Difficulty of producing different Fo in speech

Jean-Marie Hombert

[Department of Linguistics
University of California, Los Angeles]

1. Introduction.

The notion of articulatory difficulty is often used but rarely quantified in the linguistic literature. Such data are extremely valuable when explaining the naturalness of phonological processes. This study investigates the difficulty associated with the production of different fundamental frequencies in speech.

Fundamental frequency (i.e., the rate of vibration of the vocal cords) is controlled by aerodynamic and muscular forces. Although there is still some disagreement about the relative weight of these two sets of forces (Liebermann 1967, 1970; Ohala 1970, 1973, in press) it seems that muscular tension is the primary factor controlling Fo changes. Electromyographic (EMG) recordings allow us to evaluate the timing and to some extent the relative level of activity of different muscles. But these data are extremely difficult to relate to the notion of articulatory difficulty. The main problem is that, assuming that all the relevant muscles have been sampled, it is not possible to associate an "overall muscular activity" to a given Fo change by adding together the EMG activity detected in each muscle because the amplitude of each recording is a function of electrode placement and also because we have no reason to believe that equal EMG activity in two different muscles correspond to equal physiological effort and consequently to directly summable articulatory difficulty. Furthermore, even if we were able to evaluate the muscular effort involved in a given tone, we would still have to attribute relative weights to the pulmonic effort (i.e., aerodynamic forces) vis a vis the muscular effort in order to get a global measurement of articulatory difficulty.

Assuming there is a direct correlation between the difficulty involved in doing a certain task and the level of accuracy reached when doing this task, it is possible to by-pass the problems mentioned earlier and get an evaluation of the aritculatory difficulty involved in producing certain tones. Let us consider an experiment in which subjects are asked to perform Fo changes in which the beginning and the end point of the Fo contour is given. Subjects' accuracy at imitating the target tone can be considered as being related to the articulatory difficulty involved in producing this target tone. Ohala and Ewan (1973) and Sundberg (1973) ran a somewhat similar experiment although they were mainly interested in maximum speed of Fo change and consequently ask their subjects to perform their Fo changes as fast as they could between the onset and the offset frequencies (two pure tones representing these frequencies were provided to the subject just before his own production). Their results showed that naive subjects (i.e., non-singers) perform

Fo drop faster than Fo rise over the same frequency range. This type of experimental paradigm has two main shortcomings if one wants to relate these results with the articulatory difficulty associated with various tone shapes. Firstly, it is not clear that the results obtained with maximum speed of Fo change would still hold with slower rate of Fo change more commonly found in natural speech. Secondly it would be better to use a paradigm in which the frequency range of the target tone is adapted to the frequency range of the subject in order to avoid creating unwanted articulatory constraints caused by the subject's attempt at speaking outside her/his normal Fo range.

2. Experimental paradigm.

In order to remedy these shortcomings the following experiment was carried out: Five American male speakers were asked to produce the following nine tone patterns:

-3 level tones: low level, mid level and high level
-3 rising tones: from low to mid, from mid to high, and from low to high
-3 falling tones: from high to mid, from mid to low, and from high to low

The stimuli were presented visually in a word list in which each token was represented using Chao's tone numbers (1930). (i.e., 11, 33, 55, 13, 35, 15, 53, 31, 51). There were five randomized repetitions of each tone with the vowel [ɪ] and five repetitions with the vowel [a]. The stimuli were presented visually in order to avoid influencing the subjects' production by providing them with an acoustic signal of given duration and amplitude. Two vowels [ɪ] and [a] were used in order to see whether certain tone patterns were produced more accurately with high vowels than with low vowels or vice-versa. Fundamental frequency measurements were made on a PDP-12 computer using the CEPSTRUM method on the waveform digitized at 10 KHz. In order to get comparable Fo and amplitude values from tokens of different durations, each token was divided into five intervals, the average Fo and the average amplitude were computed for each interval.

3. Results.

The results from the three level tones: low level (11), mid level (33), high level (55) are presented in Figure 1. The three upper curves represent the average fundamental frequency patterns from the five subjects. The numbers indicated at the onset and offset of each tone refer to the subject number. Each curve is composed by five points representing the five intervals mentioned earlier. Each one of these points is the average of 50 measurements (5 subjects x 10 repetitions (5 repetitions of [ɪ] and 5 repetitions of [a])). The three lower curves represent the averaged intensity curves for these three Fo patterns. Three facts should be noticed from Figure 1: (1) the low tone is realized as a low falling tone, (2) the frequency spacing between low and mid is roughly equal to the spacing between mid and high, (3) there is a

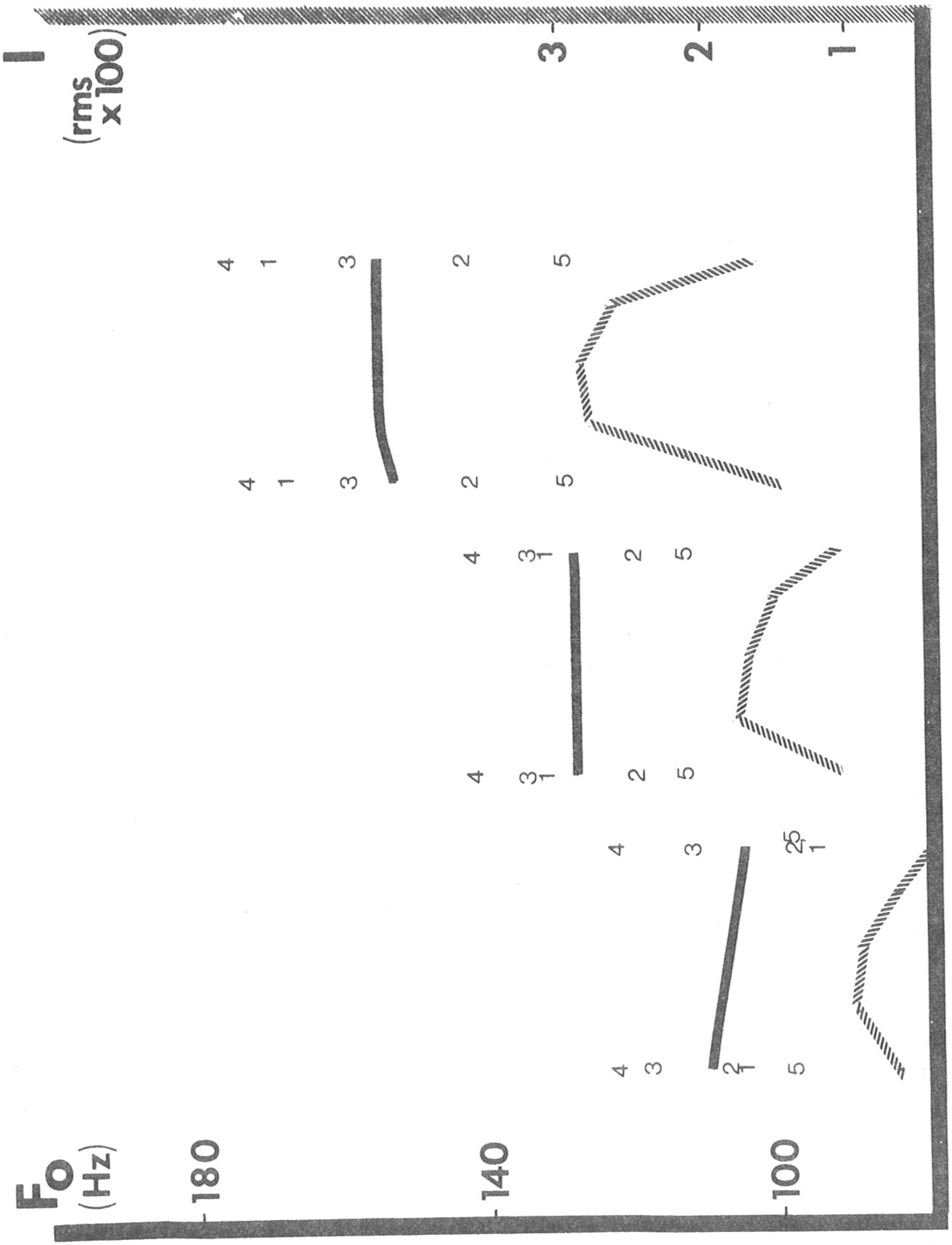

Fig. 1. Fo and amplitude values for the three level tones.

good correlation between amplitude levels and Fo levels.

Figure 2 shows the three rising tones; as can be seen:

(1) The low to mid rising tone (13) covers a smaller frequency range than the mid to high contour (35) (8.8 Hz vs. 33.6 Hz)

(2) the magnitude of the slope of 13 and 15 (i.e., the two rising tones starting at low level) is small during the first half of these tones.

(3) Here also, there is a good correlation between amplitude values and frequency values. (Notice for instance that the third interval has a greater intensity value than the second interval).

As far as the falling tones are concerned (Figure 3), we notice that:

(1) The falling tone from high to mid (53) covers a smaller range than the falling tone from mid to low (31) (20.7 Hz. vs. 29.9 Hz).

(2) The beginning point of the 53 and 51 tones (i.e., the two falling tones which are starting high) have a starting point higher than the high level tone (55).

(3) The magnitude of the slope at the onset of the 53 and 51 tones and at the offset of the 53 tone is small.

(4) There is again a good correlation between amplitude values and frequency values.

A direct correlation was obtained between frequency height and tone duration (i.e., the lower the tone, the shorter the duration). This correlation was found to be significant for three subjects ($P<0.05$); the same tendency was observed for the fourth subject ($P<0.1$) but not for the last subject. For the rising tones, the following durational hierarchy was found for all five subjects: duration of 13 <duration of 35 <duration of 15. Notice that this hierarchy corresponds to the hierarchy based on the frequency range covered by these tones.

31 is the shortest falling tone; no significant differences in duration were found between 53 and 51. It is interesting to point out that falling tones are always shorter than rising tones over comparable frequency range. I mentioned earlier that 2 vowels ([i] and [a]) were used in this experiment. Since no interesting differences were noticed between contour tones produced with these 2 vowels, they were averaged together. However, the intrinsic Fo differences between [i] and [a] exhibit an interesting pattern when produced with the three level tones. These differences are presented in Table 1.

Table 1. Intrinsic Fo differences (in Hz) between [i] and [a]

Low tone	.7	nonsignificant
Mid tone	7.9	$P<.01$
High tone	12.3	$P<.01$

Fig. 2. Fo and amplitude values for the three rising tones.

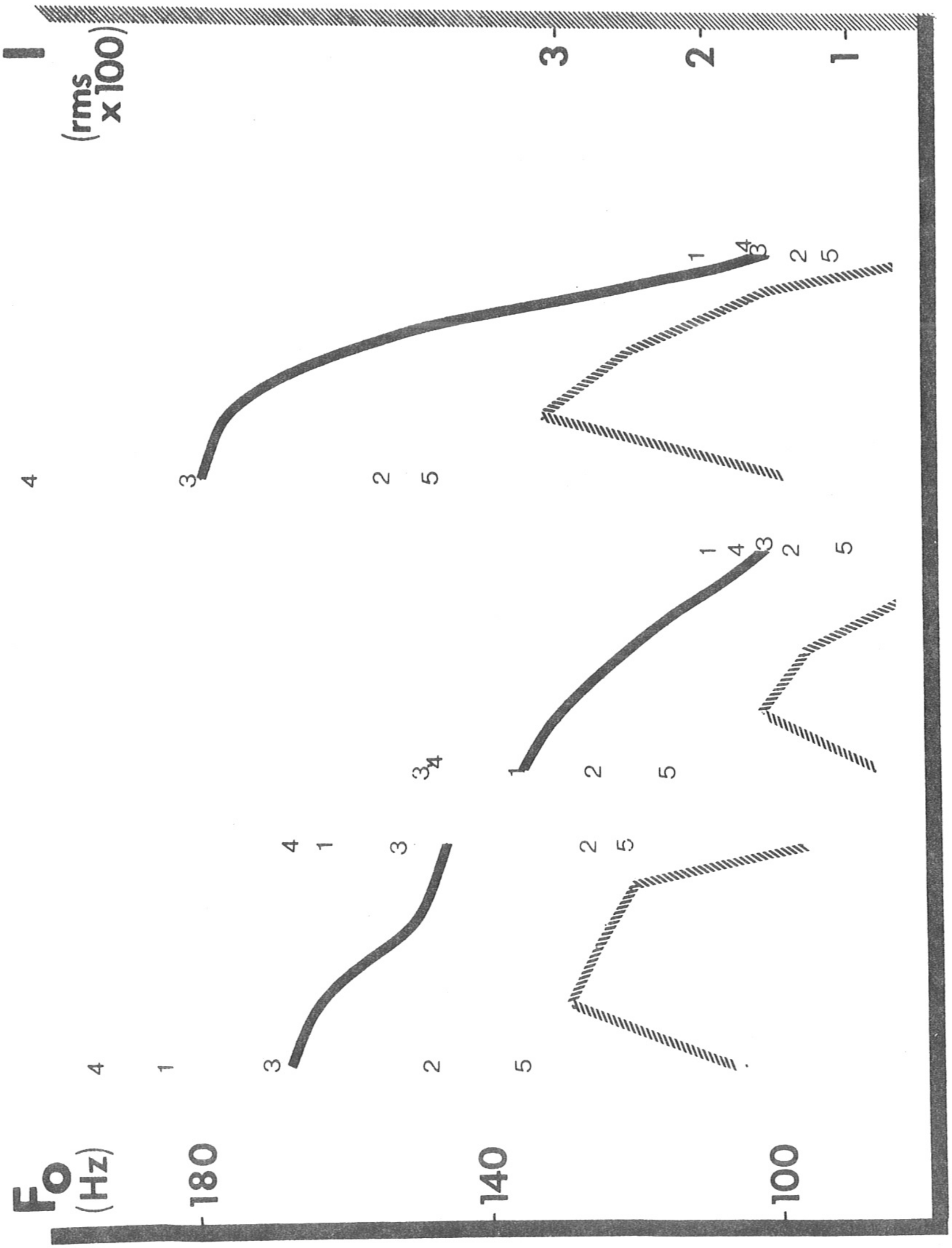

Fig. 3. Fo and amplitude values for the three falling tones.

4. Discussion

The duration data show that even for Fo change realized at a speed used in normal speech, the mechanism controlling Fo lowering is faster than the mechanism controlling Fo rising. As far as level tones are concerned, it is not clear why Fo height should be directly correlated to vowel duration.[1] The Fo data indicate that it is more difficult to control contour tones going from one extremity of the range towards the middle of the range than contour tones going from either the center of the Fo range towards one extremity or from one extremity to the other. The consistently good correlation obtained between amplitude and frequency values suggest that aerodynamic factors play a significant role in controlling Fo in natural speech. The fact that intrinsic Fo differences between high and low vowels are greater at higher Fo confirms the results of previous studies obtained from other languages such as Danish (Petersen 1976), French (Di Cristo and Chafcouloff 1976) and Yoruba (Hombert 1977).[2]

5. Conclusion.

In conclusion, these data are too preliminary to allow an accurate quantification of the articulatory difficulty associated with each tone shape. As a first approximation, it is proposed that contour tones covering a small Fo range (e.g., one stop in Chao's scale) represent a greater articulatory complexity than contour tones covering a wider frequency range. This claim is based on the observation that Fo changes are relatively slow during the initial part of the vowel. However, this constraint should not include contour tones ending at the extremity of the frequency range (i.e., 1 or 5).

Footnotes.

1. Pike (1974) claims that the opposite hierarchy is found in tone languages (i.e., low tones have a longer duration than mid tones which in turn have a longer duration than high tones).

2. Since it was found that in the case of Danish and French, subjects with higher Fo exhibited greater intrinsic differences between high and low vowels, these data suggest an acoustic rather than an articulatory explanation.

References.

Chao, Y.R. 1930. A system of tone letters. *Le Maitre Phonetique* (3), 45:24-27.

Di Cristo, A. and Chafcouloff, M. 1976. An acoustic investigation of microprosodic effects in French vowels. Paper presented at the 14th Conference on Acoustics, High Tatra, Czechoslovakia.

Hombert, J.M. 1977. Consonant types, vowel height and tone in Yoruba *Studies in African Linguistics* 8.2.

Lieberman, P. 1967. *Intonation, perception and language* MIT Press, 38.

Lieberman, P. 1970. A study of prosodic features. *Status Reports on Speech Research.* Haskins Lab., SR 23, 179-208.

Ohala, J.J. 1970. Aspects of the control and production of speech. *Working Papers in Phonetics.* 15.

Ohala, J.J. 1973. The physiology of tone. In L. Hyman (Ed.), Consonant types and tone. *Southern California Occasional Papers in Linguistics* 1. Univ. of So. Calif. 1-14.

Ohala, J.J. (in press) The production of tone. In *Tone: A linguistic survey* V. Fromkin (ed.) Academic Press.

Ohala, J.J. and Ewan, W.G. 1973. Speed of pitch change. *Journal of the Acoustical Society of America* 53, 345.

Petersen, N.R. 1976. Intrinsic fundamental frequency of Danish vowels. *Annual Report of the Institute of Phonetics* University of Copenhagen, 1-27.

Pike, E. 1974. A multiple stress system vs. a tone system. *IJAL* 40, 169-175.

Sundberg, J. 1973. Data on maximum speed of pitch changes. *STL-QPSR* 4. 39-47.