# PERCEPTION OF TONAL PATTERNS IN SPEECH: IMPLICATIONS FOR MODELS OF SPEECH PERCEPTION

DAVID HOUSE

Department of Linguistics and Phonetics
Lund University
Sweden

ABSTRACT

This paper advances a model of pitch perception in speech in which spectral changes influence the analysis of the tonal contour. This interrelationship is examined in view of certain linguistic requirements of tonal contours in the perception of spoken language. It is concluded that the perception of tonal movements is optimized when these movements occur in regions of spectral stability, that movement at the syllable level can be perceived directly as linguistic categories and that movement at the phrase level can be reconstructed from tonal levels stored in short-term memory.

## INTRODUCTION

Intonation provides listeners with important information which facilitates the perception of spoken language (1). In this paper the word intonation will be used in a wide sense, that of perceptually significant changes in fundamental frequency which have a linguistic function. The purpose of this paper is to examine how these changes and their relationships to spectral changes can be represented in the peripheral auditory system and in short-term memory, and how this representation can be used to aid and guide the speech perception process.

Information obtained from Fo movement can be greatly varied and can function on several different levels simultaneously. The type of information dealt with here concerns linguistic categories such as relative syllable importance (stress), relative word importance (focus), language specific information at the word level (word accents and tones), phrase boundaries (juncture) and connective patterns over a longer time domain (grouping). Some of the principles involved in Fo-movement perception might, however, also be applicable to other types of information such as emotions, involvement, etc.

Raw Fo movement must be transformed by the perceptual mechanism into relevant tonal categories. This transformation presupposes an analysis of frequency (pitch), direction of movement (rising, falling) and range of movement. Current psychoacoustic and physiological models of pitch perception are generally in agreement that some degree of central processing is involved, but it is still unclear as to what extent pitch analysis interacts with spectral resolution (2,3). Pitch perception in spoken language involves the additional problem of coping with rapidly changing spectral cues and a pitch contour broken up by voiceless segments. This leads to a key question. Is pitch analysis continuous, following Fo without being influenced by breaks and spectral events, or is it more selective and economical using critical portions of movement which are then stored in short-term memory and retained for decisions involving larger time domains? On the basis of two perception experiments, this paper advances a model which takes the latter view.

## PERCEPTION OF TONAL MOVEMENT AT THE SYLLABLE LEVEL

The first experiment was designed to test the influence of rapid spectral changes on the categorization of simple rise-fall and fall-rise tonal patterns at the syllable level. In this experiment, the categories were not linguistic ones but rather were presented to the listeners in the form of an ABX test design (4).

A Klatt software synthesizer and a VAX digital computer were used to synthesize a Swedish /ɑ/ vowel with formant frequencies of 600, 925, 2540 and 3320 Hz. (5,6). Vowel duration was 300 ms including 30 ms intensity onset and offset. Fundamental frequency was systematically varied to create 18 different stimuli. The Fo contour for stimulus A, designed to elicit rise-fall categories, rose from 120 Hz to a turning point of 180 Hz and then fell to an end point of 100 Hz. The Fo contour for stimulus B, designed to elicit fall-rise categories, began at 120 Hz falling to 80 Hz and then rose to 160 Hz. The difference in end-point frequency was designed to test the effect of end-point variation on the rise-fall, fall-rise categories, i.e. movement pattern versus discrete frequency analysis. The 18 stimuli were constructed by systematically varying the turning point in steps of 20 Hz from 80 Hz to 180 Hz with three different end-point configurations: 100 Hz, 160 Hz and 120 Hz. The beginning point was always 120 Hz. Listeners consistently categorized these stimuli on the basis of movement pattern and did not use end-point frequency.

To test the effects of rapid spectral changes on the categorization, three more versions of the test were made by introducing a gap, consisting of an intensity drop preceded and followed by formant transitions for /b/, into the first part, the middle part, and the final part of the vowel respectively. Figure 1 illustrates the Fo contours of the stimuli with the gap in the first part of the vowel.

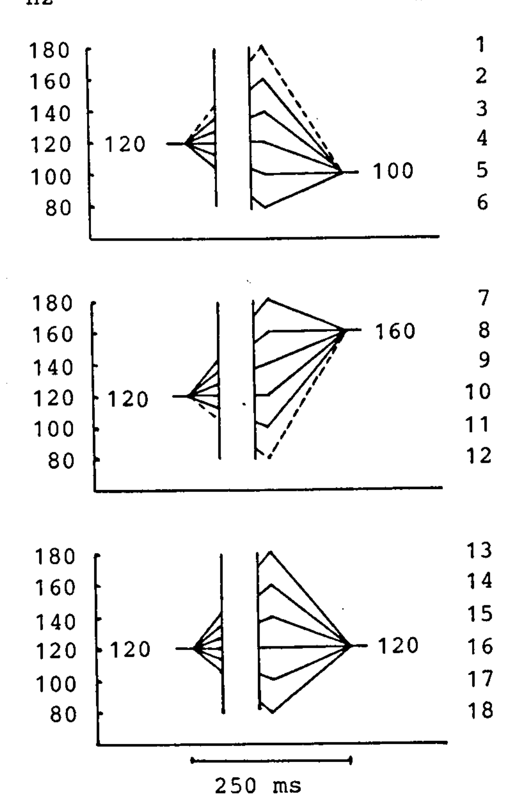


Figure 1.
Stylized tonal contours of one version of the ABX test. The dashed lines (stimuli 1 and 12) were also stimuli A and B.

Although a few listeners continued to categorize the new stimuli on the basis of tonal movement, most of the listeners' responses were altered by the intrusion of the spectral changes. When the intrusions were placed in the middle and in the last part of the vowel, categorization was more strongly based on end-point frequency. When the intrusions were placed in the beginning of the vowel, the categorizations were reversed vis-a-vis the end-point frequency but corresponded to the average frequency 40-80 ms after the intrusion.

These results seem to indicate that tonal movement is optimally perceived during portions of high spectral stability. If the perceptual load is increased by rapid spectral changes, and the duration of spectral stability is decreased, tonal movement will then be perceived and stored as tone levels. This interpretation also complies with the results obtained by Gårding, et al. (7) where perception of tone 4 (falling) in Standard Chinese was altered to tone 3 (dipping) by moving the fall backwards in time toward the CV boundary and also by increasing the steepness of the fall. These manipulations were done by means of LPC synthesis.

Languages, then, which need to manifest rising and falling Fo at the syllable level should optimally place these movements in places of spectral stability. This corresponds to Bruce's (8) production and perception data for Swedish concerning the timing of the word accent fall in non-focal position, where accent II is marked by a strong falling Fo well within the stressed vowel. This interpretation also has explanatory power concerning production data reported by Lindau (9) for Hausa (a two-tone language) where tonal turning points occur at the end of the vowel, a high being manifested as a rise and a low being manifested as a fall.

## PERCEPTION OF TONAL MOVEMENT AT THE PHRASE LEVEL

The second experiment concerns perception of phrase boundary markers and connective patterns (10,11,12). Listeners were presented with sequences of five fives (55555) and asked to judge whether the sequence was grouped 55-555 or 555-55. The fundamental frequency of a natural Scanian fem (five) was manipulated in various ways using LPC synthesis. Variations comprised fall-rise and rise-fall patterns at different frequency levels as well as rising and falling patterns having different ranges. These variations were then joined together to create the sequences. Duration was not variable as each syllable was equal in length as were the intervals between them. 36 different sequences were used as stimuli.

The results clearly showed that listeners can use a rising or a falling Fo movement having a greater range than in the surrounding syllables as a demarcative cue signalling the end of a group. The results also indicated that listeners can rely on connective Fo movement patterns encompassing the entire group. Examples of such patterns are the "hat-like" and "trough-like" intonation patterns (13). The perception of such patterns implies the use of some type of short-term memory where Fo movement is stored (either as movement patterns or as frequency levels) to be retrieved when the entire group has been heard.

Another example from the material where the use of memory seems to be important is found where listeners interpret precisely the same falling syllable in the same position (the second "five") in two different ways depending on the surrounding Fo movement. In one case the falling Fo movement of the syllable is interpreted as the end of a "hat-like" pattern signalling the end of a two-syllable group. In the other instance, the same falling Fo movement is followed by a greater fall to a lower frequency. This causes the second syllable to be interpreted as the middle "five" of a three syllable group (Figure 2).

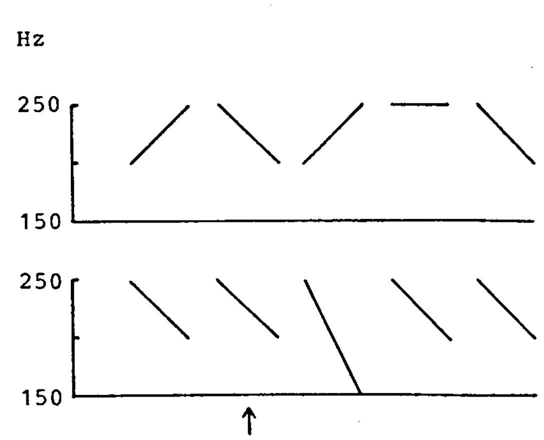


Figure 2.
Stylized tonal contours of two 55555 stimuli showing how the same falling syllable was interpreted in two ways. The top stimulus was interpreted as 55-555 and the bottom one as 555-55.

IMPLICATIONS FOR SPEECH PERCEPTION MODELS

When constructing a model of speech perception which takes into consideration fundamental frequency movement, pitch analysis is generally viewed as presupposing a first-order frequency analysis of the speech wave based on the the mechanical properties of the basalar membrane and characteristic frequencies and temporal responses of auditory-nerve fibers. This analysis provides the raw materials for a second-order analysis of pitch and timbre (14). On the basis of the data reported here, I would like to tentatively propose two different mechanisms of second-order pitch perception. The first is a direct conversion of Fo movement into linguistic categories. The second is a reconstuction of tonal movements or levels from short-term memory.

The categories of stress, word accents and tones, and in certain cases focus are likely candidates for the direct conversion of Fo movement. This movement, optimally located in the vocalic segments, is not then stored as movement, but rather as the corresponding linguistic category. This type of direct perception can be seen as corresponding to an event approach to segmental perception as proposed by Fowler (15). The rapidly perceived stressed syllables, for example, marked by tonal movement, can serve to guide perception to important areas of meaning (16).

Candidates for short-term memory based pitch analysis are juncture cues for boundaries, connective patterns for grouping and in certain cases focus. In this type of analysis, pitch could be stored first as tonal levels and then transformed into linguistic categories. Figure 3 presents a schematic diagram of the two different perceptual mechanisms.

Where the perception of intonation is seen as an important part of speech perception, the proposed division of movement perception into two mechanisms could have implications for more general models of speech perception. Although this division is tentative and speculative, it is an attempt to understand pitch perception in a linguistic frame of reference.



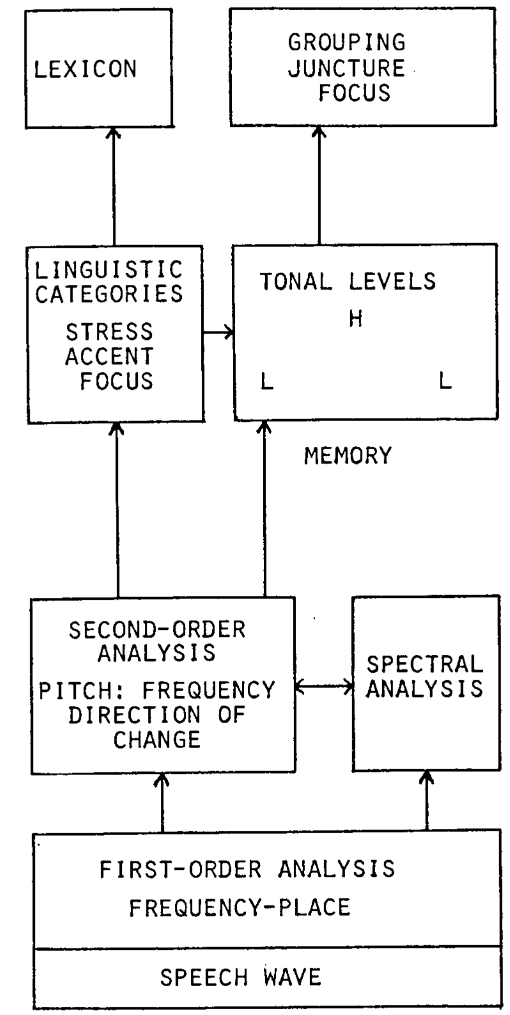


Figure 3.
Diagram illustrating two different perceptual mechanisms for pitch movement perception.

REFERENCES


(1) Lehiste, I. 1970. Suprasegmentals. MIT Press, Cambridge, MA.

(2) Plomp, R. 1976. Aspects of Tone Sensation, A Psychophysical Study. Academic Press, London.

(3) Itoh, K. 1986. A neuro-synaptic model of auditory memory and pitch perception. Annual Bulletin 20, Research Institute of Logopedics and Phoniatrics, University of Tokyo.

(4) House, D. 1985. Implications of rapid spectral changes on the categorization of tonal patterns in speech perception. Working Papers 28, Department of Linguistics and Phonetics, Lund University.

(5) Klatt, D.H. 1980. Software for a formant synthesizer. J. Acoust. Soc. Am. 67, 971-995.

(6) Fant, G. 1973. Speech sounds and features. The MIT Press. Cambridge, Mass.

(7) Gårding, E., Kratochvil, P., Svantesson, J.O., & Zhang, 1985. Tone 4 and Tone 3. Discrimination in Modern Standard Chinese. Working Papers 28, Department of Linguistics and Phonetics, Lund University

(8) Bruce, G. 1977. Swedish word accents in sentence perspective. Travaux de l'Institut de Linguistique de Lund XII. Gleerups, Lund.

(9) Lindau, M. 1986. Testing a model of intonation in a tone language. J. Acoust. Soc. Am. 80, 757-764.

(10) Gårding, E. & House, D. 1985. Frasintonation, särskilt i svenska In Svenskans beskrivning 15, eds. S. Allén et al. Göteborgs universitet, Göteborg: 205-221.

(11) Gårding, E., & House, D. 1986. Production and perception of phrases in some Nordic dialects. Working Papers 29, Department of Linguistics and Phonetics, Lund University.

(12) House, D. & Gårding, E. 1986. Phrasing in some Nordic Dialects. Paper presented at the fourth Nordic Prosody Conference in Middlefart, Denmark. In preparation.

(13) Collier, R. & t'Hart, J. 1975. The role of intonation in speech perception. In Structure and Process in Speech Perception. (Eds) Cohen, A. & Nooteboom, S.G. Berlin.

(14) Gelfand, S.A. 1981. Hearing, an introduction to psychological and physiological acoustics. Marcel Dekker, Inc. New York and Basel, Butterworths, London.

(15) Fowler, C.A. 1986. An event approach to the study of speech perception from a direct-realist perspective, Journal of Phonetics, 14, 3-28.

(16) Bannert, R. 1986. From prominent syllables to a skeleton of meaning: a model of prosodically guided speech recognition. Working Papers 29, Department of Linguistics and Phonetics, Lund University.