The size of F_0 excursions in speech production and perception

Hartmut Traunmüller and Anders Eriksson
Institutionen för lingvistik
Stockholms universitet, S-106 91 Stockholm, Sweden.

ABSTRACT

In a set of experiments, subjects had to estimate the liveliness of an utterance in which variations in the speaker's age, sex, articulation rate, voice register, and liveliness were simulated. The results showed that listeners equate F_0 intervals that are equal in semitones, as long as no variation in voice register is involved.

INTRODUCTION

A substantial amount of investigations of the voice shows the width of the F_0 range used by individual male and female speakers to be, on average, the same if it is expressed in semitones. This range, which reflects the extent of the F_0 excursions and which can be conveniently expressed in terms of the standard deviation (SD) of F_0, is affected by several linguistic, paralinguistic, and extralinguistic factors. Thus, SD is higher in tone languages than in languages which do not use tone for segmental distinctions and it varies with the type of text and discourse. Based on the description of the type of discourse investigated in a number of studies, we have assigned one of four 'liveliness classes' to each of them. Not surprisingly, we found SD to increase with 'liveliness'. The increase in SD in very lively types of discourse tends to be more pronounced in the speech of women than of men. Table 1 gives a rough overview, based on ten investigations, some of which involved several types of discourse. For details, see Traunmüller and Eriksson (1994a), where it is also shown that speakers normally achieve an increased SD by increasing the excursions of F_0 from a 'base-value' F_b that lies about 1.5 σ below their mean F_0. It is also observed that SD increases slightly with age from 20 to 70 years.

It is well known that pitch intervals in music are perceived as equivalent if the ratio between the two frequencies that define the interval is the same, i.e., if the intervals are equal in semitones. Although this would tie in nicely with the observed properties of F_0 in the speech of men and women, the production data do not tell us which pitch intervals are perceived as equivalent in speech. Except for certain types of calls, 'musical' pitch intervals are not found in speech (Lehiste and Peterson, 1961) and the musical way of perceiving pitch intervals is not the only way. Hermes and van Gestel (1991), investigated the perceptual equivalence of F_0 excursions in speech by means of letting subjects adjust the

Table 1. *Average F_0-variation (SD in semitones) as a function of the type of speech used in ten investigations with male and female speakers in each. For each case in which the SD was higher (lower) for women than for men, a "+" ("–") sign has been enterd.*

	European lang.		Chinese lang.	
Liveliness class	SD	N	SD	N
(4) Very high	4.8	+ +		
(3) High	4.0	+ – –		
(2) Moderate	2.8	– + – – –	4.0	– –
(1) Low	2.1	–		

size of F_0 excursions in resynthesized speech signals to match the perceptual prominence of the syllable marked by the excursion with that of the corresponding syllable in a fixed comparison stimulus produced in a different register. The results showed that the F_0 excursions responsible for the perception of prominence were not judged like musical intervals. Instead, listeners judged F_0 excursions to be equivalent when the excursions of the first partial of the voice source signal had approximately the same size expressed in equivalent rectangular bandwidths (ERB, Moore and Glasberg, 1983).

Since the extent of the F_0 excursions expressed in ERB is larger in the speech of women as compared with men, the speech of women would be heard as more lively than that of men, if the result obtained by Hermes and van Gestel (1991), in accordance with what they claimed, were to hold in general. Similar previous suggestions have been ardently refuted by Henton (1989), although mainly with political arguments. In this case, however, Henton's position can be defended without reliance on political reasoning: We must not take for granted that a difference in register — with unchanged formant frequencies — has the same perceptual effects as a difference in speaker sex.

In order to learn how listeners evaluate F_0 excursions in general, we performed a set of experiments in which subjects had to estimate the liveliness of utterances in which the F_0 excursions had been expanded or compressed. The stimuli used in these experiments were obtained by LPC-analysis of one natural utterance that was resynthesized with F_0, the formant frequencies, and speech rate modified in such a way as to simulate some of the natural extra- and paralinguistic variations that affect F_0 and/or liveliness, namely the speaker's age, sex, articulation rate, and voice register. In each case, the extent of the F_0 excursions was varied in 7 steps. Here follows an account of one of those experiments.

METHOD

Stimuli

The original sentence "Det finns folkstammar som äter både kattkött och hundkött" produced by a female speaker, 28 years of age, had been recorded previously for the purpose of developing the method of simulating extra- and paralinguistic variations by means of LPC-analysis and resynthesis after recalculation of the parameter values (Traunmüller *et al.*, 1989). The parameters affected in the present case included the frequency values of the fundamental and the formants. The Q-values of the formants were kept unmodified. The values of F_0 were recalculated according to the equation

$$f' = k_b\,[160 + k_e\,(f - 160)] \qquad (1)$$

where f' is the recalculated value of F_0 for a given analysis frame, f is its original value, k_e is the 'excursion factor' by which the deviation of F_0 from F_b was multiplied (k_e = 1.00 for the versions with the original F_0 modulation factor), and k_b is the 'base-value factor' that describes the relation between the actual and the original F_b.

The mean F_0 of the original utterance was 215 Hz with an SD of 38.4 Hz (3.0 semitones). Based on the analysis of the extent of the F_0 excursions in speech types that differed in intrinsic liveliness, we assumed a base-value of 160 Hz (the numerical constant in Equation 1). This corresponds to 90 Hz in the male versions, with k_b = 0.56.

The excursion factors chosen (k_e = 0.125, 0.354, 0.650, 1.000, 1.398, 1.837, and 2.315) covered a large range of variation in liveliness. The formant frequencies were transformed in accordance with the power-function approach described in Traunmüller (1988). The male/female frequency ratio was 0.85 at 300 Hz and 0.80 at 3000 Hz. The method involved a change in sampling frequency from 16.0 kHz in the female to 12.474 kHz in the male versions.

Subjects

Fourteen adults, 6 male and 8 female, served as listeners in this experiment. They were undergraduate students at the University of Stockholm and staff members at the department of linguistics. Participation was voluntary and unpaid. No subject participated in more than one experiment.

Procedure

The experiment was run in a quiet lecture room and the stimuli were presented via headphones (AKG K 25) at a comfortable loudness level. The subjects had to note their responses on answer sheets. The instruction had been recorded on a tape that also contained the stimuli. The instruction was immediately followed by an exercise consisting of 8 stimulus pairs, the ratings of which have not been evaluated.

The main part of the experiment consisted of a set of magnitude estimation tasks using pairwise comparison. In each pair the standard was presented before the comparison, with a gap of 500 ms in between and a pause of 5 seconds between successive pairs.

The subjects were asked to assign a number to the comparison stimulus expressing its perceived liveliness. They were instructed to use the number 100 for stimuli whose liveliness they perceived to be equal to that of the standard and to use 50 and 200 for stimuli perceived as 'half as lively' and 'twice as lively', respectively. The subjects were further encouraged to use any more precise number they considered suitable. The concept of 'liveliness' was not further explained and the subjects did not ask for such an explanation.

In the main part of the experiment, each of the 7 female and male comparisons was presented twice. In all pairs, the female version with $k_e = 1.00$ served as standard. The 14 comparisons with female characteristics were presented before those with male characteristics. In the first presentation (stimuli 1 to 7 and 15 to 21) the order of presentation was random. As for the repetitions, it was attempted to balance possible context effects to some degree by presenting the stimuli in the opposite k_e rank order.

At the end of the experiment, the versions with $k_e = 1.00$ were presented alone for the purpose of judging the sex and the age of the speakers.

RESULTS AND DISCUSSION

All subjects heard the stimuli presented for age and sex judgment as produced by an adult female when $k_b = 1.00$ and by an adult male when $k_b = 0.56$. In both cases, the typical age rating was 30 years.

As for the liveliness ratings, inter-rater reliability was high enough (Cronbach's $\alpha = 0.985$ for both male and female stimuli) to justify pooling the results. In Figure 1, the mean value of the liveliness ratings obtained for each stimulus is plotted against the SD of F_0 expressed in Hz and in semitones. It can be seen in these diagrams that, given the same variation of F_0 in Hz, the liveliness ratings for the male speaker were consistently higher than those for the female speaker, and the slope of a linear regression line fitted to the male data is significantly ($p < 0.002$) steeper than that of a regression line fitted to the female data. If the comparison is based on semitones, the discrepancy between the two types of speaker is almost completely eliminated, and the slopes of regression lines fitted to the two sets of data are not significantly different. This implies that if expressed in semitones, a female speaker has to increase her F_0 excursions just as much as a male speaker in order to achieve the same increase in perceived liveliness, but expressed in Hz, bark, mel, or ERB, female speakers need larger excursions.

Although the results of this experiment suggest that even in speech, listeners equate F_0 intervals that are equal in semitones, further experiments showed that this holds only as long as no variation in voice register is involved (Traunmüller and Eriksson, 1994b). When the voice register was shifted without adjustment in articulation, listeners appeared to adjust their ratings with respect to the spectral space available between F_b and some average F_1, so that increased (decreased) ratings were obtained when this space was compressed (expanded). With stimuli simulating modal and falsetto register of a male speaker, we obtained results quite similar to those of Hermes and van Gestel (1991), but the perceptual equivalence of equal ERB intervals must now be seen as accidental and specific to this particular case.

Additional findings included a strong effect of articulation rate on perceived liveliness and an unexpected and as yet unexplained but significant difference in the age ratings obtained from male and female listeners when presented with versions whose F_0, formants, and speech rate were those of a child.

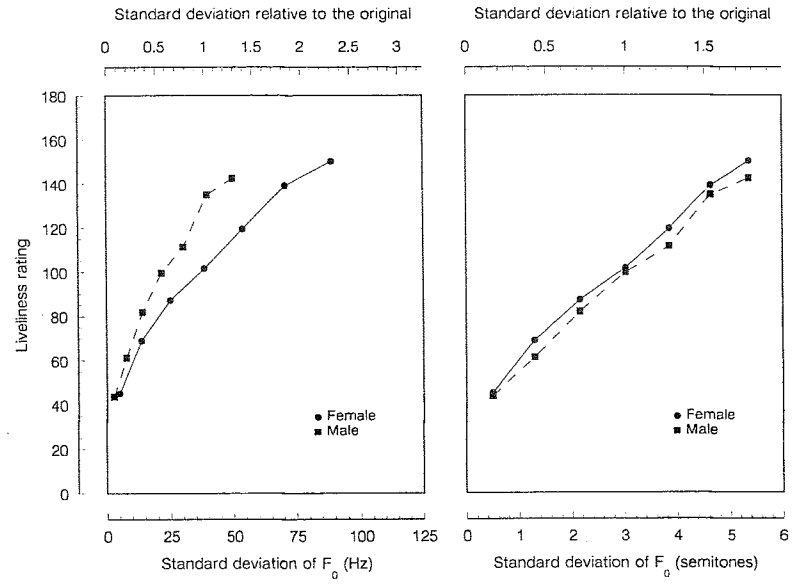

Figure 1. *Liveliness ratings obtained with male and female versions of an utterance shown as a function of the standard deviation of F_0 expressed in Hz (left) and in semitones (right).*

ACKNOWLEDGMENTS

This research has been supported, in part, by a grant from HSFR, the Council for Research in the Humanities and Social Sciences, within the frame of the Language Technology Program.

REFERENCES

Henton, C. G. (1989) "Fact and fiction in the description of female and male pitch", *Language & Communication* 9, 299—311.

Hermes, D. J. & J.C. van Gestel (1991) "The frequency scale of speech intonation", *J. Acoust. Soc. Am.* 90, 97—102.

Lehiste, I. & G.E. Peterson (1961) "Some basic considerations in the analysis of intonation", *J. Acoust. Soc. Am.* 33, 419—425.

Moore, B.C.J. & B.R. Glasberg (1983) "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns", *J. Acoust. Soc. Am.* 74, 750—753.

Traunmüller, H. (1988) "Paralinguistic variation and invariance in the characteristic frequencies of vowels", *Phonetica* 45, 1—29.

Traunmüller, H., P. Branderud & A. Bigestans (1989) "Paralinguistic speech signal transformations", in *PERILUS* 10, Institute of Linguistics, University of Stockholm, 47—64.

Traunmüller, H. & A. Eriksson (1994a) "The frequency range of the voice fundamental in the speech of male and female adults", submitted to *J. Acoust. Soc. Am.*

Traunmüller, H. & A. Eriksson (1994b) "The perceptual evaluation of F_0 excursions in speech as evidenced in liveliness estimations", submitted to *J. Acoust. Soc. Am.*