Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech

Thomas Baer and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 23 August 1993; revised 29 November 1993; accepted 2 December 1993)

In a previous study [T. Baer and B. C. J. Moore, J. Acoust. Soc. Am. **94**, 1229–1241 (1993)], a spectral smearing technique was used to simulate some of the effects of impaired frequency selectivity so as to assess its influence on speech intelligibility. Results showed that spectral smearing to simulate broadening of the auditory filters by a factor of 3 or 6 had little effect on the intelligibility of speech in quiet but had a large effect on the intelligibility of speech in noise. The present study examines the effect of spectral smearing on the intelligibility of speech in the presence of a single interfering talker. The results were generally consistent with those of the previous study, suggesting that impaired frequency selectivity contributes significantly to the problems experienced by people with cochlear hearing loss when they listen to speech in the presence of interfering sounds.

PACS numbers: 43.71.Ky, 43.71.Es, 43.71.Gv, 43.66.Dc

INTRODUCTION

It is well known that people with sensorineural hearing losses experience difficulties with the perception of speech in noise. Listeners with moderate impairments usually have little or no difficulty understanding speech in quiet, but they show reduced intelligibility at signal-to-noise ratios that cause few problems for normally hearing listeners. The reasons for the difficulties with speech in noise remain somewhat unclear, but it is possible that impairment of frequency selectivity plays a part. In a recent study (Baer and Moore, 1993—hereafter B&M), a spectral smearing technique was used to simulate some of the effects of impaired frequency selectivity so as to assess its influence on speech intelligibility. The use of such a simulation is motivated by the fact that reduced frequency selectivity is highly correlated with other deficits associated with cochlear hearing loss, and thus it remains uncertain to what extent it contributes to the communication difficulties of the hearing impaired. Simulation of impaired frequency selectivity allows it to be isolated as a possible factor affecting speech intelligibility. This issue is discussed in detail in B&M.

In the B&M study, speech stimuli in quiet and in noise were processed and presented to normal subjects. The processing was designed to evoke, in the normal subjects, auditory excitation patterns resembling those that would be evoked by the unprocessed stimuli in subjects with broadened auditory filters. Results showed that spectral smearing to simulate broadening of normal auditory filters by a factor of 3 or 6 had little effect on the intelligibility of speech in quiet, but had a large effect on the intelligibility of speech in noise at speech-to-noise ratios of 0 and -3 dB. Unsmeared speech at these speech-to-noise ratios was highly intelligible. The two degrees of broadening are in the range measured for people with mild-to-moderate and moderate-to-severe cochlear hearing loss, respectively. The pattern of results is thus consistent with the view that impaired frequency selectivity contributes significantly to the speech-perception problems of the hearing impaired.

Similar results were obtained by ter Keurs *et al.* (1992, 1993). They studied the effects of spectral envelope smearing, using methods similar to those of B&M but differing in several respects (see B&M). Where comparisons were possible, there was broad agreement between their results and those of B&M, despite differences in processing technique, speech material, and measures of performance.

Ter Keurs *et al.* (1993) also examined the effects of smearing when the interfering signal was speech rather than noise. They measured SRTs as a function of smearing bandwidth for both male and female speech in the presence of both speech-shaped noise and speech from a single interfering talker. For both types of masker, SRTs increased when the smearing bandwidth was increased. For unsmeared speech, SRTs with speech maskers were 5–7 dB lower than for corresponding noise maskers. This difference decreased as smearing bandwidth increased up to two octaves (the largest value used). Hence, the effect of spectral smearing on SRTs was greater for the speech masker than for the noise masker.

Spectral smearing might be predicted to have a smaller effect for a speech masker than for a noise masker, for the following reason. It is known that, for normally hearing subjects, speech can be understood at lower speech-to-background ratios when the background is a single talker than when it is speech-shaped noise (Duquesnoy, 1983; Festen and Plomp, 1990; ter Keurs *et al.*, 1993), although this is not always true for hearing-impaired subjects (Duquesnoy, 1983; Festen and Plomp, 1990; Moore *et al.*, 1991). The advantage gained by normal listeners with a single competing talker has usually been attributed to the ability to "listen in the dips" of the competing speech, when the signal-to-noise ratio is high. In this context, "dips" can be of two types; temporal and spectral. Speech-shaped noise has a relatively smooth spectral envelope that fluctuates little over time, while speech fluctuates in both dimensions. Our previous study showed that the effects of spectral smearing were small at high signal-to-noise ratios (or in quiet). Spectral smearing will tend to "fill in" spectral dips, but not temporal dips. If listening in the temporal dips accounts for the difference between speech and noise maskers in the unsmeared condition, spectral smearing should have smaller effects when a single competing talker is used. However, this prediction is at odds with the results of ter Keurs *et al.* (1993).

In our previous study (B&M) we used a continuous speech-shaped noise as the interfering sound. The present study was designed to extend the generality of conclusions based on the previous study by examining the effects of using speech as the interfering sound. Although ter Keurs *et al.* (1993) have already addressed this area of research, there are significant differences between their methods and ours. Most significantly, their practice of processing the target and interfering sounds separately and then combining them may not be very realistic, since the real auditory system always has to operate with the mixture of the two. In our simulation, the target and interfering speech are mixed prior to spectral smearing. Hence, we considered it worthwhile to reexamine the issue by measuring the intelligibility of spectrally smeared and unsmeared stimuli containing target speech in the presence of speech from a single competing talker.

I. METHOD

The method for processing the stimuli is described in detail in B&M. The most significant features are that short-term spectra are calculated for overlapping waveform segments and smearing is performed in the spectral domain; the aim is to produce excitation patterns in the normal subjects similar to those that would be produced in subjects with widened auditory filters. The technique was evaluated by B&M using synthetic speech stimuli, and the ideas underlying it were evaluated by Moore *et al.* (1992).

A. Stimuli and processing conditions

The target stimuli were the same as those in B&M, namely, the 18 Institute of Hearing Research Adaptive Sentence Lists (ASLs). The competing talker was a male speaker of standard Southern British English. His voice was recorded in a double-walled sound-attenuating chamber while reading a passage from a textbook, using a Sennheiser MKH 40P48 microphone and a DAT recorder. The recording was subsequently sampled as for the target sentences (16-kHz rate, 12-bit resolution). A short (about 80-s) sample was selected for further use, and the longest silent pauses were shortened with the aid of a waveform editing program. Segments of the resulting waveform, chosen at random, were then combined digitally with the target speech before processing.

Three separate sets of the ASL stimuli were processed: one in interfering speech at −3-dB signal-to-masker ratio (S/M), one at −6 dB S/M, and one at −9 dB S/M. These conditions were chosen to be in the range where normal listeners have little difficulty with these materials but hearing-impaired listeners experience reduced speech intelligibility (Plomp, 1978, 1986; Glasberg and Moore, 1989). Stimuli were spectrally smeared to simulate the effects of the same three degrees of broadening of the auditory filter as used in the B&M study: unbroadened, symmetrically broadened by a factor of 3, and symmetrically broadened by a factor of 6. These three conditions will be referred to as no smearing, ERB×3, and ERB×6, respectively. The two degrees of broadening were chosen to be representative of filter bandwidths associated with mild-to-moderate and moderate-to-severe hearing impairment (Glasberg and Moore, 1986). Each of the three masker conditions was combined with each of the three processing conditions, yielding nine conditions in all.

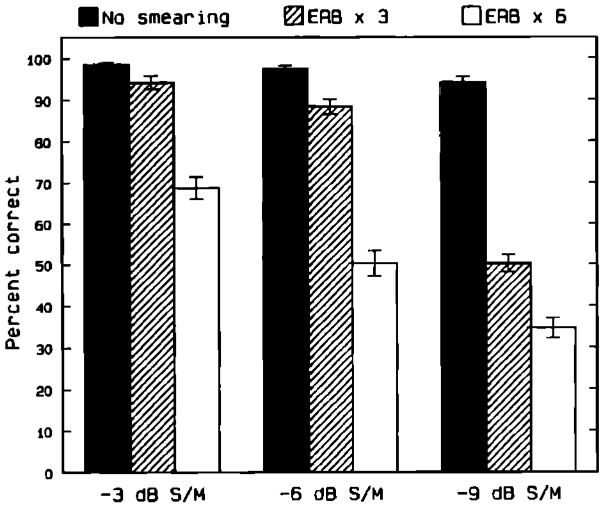

FIG. 1. The percentage of key words correct for three amounts of spectral smearing and three different S/M ratios. Error bars indicate ± one standard error.

The rms level of the target speech, before processing, was the same for all three masker conditions. Methods for performing the processing and preparing the stimulus tapes were as described in B&M, except that the silent interval between sentences was 7 rather than 9 s in duration.

B. Subjects and procedure

Nine subjects were used. All were young university students with no known hearing impairment. All were speakers of British English and were familiar with the dialect used in the ASL test. None had previously listened to the ASL sentences. They were paid for their participation. Experimental procedures were identical to those in the B&M study.

II. RESULTS

Performance was similar across subjects and across the two lists used for each subject and each condition. Hence, data were averaged across subjects and across the two lists. The results are shown in Fig. 1. Each group of bars represents averages for one of the three masker conditions.

Within a group, the three bars represent the three smearing conditions. The error bars indicate the standard error for each condition.

It is evident from the figure that scores tend to decrease both with decreasing signal-to-masker ratio and with increasing auditory-filter bandwidth, as was the case with noise masking (B&M). It is also clear that there was an interaction between these two effects. For each of the three masker-level conditions, scores decreased with increasing smearing bandwidth, but at the lowest masker level (−3 dB S/M), the decrease between no smearing and ERB×3 was relatively small. For the no smearing condition, performance was very good at all three masker levels. Scores were 98.6%, 97.5%, and 94.2% at −3, −6, and −9 dB, respectively. For the ERB×3 condition, scores decreased more rapidly with decreasing S/M (94.2%, 88.3%, and 80.1%, respectively). For the ERB×6 condition, performance was relatively low at all three masker levels (68.8%, 50.2%, and 34.7%). Although there was no condition without masking in the present experiment, the previous study showed that performance is nearly perfect in this condition for all three amounts of smearing.

An analysis of variance (ANOVA) was performed with factors S/M (3 values) and degree of smearing (3 values). Data were blocked by subject and list (order of testing) and sub-blocked by order (first or second list in a condition). The analyses were performed twice; once on the raw scores and once with an arcsine transformation, which gives the scores a more nearly Normal distribution. Results of the two analyses were similar. The patterns described below apply to both analyses, except where otherwise specified.

The main effects of S/M and degree of smearing and their interaction were all significant ($p<0.001$). The GENSTAT package used gave estimates of the standard errors of the differences between the mean scores for the different conditions. These standard errors were used to assess the significance of the differences between means (Lane *et al.*, 1987). For both main effects, all three means were significantly different from each other ($p<0.001$). For each masker level, all differences between smearing conditions were significant at the 5% level, although the difference between no-smearing and ERB×3 in the −3-dB S/M condition barely reached this level when the raw scores were analyzed. All of the other differences were highly significant ($p<0.001$). Within the no-smearing condition, the difference between −3 and −9 dB S/M was barely significant ($p<0.05$) when the raw scores were analyzed, but neither of these conditions was significantly different from −6 dB S/M. When the arcsine-transformed data were analyzed, the difference between −3 and −6 dB S/M remained insignificant, but the −9-dB S/M condition was different from the other two ($p<0.02$). For the other two smearing conditions, all differences were significant ($p<0.01$). In general, performance worsened when the masker level was increased and when the degree of smearing was increased; this effect was small for the lower masking levels and lower degrees of smearing, but was larger at the higher masker levels and higher degrees of smearing.

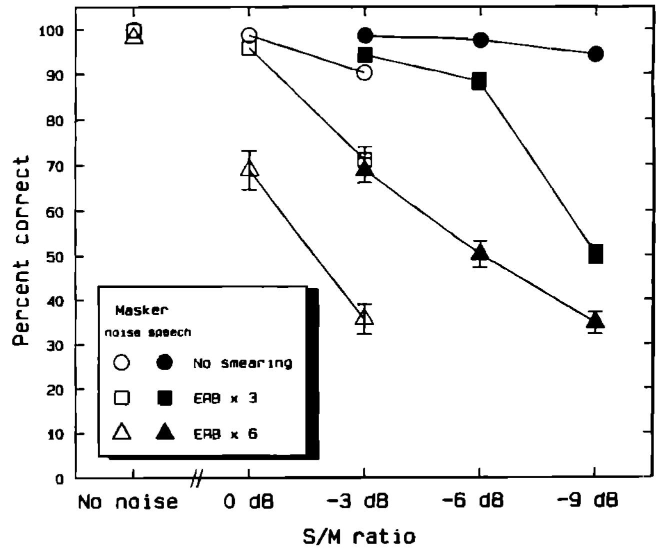

FIG. 2. Comparison of the results of the present experiment (filled symbols) with those of Baer and Moore (1993) (open symbols). The percentage of key words correct is plotted as a function of S/M ratio. Each symbol represents a different amount of spectral smearing.

III. DISCUSSION

The results are consistent with those of the previous study (B&M), in that performance was seriously degraded only when there was both interfering speech and simulated broadening of the auditory filters. With no smearing, even competing speech 9 dB more intense than the target speech had a relatively small effect on performance. However, simulated broadening of the auditory filters made performance significantly worse, this effect increasing dramatically as the signal-to-masker ratio decreased and as the degree of smearing increased. With broadening of the filters by a factor of three, typical of mild-to-moderate impairment, the most serious decline in performance was between −6 and −9 dB S/M. With broadening by a factor of 6, typical of moderate-to-severe impairment, performance declined more steadily from the no-noise condition (measured at 98.3% in the previous study) to the −3, −6, and −9-dB conditions. This pattern matches qualitatively that found in hearing impairment. People with normal hearing can tolerate a great deal of interfering sound, while those with hearing impairment, who are adversely affected by interfering sounds, still perform quite well in quiet.

To facilitate comparison across experiments, Fig. 2 replots the present results (filled symbols) along with the results of B&M, experiment 1 (open symbols), which used the same smearing conditions but a noise masker. Although the subjects were different in the two experiments, each experiment used a relatively large number of normally hearing subjects, and the standard errors of the measurements were low, so the comparison seems reasonable.

In the no-smearing condition (circles), scores are very high in all conditions, falling off only slightly at the most adverse S/M ratios. However, performance is better with a speech background at −9-dB S/M ratio than with a noise background at −3-dB S/M ratio. Thus the speech produces at least 6 dB less masking than the noise, consistent with earlier studies (Duquesnoy, 1983; Festen and Plomp, 1990; ter Keurs *et al.*, 1993). For ERB×3 smearing (squares), performance with a noise background at 0-dB S/M ratio is most nearly matched by performance with a speech background at −3-dB S/M ratio, while performance for the noise background at −3-dB S/M ratio is between that for the speech background at −6- and −9-dB S/M ratio. Thus, the difference in masking produced by speech and noise maskers is between 3 and 6 dB. For ERB×6 smearing (triangles), performance with a noise background at 0-dB S/M ratio is most nearly matched by performance with a speech background at −3-dB S/M ratio, while performance for the noise background at −3-dB S/M ratio is close to that for the speech background at −9-dB S/M ratio. Thus, again, the difference in masking appears to be between 3 and 6 dB. Overall, the results show that the difference in masking produced by the speech and noise maskers is somewhat less for spectrally smeared stimuli than for unsmeared stimuli, in agreement with the results of ter Keurs *et al.* (1993). Put another way, the deleterious effects of spectral smearing appear to be slightly greater for a speech masker than for a noise masker, at least when the results are expressed in terms of S/M ratios needed for a given level of performance.

The results do not support the expectation, described in the introduction, that the effects of spectral smearing would be smaller for a speech background than for a noise background. There are at least two ways of accounting for this. The first is that the ability of normally hearing subjects to understand speech at lower S/M ratios in a speech background than in a noise background depends more on listening in spectral dips than in temporal dips of the speech background. The spectral smearing adversely affects the ability to listen in spectral dips, but should not affect the ability to listen in temporal dips. A second possibility is that spectral smearing may have substantial effects even at high speech-to-masker ratios (as would occur during temporal dips in the masker) when listening conditions are more difficult, for example, when redundancy is reduced by masking of other parts of the speech (during peaks in the masker). Whatever the reason, it is clear that spectral smearing has substantial effects on speech intelligibility for both types of background sound, suggesting that it would play a significant role in many everyday situations.

Our finding that the difference in masking produced by the speech and noise maskers is somewhat less for spectrally smeared stimuli than for unsmeared stimuli is consistent with results found for hearing-impaired subjects; these subjects typically show small or even no differences in masking between a noise background and a background of a single talker (Duquesnoy, 1983; Festen and Plomp, 1990; Moore *et al.*, 1991).

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (UK). We thank Joe Alcántara, Sarah Hawkins, Andrew Oxenham, Michael Shailer, and Michael Stone for comments on an earlier version of this paper.

Baer, T., and Moore, B. C. J. (1993). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229–1241.

Duquesnoy, A. J. (1983). "Effect of a single interfering noise or speech source upon the perceived sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74**, 739–743.

Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Glasberg, B. R., and Moore, B. C. J. (1986). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

Glasberg, B. R., and Moore, B. C. J. (1989). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1–25.

Lane, P., Galwey, N., and Alvey, N. (1987). *Genstat 5. An Introduction* (Clarendon, Oxford).

Moore, B. C. J., Glasberg, B. R., and Simpson, A. (1992). "Evaluation of a method of simulating reduced frequency selectivity," J. Acoust. Soc. Am. **91**, 3402–3423.

Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (1991). "Optimization of a slow-acting automatic gain control system for use in hearing aids," Br. J. Audiol. **25**, 171–182.

Plomp, R. (1978). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Plomp, R. (1986). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," J. Speech Hear. Res. **29**, 146–154.

ter Keurs, M., Festen, J. M., and Plomp, R. (1992). "Effect of spectral envelope smearing on speech reception," J. Acoust. Soc. Am. **91**, 2872–2880.

ter Keurs, M., Festen, J. M., and Plomp, R. (1993). "Effect of spectral envelope smearing on speech reception. II," J. Acoust. Soc. Am. **93**, 1547–1552.