A MODEL OF OPTIMAL TONAL FEATURE PERCEPTION

D. House

Department of Linguistics and Phonetics
Lund University, Lund, Sweden.

ABSTRACT

This paper postulates an optimal perceptual model for the tonal features *High, Low, Falling* and *Rising* based on proposed perceptual constraints of the auditory system in processing tonal movement in speech. These constraints involve two critical issues concerning the perception of tonal movement: namely the relationship between perception and the timing of tonal movement in terms of segment boundaries, and the perceptual primacy of level features over movement contour features.

1. INTRODUCTION

A classic descriptive problem in tone feature analysis concerns the division of tones into level tones and contour tones [1]. The principal question relating to this problem is whether tonal features can best be described in terms of both levels (e.g. *High, Low, Mid*) and contours (e.g. *Falling, Rising*) or as only levels and combinations of levels (e.g. *High + Low* instead of *Falling*).

This paper attempts to answer the question in terms of an optimal perceptual model of tonal movement categorization. The model is based on a series of speech perception experiments where tonal contours were varied in relation to segmental boundaries (see House [6] for a full account of the experiments). Results of the experiments indicate that actual pitch movement is optimally perceived as movement when it occurs during spectrally stable portions of vowels. Pitch movement occurring through areas of spectral discontinuities with rapidly shifting intensity and spectral information (roughly corresponding to segment boundaries) tends to be recoded in terms of pitch levels. In the model, therefore, tonal movement through areas of rapid spectral change is optimally categorized as level features while constraint conditions must be fulfilled for tonal movement to be optimally perceived as contour features.

2. THE MODEL

2.1. Description and Constraints

The model proposed here is an optimal perceptual model for the tonal features *High, Low, Falling* and *Rising*. According to the model, tonal movement through areas of rapid spectral change will be optimally categorized as level features, a falling movement as the feature *Low* and a rising movement as the feature *High*. These basic level features *High* and *Low* are then perceptually associated with the vowel following the rapid spectral changes. This perceptual recoding of movement into level features is also seen as a perceptual primacy of level features over movement contour features where level features can also be manifested without tonal movement.

For the movement contour features *Falling* and *Rising* to be optimally perceived, three constraint conditions must be fulfilled. First of all, the falling or rising movement must take place through a zone of relative spectral stability during the vowel. Second, the beginning of the movement must be synchronized in relationship to vowel onset so that the beginning of the fall or rise coincides with an area of decreasing new spectral information following the rapid spectral changes associated with the transitions from consonant to vowel. This enables pitch extraction of a relative pitch frequency (high before falling and low before rising) to which the perception of pitch movement direction can be calibrated. Finally, the model proposes a duration constraint which requires a vowel duration greater than 100 milliseconds to optimize movement feature perception.

100 milliseconds is an ad hoc value chosen to illustrate the durational component of the model. Relative vowel durations vary with speech tempo and speaking style, but the basic tenet is that longer vowel duration is associated with movement features while shorter duration is associated with level features. Based upon the effect of duration on the experimental results reported in [6], 100 milliseconds is a reasonable quantification of a duration constraint.

There may also be differences between rises and falls in their influence on production and perception of vowel durations [10]. These differences could have implications for the duration constraint in the model. However, for the purpose of simplicity in the model the duration constraint is the same for both rises and falls.

When the three constraint conditions are met, tonal movement is optimally categorized in terms of the movement contour features *Falling* and *Rising*. By implication, when the conditions are not met, tonal movement is then optimally categorized in terms of the level features *High* and *Low*.

Finally, the model assumes a tonal movement of 3 to 8 semitones per 100 milliseconds. Although the size of tonal movement needed for the optimal perception of movement contour features in the context of this model has yet to be tested experimentally, this range corresponds to that used in the experiments in [6] and in many other experiments and models (e.g. [5], [13] & [15]).

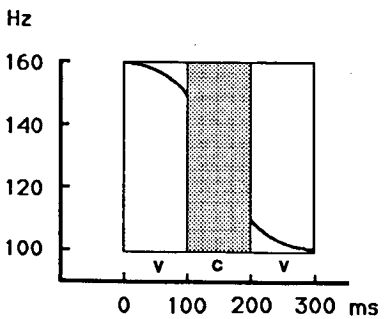

Figure 1. Illustration of the model applied to a falling tonal movement through a VCV context. The model predicts perception of the tonal features *High + Low* in the two vowels.

2.2. Illustrations

Illustrations of the model as applied to a prototypical falling fundamental frequency contour in different segmental contexts are shown in Figures 1–4.

In Figure 1, with a VCV context and most of the tonal movement occurring during the consonant, the model would predict recoding in terms of the level features *High* and *Low*. In this example, none of the three movement feature conditions is met.

Figure 2 presents a CVC context in which only one of the three movement feature conditions is met. Although the falling movement does occur during spectral stability, the beginning of the fall occurs during spectral change and is not synchronized with the area of decreasing new spectral information following vowel onset. Finally duration does not exceed 100 milliseconds. Here, the model would predict categorization in terms of the level feature *Low*.

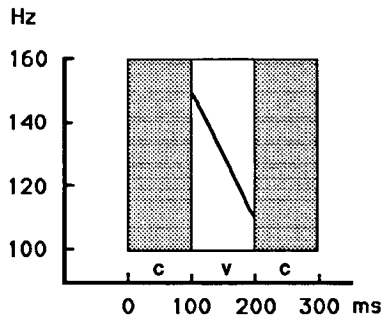

Figure 2. Illustration of the model applied to a falling tonal movement through a CVC context. The model predicts perception of the tonal feature *Low* in the vowel.

In Figure 3, a VC context is presented. Here, all three movement conditions are met. The model would therefore predict optimal coding in terms of the movement feature *Falling*.

Finally, in Figure 4, a CV context is presented. In this example, two of the three movement conditions are met. Tonal movement occurs during spectral stability in the vowel and vowel duration exceeds 100 milliseconds, but the beginning of the tonal movement is not synchronized with the end of maximum new spectral

information after vowel onset. Therefore the model would predict optimal recoding in terms of the level feature *Low*.

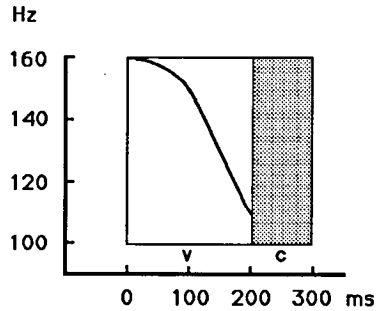

Figure 3. Illustration of the model applied to a falling tonal movement through a VC context. The model predicts perception of the tonal feature *Falling* in the vowel.

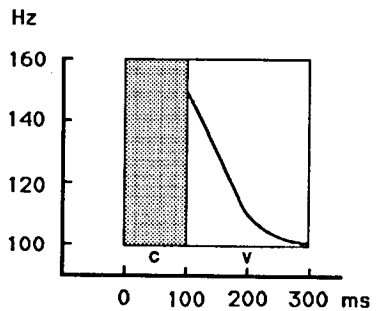

Figure 4. Illustration of the model applied to a falling tonal movement through a CV context. The model predicts perception of the tonal feature *Low* in the vowel.

In the illustrations above, a prototypically *falling* fundamental frequency contour was used. The model would deal with a *rising* contour in the same way with the same constraint conditions applying for optimal perception of the feature *Rising*.

Finally, it must be pointed out that the model described here is preliminary and does not claim to account for all possible tonal contrasts as it contains only four tonal features. More production and perception data from various languages could help expand the model as feature contrast requirements in different languages might alter the basic constraint conditions, especially where more features are needed.

3. IMPLICATIONS OF THE MODEL

3.1. Pitch Perception

The main implication of the model in terms of pitch perception theories is that it takes into consideration the constraints imposed on the perceptual system by the dynamic nature of speech. In the model, pitch movement sensitivity is strongly related to the speed of spectral change. In areas of rapid spectral change and spectral discontinuities, sensitivity is lower. In areas of spectral stability, sensitivity is higher.

The model assumes spectral analysis to be an important component in pitch perception. The use of resolved lower harmonics in pitch perception is also crucial for the model. It is the disruption and changes in the lower harmonics brought about by spectral discontinuities which give rise to the optimal recoding of tonal movement into level and movement contour features.

Thus the model supports central processing theories of pitch perception where a first order analysis of harmonic frequencies is crucial (e.g. [7]). Important to these theories is an interaction between spectral analysis and pitch extraction. This interaction is also crucial to the model.

3.2. Tone and Accent Features

Returning to the descriptive problem of contour tones versus level tones, the model clearly differentiates between contour and level features from a perception point of view. The model can be used to provide a framework for the assignment of features on a universal level. The model also implies perceptual primacy of levels over contours. These perceptual constraints can be used to explain certain aspects of universals of tone such as those reported by Maddieson [11] where it is claimed that languages do not have contour tones unless they have at least one level tone. Thus the features *High* and *Low* would be more perceptually salient and more frequent than the features *Rising* and *Falling*.

Perceptual constraints and the synchronization between tonal movement and segmental boundaries appear to be important in word accent and tone languages such as Swedish, Chinese and Thai 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 movements by means of the critical timing of tonal movement (cf. [2], [3], [4], & [8]).

In other languages where the use of movement features is less clear, this type of synchronization may not be as important. However, data from German [9] and English [12] seem to indicate that level and movement features related to critical timing and alignment of tonal movement may play an important perceptual role for these languages as well.

3.3. Speech Perception Theories

An additional factor of importance for the model is the load on the perceptual system at vowel onset. Spectral cues at vowel onset have been shown to be crucial for segment perception in speech [14]. An implication of the model could thus be separate perceptual mechanisms for segmental cues and tonal cues. Segmental perception mechanisms would then favor rapid spectral changes and discontinuities while tonal perception mechanisms would favor spectral stability. Following this line of argument, Fo differences at vowel onsets can function primarily as discriminatory cues for consonant features while Fo differences during spectral stability function as cues for tonal features.

4. CONCLUSIONS

In the model presented in this paper, level features are given perceptual priority over movement contour features. For the optimal perception of contour features, constraint conditions are proposed and illustrated. Although the details of these constraint conditions may vary between languages, the principles of perceptual constraints should be applicable to many different languages on a universal level. These principles provide a framework for the assignment of tonal and intonational features from a perceptual point of view.

Finally, in view of the importance of tonal features for the overall speech perception process, the addition of a tonal component can be seen as a necessary enrichment of general models and theories of speech perception. The tonal perception model presented here is an example of such a tonal component.

5. REFERENCES

[1] ANDERSON, S.R. (1978), "Tone Features", in V.A. Fromkin (ed.) *Tone: A linguistic survey*, 133-175, New York: Academic Press.

[2] BRUCE, G. (1977), "*Swedish Word Accents in Sentence Perspective* ", Lund: Gleerups.

[3] GANDOUR, J.T. (1983), "Tone perception in Far Eastern Languages", *Journal of Phonetics 11*, 149-175.

[4] GÅRDING, E., P. KRATOCHVIL, J.O. SVANTESSON, & J. ZHANG (1986), "Tone 4 and Tone 3. Discrimination in Modern Standard Chinese", *Lang. and Speech 29*, 281-293.

[5] HART, J. 'T. & A. COHEN (1973), "Intonation by rule: a perceptual quest", *Journal of Phonetics 1*, 309-327.

[6] HOUSE, D. (1990), "*Tonal Perception in Speech* ", Lund: Lund University Press.

[7] HOUTSMA, A.J.M. & J.G. BEERENDS (1987), "An Optimum Pitch Processing Model for Simultaneous Complex Tones", in U. Viks (ed.) *Proc. 11th ICPhS*, 4:325-330. Tallin.

[8] HOWIE, J.M. (1974) "On the Domain of Tone in Mandarin", *Phonetica 30*, 129-148.

[9] KOHLER, K.J. (1987), "Categorical Pitch Perception", in U. Viks (ed.) *Proc. 11th ICPhS*, 5:331-333. Tallin.

[10] LEHISTE, I. (1976), "Influence of fundamental frequency pattern on the perception of duration", *Journal of Phonetics 4*, 113-117.

[11] MADDIESON, I. (1978), "Universals of Tone", in J. Greenberg (ed.) *Universals of Human Language, Volume 2, Phonology*, 335-365. Stanford, CA: Stanford University Press.

[12] PIERREHUMBERT, J.B. & S.A. STEELE (1989), "Categories of Tonal Alignment in English", *Phonetica 46*, 181-196.

[13] ROSSI, M. (1978), "La perception des glissandos descendants dans les contours prosodiques", *Phonetica 35*, 11-40.

[14] STEVENS, K.N. & S.E. BLUMSTEIN (1981), "The Search for Invariant Acoustic Correlates of Phonetic Features", in P. Eimas & J. Miller (eds.) *Perspectives in the Study of Speech*, Hillsdale, NJ: Lawrence Erlbaum Associates.

[15] WILLEMS, N., R. COLLIER & J.'T HART (1988), "A synthesis scheme for British English intonation", *Journal of the Acoust. Soc. of America 84*, 1250-1261.