

# Size and physiological effort in the production of signed and spoken utterances

Hartmut Traunmüller

Dept. of Linguistics, Stockholm University  
hartmut@ling.su.se

## Abstract

*It is shown that the energy required in order to articulate manual and oral gestures at a given rate varies in proportion with the fifth power of linear body size, while the energy supply varies with its second power. This provides for a better understanding of the differences in peripheralness observed in the formant frequencies of vowels in speech articulated more or less forcefully by men, women and children.*

## 1 Introduction

It is well known and well understood that all formant frequencies of vowels vary between speakers in rough proportion with the inverse of the length of the vocal tract, so that the relationship can be described, in approximation, by uniform scaling factors (the same for all formants of all vowels). However, the female-male difference observed in the formant frequencies of the same vowels can not be fully accounted for by a uniform scaling factor. The deviation from uniformity was observed to be similar in speakers of various languages: After normalization with a uniform scaling factor, the vowels produced by women tend to stand out as more peripheral than those produced by men.  $F_1$  and  $F_2$  both show more dispersion (more extreme values) in vowels produced by women.

The reason for the non-uniformity can be seen in the fact that the male vocal tract is not a proportionally upscaled version of the female (Fitch & Giedd 1999). While it is somewhat larger in all its dimensions, the major difference resides in its disproportionately elongated pharynx. In a model simulation of the changes that occur in the male during puberty, and which were assumed to account for the major part of the female-male non-uniformity, it was shown that the more peripheral values of  $F_2$  may result from this physiological difference (Traunmüller 1984).  $F_3$  was also well predicted, but the model failed for  $F_1$ . However, this simulation was based on anatomical data that described only the growth of the skeleton. The growth of the tongue was assumed to be more uniform, which was crucial for the resulting values of  $F_2$ . If the changes that boys experience during puberty are responsible for the non-uniformity, vowel contrast should be reduced in the speech of men as compared with that of children and women. However, an analysis of formant frequency data obtained from Japanese children, adolescents and adults suggested, instead, that the vowels of adult women deviate from the general pattern in their increased contrast (Traunmüller 1988). The "general pattern" showed a steady decrease in the dispersion of  $F_1$  (expressed in barks) and an increase in that of  $F_2$  and  $F_3$  as a function of age.

A different explanation for the increased peripheralness of women's vowels had been first suggested by Ryalls & Lieberman 1982 and was considered again by Diehl, Lindblom, Hoemeke & Fahey 1996. Their idea was that women use more peripheral and, thus, more contrastive vowels in order to compensate for a loss in distinctness due to their higher  $F_0$ .

This suggestion is based on the fact that the envelope of the spectrum, which reflects the resonance properties of the vocal tract, is sampled by the partials of the voice, and that these samples are taken more sparsely when  $F_0$  is higher. This provides an explanation for the difference if male speech is taken as the norm or, more reasonably, if combined with the laziness principle, which says that people limit their efforts to the necessary minimum. If this principle holds, men do not use so peripheral vowels since this is not required in order for their speech to be intelligible. This explanation appears to suggest that speaking requires a larger effort of women than of men. In the following, we shall consider whether this is true and take vowels produced by children into account.

If we compare the formant frequencies of vowels in stressed syllables with those of the same vowels in unstressed position (Koopmans-van Beinum 1980), we can observe differences that are quite similar to those between adult female and male productions of vowels. Vowels in unstressed position are less contrastive. These variations in vowel contrast appear to result from the same commands being executed more forcefully and more persistently in stressed syllables. This may also account for the longer segment durations we can observe in stressed syllables, especially in the vowels.

However, the laziness principle does not appear to give us any reason to expect men's vowels to be less peripheral than those of women when pronounced in isolation and sustained for a sufficient time. We shall consider some data relevant to this point.

## 2 Energetics

Basic physics tells us that  $E = m v^2 / 2$ . This is the energy  $E$  required in order to accelerate a body with mass  $m$  to the velocity  $v$ . Thus, the energy needed in order to accelerate an articulator such as the arm (in signed language and in applauding) or the jaw (in spoken language) to a given velocity is proportional with its mass. Let us now compare two articulators that are proportional but different in linear size by a ratio  $\alpha$ , their volumes and masses subsequently being different by a ratio of  $\alpha^3$ . If  $v$  is given, we see that the energy requirement also varies in proportion with  $\alpha^3$ . However, if all the dimensions of the body are increased by  $\alpha$ , the distance that the articulator has to move between its goals is also increased by  $\alpha$ . Therefore, we need a larger velocity in order to reach a certain goal in a given time, i.e., in order to articulate at a given rate. Since  $E$  for a given  $m$  increases in proportion with  $v^2$ , we now see that the energy requirement for signing or speaking at a given rate varies no less drastically than in proportion with  $\alpha^5$ . In principle, this holds for all articulators, including jaw, tongue body and tongue tip as well as for the limbs involved in signing. However, while the jaws of men and women can be considered as roughly proportional in all dimensions, this does not hold for the vocal tracts. By the descent of the larynx, the vocal tract of men is not only elongated but also widened in the velo-pharyngeal region. Therefore, the distance that the tongue body has to move in order to reach the velum, the uvula, or the pharyngeal wall from a neutral position (or vice versa) is increased by a larger factor than the overall average size ratio of the organs of speech. This can be seen in the tracings of tongue movements obtained by Simpson (subm.).

We conclude that the energy required for speaking or signing or applauding at a given rate increases in proportion with the 5th power of the linear size of the organs involved, and in male speech production the energy requirements are augmented even more for tongue movements in the velo-pharyngeal region.

Now, we must also take the difference in energy supply into consideration. The potential of muscles to do work is generally considered to be proportional to the cross sectional area of the muscle. Thus, the energy supply increases in proportion with  $\alpha^2$ . If we understand "physiological effort" in relation to the available potential, the muscles of a larger organ of speech can be said to provide  $\alpha^2$  times more energy at the same physiological effort.

Since the energy supply only increases in proportion with  $\alpha^2$ , it does not match the demand, which increases in proportion with  $\alpha^5$  or more. The "physiological effort" required in order to speak at a given rate and degree of distinctness increases approximately in proportion with  $\alpha^3$ . Therefore it would take a substantially greater physiological effort of men to match women in the articulation of speech. Thus, it is not true that speaking requires a larger effort of women than of men. On the contrary, the reverse would be true if men had to speak like women. Now, we know that they do not speak like women, and it may be that their "physiological effort" just matches that of women.

All the energy supplied to the articulators will eventually be transformed into heat when the articulator is decelerated. In speech, this happens predominantly at the onset of occlusions in which the tongue or the lower lip makes contact with the opposite surface of the vocal tract, just as it happens in applauding when the palms meet each other. The energy that has to be transformed into heat at these instances is still proportional with  $\alpha^5$ , but the limitation in energy supply is not relevant here. It takes no muscular effort to stop the motion of an articulator in such a collision, but it does take an effort and some extra time to decelerate an articulator by the action of muscles.

### 3 Discussion

It is fairly clear that the acoustic differences between the formant frequency trajectories in stressed syllables compared with those in the same syllables in unstressed position (Koopmans-van Beinum 1980) can be understood as due to the difference in physiological effort made in their production. A comparison of the formant frequency trajectories in the speech of women and men reveals differences similar to those between stressed and unstressed syllables, but now we understand that they are not due to a similar difference in physiological effort. Women do not make a larger effort. On the contrary, our considerations of the energy balance suggest that speaking is a more demanding task for men.

If men used no more physiological effort than women do, then their articulators would not accelerate as fast as in the speech of women, while they have to move a longer distance. In order to approach the articulatory goals as closely as in the speech of women (in relative terms), men would need a longer time for performing their speech gestures. This would show itself more in the transitory than in the stationary sections of the speech signal. However, men do not appear to speak more slowly than women do. Lee, Potamianos & Narayanan 1999 observed no significant sex difference in vowel durations. Some investigations even revealed a slightly higher average articulation rate for men. Instead, men allow themselves to be more sloppy, not approaching the articulatory goals as closely as women do. Simpson (subm.) observed the stationary sections of vowels to be shorter in the speech of men although their speech rate was the same. The sloppy behavior of men is compatible with the hypothesis advocated by Diehl et al. 1996. Men do not need to be more distinct in order for their speech to be intelligible enough.

So far, this account appears to be incontestable. But consider sustained vowels pronounced in isolation. In these, the articulators have simply to be held in a stationary position, which takes very little effort. In such cases, there appears to be no reason for men not to use a more distinct articulation, but it may be that speakers stick to their habits regardless of the needs of the particular situation. In this question, we do not have unequivocal evidence. An investigation by Eklund & Traunmüller 1997, in which vowels without a consonantal context were used, revealed more peripheral positions of  $F_2$  in the women's vowels, but there was no such tendency in  $F_1$ . However, in Fant's 1959 data obtained with sustained vowels, there was such a tendency in  $F_1$  as well, although it showed itself where the need of improved distinctness was the least.

Now, consider children whose  $F_0$  is higher than that of women. These would seem to be in need of being more distinct, and according to our considerations, it would not take them a large physiological effort to produce more peripheral vowels. However, the data do not suggest an everywhere consistent relationship between  $F_0$  and formant frequency dispersion nor between  $F_3$  (as a measure of size) and the dispersion of  $F_1$  and  $F_2$  (Traunmüller 1988, Lee, Potamianos & Narayanan 1999), but it is conspicuous that children below an age of about 7 years do have more extreme values of  $F_1$  as compared with women while their  $F_2$  values are *less* extreme.

We can conclude that considerations of physiological effort weighted against the need of distinctness can explain most of the tendency in  $F_1$ , but not that in  $F_2$ . It appears quite reasonable that speakers are concerned more with the distinctness of  $F_1$  than with that of  $F_2$ . With an uncertainty of  $F_0/2$ ,  $F_2$  is relatively less sensitive to spectral underdefinition due to a high  $F_0$  than  $F_1$  is. Moreover, languages tend to make distinctive use of more levels in vowel height ( $F_1$ ) than in backness and roundness ( $F_2$ ). The larger  $F_2$  dispersion in the speech of adults as compared with children can not be explained on the basis of the sufficient-distinctness-by-least-effort hypothesis. The age- and sex-related variation in  $F_2$  dispersion may, instead, be explained on anatomical grounds. However, we can not rule out a possible reinforcement of the male-female difference by sociolinguistic factors.

### Acknowledgments

The ideas presented here emerged from research projects financed by HSFR. I am grateful to Randy Diehl and Björn Lindblom for stimulating comments.

### References

- Diehl, Randy L., Björn Lindblom, Kathryn A. Hoemeke & Richard P. Fahey. 1996. 'On explaining certain male-female differences in the phonetic realization of vowel categories'. *Journal of Phonetics* 24, 187-208.
- Fant, Gunnar. 1959. 'Acoustic analysis and synthesis of speech with applications to Swedish'. *Ericsson Technics* 1.
- Fitch, W. Tecumseh & Jay Giedd. 1999. 'Morphology and development of the human vocal tract: A study using magnetic resonance imaging'. *Journal of the Acoustical Society of America* 106, 1511-1522.
- Koopmans-van Beinum, Florian J. 1980. *Vowel Contrast Reduction: An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions*. Thesis, University of Amsterdam.
- Lee, Sungbok, Alexandros Potamianos & Shrikanth Narayanan. 1999. 'Acoustics of children's speech: Developmental changes of temporal and spectral parameters'. *Journal of the Acoustical Society of America* 105, 1455-1468.
- Ryalls, J.H. & Philip Lieberman. 1982. 'Fundamental frequency and vowel perception'. *Journal of the Acoustical Society of America* 72, 1631-1634.
- Simpson, Adrian. (subm.) 'Dynamic consequences of differences in male and female vocal tract dimensions'. *Journal of the Acoustical Society of America*.
- Traunmüller, Hartmut. 1984. 'Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels'. *Speech Communication* 3, 49-61.
- Traunmüller, Hartmut. 1988. 'Paralinguistic variation and invariance in the characteristic frequencies of vowels'. *Phonetica* 45, 1-29.
- Traunmüller, Hartmut & Ingegerd Eklund. 1997. 'Comparative study of male and female whispered and phonated versions of the long vowels of Swedish'. *Phonetica* 54, 1-21.