2pPPb. Talker effects in speech band importance functions

Eric Healy*, Sarah E. Yoho, Carla L. Youngdahl and Frederic Apoux

*Corresponding author’s address: Speech and Hearing Science, The Ohio State University, 1070 Carmack Rd, Columbus, OH 43210, healy.66@osu.edu

The literature is somewhat mixed with regard to the influence of (a) the particular speech material (e.g., sentences or words) versus (b) the particular talker used to create the recordings, on band-importance function (BIF) shape. One possibility is that previous techniques for creating BIFs are not sensitive enough to reveal these influences. In the current investigation, the role of talkers was examined using the compound technique for creating BIFs. This technique was developed to account for the multitude of synergistic and redundant interactions that take place among various speech frequencies. The resulting functions display a complex microstructure, in which the importance of adjacent bands can differ substantially. It was found that the microstructure could be traced to acoustic aspects of the particular talkers employed. Further, BIFs for IEEE sentences based on ten-talker recordings displayed less microstructure and were therefore smoother than BIFs based on one such talker. These results together suggest that the compound technique is sensitive enough to reveal acoustic aspects of the particular talker employed. It is further suggested that multiple talkers, rather than smoothing of the functions, be used if the goal is to describe speech more generally.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

The Speech Intelligibility Index (SII, ANSI, 1997) provides a method for estimating intelligibility based on acoustic measurement of speech and noise level. At the heart of the SII is the band-importance function, which represents the proportion of total speech information provided by each spectral region. The SII value is simply the sum of these individual importance values, each scaled to reflect its availability in a particular environment.

In addition to this practical significance, the SII also reflects much of our knowledge of speech cues and their use. However, one criticism of the SII is that it cannot adequately account for the multitude of synergistic and redundant interactions that take place across speech frequencies (e.g., Breeuwer and Plomp, 1984, 1985; Warren et al., 1995; Lippmann, 1996; Müsch and Buus, 2001; Healy and Warren, 2003; Healy and Bacon, 2007). This is because the traditional method for calculating band importance involves successive low-pass and high-pass filtering. The importance of a speech band is then determined by the difference in performance as that band is eliminated. The limitation of this technique is that a given speech band is assessed only when information either below it or above it is entirely intact. But we are now more aware that the contribution of a speech band may depend largely on the extent to which it interacts with information elsewhere in the spectrum, as the importance of a given band may change dramatically as information elsewhere in the spectrum changes.

To address this limitation, Apoux and Healy (2012) developed the “compound method” of assessing band importance. It is similar to the traditional method in that the importance of a speech band is determined by the reduction in performance when that band is eliminated. In the compound method a target speech band is presented along with n “other” speech bands, such that only a subset of the total number of speech bands is present in any given trial. Critically, the locations of the n other speech bands is determined randomly between each trial. As a result, the target speech band is assessed as information elsewhere in the spectrum varies.

Healy, Yoho, and Apoux (2013) measured band importance for the CID W-22 words and the SPIN sentences, using recordings identical to those on which the SII functions are based. Two main characteristics are apparent in these data, which are shown in Fig. 1: (1) The compound-method functions differ from the SII functions, primarily due to the existence of a pronounced “microstructure,” in which the importance of one band often differs substantially from that of adjacent bands, and (2) the functions for words differs substantially from that for sentences. One interpretation for both these phenomena is that the compound method is highly sensitive to the particular acoustic characteristics of the talkers used to create the speech recordings. The upper panels of Fig. 1 show that the observed microstructure does not simply reflect noise in the data, as it is consistent across independent groups of listeners.

If microstructure in the band-importance function results from an increased sensitivity to acoustic aspects of the particular talker employed, then maximum microstructure should be observed when a single-talker recording is used and reduced microstructure should be observed when a multiple-talker recording is employed. The current study examines this role of talker in band-importance functions created using the compound method by comparing functions for identical speech materials using single- versus multitalker recordings.
FIGURE 1. Band-importance functions created using the compound method. The left panels (a and c) show functions for the CID W-22 words, and the right panels (b and d) show functions for the SPIN sentences. The upper panels display functions based on data from the first 10, 15, and 20 subjects run, and from the last 10 subjects run. The lower panels display functions based on data from all 20 subjects, along with functions from the ANSI standard SII, for the identical speech recordings. From Healy, Yoho, and Apoux (2013).

METHOD

Band-importance functions were generated using the 21 spectral bands specified in the SII and the procedures employed by Healy, Yoho, and Apoux (2013). Because details of those procedures are reported there, the following is focused on methodology unique to the current study.

Subjects

A total of 60 normal-hearing listeners participated (20 dB HL or better from 250 to 8000 Hz; ANSI 2004, 2010). They ranged in age from 19 to 37 years (mean = 21.9) and 55 were female. There were recruited from undergraduate courses at The Ohio State University and received course credit or a monetary incentive for participating. None had prior exposure to the sentences used.
Stimuli and Procedure

Two hundred and eighty sentences, plus 20 for practice, were selected from the IEEE database (IEEE, 1969). In a single-talker condition, a single male-voice recording was used. In a multi-talker condition, recordings from 10 talkers (half from each gender) were employed. The single male talker was one of the 10 multi talkers. The recordings were divided into 21 spectral bands using high-order filtering that produced extremely steep filter slopes. The subjects were divided randomly and equally into three groups and each group was assigned to one of three target-band sets (bands 1 - 7, 8 - 14, or 15 - 21). The target band was presented along with four other bands, having spectral locations selected randomly for each band-present/band-absent trial pair. Each listener heard 140 sentences in the single-talker condition (10 sentences x 1 talker x 2[target band present/band absent] x 7 spectral bands) and 140 sentences in the multi-talker condition (1 sentence x 10 talkers x 2[target band present/band absent] x 7 spectral bands). Thus, the single- versus multi-talker comparison was performed within subjects. The following randomization/balancing was performed: (i) Half of the subjects heard the single-talker condition followed by the multi-talker condition, and the other half heard the opposite order of conditions. (ii) The conditions were blocked by target band, such that conditions within one target band were completed prior to moving on to the next. (iii) Band-present/band-absent conditions were paired and presented in random-order succession. (iv) The sentence-to-condition correspondence was randomized for each listener. And (v) the talker-to-sentence correspondence was randomized in the multi-talker condition. Testing began with a brief practice consisting of 20 multi-talker sentences, five heard as all 21 bands, five heard as 11 randomly-selected bands, and ten heard as 4 randomly-selected bands. Listeners were tested individually in a double-walled booth, seated with an experimenter who controlled the presentation of sentences and recorded responses.

RESULTS

Figure 2 displays band-importance functions based on the single- and multi-talker recordings. As predicted, the function representing the single talker is more irregular (has more microstructure) and is less smooth than that representing the multi talkers. Regression analysis indicated that the fit provided by a quadratic equation was poorer for the single-talker function (R^2 = 0.60) than for the multi-talker function (R^2 = 0.73). The fact that a greater proportion of variance was accounted for in the multi-talker condition reflects the smoothness of that function.

FIGURE 2. Band-importance functions based on the IEEE sentences and created using the compound method. The filled symbols represent a condition in which a single male talker produced all sentences and the open symbols represent a condition in which a group of ten talkers produced the sentences.
DISCUSSION

Healy, Yoho, and Apoux (2013) described two effects that could potentially influence the shape of the band-importance function. The first is a “speech-material” effect and involves differences in the shape of the function due to differences in the nature of the speech materials (e.g., words versus sentences). The existence of this effect is supported by the fact that the SII has differently shaped band-importance functions for different speech tests, and that a common assumption seems to exist that importance functions are generally different for different speech materials. It is reasonable to assume that the type of speech material will have some effect on function shape -- certainly consonant functions (e.g., aCa) might be expected to differ from vowel functions (e.g., hVd). However, it might also be reasonable to expect that this influence will diminish once phonetic composition becomes sufficiently complex. For instance, there should be little difference between the low-context SPIN sentences and the W-22 words in a carrier phrase. Indeed, both sets possess substantial phonetic complexity in the final scoring words, neither possesses semantic predictability, and both possess coarticulation. The only difference is that the carrier phrase preceding the final word differs across sentences in the SPIN test, whereas it is constant in the W-22 test. Taken together, it is suggested that the influence of speech material on the band-importance function should be limited.

The second effect described by Healy, Yoho, and Apoux (2013) was termed a “talker effect.” In contrast to the speech-material effect, there is reason to believe that the compound method is highly sensitive to acoustic characteristics of the particular talker used to create the speech recordings -- Reasonable correspondence was observed between prominent peaks in the band-importance functions and the frequencies of the individual talkers’ first three formants in Healy, Yoho, and Apoux (2013). The current experiment supports the existence of a talker effect by indicating that microstructure is reduced when a single talker is replaced by a multi-talker set. This result can be interpreted as reflecting the fact that the acoustic characteristics of the talkers employed average across a range of frequencies in the multi-talker condition, eliminating the pronounced microstructure in the resulting functions. The fact that SII band-importance functions have different shapes for different speech materials may potentially be attributed not to differences in speech materials, but instead to the fact that these different materials were created using different talkers.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Institute on Deafness and other Communication Disorders (Grant No. DC08594 to EWH and DC09892 to FA).

REFERENCES