INTRODUCTION

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 perception of spoken language. This interrelationship is the perception of spoken language (1).

In this paper the word intonation will be important information which facilitates 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.

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 CDC digital computer were used to synthesize a Swedish /v/ vowel with formant frequencies of 600, 925, 2560 and 3320 Hz. The vowel duration was 300 ms including 30 ms intensity onset and offset. Fundamental frequency was systematically varied to create 18 different stimuli. The 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 100 Hz then rose to 180 Hz. The difference in end-point frequency was designed to test the effects of end-point variation on the rise-fall, fall-rise categories, i.e. movement patterns 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 an intensity drop preceded and followed by formant transitions for /v/, 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.

Types of curves (A) and (B) were used in the test design (3). 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 Fo. In addition, rapid spectral changes were introduced in the middle of the vowel, and the categorization was more strongly based on end-point frequency. When the intrusions were placed in the beginning of the vowel, the categorizations were reversed via-via the end-point frequency, and the listeners responded to the average frequency 40-80 ms after the intrusion.

The difference in categorization between stimuli A and B was thus strongly influenced by the fo movement pattern, and the categorization was based on the perception of phrase boundary markers.

Languages, then, which need to manifest rising and falling Fo at the syllable level should optimally place these movements in places of spectral stability. The categorization of high Fo levels as each syllable was classified as a rise.

The second experiment concerns perception of phrase boundary markers and connectsive patterns (10,11). Listeners were presented with sequences of five fives (55555) and asked to judge whether the sequence was legal or not. The fundamental frequency of a natural Standard Chinese was altered to tone 3 (dipping) by moving the F0 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.

The first experiment was repeated with the same stimuli as above, and the sequence was grouped 55-555 or 555-55. Listeners were presented with sequences of five fives (55555) and asked to judge whether the sequence was legal or not. The fundamental frequency of a natural Standard Chinese was altered to tone 3 (dipping) by moving the F0 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.

Variations in communication skills and non-verbal cues 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 a variable as each syllable was equal in length.

REFERENCES


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 mechanical properties of the basilar 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 reconstruction 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 boundaries, connective patterns for groupings and in certain cases focus. In this division is tentative and speculative, it is an attempt to understand pitch perception in a linguistic frame of reference.

**IMPLICATIONS FOR SPEECH PERCEPTION MODELS**

- **LEXICON**
- **GROUPING JUNCTURE**
- **FOCUS**

**LINGUISTIC CATEGORIES**

- **STRESS ACCENT**
- **FOCUS**

**TONAL LEVELS**

- **L**
- **L**

**MEmory**

**SECOND-ORDER ANALYSIS**

- **PITCH: FREQUENCY**
- **DIRECTION OF CHANGE**

**FIRST-ORDER ANALYSIS**

- **FREQUENCY-PLACE**

**SPEECH WAVE**

**REFERENCES**