Proceedings of the Annual Meeting of the Berkeley Linguistics Society

Title

The Perception of Contour Tones

Journal

Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 1(1)

ISSN

2377-1666

Author

Hombert, Jean-Marie

Publication Date

1975

The Perception of Contour Tones
Author(s): Jean-Marie Hombert
Proceedings of the First Annual Meeting of the Berkeley Linguistics Society (1975), pp. 221-232

THE PERCEPTION OF CONTOUR TONES

Jean-Marie Hombert
University of California, Berkeley

0. Introduction

In this paper, I shall address myself to the perception of speech-like acoustic signals with changing frequency contours. The paper is divided into three parts: 1) a brief discussion of the relationship between pitch and fundamental frequency; 2) the presentation of my data from a perception experiment using synthesized speech; and 3) the linguistic relevance of these data for a proper understanding of tone languages.

1. Pitch and Frequency

Pitch and frequency are often used interchangeably in the linguistic literature. In fact, however, they refer to two different entities and there is not a one to one relationship between them. Pitch is a psychological attribute of tone, referring to the perception of a stimulus by a subject. Fundamental frequency, on the other hand, can be measured by instruments without any reference to our auditory mechanism. In the case of pure tones, it has been shown by Stevens[1] (1935) that at high intensity levels low tones (frequencies up to 1000 Hz) are perceived lower and high tones (above 3000 Hz) are perceived higher. More recent studies (Cohen 1961) show that pitch is affected by loudness in the direction indicated by Stevens, but to a much smaller extent (less than 2%) than originally claimed. Because of the spectral characteristics of speech sounds, the effect of intensity is even lower than for pure tones and consequently can be judged as irrelevant for the determination of pitch. For speech, the main correlate of pitch for voiced sounds is fundamental frequency (which corresponds to the rate of vibration of the vocal cords), although we have evidence (Schouten 1940, Licklider 1956) that higher frequency regions (the fourth and higher harmonics; Plomp 1967) are used for this pitch perception.[2] Klatt (1973) indicates in fact that pitch perception is improved when the fundamental frequency region is filtered out of the acoustic signal. Unfortunately, most of the experiments dealing with pitch perception have been done with pure tones. Because of the acoustic characteristics of the stimuli generally used in these experiments (wide frequency range, steady state frequencies) the results are difficult to extrapolate to speech.

2. Perception Experiment

In order to obtain data more closely related to speech in general and to tone languages in particular, the following experiment was conducted. Ten American subjects participated in this experiment which was divided into two parts.

2.1. Part I

2.1.1. Experimental Paradigm. The stimuli consisted of 30 instances of the vowel [i] synthesized with different fundamental frequency patterns. As seen in Fig. 1 the stimulus was composed by a slope followed by a level tone maintained constant at 120 Hz. The onset frequency was one of the following values: 70, 100, 110, 130, 140 and 190 Hz; that is, 3 values below and 3 values above 120 Hz. The duration of the slope (Δt) was varied at 40, 60, 100, 150 and 250 msec. In other words, 15 stimuli (3 F_0 onset X 5 slope duration) had a falling fundamental frequency and 15 stimuli had a rising fundamental frequency. The overall duration of the stimulus was fixed at 250 msec. Each time the stimulus was presented, it was followed by a 500 msec pause and a second vowel [i] with a steady state fundamental frequency. The duration of this vowel was also 250 msec. The level of its fundamental frequency was adjustable by a knob controlled by the subject. The task of the subjects was to match the pitch of the second vowel to the beginning of the first vowel. The rate of stimulus presentation as well as the number of trials for a given presentation were controlled by the subject. Each one of the 30 stimuli was presented 3 times in a randomized order. The subjects heard the stimuli through earphones at a comfortable level (about 70 dB).

2.1.2. Results. The results of this part are presented in Fig. 3. On this graph, the subjects' responses (i.e. the steady state fundamental frequencies of the second vowel adjusted by the subjects) are plotted as a function of the duration of the slope. The thin straight lines represent the ideal responses; the wider lines with corresponding symbols represent the subjects' actual responses for this fundamental frequency onset.

This graph suggests that 1) falling patterns (i.e. vowels with fundamental frequency onset above 120 Hz) are perceived more accurately than rising patterns (i.e. vowels with fundamental frequency onset below 120 Hz); 2) the longer the slope, the more accurate the matching, but correlation between slope duration and accuracy of matching is not linear.

FIGURE 1. Stimulus Format (Part I)

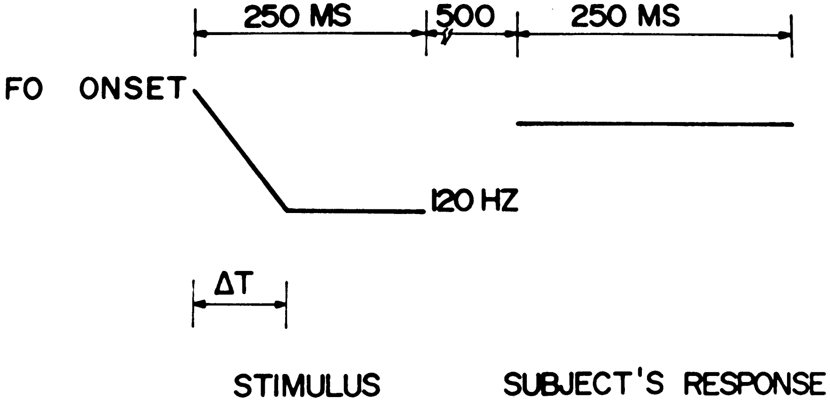


FIGURE 2. Stimulus Format (Part II)

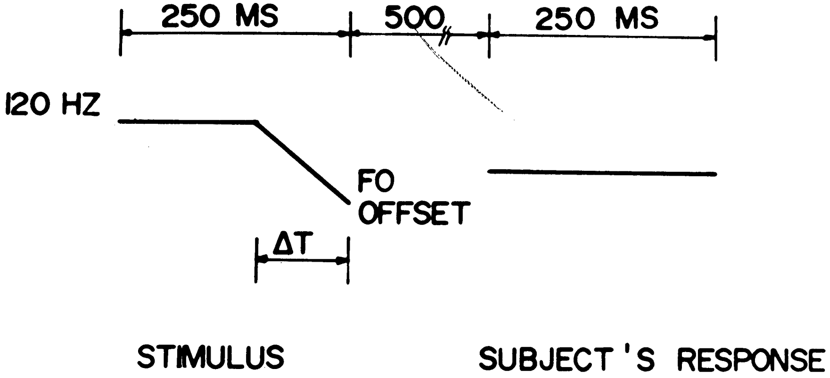


FIGURE 3. Subjects' Responses (Part I)

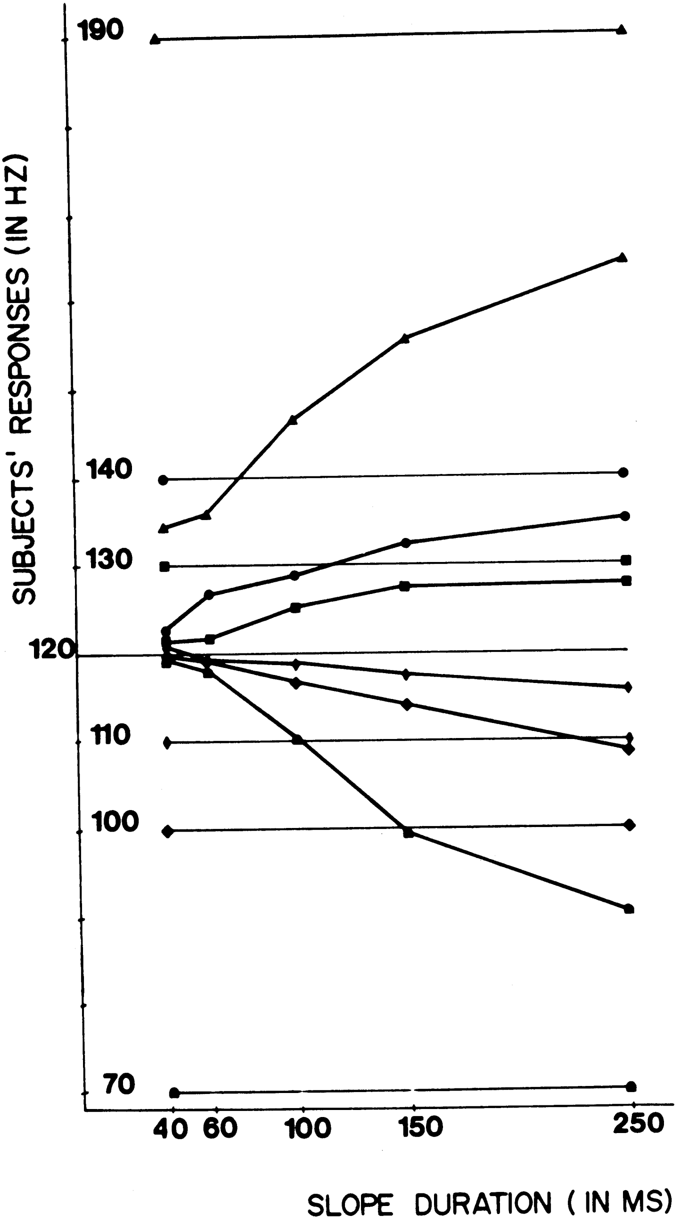
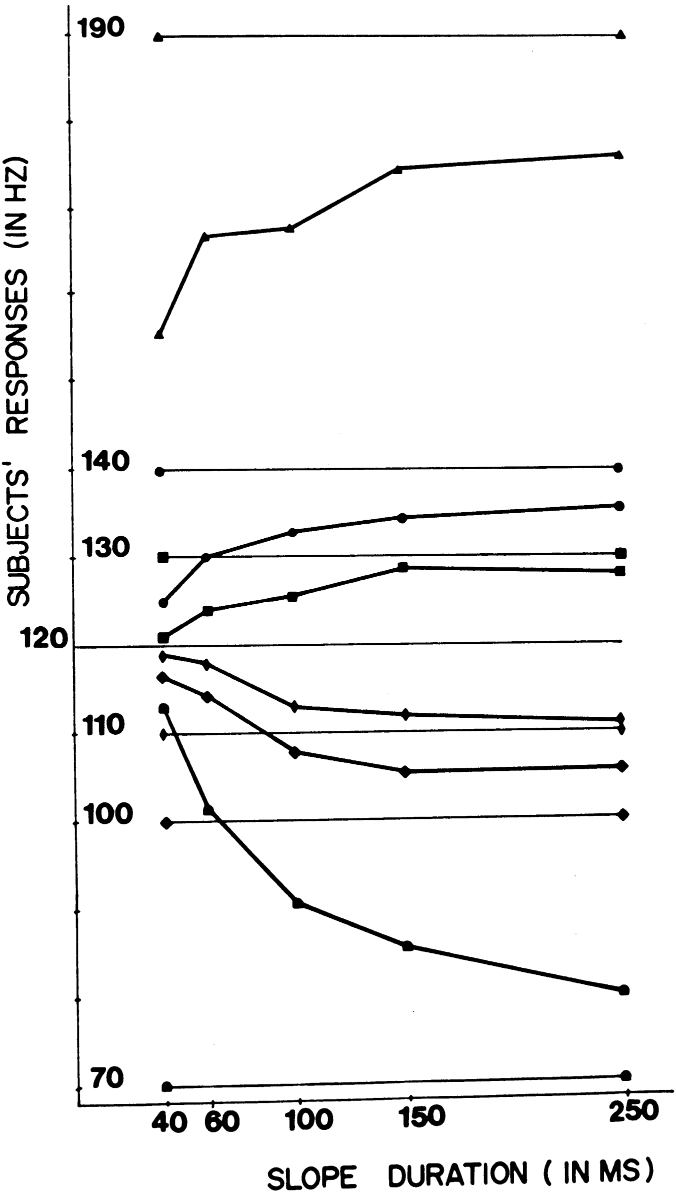


FIGURE 4. Subjects' Responses (Part II)



2.1.3. *Discussion*. These results can be explained by forward masking. Masking is the obscuring of a sound by another. Forward masking is the masking that occurs when the masker tone is terminated before the test tone (maskee) is begun. Forward masking is the most effective when the frequency of the earlier tone is lower than but close to the frequency of the second tone (Lüsher and Zwislocki 1949, Licklider 1951, Jeffress 1970). If we extrapolate the results obtained with steady state tones to contours, we can understand why the onset region of the rising ramp was not accurately perceived, since each frequency was masked by the previous lower frequency (since the frequency is going up). This is not the case for the falling tone in which each frequency is followed by a lower frequency.[3] Data from the psychoacoustic literature (Brady et al 1961, Heinz et al 1968, Pollack 1968, Nabelek and Hirsh 1969, Nabelek et al 1970, Tsumura 1973) can be interpreted as supporting my claim concerning the role of masking in the perception of changing frequency contours.

In these experiments subjects were asked to match the pitch of a steady state signal with a changing frequency signal. They consistently adjusted their steady state tone closer to the final point of the contour. This fact already shows the role of masking which attenuates the effect of the onset region in favor of the offset region, but furthermore there is a tendency to match closer to the final point when the stimulus is a rising contour (as opposed to a falling contour). This indicates, as I have suggested, that the masking of the onset region is more effective in the case of rising contours (as opposed to falling contours), and consequently, this leads to the perception of an averaged pitch closer to the offset frequency.

These data are also in agreement with the study of Brady et al (1961) with respect to the role of the rate of frequency change. They found that the matching of a steady state frequency with a contour frequency is closer to the end point of the contour when the rate of change is high; in other words, the onset region is less salient at high glide rate. This is shown on Fig. 3 by responses close to 120 Hz when the slope duration of the stimulus is short (i.e. the rate of frequency change is high). This is also in agreement with Pollack (1968) and Nabelek and Hirsh (1969), whose results indicate that optimum discriminability of relatively small frequency changes is obtained at relatively slow glide rates.

2.2. *Part II*

2.2.1. *Experimental Paradigm*. The experimental paradigm used here is similar to the one used in Part I, but instead of having a slope followed by a level tone, now the level tone precedes the slope (see Fig. 2). As in Part I, the level tone was maintained constant at 120 Hz; the slope duration had one of the following values: 40, 60, 100, 150 or 250 msec, and the values for fundamental frequency were identical to the values used for fundamental frequency onsets in the previous part of the experiment (i.e. 70, 100, 110, 130, 140 and 190 Hz). The subjects were asked to match the adjustable steady state fundamental frequency of the second vowel to the end point of the first vowel.

2.2.2. Results. The results are presented in Fig. 4. As on the previous graph, the subjects' responses are plotted as a function of the length of the slope. The thin straight lines represent the ideal responses, and the wider lines with similar symbols represent the subjects' actual responses.

Three points should be emphasized: 1) The matching is more accurate with the end point of the stimulus vowel than with the onset of the stimulus vowel (i.e. Part I of the experiment). This can be easily visualized by the fact that the subjects' actual responses are closer to the ideal response line (offset frequency) in Part II than to the onset frequency in Part I. 2) It was mentioned in Part I that the onset of a falling pattern was more accurately perceived than the onset of a rising pattern. This is not the case for the offset point. 3) As before, the longer the slope, the more accurate the matching, but it is obvious on this graph that there is an upper limit in duration above which the accuracy of matching does not improve; in particular, the change of slope duration from 150 to 250 msec has very little effect on the subjects' responses.

2.2.3. Discussion. The explanation based on forward masking proposed in the first part of the experiment to account for a different degree of accuracy in the perception of onset frequencies of falling vs. rising contours cannot be applied here (otherwise one would predict a higher degree of accuracy for the matching of the offset falling ramp). Rather, it seems that the final auditory image (the offset frequency) can be processed equally well no matter what the preceding frequencies are. The fact that the matching is more accurate than in Part I can be accounted for by considerations based on memory. The effect of glide rate on pitch perception is consistent with the first part of the experiment.

3. Linguistic Implications and Directions of Further Research

These data are of potential interest in the understanding of tone languages in two respects: 1) the development of tones; and 2) the distribution of tones.

3.1. The Development of Tones

The historical development of tone systems from the reinterpretation of intrinsic perturbations caused by neighboring consonants on the pitch of the preceding or following vowel has been suggested for a number of genetically unrelated languages (Beach 1938, Haudricourt 1954, 1961, 1972; Matisoff 1973, Ohala 1973, Purcell 1972, 1974; Purcell et al 1975). In order to validate such historical developments, two sets of data are necessary: first, data from production tests showing that these intrinsic perturbations caused by pre- and postvocalic consonants do exist (House and Fairbanks 1954, Lehiste and Peterson 1961, Mohr 1969, Löfqvist 1973, Hombert 1974); and second, perceptual data showing that these intrinsic perturbations can be perceived (Haggard et al 1970, Fujimura 1971, Abramson 1974, Hombert 1974, 1975a).

3.2. The Distribution of Tones

Data such as those presented in this paper represent a preliminary step in our understanding of the distribution of tone shapes in tone languages. For instance, the fact that languages have a tendency to favor falling tones rather than rising tones (Cheng 1973, Hyman 1973b) has been explained in articulatory terms (Ohala and Ewan 1973, Sundberg 1973). It was suggested earlier in this paper that falling tones are probably more salient perceptually than rising tones because they cover a wider frequency range (since the onset is more accurately perceived). This indicates a perceptual correlate to the articulatory constraint generally proposed to account for the assymmetry between the number of rising tones vs. the number of falling tones in languages.

Our understanding of the distribution of tones in tone languages would be significantly improved if we would be able to develop a theory of tone space based on perceptual distance analogous to the model proposed for vowel space by Liljencrants and Lindblom (1972). For the construction of such a model data from three sources seem to be relevant: 1) acoustic descriptions of tones from tone languages; unfortunately, only a limited number of such studies have been done (Howie 1972, Abramson 1962, Chuang 1972, Han 1969, Gandour 1974, LaVelle 1974); 2) confusion matrices from tone languages showing how similar or dissimilar different frequency contours are judged; 3) controlled investigation of the different parameters involved in pitch perception: fundamental frequency onset, speed of pitch change, fundamental frequency offset.... It is in this area that the perception data presented in this paper will be useful.

4. Conclusion

The data presented in this paper indicate the following: 1) The onset of falling tones is more accurately perceived than the onset of rising tones. This asymmetry can be explained by forward masking. This explanation accounts for most of the data reported in the psychoacoustic literature. 2) There is no difference in the perception of the offset of a falling vs. a rising tone. The explanation based on masking does not hold for offset frequencies. It seems that the final auditory image is processed equally accurately independantly of the preceding frequencies. 3) The perception of onset and offset frequencies is more accurate for slow rate of frequency change. Finally, the relevance of these data for the development and distribution of tones in tone languages is mentioned.

5. Acknowledgements

I would like to thank John Ohala for his constant, helpful guidance throughout my experimental investigations of the perception of tone. In addition, I am grateful to Michelle Mieusset for her aid in preparing the photographs and to Harold Clumeck, Hector Javkin, Larry Hyman, C.K. Chuang, Bob Krones and Steve Greenberg for their comments on an earlier draft of this paper.

6. Footnotes

[1] For low frequencies, Stevens' results are in agreement with Zurmühl (1930) and Snow (1936).

[2] For a review of theories of pitch perception and their relevance for speech, see Hombert (1975b).

[3] Obviously this explanation holds only if we consider forward masking. Backward masking would make the opposite prediction (Massaro 1968).

7. References

Abramson, A.S. 1962. "The vowels and tones of Standard Thai: acoustical measurements and experiments". IJAL, 28.2, Part II.

Beach, D.M. 1938. The Phonetics of the Hottentot Language. Cambridge.

Brady, P.T., A.S. House and K.N. Stevens. 1961. "Perception of sounds characterized by a rapidly changing resonant frequency." JASA 33.1357-1362.

Cheng, C.C. 1973. "A quantitative study of Chinese tones". Journal of Chinese Linguistics 1.

Chuang, C.K. 1972. Acoustic Study of Mandarin Tones. Unpublished Ph.D. Dissertation. Tohoku University, Japan.

Cohen, A. 1961. "Further investigation of the effects of intensity upon the pitch of pure tones". JASA 33.1363-1376.

Fujimura, O. 1971. "Remarks on stop consonants: Synthesis experiments and acoustic cues". In Form and Substance, 221-232. Akademisk Forlag.

Gandour, Jack. 1974. "Consonant types and tone in Siamese". Journal of Phonetics 2.

Haggard, M., S. Ambler and M. Callow. 1970. "Pitch as a voicing cue". JASA 47.613-617.

Han, M.S. 1969. Studies in the Phonology of Asian Languages VIII: Vietnamese Tones. Acoustic Phonetics Research Laboratory, University of Southern Calif.

Haudricourt, A.G. 1954. "De l'origine des tons en vietnamien". Journal Asiatique 242.69-82.

Haudricourt, A.G. 1961. "Bipartition et tripartition des systemes de tons dans quelques langues d'extreme orient". BSLP 56.163-180.

Haudricourt, A.G. 1972. "L'apparition des registres des langues a tons ponctuels". Proceedings of the Seventh International Congress of Phonetic Sciences, Montreal.

Heinz, J.M., B. Lindblom and J. Lindqvist. 1968. "Patterns of residual masking for sounds with speech like characteristics". IEEE Transactions on Audio and Electroacoustics, Vol. AU-16, 107-111.

Hombert, J.M. 1974. "Production and perception: their respective roles in one case of tonal development". Paper presented at Annual L.S.A. Meeting, New York.

Hombert, J.M. 1975a. Towards a Theory of Tonogenesis: An Empirical, Physiologically and Perceptually-Based Account of the Development of Tonal Contrasts in Languages. Unpublished Ph.D. Dissertation, University of California, Berkeley.

Hombert, J.M. 1975b. "Development of tones from final consonants: an experimental investigation". Paper to be presented at the Eighth International Congress of Phonetic Sciences, Leeds.

House, A.S. and G. Fairbanks. 1953. "The influence of consonant environment upon the secondary acoustical characteristics of vowels". JASA 25, No. 1.

Howie, J.M. 1972. "Some experiments on the perception of Mandarin tones". Proceedings of the Seventh International Congress of Phonetic Sciences, Montreal. The Hague: Mouton.

Hyman, Larry M. 1973a. Consonant Types and Tone (ed.). Southern California Occasional Papers in Linguistics No. 1. University of Southern California.

Hyman, Larry M. 1973b. "The role of consonant types in natural tonal assimilations". In Hyman (1973a), 151-179.

Jeffress, L.A. 1972. "Masking". In Tobias, Vol. 1. New York and London: Academic Press.

Klatt, D.H. 1973. "Discrimination of fundamental frequency contours in synthetic speech: implications for models of speech perception". JASA 53.8-16.

LaVelle, C.R. 1974. "An experimental study of Yoruba tone". Studies in African Linguistics, Supplement 5, 185-194.

Lea, W. 1973. "Segmental and suprasegmental influences on fundamental frequency contours". In Hyman (1973a), 15-70.

Lehiste, I. and G. Peterson. 1961. "Some basic considerations in the analysis of intonation". JASA 33.419-425.

Licklider, J.C.R. 1951. "Basic correlates of the auditory stimulus". In S.S. Stevens (ed.), Handbook of Experimental Psychology. New York: Wiley.

Licklider, J.C.R. 1956. "Auditory frequency analysis". Proceedings of the Third London Symposium on Information Theory. London.

Liljencrants, J. and B. Lindblom. 1972. "Numerical simulation of vowel quality systems: the role of perceptual contrast". Language 48.839-862.

Löfqvist, A. 1973. "On intrinsic and extrinsic F_o variations in Swedish tonal accents". Monthly Internal Memorandum, Phonology Laboratory, Berkeley. March, 19-32.

Lüsher, E. and Zwislocki, J. 1949. "Adaptation of the ear to sound stimuli". JASA 21.135-139.

Matisoff, J. 1973. "Tonogenesis in Southeast Asia". In Hyman (1973a), 71-95.

Mohr, B. 1971. "Intrinsic variations in the speech signal". Phonetica 23.69-93.

Nabelek, I. and I.J. Hirsh. 1969. "On the discrimination of frequency transitions". JASA 45.1510-1519.

Nabelek, I., A.K. Nabelek and I.J. Hirsh. 1970. "Pitch of tone bursts of changing frequency". JASA 48. 536-553.

Ohala, J.J. 1973. "The physiology of tone". In Hyman (1973a), 1-14.

Ohala, J.J. and W. Ewan. 1973. "Speed of pitch change" (abstract). JASA 53.345.

Plomp, R. 1967. "Pitch of complex tones". JASA 41. 1526-1533.

Pollack, I. 1968. "Detection of rate of change of auditory frequency". Journal of Experimental Psychology 77.535-541.

Purcell, E.T. 1972. "Acoustic differentiation of Serbo-Croatian Word-Tones in Statement Environments". Proceedings of the Seventh International Congress of Phonetic Sciences, Montreal. The Hague: Mouton.

Purcell, E.T. 1974. "From dynamic stress to word-tone: the consonantal origins of Balto-Slavic tone". In M. LaGaly et al (eds.), Papers from the Tenth Regional Meeting of the Chicago Linguistic Society, 528-538.

Purcell, E.T. et al. 1975. "Acoustics of Tonogenesis". To appear in the Proceedings of the First Annual Meeting of the Berkeley Linguistics Society.

Schouten, J.F. 1940. "The residue and the mechanism of hearing". Proc. K. Ned. Akad. Wet. 43.991-999.

Snow, W.B. 1936. "Change of pitch with loudness at low frequencies". JASA 8.14-19.

Stevens, S.S. 1935. "The relation of pitch to intensity". JASA 6.150-154.

Sundberg, J. 1973. "Data on maximum speed of pitch changes". STL-QPSR 4.39-47.

Tobias, J. 1972. Foundations of Modern Auditory Theory, Vol. I and II. New York and London: Academic Press.

Tsumura, T., T. Sone and T. Nimura. 1973. "Auditory detection of frequency transition". JASA 53.17-25.

Zurmühl, G. 1930. "Abhängigkeit der Tonhöhenempfindung von der Lautstärke und ihre Beziehung zur Helmholzschen Resonanztheorie des Hörens". Z. Sinnesphysiol. 61.40-86.