceived in terms of average fundamental frequency over time. This would involve a recoding of tonal movement into tonal levels and would be in keeping with the results of the above-mentioned perception experiment where tonal movement perception is optimized during areas of spectral stability and tonal level perception is optimized during areas of large spectral change.

The influence of spectral change on tonal perception can be further illustrated by results from a listening test where the synthesized, nonsense utterance AMAMA with varying tonal contours was presented to Swedish listeners. The test was an ABX configuration where stimulus A (contour no. 1 from the left in Figure 3) consisted of a high F_0 in the first vowel followed by a falling F_0 through the first consonant and then a low F_0 through the rest of the utterance. Stimulus B (contour no. 9 from the left in Figure 3) consisted of a high F_0 through the first VCV portion followed by a falling F_0 through the second consonant with a low F_0 on the final vowel. Nine stimuli were then created by shifting the fall in increments of 25 milliseconds from stimulus A to stimulus B using a modified version of the Klatt software synthesizer (Klatt 1980). The tonal variations for the stimuli are presented in Figure 3.

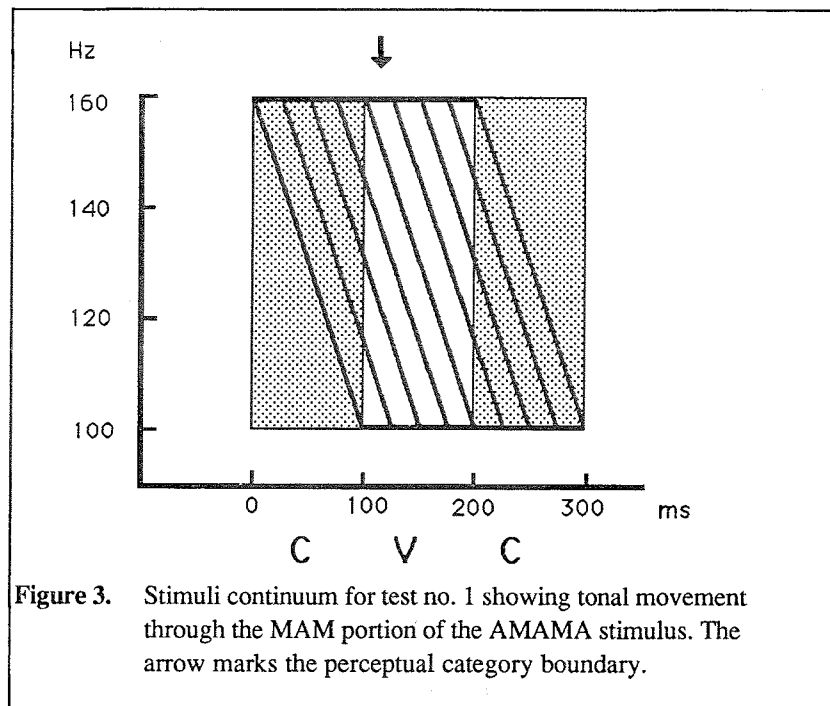


Figure 3. Stimuli continuum for test no. 1 showing tonal movement through the MAM portion of the AMAMA stimulus. The arrow marks the perceptual category boundary.

Eleven listeners were then given the task of choosing for each stimulus if it was most like stimulus A or stimulus B. The stimuli were randomized with each one being presented six times, three times as ABX and three times as BAX. Listeners grouped the first five falling contours as most like stimulus A and the last four contours as most like stimulus B. This category boundary, with reference to the beginning of the F_0 fall, is shown by the arrow in Figure 3. This boundary corresponds to the area of maximum amount of new spectral information as illustrated in Figure 2. It also, however, corresponds to the point in time nearly midway between the two endpoint stimuli. Therefore, a similar test was presented to the listeners where the two endpoint stimuli (A and B) were comprised of a falling F_0 contour through the first and second vowel respectively, i.e. the initial AMA of the AMAMA stimulus (see Figure 4).

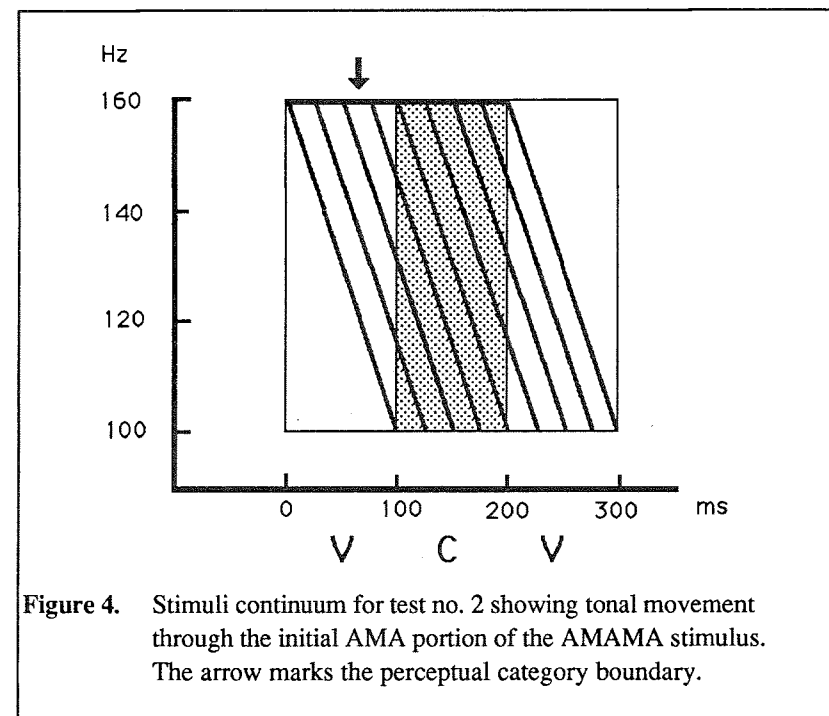


Figure 4. Stimuli continuum for test no. 2 showing tonal movement through the initial AMA portion of the AMAMA stimulus. The arrow marks the perceptual category boundary.

In these new stimuli, the area of maximal spectral change does not coincide with the temporal midpoint in the stimuli continuum but instead occurs near

the beginning of the F_0 fall for stimulus no. 4 (counting from the left). In this test, listeners grouped the first three stimuli as most like stimulus A and the last six as most like stimulus B. Here, the perceptual boundary corresponds not to the temporal midpoint but to the area of maximal spectral change during the continuum. A comparison of the results of the two different tests shows a category boundary shift of two stimuli or 50 milliseconds. This boundary shift represents the influence of the spectral configuration upon the perceptual categorization of tonal movement and demonstrates that listeners did not simply categorize on the basis of temporal distance from endpoint stimuli.

One interpretation of these results would be characterized by the perceptual recoding of movement to levels. In the first experiment (Figure 3), stimulus A (no. 1 from the left) can be said to have a tonal low in the vowel while stimulus B (no. 9 from the left) has a tonal high in the vowel. The tonal movement in the middle stimulus (no. 5 from the left) is perceptually most like stimulus A, i.e. a tonal low in the vowel. This could mean that the falling movement is recoded as a low in the vowel since the beginning of the fall occurs at the point of maximum new spectral information. The next stimulus (no. 6 from the left) is perceptually most like stimulus B, i.e. a tonal high in the vowel. We can speculate that this movement is recoded as a high in the vowel since it maintains a high level during the area of spectral change. The perceptual category shift is therefore from High-Low-Low in the three vowels of AMAMA to High-High-Low, the boundary occurring shortly after vowel onset.

This interpretation corresponds well to results from perceptual tests using linguistic categories in different languages. Bruce 1977 presents perception data for Swedish concerning the timing of the word accent fall in non-focal position. In his data, the perceptual boundary between accent I and accent II occurs when the midpoint of a falling contour is 30 milliseconds into the vowel. In an experiment involving the perception of Tone 3 and Tone 4 in Modern Standard Chinese, Gårding et al. 1986 demonstrated the influence of the CV boundary on the perceptual category shift. Here the category shift from Tone 3 to Tone 4 occurs when the tonal peak is moved from the VC boundary to a point 32 milliseconds into the vowel. Similar results for German have been reported by Kohler 1987 where perceived changes in meaning from 'established' to 'new' information result from a tonal peak shift through the CV boundary 60 milliseconds into the stressed vowel.

These results point towards a perceptually significant area some 30-60 milliseconds following vowel onset after a CV boundary. This area can be seen as a crucial point in time at which tonal level functions as a perceptually optimal feature. Tonal movement during this area is, therefore, suitable for conversion into a tonal level.

A MODEL OF OPTIMAL TONAL FEATURE PERCEPTION

On the basis of these results, we can now postulate an optimal perceptual model for the tonal features High, Low, Falling and Rising (features being capitalized in this text to distinguish them from absolute tonal movement and levels). Tonal movement through areas of spectral change will be optimally categorized as levels, a falling movement as Low and a rising movement as High. The movement features Falling and Rising need spectral stability during the movement. Additionally, they need reference levels at vowel onset preceding the movement (cf. Gårding et al. 1986). The feature Falling, for example, would be characterized by a tonal high extending some 30 to 60 milliseconds into the vowel followed by a tonal fall through an area of spectral stability. The feature Rising would be the reverse, a tonal low followed by a tonal rise. Perception of movement features is therefore strongly tied to the timing of tonal movement with reference to CV boundaries.

Duration, in addition to the reference level and spectral stability, is also most probably involved in the optimal configuration for perception of the movement features. Longer vowel duration would be associated with movement features while shorter duration would be associated with level features.

This model, with perceptual constraints on tonal movement features, can provide a framework for the assignment of features on a universal level. Although language specific tonal contrast requirements may sometimes override an optimal perceptual structuring of tonal movement, these perceptual constraints could explain certain aspects of universals of tone such as that reported by Maddieson 1978 where languages do not have contour tones unless they have at least one level tone. The features High and Low would therefore be more perceptually salient and more frequent than the features Rising and Falling.

These perceptual constraints and the synchronization between tonal movement and segmental boundaries appear to be important in languages such as Swedish and Chinese which make use of lexical movement features. In

these languages tonal production can be seen to make optimal use of the perceptual contrast between levels and movement by means of critical timing of tonal movement. In other languages where the use of movement features is less clear, this type of synchronization may not be as important. However, the tonal features Low, Falling, and High would seem to fit the data for German presented by Kohler 1987 where changes in pitch peak location through the stressed vowel result in categorical pitch perception of 'established', 'new' and 'emphatic' information respectively.

This model of optimal tonal feature perception is preliminary and somewhat simplified as it contains only four tonal features. An examination of production data from different languages could help expand the model and lead to further perception tests. An awareness of perceptual constraints may facilitate the analysis of production data and improve our description of the contrasts used to distinguish between different tonal features.

REFERENCES

- Anderson, S.R. 1978. 'Tone features'. *Tone: a linguistic survey*, ed. V.A. Fromkin, 133-175. New York: Academic Press.
- Bruce, G. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerups.
- Gårding, E., P. Kratochvíl, J.-O. Svantesson & J. Zhang. 1986. 'Tone 4 and tone 3 discrimination in modern standard Chinese'. *Language and Speech* 29, 281-293.
- House, D. 1987. 'Perception of tonal patterns in speech: Implications for models of speech perception'. *Proceedings of the Eleventh International Congress of Phonetic Sciences*, ed. Ü. Viks, 1, 76-79. Tallinn: Academy of Sciences of the Estonian S.S.R.
- Klatt, D.H. 1980. 'Software for a formant synthesizer'. *Journal of the Acoustical Society of America* 67, 971-995.
- Kohler, K.J. 1987. 'Categorical pitch perception'. *Proceedings of the Eleventh International Congress of Phonetic Sciences*, ed. Ü. Viks, 5, 331-333. Tallinn: Academy of Sciences of the Estonian S.S.R.
- Maddieson, I. 1978. 'Universals of tone'. *Universals of Human Language, Volume 2, Phonology*, ed. J. Greenberg, 335-365. Stanford, CA: Stanford University Press.
- Stevens, K.N. & S.E. Blumstein. 1981. 'The search for invariant acoustic correlates of phonetic features'. *Perspectives on the Study of Speech*, eds. P. Eimas & J. Miller, 1-38. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mora and Temporal-Tonal Interaction in Japanese

Yasuko Nagano-Madsen

Abstract

When seeking the phonetic realization of the mora, it has been a common practice to measure the acoustic duration on the assumption that it reflects the concept of quantity. It is argued that this duration-only approach is inadequate in certain cases. An alternative approach considering the interaction between temporal and Fo events is suggested to explain the abstract nature of the mora. It is also pointed out that words contrasting in vowel length have a strong tendency to accompany different pitch patterns in those dialects of Japanese that have pitch accent contrast.

INTRODUCTION

When seeking the phonetic reality of the mora, most works study only one of the acoustic dimensions, duration or fundamental frequency contour (henceforth Fo). This paper suggests that in certain cases this one-way approach is inadequate and that looking at the interaction between the two may explain the abstract nature of the mora better.

The phonetic research on the mora has been almost exclusively concentrated upon Japanese. In most dialects of Japanese, the mora is the smallest prosodic unit of tone assignment. The central issue in the description of the Japanese mora is its function as a temporal unit, which has long been controversial. It has been claimed that Japanese moras are pronounced with roughly equal length of time regardless of their internal moraic structure (Bloch 1950, Hockett 1955, Han 1962, Ladefoged 1975). Thus in deliberate speech (e.g. teaching children or foreigners), the words *nippon* 'Japan', *amaoto* 'the sound of rain', and *kjookai* 'church' will be pronounced as [ni-p-po-n], [a-ma-o-to], and [kjo-o-ka-i] ('-' equals mora boundary) respectively, each mora taking about the same time. This principle of pronunciation might have something to do with the moraic writing system in which, except for CjV mora, one *kana* character is assigned to one mora.

Some have recognized the mora as a relevant unit in regulating the Japanese utterances and have incorporated it in their production-oriented timing model (Port et al. 1980, 1987) while others have denied that the mora