Working Papers 25 1983
Linguistics-Phonetics
Lund University, 67-74

PERCEPTUAL INTERACTION BETWEEN FO EXCURSIONS AND SPECTRAL CUES

David House

ABSTRACT

Questions and problems are formulated concerning perception of Fo movement in the frequency domain and possible perceptual relationships between such movement and spectral cues, particularly vowel formant transition place-of-articulation cues for stops. These issues are to provide the basis for experimental thesis work in speech perception involving both hearing impaired listeners and listeners with normal hearing. A preliminary experiment using band-pass filtered speech and normal listeners where the task was to identify VCV words with varying Fo movement indicated the possibility that a certain improvement in stop identification might be related to increased movement of Fo.

INTRODUCTION

Fundamental frequency excursions in running speech are generally considered to play an important role in signaling sentence focus, i.e. key words or phrases which contain new information important for sentence comprehension. Listeners can be forced to rely heavily upon "bottom-up" speech processing when perceiving segmental characteristics of words in focus where little or no contextual information is present to assist in presupposive "top-down" speech processing (Marslen-Wilson & Tyler, 1980). This type of new information processing therefore places greater demands on the segmental resolution of words in focus than does the processing of words outside focus. How then does the perception mechanism process the Fo excursions related to focus? Is it simply a matter of contrasting the focus word to the other words and, in effect, causing a sharpening of attention, or can we find relationships between

Fo movement and spectral cues which could provide direct assistance to the perceptual mechanism in performing the task of segmental resolution?

Fo excursions can be thought of as perceptually significant from the point of view of dynamic perception in that they comprise a movement or change in frequency over time. This movement could be registered as an event which would sharpen attention and aid in short-term memory retrieval of spectral cues. Hypothetically, Fo movement could also directly interact with spectral cues such as a timing relationship between V.O.T. and an Fo rise or an interaction between Fo movement and vowel formant transitions. Also, can Fo movements alter the spectrum, e.g. causing F_1 to rise, and if so, what perceptual significance is found in such alteration? Finally, is there an optimum Fo excursion in scope and range for facilitating segment perception?

The purpose of this report is not to provide answers to these questions but rather to formulate some problems which are to at least serve as the initial basis for my thesis work in speech perception and to report the results of a preliminary listening test.

FO AND FORMANT TRANSITIONS

As both Fo movement and vowel formant transitions involve a change in frequency over time, these two aspects of the speech wave can serve as a point of departure in the investigation of perceptual interaction. Resonance induced vowel formant transitions are generally considered to be important perceptual cues for stop consonant place-of-articulation identification. Formant frequency movement can be seen as rising or falling shifts in intensity through successive layers of harmonics of the fundamental. As long as a steady fundamental frequency is maintained, the formant transition patterns comprise a single movement. During an Fo excursion, however, the formant transitions will be altered by rising and falling harmonics. We are presented then, during focus, with a more complex frequency movement, the question being whether this added complexity is perceptually significant and if so, can this movement facilitate stop consonant identification?

FREQUENCY MOVEMENT AND HEARING LOSS

If we can establish some type of perceptual interaction between Fo and formant transitions, what could the possible implications be for listeners with hearing disabilities? It is well documented that place-of-articulation identification causes difficulties for individuals with moderate sloping sensorineural hearing losses. These difficulties, however, do not necessarily relate directly to specific frequency intensity attenuation. Van de Grift Turek, et al. (1980) demonstrated that subjects having similar audiometric configurations differed radically in their performance in synthetic stop identification tests.

Similar performance differences have also been reported by Risberg and Agelfors (1978) concerning the identification of intonation contours. This brings us back to the question of frequency movement processing. Can individual differences be accounted for by variations in capacity to perceive and process various types of movements in frequency and their possible interactions? If so, can this information be of any help to individuals with hearing disabilities?

PRELIMINARY TEST METHOD

To begin testing some of these questions, a preliminary perception experiment was devised using band-pass filtered speech in an attempt to roughly simulate a hearing loss and to essentially confine formant transition information to F_1 while varying Fo excursions by using different forms of presentation. The following test material was randomized and recorded by a native speaker of American English. Three different presentation types were used to vary Fo: 1) neutral, 2) question, and 3) emotional-emphatic. The capitalized words indicate sentence focus. Each sentence was read twice, once with "fifteen" and once with "fifty".

Test material:

She said to KEEP fifteen. (fifty)
She said to TUCK fifteen. (fifty)
She said to PIT fifteen. (fifty)
She said to put FIFTEEN. (FIFTY)
She said to tip FIFTEEN. (FIFTY)
She said to cut FIFTEEN. (FIFTY)

He said to DUB fifteen. (fifty)
He said to BAG fifteen. (fifty)
He said to GOAD fifteen. (fifty)
He said to dab FIFTEEN. (FIFTY)
He said to bid FIFTEEN. (FIFTY)
He said to dig FIFTEEN. (FIFTY)

The test was presented through a Fonema filter (390-820Hz, -36dB/octave) to two listeners with normal hearing and no prior knowledge of the test material. They were instructed to write the word and number which followed "He said to". They were also instructed to indicate if they heard "She" instead of "He" at the beginning. This was to provide a rough test of the filter function.

Selected spectrograms and mingograms were then made from the material (Fig. 1, 2).

RESULTS

Sentence focus had no significant effect upon word identification in neutral and question presentation. Correct identification remained about 1 in 3 (Fig. 3). With emotional-emphatic presentation, however, word identification was greater in focus (1 in 2) than outside focus (1 in 3) (Fig. 3). The numbers, "fifteen, fifty" were correctly identified in all cases, and the sibilant in "she" was effectively filtered.

DISCUSSION

Although this preliminary experiment contains a number of uncontrolled variables such as intensity and task word frequency, a relationship can be seen between increased word identification and Fo movement. Fo movement is greatest in the task word

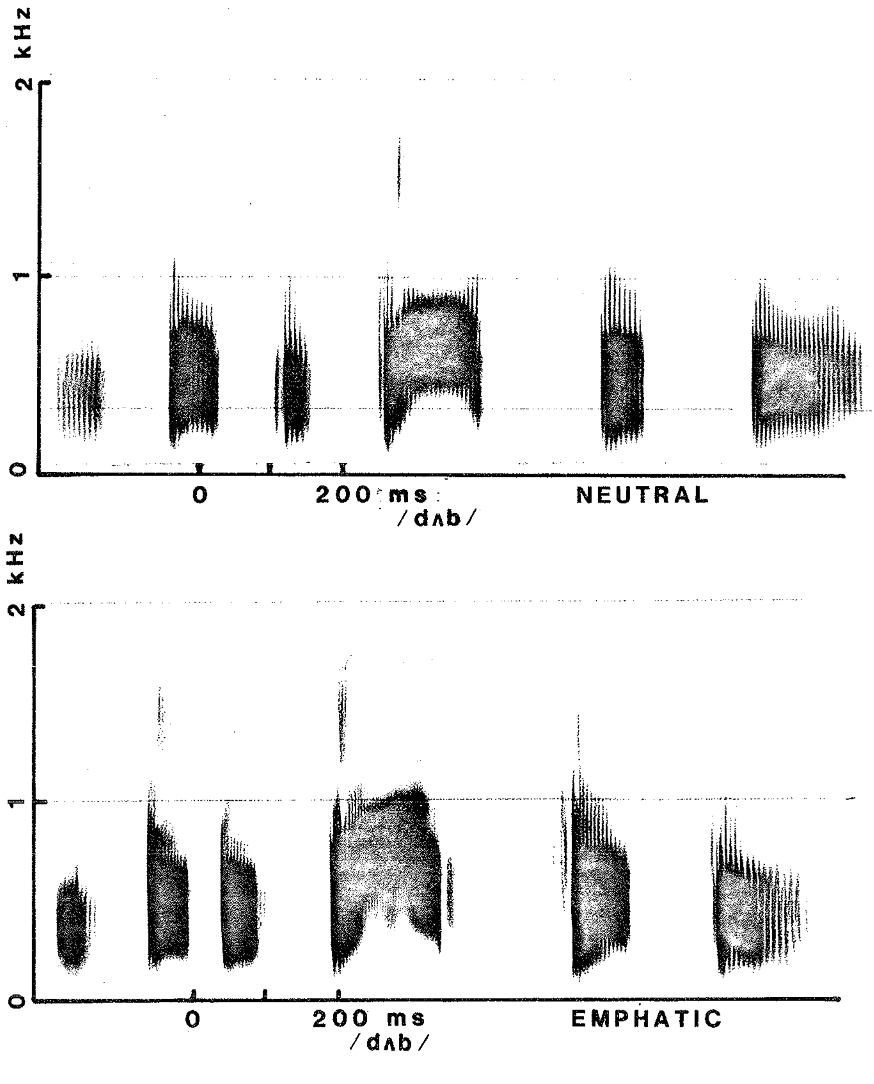

Figure 1. Spectrograms of the filtered test sentences, "He said to DUB fifteen," in neutral presentation (top) and "He said to DUB fifty," in emphatic presentation (bottom).

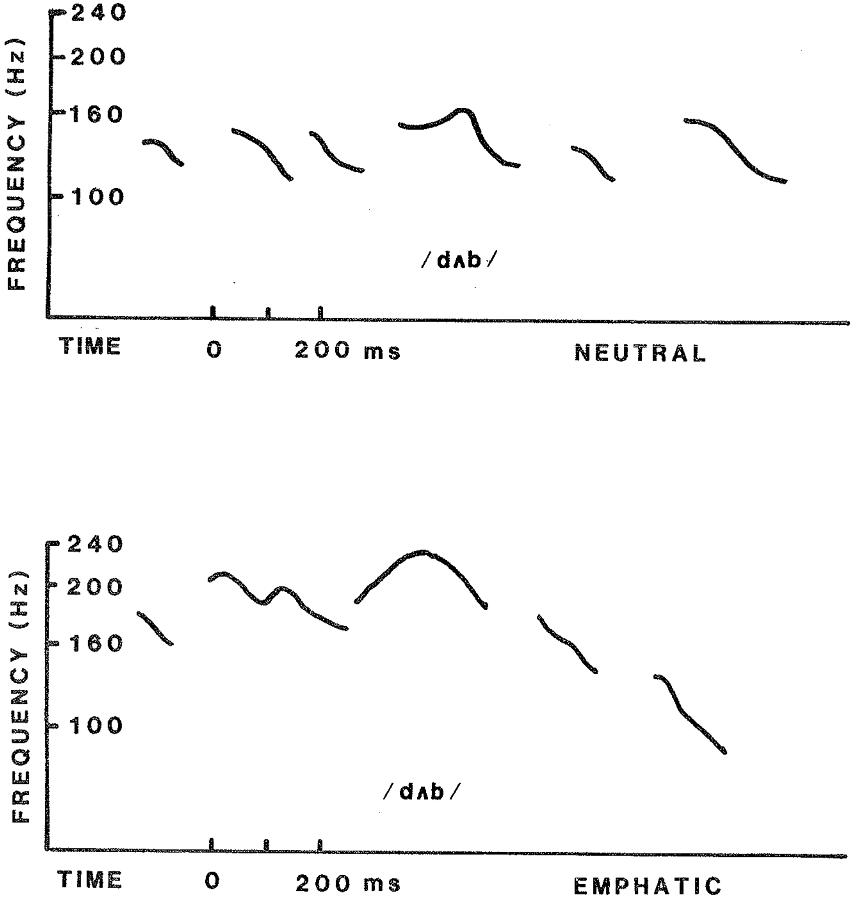

Figure 2. Fo tracings of the test sentences, "He said to DUB fifteen," in neutral presentation (top) and "He said to DUB fifty," in emphatic presentation (bottom).

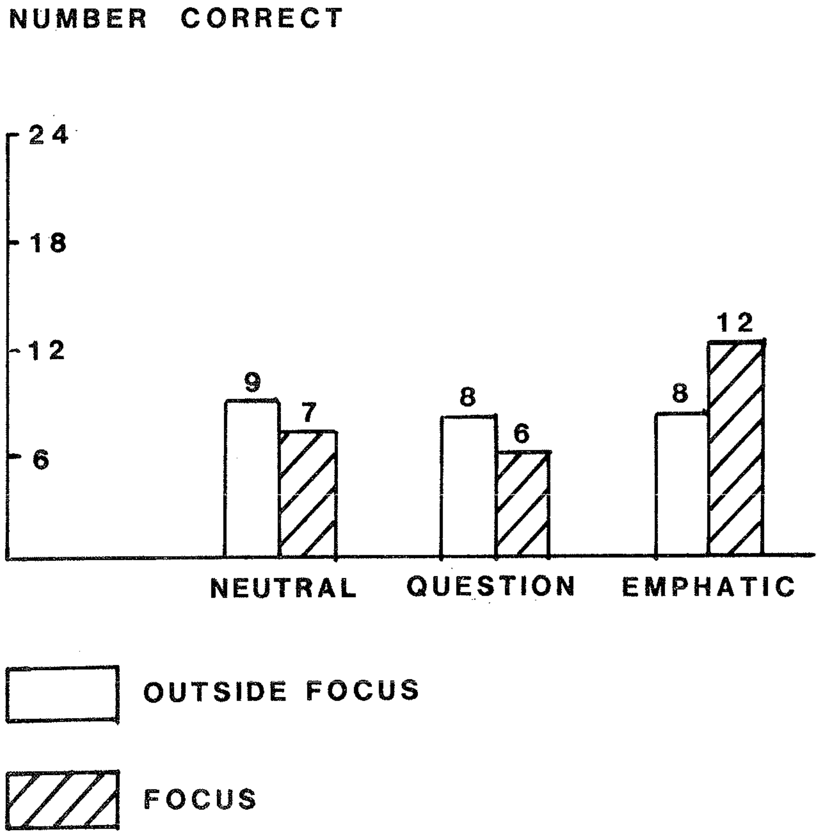

Figure 3. Results of the listening test showing number of task words correctly identified in the three presentation types divided into focus and outside focus.

in focus in the emotional-emphatic presentation (Fig. 2) where identification was also greatest (12 of 24). Intensity, of course, may have been a major factor although length was not appreciably different between words in and outside of focus or between words in different presentations. There was also a spectral raising in the emphatic presentation (Fig. 1) which could have contributed to identification.

The correct identification of "fifty-fifteen" can be attributed to length and intonation differences which are readily perceived since the same numbers are used throughout the test. This demonstrates the power of these cues in primarily "top-down" processing.

Additional experiments are planned enlisting the help of speech synthesis and listeners with hearing disabilities in an attempt to improve our understanding of cue interaction in speech perception.

REFERENCES

Marslen-Wilson, W. & Tyler, L. K. 1980. The temporal structure of spoken language understanding. Cognition 8: 1-71.

Risberg, A. & Agelfors, E. 1978. On the identification of intonation contours by hearing impaired listeners. Quarterly Progress and Status Report of the Speech Transmission Laboratory, Royal Institute of Technology, Stockholm 2/3: 51-61.

Van de Grift Turek, S., Dorman, M. F., Franks, J. R., & Summerfield, Q. 1980. Identification of synthetic /bdg/ by hearing-impaired listeners under monotic and dichotic formant presentation. Journal of the Acoustical Society of America 67: 1031-1039.