HIGH FREQUENCY SPEECH PERCEPTION: PHONETIC ASPECTS AND APPLICATION

NÁRIA GÖGY
Department of Phonetics, Linguistics Institute of HAS
Budapest 1520 Pr. 15, Hungary

ABSTRACT

The so-called speech frequencies (1000–3000 Hz) seem to be both necessary and sufficient for perception and understanding. The role of speech elements occurring 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 ([s]) - 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 so-called speech frequencies (1000–3000 Hz) seem to be both necessary and sufficient 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.

The so-called speech frequencies (1000–3000 Hz) seem to be both necessary and sufficient 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.
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).

One of the questions to be asked in this respect is: why can we not use the whole information of the high frequencies in perception, why do they seem to cause difficulties? 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 [a:]a [a:] 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: [ðːma, tɛðu, tɛðːr, tʊːdːl]. Instead of [a:] dominantly [ɛ] 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 differences.

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 identifications 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. [sɛːˈplikɛː] ('piecle' (705) and sim 'bull' (205). 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 2200-2700 Hz. 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.

Table 2

<table>
<thead>
<tr>
<th>Original</th>
<th>Responses of a hearing-impaired word</th>
<th>paired adult</th>
<th>normal</th>
<th>after pass-band filtering (2200-2700 Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[nɒ]</td>
<td>[nɒ]</td>
<td>[nɒ]</td>
<td>[nɒ]</td>
<td></td>
</tr>
<tr>
<td>[hɒ]</td>
<td>[hɒ]</td>
<td>[hɒ]</td>
<td>[hɒ]</td>
<td></td>
</tr>
<tr>
<td>[mɛː]</td>
<td>[mɛː]</td>
<td>[mɛː]</td>
<td>[mɛː]</td>
<td></td>
</tr>
<tr>
<td>[lɛːb]</td>
<td>[lɛːb]</td>
<td>[lɛːb]</td>
<td>[lɛːb]</td>
<td></td>
</tr>
<tr>
<td>[ʃɛːb]</td>
<td>[ʃɛːb]</td>
<td>[ʃɛːb]</td>
<td>[ʃɛːb]</td>
<td></td>
</tr>
<tr>
<td>[hɪːd]</td>
<td>[hɪːd]</td>
<td>[hɪːd]</td>
<td>[hɪːd]</td>
<td></td>
</tr>
<tr>
<td>/sɛːp/</td>
<td>/sɛːp/</td>
<td>/sɛːp/</td>
<td>/sɛːp/</td>
<td></td>
</tr>
<tr>
<td>[fɔːl]</td>
<td>[fɔːl]</td>
<td>[fɔːl]</td>
<td>[fɔːl]</td>
<td></td>
</tr>
</tbody>
</table>

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 fulfill 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