Frequency-weighting functions for speech in young and older listeners

43.71.Lz

Gilbert Gaëtan; Micheyl Christophe; Berger-Vachon Christian; Collet Lionel
UMR CNRS 5020. Université Claude Bernard Lyon I
50, Av. Tony Garnier. Lyon. France

ABSTRACT

The aim of the present study was to test whether slight hearing losses that occur at high-frequencies with ageing can influence spectrally-based perceptual strategies for speech recognition. Perceptual strategies were assessed using the correlational method developed by Turner et al. [JASA 104, 1580-1585]. This method involves the computation of correlation coefficients between the percentage of correct recognition of natural vowel-consonant-vowel (VCV) stimuli and the signal-to-noise ratio in different frequency bands. The estimated correlations are taken to represent the perceptual weights assigned to the different spectral bands in a given subject. Twenty subjects took part in the study. They were divided equally into two groups: one consisting of young (22 to 27 years old) listeners with pure-tone detection thresholds <=20 dB HL, and one consisting of elderly subjects (48 to 61 years old), most of whom (7 out of 10) had a slight hearing loss at high frequencies (mean absolute threshold between 4 and 8 kHz = 20.4 dB HL versus 2.6 dB HL in the first group). Signal processing was carried out in the digital domain and involved splitting of the signal spectra into 5 frequency bands (band 1: 100-250 Hz; band 2: 250-750 Hz; band 3: 750-1750 Hz; band 4: 1750-3750 Hz; band 5: 3750-7750 Hz), and adding a pseudo-randomly selected amount of noise independently in each band, on each presentation. The stimuli were presented under headphones in a quiet room. The subject's task was simply to identify the phonemes heard. Despite large across-subject variability, statistically significant differences were observed between the two groups. Notably, older subjects put on average less weight than younger subjects on the speech information present in the fourth frequency band but, conversely, they relied more on speech information in the third band. Furthermore, a significant relationship was observed between the perceptual weights of these two spectral bands and the hearing thresholds at high frequencies. Altogether, these results indicate that a slight degree of hearing loss may have a substantial influence on spectrally-based perceptual strategies for speech recognition.

INTRODUCTION

Over the past fifty years, a number of researches have been devoted to unravelling acoustic cues as to phonemic identity in speech signals. A large number of cues have been characterized, both in the spectral and in the temporal domain. However, which of these cues are effectively used by the auditory system for the purpose of acoustic-phonemic decoding remains largely unknown. An important aspect of the problem relates to the fact that the acoustic cues as to phomenic identity are scattered throughout the spectrum. Therefore, it is important to try and determine the importance of different frequency bands for speech understanding.

A first approach traditionally used to this aim is to measure systematically the intelligibility of filtered speech signals. This approach has led to the elaboration of the articulation index (AI), subsequently transformed into the speech intelligibility index (SII) (ANSI S3.5., 1997). One problem with this approach, however, is that narrow-band filtered speech signal are unnatural, and it is possible that speech recognition on the basis of such unnatural signals involves different mechanisms from those involved in the perception of broadband speech signals. In order to overcome this limitation, Doherty and Turner (1996) devised another approach which aims to measure the importance of different frequency bands with unfiltered speech. This method is inspired from the conditional-on-a-single-stimulus (COSS) method that was initially developed by Berg (1989) in order to study the micro-mechanisms of spectral profile analysis with synthetic complex sounds. The general principle of the COSS method consists in imposing random level variations on the frequency components of the complex signals that listeners must discriminate or identify, and to measure how these random variations influence performance. The stronger the correlation between the level of a given component and the perceptual performance, the more the component in question contributed to the listener performance. Positive correlations indicate a positive contribution of the considered component; negative correlations reveal a detrimental influence. The correlationnal method developed by Doherty and Turner (1996) is a generalisation of the COSS method, which consists in calculating the correlation coefficients between phoneme identification performance (a binary variable) and the signal-to-noise ratio in each of different frequency bands. The stronger the correlation, the more the considered frequency band is important for speech understanding. The representation of the correlation coefficients normalized to 1 as a function of the frequency band, which is called a frequency weighting function, provides a global picture of how the listener weighs the information contained in the different frequency bands.

Besides their theoretical interest for the comprehension of the mechanisms underlying speech perception, the measure of frequency weighting function for speech recognition has a potential practical interest for the fitting of hearing-aids in hearing-impaired subjects. Most hearing-aids (be they external acoustic prostheses or cochlear implants) carry out a decomposition of acoustic signals into a number of frequency bands, and allow to set different fitting parameters (like compression ratios and compression delays) individually for each frequency band. Unfortunately, in most cases, the actual importance of these different frequency bands for speech understanding in the patient being tested remains largely, if not completely undetermined. By providing a systematic and statistically reliable way of characterizing the importance of different frequency bands for speech udnerstanding in a given individual, the correlationnal method might prove very useful for the adjustment of optimal fitting parameters of hearing-aids and cochlear implants.

However, before this can be achieved, it is first necessary to gather further information on the capacity of the correlational method to provide reliable information on the perceptual strategies of both normal hearing and hearing-impaired listeners. So far, only three studies have been published on the use of the correlational method for the measurement of frequency-weighting functions for speech. Two of these studies employed normal-hearing listeners; the third was performed in cochlear implantees. Thus, the amount of data available at present on this method remains very limited, and further study is undoubtedly in order.

In the present study, we first measured frequency weighting functions for speech recognition in 10 young normal-hearing subjects and 10 elderly subjects having normal hearing for their age. These two groups exhibited significant differences in their weighting functions, suggesting that they rely differently on the information contained in different frequency bands in order to understand speech. In order to try and determine what were the factors responsible for these differences, we explored the relationships between the weighting functions and the absolute hearing thresholds measured in these subjects. Finally, we also carried out detailed spectral and temporal analyses of the speech signals used in this study, as well as simulations of the responses of the peripheral auditory system to these signals, in order to try and determine what factors where responsible for the general shape of the frequency weighting functions for speech that were measured in this study.

II. MATERIAL AND METHODS

II.1. Subjects

Twenty subjects took part in the study. They were divided equally into two groups: one consisting of young (22 to 27 years old) listeners with pure-tone detection thresholds <=20 dB HL (per octave from 0.5 to 8 kHz), and one consisting of older subjects (48 to 61 years old), most of whom (7 out of 10) had a slight hearing loss at high frequencies (mean absolute threshold between 4 and 8 kHz = 20.4 dB HL versus 2.6 dB HL in the first group).Each group contained as many female subjects than male subjects.

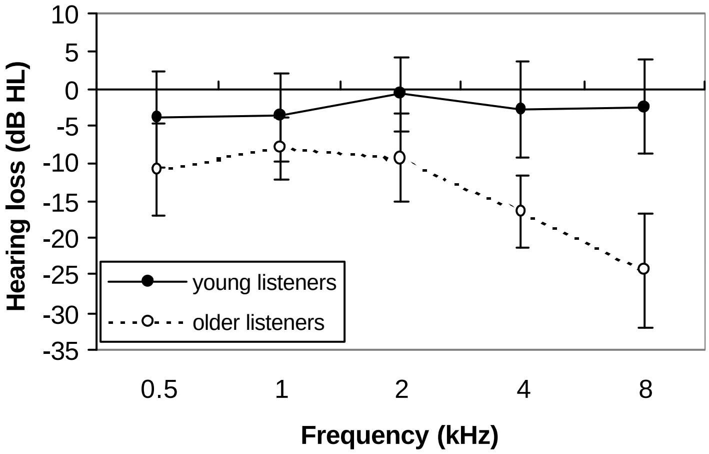

Figure 1

II.2. Stimuli

The stimuli were VCV (vowel consonant vowel) signals. On each presentation, the vowels were chosen randomly among a set of three (/a/,/i/ and /u/) and the consonant among a set of 17 consonants of the French language. This led to 51 different stimuli. Each stimulus was pronounced 4 times by 4 different speakers (2 male, 2 female), leading to a corpus of a total of 816 stimuli. Acquisitions were made on a Pentium computer equipped with a 16-bit sound card using a sampling frequency of 44.1 kHz; the stimuli were stored on an audio-CD. Signals were filtered between 100 and 7750 Hz (IIR digital filter, 4^{th} order). The perceptual tests involved the presentation of 1000 stimuli, drawn at random from the above corpus, to each subject. Before presentation, the signal was added with noise in 5 adjacent frequency bands, which had the following upper and lower cutoff frequencies 100-250, 250-750, 750-1750 Hz, 1750-3750 Hz, and 3750-7750 Hz. The bands were numbered consecutively, from 1 to 5, by order of increasing frequency. The level of the noise in each band was chosen pseudo-randomly in such a way that the signal-to-noise within the considered band was comprised between a 24 dB range, in 2-dB steps. The signal-to-noise ratios in the different bands were independent.

II.3. Procedure

Signal processing and stimulus presentation was achieved using a dedicated software on a Pentium III 350 MHz computer. Stimuli were presented at a 44.1 kHz sampling rate via a Roland UA30 soundcard connected to circumaural Sennheiser HD 265 Linear II headphones. Subjects were seated in a quiet room and their responses after each stimulus presentation were entered by the experimenter. The experiment began with a familiarization regimen which had two aims: familiarize the subjects with the stimuli and procedure and collect preliminary data that were used to set the nominal signal-to-noise ratio around which the actual signal-to-noise ratios varied in the considered subject during the experiment proper. This nominal SNR, which determined the difficulty of the test, was set so as to produce an error rate comprised between 25% and 35% in each subject. Once the test was finished, weights for each band were obtained by computing multiple correlation coefficients between the SNRs in the different bands and the binary identification variable (correct-incorrect) (Richards and Zhu, 1994).

III. RESULTS

III.1. General shape and reproducibility of the weighting functions

Figure 2 shows the average frequency weighting functions measured in the 20 subjects, sorted by test session (test and re-test) and by ear (left and right). These data were analysed using a four-way ANOVA, with the relative weights as dependent variable, the ear (left or right) and frequency bands (1 to 5) as within-subject factors, and the testing order (left ear first, right ear second, or the converse) as across-subjects factors. No statistically-significant difference was observed between the two sessions ($F(1,18)=1.014$, $p=0.327$), nor between the two ears ($F(1,18)=0.189$, $p=0.669$); no interaction between these two factors was obtained either ($F(1,18)=1.755$, $p=0.395$). The frequency-band factor had a significant main effect ($F(4,72)=42.753$, $p<0.001$). Planned comparisons revealed that the weights were significantly larger on band 4 than in any other band ($p<0.001$).

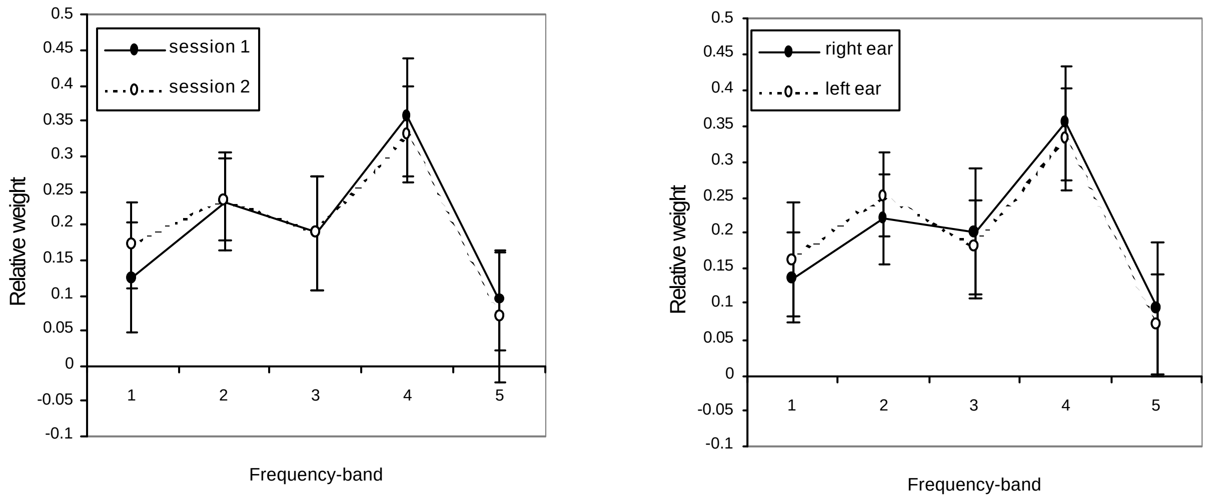

Figure 2

III.2. Intersubject differences

As indicated by the rather error bars in Figure 2, the weights, whilst largely similar across the two ears and test sessions, varied substantially across subjects. In order to try and understand the source of this across-subjects variability, we looked at different things. First of all, we considered the influence of the sex and age of the subjects on the measured weights. Figure 3 illustrates the influence of these two factors. Introducing these two factors in the ANOVA, we found no effect of sex (P*SEXE, $F(4,13)=0.547$, $p=0.705$) but a significant effect of age (P*AGE, $F(4,13)=6.271$, $p<0.01$). Post-hoc comparisons revealed that older subjects put significantly more weight than younger subjects on band 3, and at the same time ($t=-3.512$, $p<0.01$ (Bonferroni adjusted probability), $df=34.6$), significantly less weight on band 4 ($t=2.930$, $p<0.05$ (Bonferroni adjusted probability), $df=33.8$).

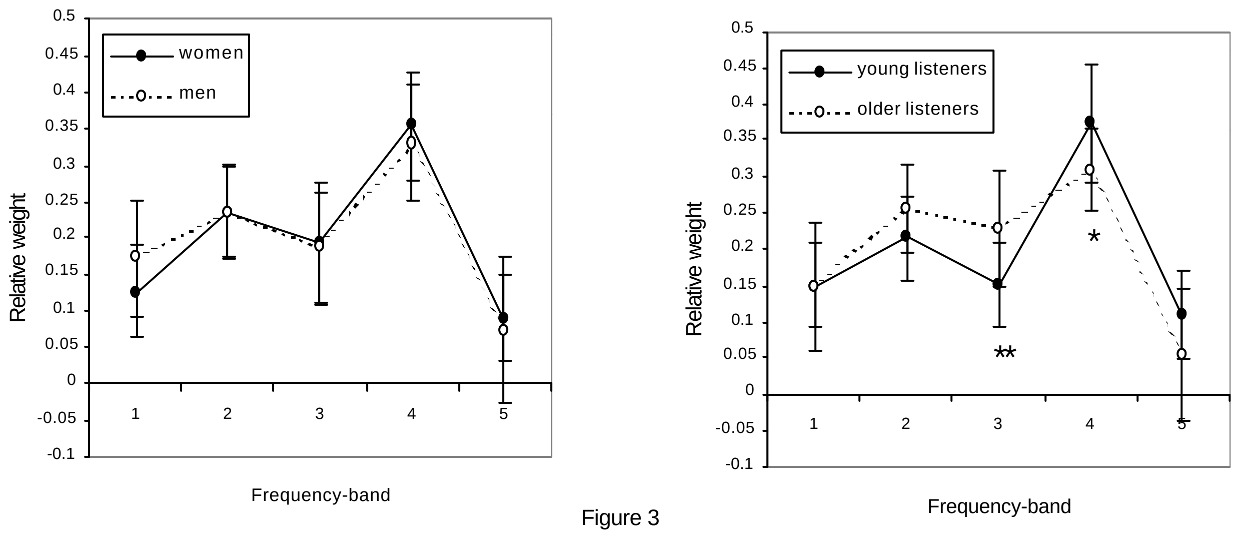

Figure 3

A possible hidden factor of the observed differences in weighting functions between young and older listeners was that, although all subjects had normal hearing for their age, the older listeners had on average slightly elevated hearing thresholds toward high frequencies. This difference is illustrated in Figure 1. In order to test this possibility, we computed the correlations between frequency weights and pure-tone hearing thresholds at the different test frequencies. The resulting correlation coefficients are shown in Table I. (Note that two subjects, whose absolute thresholds were not normal for their age, and who were therefore not included at first in the study, were however included in the calculation of these correlation coefficients).

	P1	**P2**	**P3**	**P4**	**P5**
500 Hz	*0.24406074*	*-0.06564324*	**-0.41215912**	*0.04801322*	*0.17632693*
1 kHz	*0.1753342*	*-0.03389003*	**-0.34916529**	*0.12955877*	*0.07889341*
2 kHz	*0.04470082*	*-0.17557732*	*-0.16716641*	*0.14979594*	*0.11530457*
4 kHz	*-0.03466038*	*-0.18381223*	**-0.29909548**	**0.36183619**	*0.12183702*
8 kHz	*0.03101364*	*-0.2883872*	*-0.27937393*	**0.32886674**	*0.15563648*

Table 1

Weak but significant (at the 0.05 level) correlations (indicated by bold characters in the Table) were observed between the absolute pure tone thresholds and the perceptual weights for bands 3 and 4. Negative correlations were observed between band-3 weights and hearing loss at 500 Hz ($p<0.01$), 1kHz, and 4kHz ($p<0.05$). Positive correlations were observed between band-4 weights and hearing loss at 4 and 8 kHz ($p<0.05$). This suggests that hearing loss at high frequencies, even if only slight, contributes to reduce the importance of band 4 for speech understanding.

IV. Commentary

A general finding of the present study, which can be paralleled with the earlier results of Doherty and Turner (1996) and Turner et al. (1998), consists in the jagged shape of the frequency weighting functions. Turner and colleagues attributed this result to a self-masking effect related to upward spread of masking: spectral peaks corresponding to formants at a low frequency mask peaks located just above in frequency. In order to test this interpretation, we computed the auditory excitation patterns of our stimuli using the model devised by Moore & Glasberg (1990). Figure 4 shows the long-term spectrum and simulated auditory excitation patterns averaged across all stimuli. Generally, spectral peaks that were present in the spectra were also present on the AEPs. This outcome appears to invalidate an explanation based on the upward spread of masking.

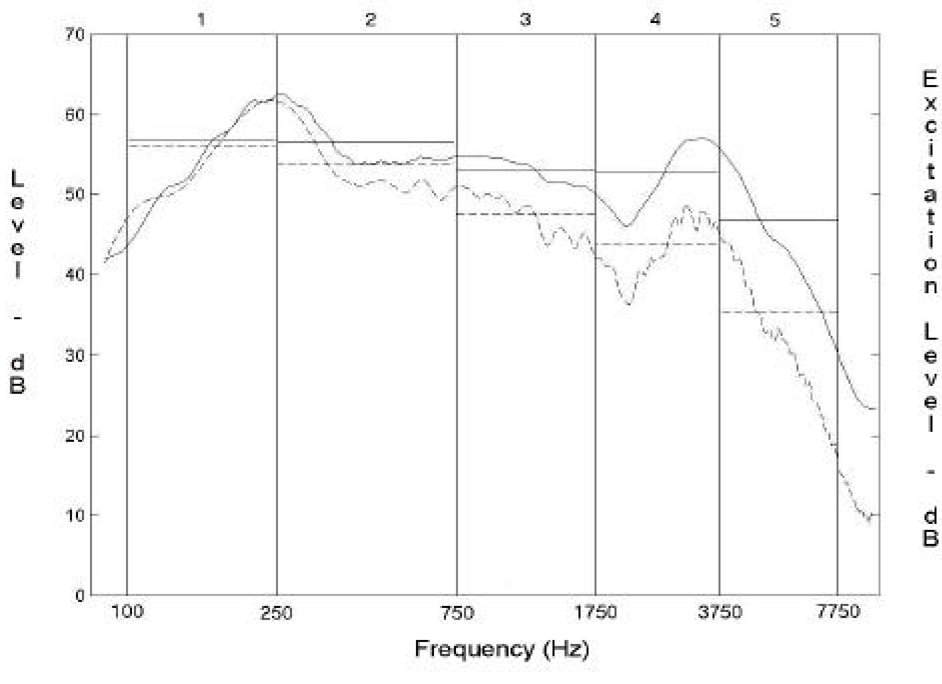

Figure 4

As suggested by several earlier studies (e.g. Shannon et al., 1995), temporal envelope cues may play a crucial role in speech recognition. Consequently, another idea to explain the particular shape of the frequency weighting functions involves quantifying the amount of temporal envelope fluctuation conveyed by each frequency band. Figure 5 shows envelope fluctuations (computed as the standard deviation of the mean envelope amplitude obtained using the Hilbert Transform) at the output of the different frequency bands used here, for three different degrees of amplitude compression (k=1/3, 0.6 or 1 in $y=x^k$). The third frequency band was found to contain generally less envelope fluctuations than the other bands. Maximal envelope fluctuations were found in the 2nd and 4th frequency bands, which are also those to which subjects attributed the largest perceptual weights for speech recognition on average.

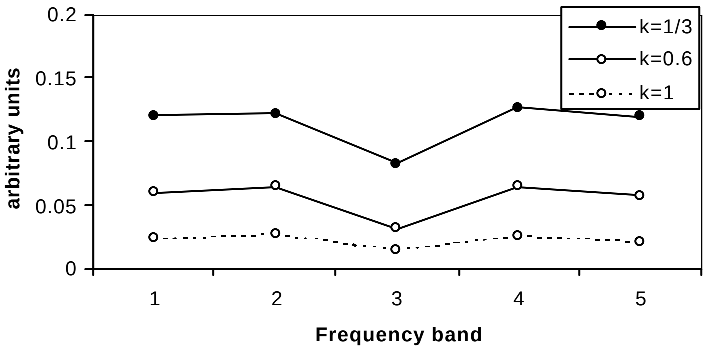

Figure 5

V. DISCUSSION

The results of this study indicate, firstly, that although they are substantially variable across subjects, frequency-weighting functions for speech are on average very similar between the two ears and relatively constant over time. Secondly, subjects appear to rely mostly on the fourth frequency band (corresponding to frequencies between 1750 and 3750 Hz), and to a smaller extent, to the second frequency band (250 – 750 Hz), in order to recognize speech. The weights for the two extreme frequency bands i.e., the lowest and the highest, which encompassed frequencies between 100 and 250 Hz and 3750 and 7750 Hz, were on average very small and in most cases non significantly different from 0. Thirdly, a significant difference was observed between young and older listeners. These differences were significantly correlated with hearing thresholds at 500 Hz, 1 kHz, 4 kHz and 8 kHz. Finally, a relationship was found between the frequency weighting functions and the temporal energy fluctuations in the different frequency bands.

REFERENCES

[1]. ANSI (1997), ANSI S3.5-1997, (ANSI, New York).

[2]. B. G. Berg, "Analysis of weights in multiple observation tasks",J. Acoust. Soc. Am. **86**, 1743-1746 (1989).

[3]. B. G. Berg, D. M. Green, "Observer efficiency and weights in a multiple observation task", J. Acoust. Soc. Am. **88**, 758-766 (1990).

[4]. K. A. Doherty, R. A. Lutfi, "Spectral weights for overall level discrimination in listeners with sensorineural hearing loss", J. Acoust. Soc. Am. **99**, 1053-1058 (1996).

[5]. K. A. Doherty, C. W. Turner, "Use of a correlationel method to estimate a listener's weighting function for speech" J. Acoust. Soc. Am. **100**, 3769-3773 (1996).

[6]. R. P. Lippman, "Intelligibility of bandpass speech: Effects of truncation or removal of transition bands", IEEE Trans. Speech Audio. Process. **4**, 66-69 (1996).

[7]. R. A. Lutfi, "Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks", J. Acoust. Soc. Am. **97**, 1333-1334 (1995).

[8]. M. A. Mehr , C. W. Turner, A. Parkinson, "Channel weights for speech recognition in cochlear implant users", J. Acoust. Soc. Am. **104**, (2001).

[9].B.R.Glasberg, B.C.J.Moore, "Derivation of auditory filter shapes from notched-noise data", Hear.res. **47**, 103-138 (1990).

[10]. C. V. Pavlovic, G. A. Studebaker et R. L. Sherbecoe, "An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals" J. Acoust. Soc. Am. **80**, 50-57 (1986).

[11].V. M. Richards, S. Zhu, "Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients, J. Acoust. Soc. Am. **95**, 423-434 (1993).

[12].R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, M. Ekelid, "Speech recognition with primarily temporal cues", Science **270**, 303-304 (1995).

[13].C. W. Turner, B.J. Kw on, C. Tanaka, J. Knapp, J.L. Hubbartt et Karen..A. Doherty, "Frequency weighting functions for broadband speech as estimated by a correlational method", J. Acoust. Soc. Am. **104**, 1580-1585 (1998).