HIGH FREQUENCY SPEECH PERCEPTION: PHONETIC ASPECTS AND APPLICATION

MÁRIA GÓSY

Department of Phonetics, Linguistics Institute of HAS
Budapest 1250 Pf. 19. Hungary

ABSTRACT

The so-called speech frequencies (100-3000 Hz) seem to be both necessary and sufficient for perception and understanding. The role of speech elements occuring above 3000 Hz is unclear. They might be totally unnecessary, on the one hand or, on the other, they might have a secondary acoustic cue function which is demonstrated experimentally by removing the lower frequencies. Experiments were carried out with Hungarian native listeners both with normal and with impaired hearing. The results are given in detail.

INTRODUCTION

The so-called speech frequencies seem to be both necessary and sufficient for the perception of vowels and consonants. The acoustic information in this frequency range is generally suitable for understanding running speech. However, a lot of comprehension problems arise if only these frequencies can be used. This can be demonstrated with the telephone where general conversation can easily be carried out without any problems in understanding. However, identification of names or comprehension of suddenly changed topic of dialogue can cause difficulty. It is known that people with hearing loss at high frequencies (above 3000 Hz) suffer from perceptual and understanding difficulty.

There is no doubt that the first two energy maximums, the formants, contain the main information for the identification of vowels and certain consonants. Moreover, components of some other consonants - like [s] or [ts] - occurring below 3000 Hz are sufficient for their identification. The role of high frequencies (above 3000 Hz) in perception, however, has been little investigated [1]. The acoustic information contained in the high frequencies may be purely supplementary; alternatively it may play an independent and special role in perception. To bring this problem a little closer to a solution, experiments were carried out with Hungarian-speaking native listeners.

METHOD AND MATERIAL

The material used consisted of (i) 25 sound-sequences without meaning and (ii) 102 monosyllabic, phonetically balanced Hungarian words. The bisyllabic sound-sequences contain almost all Hungarian speech sounds. The acoustic structure of part of them corresponds to Hungarian phonotactic rules while that of another part of them contradicts them. All the words consist of three sounds: a vowel between two consonants. The words range from well-known ones, in everyday use, to ones very rarely used. They belong to different grammatical categories. Attempts were made to choose booth the sound-sequences and words containing consonants and vowels in different phonetic positions and in different environments. The speech material was recorded by a male announcer who pronounced it as isolated statements in random order. The recording was made with a professional tape recorder and microphone under laboratory conditions. An 8 s pause was left between the sound-sequences/words. The intensity level of sound-sequences and the words varied within ± 6 dB. Two types of filtration method were used for testing: pass-band and high-pass filtering by an Audio Filter. The filter slope was always 36 dB/octave. The cut-off-frequencies were for words 2200, 2700, 3300, 3900 Hz and 2200-2700, 2700-3300, 3300-3900, 3900-4700 Hz; for sound-sequences 2200, 2700 Hz and 2200-2700, 2700-3300 Hz. These values were chosen in view of the fact that the highest acoustic cue for Hungarian vowels appears in general to be about 2200 Hz; it is the second formant for the [i] sound. There are 8 different materials for the words and 4 for sound-sequences. In order to examine the role of the upper frequencies, those below 2200 Hz were removed. The frequency analyses were made of filtered material by the Sound Spectrograph (Type 700 of Voice Identification). Each of the 12 test materials was administered to 10 adult normal-hearing subjects, totally 120 subjects, half of them females and half males. The experiments were conducted in a silent room. The listeners' task was to write down the sound-sequences or words they could perceive/understand. In order to obtain statistically significant results, we used our own Psychotest program.

RESULTS AND DISCUSSION

The experimental data for sound-sequences and for words are summarized in Table 1. These show that (i) the perception/understanding of sound-sequences/words was better under pass-band filtering than under high-pass filtering; (ii) perception/understanding decreased under high-pass filtering according to the change of the cut-off-frequency; (iii) a frequency band seems to occur with the highest perception and understanding ratio: 2200-2700 Hz. The differences between the filtered groups proved to be significant at the .01 level.

Table 1

Cut-off-frequencies of filtering (Hz)	Correct identification (%) words	sound-seq.
2200 h.p.	67	49
2200-2700 p.b.	98	78
2700 h.p.	72.5	35
2700-3300 p.b.	95	75
3300 h.p.	74	
3300-3900 p.b.	95	
3900 h.p.	46	
3900-4700 p.b.	95	

The abbreviations mean high-pass and pass-band filtering.

These results led us to the conclusion that there are frequency bands in which more acoustic information about the same word/sound-sequence seems to disturbing to the decoding processes [2]. The supposed idea is that the upper part of the acoustic structure of certain speech sounds does not remain characteristic for them when the lower part is lost. In other words: these high frequencies do not contain unambigous information about the sounds or cannot be acoustic cues used for identification. The components appearing at these frequencies have been thought to play a supplementary role in recognition. Results obtained from examinations using the low-pass filtration method confirm this [3]. If this were the case, the high elements would have been redundant. Our new results have not confirm this assumption, and, indeed, they seem to contradict it. The data have supported an alternative

hypothesis, namely that certain speech sounds and sound combinations have special 'cue-like' components above 2200 Hz. This 'secondary-cue' hypothesis was further investigated by means of spectrographic analyses. These showed that, as expected, the main difference in acoustic structure between pass-band and high-pass filtered groups lies in the presence or absence of the higher frequencies.
By way of illustration let us look at the bilabial nasal consonant [m]. The original acoustic structure of [m] contains cues at about 500 and 1500 Hz. In the absence of these frequencies it is not possible to identify [m] without elements above 2000 Hz. Spectrographic analysis of [m] shows further components at about 2800 and 3700 Hz. The word mos 'washes' was understood accurately when frequencies below 2000 Hz were removed by filtration. When the component at 2800 Hz was reduced in intensity by further filtration, identification of the consonant became impossible (Fig. 1).

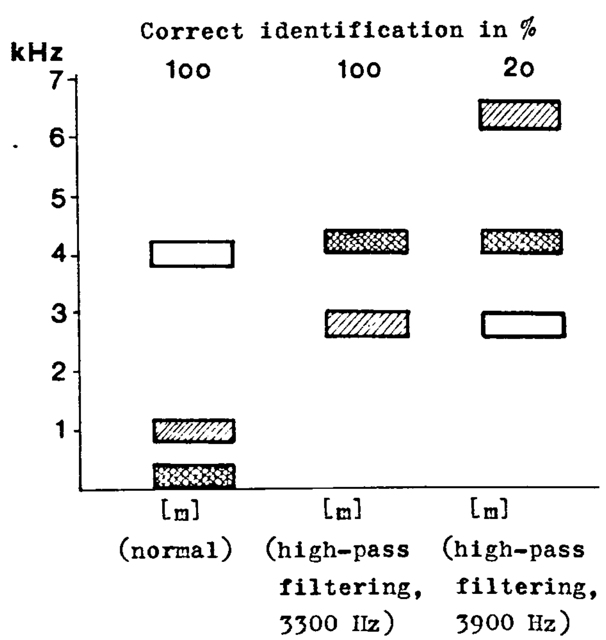

Figure 1. Consonant [m] in word mos 'washes' and its identification in %

One of the questions to be asked in this respect is: why can we not used the whole information of the high frequencies in perception, why do they seem to cause difficulty? Moreover, why do these perceptual problems disappear when there are only frequency bands? This suggests that, in contrast to the main acoustic cues (below 2200 Hz), the secondary cues act alone and independently of the disturbing higher components. As to the explanation of 'disturbing higher components' let us document it with an example. The perception of the Hungarian long [a:] vowel was analyzed. The correct identification of this sound in the sound-sequence tádó [ta:do:] after high-pass filtering is 0% and after pass-band filtering is 100% (with the cut-off-frequencies of 2200 and 2200-2700 Hz). The false responses in the first case were: [tøldu, tødu, tøldo:, tɔdu]. Instead of [a:] dominantly [ø] was perceived. In the acoustical structure of [a:] there are components between 2000 and 4000 Hz with very different intensities. This is assumed to cause the perceptual differencies. The spectrographic analyses show, however, an important difference depending upon the context (sound combination). The [a:] mentioned above was perceived more correctly when it occurred between fricative or fricative and nasal consonants. This can be explained by the transition phases which act as acoustic cues in this respect.
Finally the role of meaning should be taken into consideration. If we compare the frequency of use and the correct identification percentages of the test words, it seems clear that meaning generally does not play as important a role in this case as is supposed in literature. There are frequent words with low understanding ratio, and rarely used words with high percentage values. There are a lot of items which cannot be used in isolation in Hungarian, e.g. pác 'pickle' (70%) and zöm 'bulk' (20%). There are words with similar meaning or frequency and their understanding is quite different; and words with similar acoustic structure and different percentage values. The grammatical category of the words seems to be of lesser importance as well.
By way of final conclusion the following idea will be presented. All the results have supported that perception and understanding are better in certain high frequency bands especially in 2200-2700 Hz. This finding led us to the hypothesis that hearing-impaired people with special hearing losses can perceive/understand speech in the 2200-2700 Hz range better than in a wider band which also contains the 'disturbing' elements.
A supplementary experiment was carried out with the participation of 10 hearing-impaired adults having hearing losses of different types and extents. Table 2 shows the responses of a mixed-type hearing-impaired woman for the words with their normal acoustic structure and after pass-band filtering.

Table 2

Original words	Responses of a hearing-impaired adult: normal sounding	Responses of a hearing-impaired adult: after pass-band filtering (2200-2700 Hz)
moʃ	moʃ	moʃ
kør	kol	kør
meɲ:	-	meɟ
la:b	ɔd	la:b
ʒeb	ʒe	ʒeb
hi:d	hi:g	hi:d
si:n	sẹ:p	si:n
fɔl	-	fɔl

The results confirm that the secondary acoustic cues can, indeed, ensure the perception/understanding of speech in case the normal decoding process cannot work because of hearing problems.

What criteria should the high frequency components fulfil in order to act as acoustic cues? (i) Identification should reach a significant level and, (ii) frequency values should be defined for correct perception. On the basis of our data it can be supposed that the components appearing in certain frequency bands correspond to the above-mentioned expectations.
Further research should show how these findings can be applied in audiological examinations, phoniatric work and in speech therapy.

REFERENCES

[1] Fant, G.: Speech Sounds and Features. Cambridge, Massachusetts, and London 1973.
[2] Rosen, S. - Fourcin, A.J.: When less is more - Further work. Speech Hearing and Language No. 1. 1983, 1-27.
[3] Gósy, M.: Magyar beszédhangok felismerése, a kísérleti eredmények gyakorlati alkalmazása./Identification of Hungarian speech sounds, the application of experimental results. Hungarian Papers in Phonetics No. 15. 1986, 3-100.