Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio

Joshua G. W. Bernstein[a)] and Douglas S. Brungart

Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC 20307

Normal-hearing listeners receive less benefit from momentary dips in the level of a fluctuating masker for speech processed to degrade spectral detail or temporal fine structure (TFS) than for unprocessed speech. This has been interpreted as evidence that the magnitude of the fluctuating-masker benefit (FMB) reflects the ability to resolve spectral detail and TFS. However, the FMB for degraded speech is typically measured at a higher signal-to-noise ratio (SNR) to yield performance similar to normal speech for the baseline (stationary-noise) condition. Because the FMB decreases with increasing SNR, this SNR difference might account for the reduction in FMB for degraded speech. In this study, the FMB for unprocessed and processed (TFS-removed or spectrally smeared) speech was measured in a paradigm that adjusts word-set size, rather than SNR, to equate stationary-noise performance across processing conditions. Compared at the same SNR and percent-correct level (but with different set sizes), processed and unprocessed stimuli yielded a similar FMB for four different fluctuating maskers (speech-modulated noise, one opposite-gender interfering talker, two same-gender interfering talkers, and 16-Hz interrupted noise). These results suggest that, for these maskers, spectral or TFS distortions do not directly impair the ability to benefit from momentary dips in masker level.
[DOI: 10.1121/1.3589440]

I. INTRODUCTION

Normal-hearing (NH) listeners generally demonstrate better intelligibility when target speech is presented in a fluctuating masker than when it is presented in a stationary-noise masker with the same long-term average spectrum (e.g., Festen and Plomp, 1990). This fluctuating-masker benefit (FMB) is thought to reflect the ability to make use of temporal or spectral gaps in the masker that would ordinarily be masked by a stationary noise. In contrast, hearing-impaired (HI) listeners generally show little or no improvement in intelligibility for fluctuating maskers (e.g., Carhart and Tillman, 1970; Festen and Plomp, 1990; Peters *et al.*, 1998; George *et al.*, 2006).

In part, the reduction of the FMB for HI listeners can be explained by limitations in audibility that prevent them from detecting some parts of the target signal that are revealed by dips in the masker (Takahashi and Bacon, 1992; Summers and Molis, 2004). Other suprathreshold psychoacoustic deficits might also limit the FMB for HI listeners. For example, the limited spectral (e.g., Glasberg and Moore, 1986) or temporal resolution (e.g., Oxenham and Moore, 1997; Nelson *et al.*, 2001) associated with hearing loss might reduce a listener's ability to make use of short-duration or narrow-bandwidth gaps in the masker to obtain additional target-speech information. A reduced ability to use temporal fine-structure (TFS) information (i.e., the fast temporal fluctuations in the stimulus waveform) may also accompany sensorineural hearing loss (Buss *et al.*, 2004; Lacher-Fougère and

Demany, 2005; Lorenzi *et al.*, 2006; Moore *et al.*, 2006; Hopkins and Moore, 2007, 2010; Hopkins *et al.*, 2008). It has been suggested that TFS information might signal the presence of the target sound within the amplitude dips of a fluctuating masker (Hopkins and Moore, 2009, 2010). Thus, a distorted encoding of TFS information for HI listeners could reduce their ability to benefit from masker fluctuations.

It is difficult to directly evaluate the role that spectral or temporal resolution or the use of TFS cues has on the FMB for HI listeners, because there is no way to directly compare the magnitude of the FMB for HI listeners in cases with and without these particular suprathreshold distortions caused by the hearing loss. One approach has been to relate speech and psychoacoustic measures across a group of HI listeners (e.g., Dubno *et al.*, 2003; George *et al.*, 2006; Jin and Nelson, 2006; Lorenzi *et al.*, 2006). However, with this approach, it is difficult to rule out the influence of audibility on the FMB for HI listeners. Another approach to investigate the possible role of suprathreshold deficits in limiting the FMB is to use signal processing to simulate individual suprathreshold deficits thought to be associated with hearing loss, and to determine the effect of this processing on the FMB for NH listeners. In particular, the effects of reduced frequency selectivity have been simulated using spectral smearing (ter Keurs *et al.*, 1993; Baer and Moore, 1994; Gnansia *et al.*, 2009), and the effects of impaired TFS processing have been simulated using tone- or noise-excited vocoding (Qin and Oxenham, 2003; Gnansia *et al.*, 2008, 2009; Hopkins and Moore, 2009). In each of these cases, the simulated suprathreshold distortions in the target speech resulted in a reduction in the FMB for NH listeners.

[a)]Author to whom correspondence should be addressed. Electronic mail: joshua.g.bernstein@us.army.mil

To the extent that these processing schemes accurately simulate psychoacoustic deficits experienced by HI listeners, these results support the idea that reduced frequency selectivity or ability to use TFS information underlie the reduced FMB observed for HI listeners. However, to conclusively determine the impact that degraded spectral and TFS cues have on the magnitude of the FMB, it is first necessary to account for the impact that changes in the overall signal-to-noise ratio (SNR) might have on the release from masking that occurs with a fluctuating masker. Most previous studies that have looked at the impact of simulated suprathreshold distortion on the FMB have calculated the FMB by comparing the speech reception threshold (SRT, i.e., the SNR required to reach a certain threshold level of performance) for stationary-noise maskers to the SRT for a fluctuating masker. Because the conditions with simulated suprathreshold distortions are more difficult overall than those with unprocessed speech, they tend to require substantially higher SNRs to reach the same level of performance in the baseline stationary-noise masker conditions. Consequently, the comparisons between the fluctuating and nonfluctuating masking conditions tend to be made at a substantially higher overall SNR value for the conditions that simulate degraded speech perception. For example, one octave of spectral smearing in the ter Keurs *et al.* (1993) study yielded about a 3-dB increase in the stationary-noise SRT.

This difference in baseline SNR may create a serious problem in interpreting the differences in FMB across the processed and unprocessed conditions, because recent results have shown that changes in SNR alone can result in a substantial change in the FMB that occurs with a target speech signal and a fluctuating masker. Specifically, these results have shown that the magnitude of the FMB systematically increases with decreasing SNR (Bernstein and Grant, 2009; Oxenham and Simonson, 2009). This systematic shift in the magnitude of the FMB corresponds to a difference in the slopes of psychometric functions relating SNR to intelligibility for stationary noise and for fluctuating maskers. A more gradual slope for the fluctuating-masker case yields a larger difference between the psychometric functions, and hence a larger apparent FMB, at lower SNR values. This difference in slopes—and the associated SNR dependence of the FMB—might derive from an interaction between masker level and the shape of the intensity-importance function describing the distribution of speech information across the dynamic range. The portion of the dynamic range unmasked by dips in the fluctuating-masker level might contribute more to overall intelligibility at lower SNRs than at higher SNRs (Freyman *et al.*, 2008; Bernstein and Grant, 2009), resulting in a larger FMB at low SNR values.

In order to fully control for the effect of SNR on the change in FMB that occurs when suprathreshold distortions are added to a speech signal, it is necessary to find a way to measure performance in the processed and unprocessed speech conditions for stationary-noise and fluctuating-masker conditions without varying the SNR values used in the baseline steady-state noise conditions. This goal is very difficult to achieve using the conventional technique of estimating speech intelligibility at fixed SNR, because of the difficulty in identifying an appropriate fixed SNR for stationary-noise and fluctuating-masker conditions. The difference in difficulty between the stationary and fluctuating-masker conditions is often so large that for a SNR that is sufficient to yield performance in the stationary-noise condition substantially above floor (chance) level, performance in the fluctuating-masker condition approaches its ceiling level. In this study, we attempted to address this problem by using changes in the size of the response set in a word-identification task (Miller *et al.*, 1951) to equate performance in the baseline processed and unprocessed speech conditions with a stationary-noise masker. It was hoped that this technique would allow us to evaluate the impact that reduced frequency selectivity or a reduced ability to use TFS information has on the ability to listen in the dips in a fluctuating masker without any possible confounding changes in FMB due to differences in the SNR for the baseline stationary-noise conditions.

II. EXPERIMENT 1: WORD IDENTIFICATION FOR PROCESSED STIMULI IN STATIONARY AND FLUCTUATING MASKERS

A. Methods

Word-identification performance was measured as a function of SNR for speech presented in stationary noise, a speech-modulated noise, or interfering speech from a talker opposite in gender from the target talker. The stimuli were presented in three ways: unprocessed, noise-vocoded to remove TFS information (Hopkins *et al.*, 2008), or spectrally smeared to simulate a reduction in frequency selectivity (Baer and Moore, 1993, 1994). For each of these three processing conditions, word-identification scores were collected in a closed-set paradigm, where subjects selected their responses from a set of 72 monosyllabic words presented on a touch screen. In addition, the stimuli in the unprocessed condition were presented in an open-set paradigm intended to reduce stationary-noise performance to be similar to that of the closed-set processed conditions. Target words were selected from a set of 1000 monosyllabic words and the subjects were required to type their responses on a keyboard.

1. Target speech materials

The target words were monosyllabic keywords chosen from the first 45 ten-word lists of sentences from the IEEE (1969) speech corpus. Each IEEE sentence contains four monosyllabic keywords, for a total of 1800 words in these 450 sentences. Removing duplicate words, homophones, and contractions (e.g., "don't") from this set yielded a total of 1000 unique target words. The target words were spoken in isolation by a single male talker (Grant and Seitz, 2000). The target words were recorded with a sampling frequency of 20 kHz, then upsampled to 24.414 kHz for playback through a Tucker-Davis System III digital/analog (D/A) converter.

2. Maskers

Three different maskers were chosen: a stationary speech-spectrum shaped noise, a speech-modulated noise,

and an interfering female talker. Long (94 s) samples of each masker were recorded on hard disk. For each stimulus trial, a short-duration masker was obtained by randomly selecting a segment of long-duration masker. The stationary-noise masker was spectrally shaped to match the long-term average spectrum of the target speech. The interfering-talker masker was derived from a recording of a female speaker of American English reading the "The Unfruitful Tree" by Freidrich Adolph Krummacher (translated from German). To remove pauses between words, the amplitude of the speech was calculated using a 30 ms moving average window. Segments that were more than 20 dB below the long-term average level of the speech for more than 150 ms were removed, with 2.5 ms raised-cosine ramps applied to the speech offset and onset on either side of the removed segment. The resulting speech masker waveform was then spectrally shaped to match the long-term average spectrum of the target speech. The speech-modulated noise masker was generated by modulating the speech-shaped stationary noise with envelopes derived from the interfering-talker masker as described by Festen and Plomp (1990). The interfering-talker masker was filtered (sixth-order Chebychev) into two bands (above and below 1 kHz). Envelopes were derived in each band via half-wave rectification and low-pass filtering (fourth-order Butterworth with a 40-Hz cutoff frequency). These envelopes were used to modulate the stationary-noise masker filtered into the same two bands. The resulting signals were rescaled to the original long-term average noise levels for each band (i.e., before the modulation was applied), and then summed together.

3. Stimuli

Target words were presented at a level of 60 dB sound pressure level (SPL). For each target-word presentation, a segment of the appropriate long-duration masker was chosen at random, adjusted in level to yield the desired SNR, and used to generate an acoustic mixture. The masker was ramped on 500 ms prior to the onset of the target speech and ramped off 250 ms after the offset of the target speech (20-ms raised-cosine ramps). In the processed conditions, this combined target and masker mixture was either noise vocoded to remove TFS information or spectrally smeared to simulate a reduction in frequency selectivity.

4. Signal processing

Vocoding was applied as described by Hopkins and Moore (2009), except that noise carriers were used instead of tone carriers to increase the negative impact of the processing on speech intelligibility. The poorer performance for noise carriers than for tone carriers might result from the spurious modulations produced by narrowband filtering of the noise carrier (Whitmal et al., 2007). The combined target and masker stimuli were filtered into 32 frequency bands, with center frequencies ranging from 100 to 10000 Hz. The bandwidth for each finite impulse response (FIR) filter was set to the equivalent rectangular bandwidth of the NH auditory filter at a low stimulus level, as defined by Glasberg and Moore (1990). This was done to limit the effect of the proc-

essing on the spectral information available to the auditory system after cochlear filtering. Time delays were applied to the output of each channel to compensate for the different phase shifts associated with the filtering process. The envelope was extracted from the signal in each channel by taking the absolute value of the analytic portion of the Hilbert transform. Each envelope was used to modulate a white noise, and the resulting signals were then filtered by the same FIR filters used to derive the stimulus envelope. The 32 signals were then combined and the resulting signal rescaled to the same overall preprocessing level.

Spectral smearing was applied as described by Baer and Moore (1993, 1994). Spectrograms of the input waveform were calculated in 5-ms time frames with 50% overlap. Broadening of auditory filters by a factor of four times the NH filter bandwidth was simulated by computing a weighted average of the magnitude spectrum at each frequency point based on the relationship between the shape of the broadened auditory filter and an estimate of the shape of the NH auditory filter. The factor of 4 was selected in pilot tests to yield a similar effect on speech intelligibility to the noise vocoding described previously. The resulting smeared magnitude spectrum in each time frame was then recombined with the original phase spectrum, such that processing mainly distorted the spectrum of the signal while minimizing the change to the stimulus TFS. The resultant spectrograms were converted back into time-overlapping waveform segments that were added together to produce the spectrally smeared stimulus. The resulting stimulus was rescaled to the same overall level as the combined target and masker before processing.

5. Apparatus

Prior to being presented to the listener, the stimuli were filtered with a 128-point minimum-phase FIR filter to compensate for the nonflat headphone magnitude response in the 125–8000 Hz frequency range. The resulting digital signal was sent to a D/A converter (TDT RP2.1) where it was stored in a buffer. A software command initiated digital-to-analog conversion of the stimulus, which was passed though a headphone buffer (TDT HB7) before being presented to a listener seated in a double-walled sound-attenuating chamber through the left earpiece of a Sennheiser HD580 headset.

6. Procedure

The closed-set conditions used 72-word lists that were selected at random from the larger 1000-word list. Possible responses were arranged in alphabetical order on virtual buttons on a touch screen in nine columns of eight words. Listeners were instructed to respond following each word presentation by pressing the button associated with their best guess among the response options. Listeners were allowed to change their response as many times as they wanted before confirming their choice by pressing a button at the bottom of the screen labeled "OK." Following the confirmation of their choice, listeners were given feedback whereby the button associated with the correct response lighted up before the next trial was presented.

For the 1000-word open-set conditions, no choices were listed on the screen. Listeners responded by typing their response on the keyboard. Only the letters, backspace, and enter keys were enabled on the keyboard. Corrections could be made using the backspace key. The listener confirmed their response by pressing the enter key on the keyboard or an OK button on the touch screen. Feedback was given following the response with the correct answer displayed on the screen. Listeners were told that alternative/homonym spellings would be marked as correct, even if the spelling entered by the user was different from the correct response displayed on the screen.

The experiment was divided into a series of blocks. Each block consisted of a set of training runs and/or test runs for one of the set-size conditions (72-word closed or 1000-word open). Each training or test run consisted of 10 consecutive trials where the masker type, processing type, and SNR remained fixed. The first block was an initial training block intended to familiarize the listener with the test procedure, masking conditions, and processing types. This block consisted of 18 training runs, with each run reflecting a particular combination of one of the various masker types, processing conditions, and SNRs (−6 or 0 dB). This initial training block was conducted using a 72-word set selected at random for each listener.

Following the initial training block, each listener then completed two 72-word blocks and two 1000-word blocks, in pseudorandom order, with one block for each set size completed before the second block for either set size began. For each 72-word block, a list of words was chosen at random from the larger (1000-word) set and then held fixed for the duration of the block. Each 72-word block began with a set of training runs to familiarize the listener with this particular list of words. One training run was presented for each combination of processing condition and masker type, for a total of nine training runs. For the training runs, the stationary-noise masker conditions were presented at a SNR of 3 dB and the fluctuating-masker conditions were presented at a SNR of 0 dB. This was followed by a set of 45 test runs, with one run for each combination of masker type and processing condition presented at five different SNRs (stationary-noise conditions: −24 to 0 dB in 6 dB steps; fluctuating-masker conditions: −9 to +3 dB in 3 dB steps). This resulted in 20 words presented to each listener for each combination of masker type, processing condition, and SNR, with 10 words from each of the two randomly selected 72-word lists.

The 1000-word open-set condition was intended to offset stationary-noise performance differences between processed and unprocessed conditions by increasing the difficulty of the unprocessed conditions. Therefore, only the unprocessed conditions were presented in the 1000-word case. The first 1000-word block began with of a set of training runs, intended to familiarize the listener with the different test procedure associated with the 1000-word conditions that involved entering responses via a keyboard. There were six training runs: two runs for each masker type, presented at a SNR of 3 dB (stationary-noise masker) or 0 dB (fluctuating maskers). This was followed by a set of 15 test runs, one for each combination of masker type and the same five SNRs

that were used for each masker type in the 72-word conditions. The second 1000-word block contained only the 15 test runs (with no additional training runs). In total, 20 test words were presented to each listener for each combination of masker type and SNR in the 1000-word conditions.

7. Listeners

Eight NH listeners (age range 32–60 yr, five female) participated in the experiment. All listeners had pure-tone audiometric thresholds of 20 dB hearing level (HL) or better in both ears for octave frequencies between 250 and 4000 Hz, as well as at 6000 Hz.

B. Results

Figure 1 plots the mean proportion of words correctly identified (±1 standard error across the eight NH listeners) for each masking, processing and set-size condition tested in the experiment. Figures 1(a)–1(d) group the data by common processing condition and set size. The same data are replotted in Figs. 1(e)–1(g) with the data grouped instead by common masker condition.
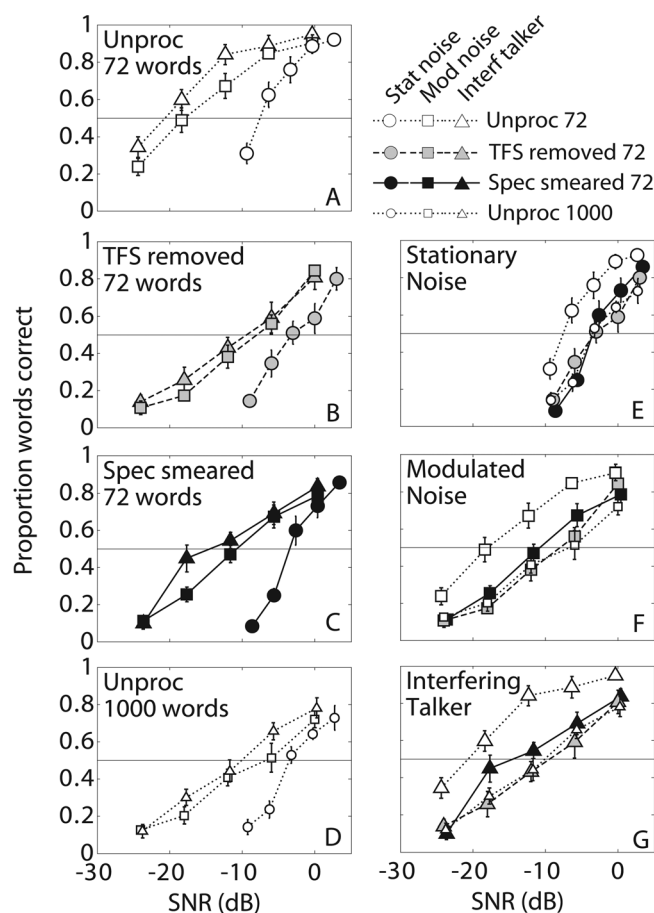

FIG. 1. Mean psychometric functions describing speech intelligibility as a function of SNR for experiment 1. (a)–(d) Conditions grouped by common processing conditions and word-set sizes to show relative performance among masker conditions. (e)–(g) The same data, but with the conditions grouped by common maskers to show relative performance among processing conditions and word-set sizes. The horizontal lines in each plot represent the 50% correct performance level used to derive the FMB estimates shown in Fig. 2. Error bars indicate ±1 standard error across listeners.
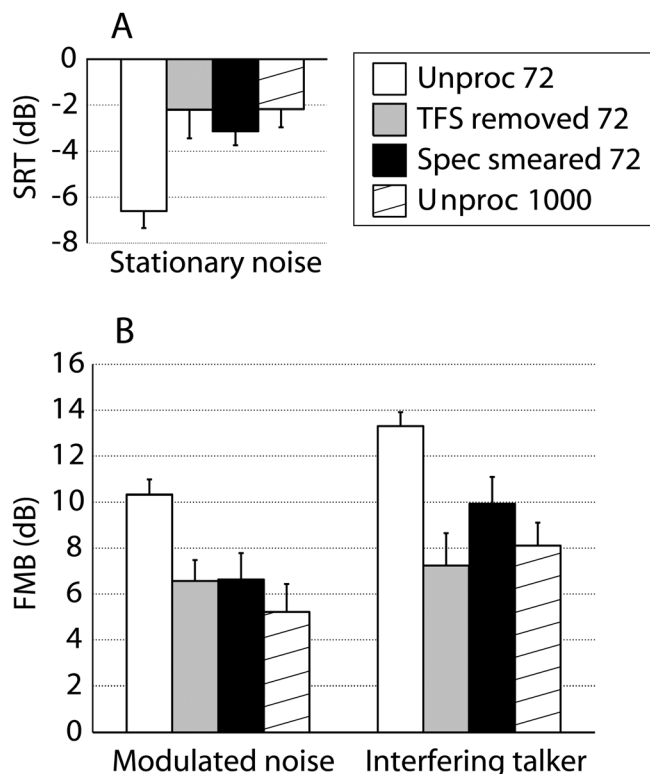
FIG. 2. (a) The mean SRT for the stationary-noise conditions in experiment 1 derived from the psychometric functions (Fig. 1) for individual listeners. (b) The mean FMB, calculated by subtracting the SRT for each fluctuating masker from the stationary-noise SRT. Error bars indicate ±1 standard error across listeners.

To estimate the magnitude of the benefit received from masker fluctuations, the SRT was estimated for each listener, processing condition, masker type and set size. A sigmoidal function with two free parameters (describing the slope and y-intercept) was fit to each psychometric function relating SNR to percent correct words identified. The fitted functions had fixed maximum (100% correct) and minimum values (chance performance: 1.4% correct for the 72-word conditions and 0% correct for the 1000-word conditions). The SRT was taken to be the SNR required in each condition to yield 50% correct performance. The FMB was calculated to be the SRT difference between each fluctuating-masker condition and the associated stationary-noise condition [i.e., the horizontal distance between the stationary- and fluctuating-masker curves in Figs. 1(a)–1(d)]. Group-mean SRTs for the stationary-noise conditions are shown in Fig. 2(a). More negative SRT values indicate better performance, with a lower SNR required to yield 50% correct word-identification performance. Mean FMBs for each of the fluctuating-masker conditions are shown in Fig. 2(b). More positive FMB values indicate a greater benefit from masker fluctuations relative to the associated stationary-noise condition.

In previous studies examining the effects of stimulus processing, the FMB was compared for unprocessed and processed conditions with the same speech identification task employed in both cases (ter Keurs et al., 1993; Baer and Moore, 1994; Qin and Oxenham, 2003; Gnansia et al., 2008, 2009; Hopkins and Moore, 2009). For a direct comparison of these studies, only the 72-word conditions are considered

first. For both fluctuating-masker types, the FMB was significantly reduced for both the TFS-removed and spectrally smeared conditions relative to the unprocessed condition. This effect is apparent in Figs. 1(a)–1(c), where the horizontal distance at the 50%-correct point (horizontal line) between stationary noise (circles) and each fluctuating masker (squares and triangles) is larger for the unprocessed condition [Fig. 1(a)] than for the processed conditions [Figs. 1(b) and 1(c)]. Figure 2(b) indicates that stimulus processing reduced the FMB by 3–6 dB, from 10–13 dB for unprocessed stimuli (white bars) to 6–10 dB for the processed stimuli (gray and black bars). To statistically examine the effect of stimulus processing on the FMB, a repeated-measures analysis of variance (RM-ANOVA) with two within-subject factors (fluctuating-masker type and processing condition) was computed on the FMB data for the 72-word conditions [Fig. 2(b), white, gray, and black bars]. (The reported degrees of freedom reflect a Huynh–Feldt correction for sphericity applied whenever necessary for all RM-ANOVA results presented in this manuscript.) A significant main effect of processing condition [$F(2,14) = 12.3$, $p < 0.005$] confirmed the negative impact of stimulus processing on the FMB. This result corroborates earlier findings of a reduction in the FMB for the processed conditions when performance differences in the stationary-noise condition are not controlled. There was also a significant main effect of fluctuating-masker type [$F(1,7) = 15.9$, $p < 0.01$], reflecting the larger FMB observed for the interfering-talker than for the modulated-noise condition.

The hypothesis investigated in this study is that the reduced FMB for the processed conditions might be attributable to the confounding effects of performance differences in the baseline stationary-noise condition. Results for the stationary-noise conditions are shown for the raw data in Fig. 1(e) and for the SRTs derived from these data in Fig. 2(a). A comparison of the processed and unprocessed 72-word conditions clearly indicates that spectral smearing and TFS removal negatively impacted speech-reception performance in stationary noise, reducing overall performance [Fig. 1(e), compare large white circles to large black and gray circles] and increasing the SRT [Fig. 2(a), compare white, black, and gray bars]. RM-ANOVAs computed on the raw 72-word stationary-noise data and corresponding SRT data confirmed this observation, with significant main effects of processing type in both cases [raw data: $F(1.65,11.5) = 29.1$, $p < 0.0005$; SRT data: $F(1.37,9.60) = 23.8$, $p < 0.005$]. Because the performance function is steeper for the stationary-noise condition than for the fluctuating-masker conditions [Figs. 1(a)–1(d)], the magnitude of the FMB (the horizontal distance between the curves) decreases with increasing SNR. A smaller FMB would therefore be expected in the processed conditions due to the higher SNR required to yield 50% correct performance in stationary noise, even if dip listening ability were not directly impacted by the processing.

To offset performance differences in the stationary-noise conditions, the unprocessed condition was also tested with the more difficult 1000-word open-response testing procedure. As would be expected, the increased difficulty of the open-set response procedure resulted in a substantial

reduction in performance in the unprocessed condition for all three masker types [compare Figs. 1(a) and 1(d)]. In the stationary-noise condition [Fig. 1(e)], the increased difficulty caused by the larger response set brought performance in the 1000-word unprocessed condition (small white circles) down to a level that was comparable to performance in the 72-word processed conditions (gray and black circles). A RM-ANOVA with two within-subjects factors [five SNRs and three processing conditions (72-word smeared, 72-word TFS-removed, and 1000-word unprocessed)] was computed on the stationary-noise data. There was a main effect of SNR [$F(4,28) = 140$, $p < 0.0005$], but not of processing type [$F(1.8,12.3) = 1.18$, $p = 0.33$] confirming that the set-size manipulation yielded similar stationary-noise performance for the processed and unprocessed conditions. There was a significant interaction between processing type and SNR [$F(7.1, 49.9) = 2.28$, $p = 0.042$], suggesting that performance was not precisely equalized at all SNRs. Nevertheless, *post hoc* tests showed no significant differences between 72-word processed and 1000-word unprocessed conditions for any SNR. The SRT data shown in Fig. 2(a) also shows that the set-size manipulation approximately equalized performance. A RM-ANOVA applied to the 72-word processed conditions (gray and black bars) and the 1000-word unprocessed condition (striped bar) indicated no significant main effect of processing type ($p = 0.38$).

The central question in this study is whether there is any performance deficit in the fluctuating-masker conditions once the stationary-noise performance has been equalized using the set-size manipulation. Figures 1(f) and 1(g) show that fluctuating-masker performance was not worse for the 72-word processed conditions (gray and black symbols) than for the 1000-word unprocessed conditions (small white symbols). In fact, performance was slightly better for the 72-word spectrally smeared conditions than for the 1000-word unprocessed conditions. A RM-ANOVA with three factors (fluctuating-masker type, SNR, and processing condition) was conducted on these fluctuating-masker data. There was a significant main effect of processing condition [$F(1.8,12.3) = 9.9$, $p = 0.003$], consistent with the better performance observed for the 72-word spectrally smeared than for the 1000-word unprocessed fluctuating-masker condition [Fig. 1(g)]. There was a significant main effect of fluctuating-masker type [$F(1,7) = 9.14$, $p = 0.019$], consistent with the generally better performance for an opposite-gender interfering-talker than for a speech-modulated noise condition observed here [compare triangles and squares in Figs. 1(b)–1(d)] and elsewhere (e.g., Festen and Plomp, 1990; Qin and Oxenham, 2003). There was also a significant main effect of SNR [$F(2.9,20.4) = 231$, $p < 0.0005$], reflecting the general improvement in performance with increasing SNR. There was a significant interaction between processing condition and SNR [$F(6.23,43.1) = 2.31$, $p = 0.049$], suggesting that the better performance observed for the 72-word spectrally smeared condition [Fig. 1(g)] did not occur at all SNRs. No other interactions were significant ($p > 0.05$).

To determine whether TFS removal or spectral smearing was responsible for the main effect of processing condition on fluctuating-masker performance [Figs. 1(f) and 1(g)], two

separate *post-hoc* three-factor RM-ANOVAs (processing condition, fluctuating-masker type, and SNR) were conducted comparing performance for each of these processing conditions to the unprocessed conditions. For the RM-ANOVA comparing 72-word TFS-removed to 1000-word unprocessed performance, there were no significant main effects or interactions ($p > 0.05$) other than a main effect of SNR [$F(4,28) = 158$, $p < 0.0005$]. This suggests that there were no differences in performance between the 72-word TFS-removed and 1000-word unprocessed fluctuating-masker conditions [compare gray and small white symbols in Figs. 1(f) and 1(g)]. In contrast, there was a significant main effect of processing [$F(1,7) = 13.2$, $p = 0.008$] for the RM-ANOVA comparing 72-word spectrally smeared and 1000-word unprocessed conditions, providing statistical support for the observation that performance was better for the 72-word spectrally smeared than for the 1000-word unprocessed fluctuating-masker conditions [compare black and small white symbols in Figs. 1(f) and 1(g)]. As before, significant main effects of SNR [$F(2.5,17.8) = 164$, $p < 0.0005$] and fluctuating-masker type [$F(1,7) = 15.6$, $p = 0.006$] confirmed the general trends observed for these variables. There was also a significant interaction between fluctuating-masker type and SNR [$F(4,28) = 2.8$, $p = 0.043$].

The effect of stimulus processing on the FMB showed similar trends when calculated in terms of the SRT difference between stationary and fluctuating-masker conditions [Fig. 2(b)]. Calculated in this way, the FMB was never reduced for the processed 72-word conditions relative to the unprocessed 1000-word condition. In fact, as was observed in the raw data, the FMB appeared to be larger for the spectrally smeared 72-word interfering-taker condition than for the corresponding unprocessed 1000-word case. A two-way (processing condition and fluctuating-masker type) RM-ANOVA calculated on the FMB data [Fig. 2(b)] for the 72-word processed (gray and black bars) and 1000-word unprocessed conditions (striped bars) identified a significant main effect of masker type [$F(1,7) = 16.3$, $p < 0.01$] reflecting the larger FMB associated with the interfering-talker condition, but no significant main effect of processing condition ($p = 0.24$) or interaction between the two variables ($p = 0.28$). Thus, the difference between the unprocessed and spectrally smeared FMB was not significant. Nevertheless, the fact that this difference was statistically significant in the raw data suggests that it might be a real effect.

The use of the set-size manipulation to estimate the FMB at comparable performance levels under different processing conditions relies on the assumption that this manipulation does not directly affect a listener's ability to make use of the gaps in the fluctuating masker to extract speech information. This is based on a fundamental assumption of the Articulation Index (AI) (ANSI, 1969), whereby the signal characteristics of the speech and masking noise determine the amount of speech information audible to the listener and the nature of the task determines the transformation from AI to speech reception score. In this framework, the AI transmitted for a particular processing condition and masker type should be unaffected by the set size, as the signals and maskers were identical in the two set-size conditions. Only

the transformation to percent correct should be affected by the set-size manipulation. Nevertheless, this view of speech intelligibility might not be accurate. Although the AI was designed to apply to only stationary-noise conditions, Rhebergen and Versfeld (2005) and Rhebergen *et al.* (2006) showed that the related Speech Intelligibility Index (SII) (ANSI, 1997) can be extended to fluctuating-masker conditions. However, the SII questions the notion that the underlying amount of speech information available is determined by signal statistics and is independent of the speech identification task. Instead, the SII provides different frequency-band importance functions depending on the nature of the task. In the current experiment, it is possible that with a different set size, the nature of speech cues that, are needed to obtain a correct answer might somehow change. If these particular cues were audible in the gaps of a fluctuating masker to a different extent than the cues needed to correctly identify speech with a different word-set size, the assumption could break down. For example, the relative importance of vowel versus consonant identification in correctly identifying the target word might change depending on the set-size context in which the target word is presented. Because vowels and consonants have different spectral and temporal characteristics, the acoustic cues required for their identification might be released to a different extent by gaps in a fluctuating masker (e.g., Phatak and Grant, 2009).

Performance with the different set sizes for the unprocessed speech conditions was examined to determine whether the set-size manipulation had the same effect on performance for all masker types. If so, this would validate the assumption that the set-size manipulation affected only the transformation to percent correct and did not differentially affect the effective amount of acoustic speech information available in the different masker conditions. Figure 3 plots performance for the 72-word unprocessed conditions [i.e., the data in Fig. 1(d)] against performance for the corresponding 1000-word unprocessed conditions [i.e., the data in Fig. 1(a)]. Each data point in Fig. 3 represents the mean proportion correct for a given SNR and masker type for the 72-word (vertical coordinate) and 1000-word conditions (horizontal coordinate). All three masker types yielded similar

curves delineating the relationship between 72- and 1000-word performance, suggesting that the set-size manipulation had a similar effect on performance for each masker type. Because both axes in Fig. 3 represent dependent variables, a statistical analysis could not be performed on the raw data to determine whether the curves for each masker type differ from one another. It was therefore necessary to fit the data for each masker type and statistically compare the fitted curves.

Boothroyd and Nittrouer (1988) defined a relationship between the probability of correct identification for speech units presented without context (p_i) and that for speech units presented with context (p_c) as

$$p_c = 1 - (1 - p_i)^k, \qquad (1)$$

where k is a constant describing the relative strength of contextual cues. In the current experiment, the 1000-word free-response conditions can be considered the without-context situation, whereas the limited set of response choices available in the 72-word conditions is analogous to providing contextual cues. For each of the masker types, the group-mean 72-word and 1000-word performance data were fit to Eq. (1). The best fitting values of k were very similar for the three masker types (stationary noise, $k = 2.55$; modulated noise, $k = 2.39$; and interfering talker, $k = 2.78$). Bootstrap re-sampling was used to make paired comparisons based on confidence intervals for the fitted k values. The k values for each of the fluctuating-masker conditions did not differ significantly from the k value for the stationary-noise condition (modulated noise, p = 0.62; interfering talker, p = 0.55) or from each other (p = 0.15). Thus, the relationship between performance for the 72- and 1000-word conditions did not depend on masker type, supporting the assumption that the set-size manipulation affected only the transformation from audible speech information to performance.

C. Discussion

The results of this experiment suggest that differences in the SNR values used in the reference stationary-noise condition are largely responsible for the decrease in FMB that has been previously reported for stimuli with degraded TFS or smeared spectra. When the experiment was run in the conventional way, where the set size was kept constant across all stimulus conditions, the FMB was reduced for both types of stimulus processing (TFS removal and spectral smearing) and both masker types (modulated noise and interfering talker), which is consistent with the results of previous studies. However, because of differences in the overall difficulty of the three conditions, the baseline SRT for 50% performance with a stationary-noise masker was roughly 4 dB lower for the unprocessed condition than it was for the two processed conditions. When the response-set size for the unprocessed stationary-noise condition was adjusted to yield stationary-noise performance comparable to that for the processed conditions, the stimulus processing no longer produced a decrease in the FMB.
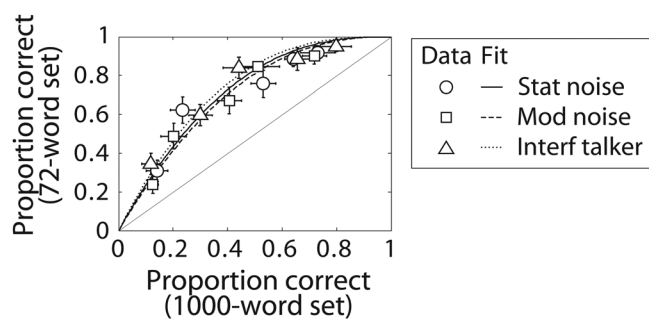

FIG. 3. The effect of word-set size (72 versus 1000 words) on group-mean performance for the three masker types tested in experiment 1 in the unprocessed case (symbols). Error bars indicate ±1 standard error across listeners. Curves represent the best single-parameter fit to the mean data of Eq. (1) describing the relationship between speech-recognition performance and the amount of context available (Boothroyd and Nittrouer, 1988).

These results suggest that it may be necessary to reinterpret the conclusions of previous studies that have used signal processing to determine the effect on the FMB for NH listeners caused by simulations of reduced TFS processing ability (Gnansia *et al.*, 2008; Hopkins and Moore, 2009), reduced frequency selectivity (ter Keurs *et al.*, 1993; Baer and Moore, 1994), or a combination of the two (Qin and Oxenham, 2003; Gnansia *et al.*, 2009). These studies have all found that stimulus processing negatively impacts performance more for fluctuating maskers than for a stationary-noise masker, leading the authors to conclude that spectral contrast discrimination (i.e., frequency selectivity) and TFS processing are essential components in the process of extracting speech information from the momentary dips in the level of the background noise. However, this interpretation is brought into question by the current study, which has shown that degraded TFS and spectral cues do not decrease the FMB when a change in the size of the response set, rather than a change in SNR, is used to equate performance in the processed and unprocessed stationary-noise reference conditions.

This new result suggests that the decreased FMB that has been reported in previous studies can be largely explained by the fact that the FMB for unprocessed speech has been measured relative to a stationary-noise reference point with a substantially lower SNR than the reference points used for the processed stimuli. This suggests that there is nothing particularly important about the TFS-removal or spectral-smearing signal-processing algorithms in terms of their effects on the ability to listen in the dips of a fluctuating masker. Rather, any manipulation that makes the task more difficult and thus reduces performance in the stationary-noise case will reduce the measured FMB simply by changing the stationary-noise SRT that forms the baseline for the FMB calculation.

These results also quantify the effect that changes in the stationary-noise SRT (i.e., the baseline for the FMB calculation) have on the estimated FMB. Increasing the set size yielded a 4-dB increase in the stationary-noise SRT for unprocessed stimuli (from -6 to about -2 dB) [Fig. 2(a)] and about a 5-dB reduction in the FMB (from 10 to 5 dB for modulated noise and from 13 to 8 dB for an interfering talker) [Fig. 2(b)]. In other words, the FMB/SRT ratio was about -1.25 dB/dB. This relationship between the baseline stationary-noise SRT and the FMB might help to explain the surprising result that performance was slightly better for the 72-word spectrally-smeared fluctuating-masker conditions than for the 1000-word unprocessed condition. This difference might have occurred if the set-size manipulation did not precisely equalize processed and unprocessed stationary-noise performance. In Fig. 2(a), the stationary-noise SRT for the 72-word spectrally smeared condition (black bar) was about 1 dB better (lower) than for the 1000-word unprocessed condition (hatched bar). (Although this SRT difference was not significant, there was a significant SNR-by-processing interaction in the corresponding RM-ANOVA computed for the raw stationary-noise data.) This SRT difference would be expected to yield a FMB about 1.25 dB larger for the spectrally smeared 72-word condition than for

the unprocessed 1000-word condition, which is qualitatively consistent with the observed 2-dB FMB difference between these conditions [Fig. 2(b)].

III. EXPERIMENT 2: ADDITIONAL FLUCTUATING-MASKER TYPES

A. Rationale

The results of experiment 1 suggest that the reduction in the FMB experienced by HI listeners for a fluctuating masker is unlikely to be directly attributable to deficits in TFS processing ability or frequency selectivity. This would lead to the conclusion that audibility limitations or some other suprathreshold deficit, such as reduced temporal resolution (Dubno *et al.*, 2003; Jin and Nelson, 2006), are likely to be responsible for the reduced FMB for HI listeners. However, the scope of the findings of experiment 1 is somewhat limited because only opposite-gender interfering-talker and speech-modulated noise maskers were tested. It is possible that the spectral-smearing and TFS-removal algorithms might have more of an effect for other types of fluctuating maskers. One possibility is an interrupted noise, for which HI listeners have been shown to receive less FMB than NH listeners (Jin and Nelson, 2006). Another possibility is a situation involving multiple interfering talkers of the same gender as the target talker, where TFS information might be important for providing pitch cues to segregate simultaneous speech sources. For example, Stickney *et al.* (2007) showed in a cochlear-implant simulation that the addition of TFS-containing low-frequency information provided the most benefit to speech-recognition performance in situations where the target and interfering talker had similar voice fundamental frequencies (F0s). To examine these possibilities, a second experiment was conducted that extended the set-size approach of experiment 1 to test the effects of TFS removal and spectral smearing on the FMB for two additional fluctuating-masker conditions: an interrupted-noise and an interfering-talker condition involving two simultaneous maskers of the same (male) gender as the target talker.

Although the procedures used in experiment 2 were based on those used in experiment 1, they were modified to incorporate an adaptive set-size procedure to make it easier to expand the paradigm to new masking conditions. There were several limitations of the method used in experiment 1 to adjust the word-set size. First, the two set sizes (72 and 1000 words) were chosen in pilot tests without systematically varying the set size to establish equivalent performance. Second, the 72-word sets were chosen at random from the larger set, allowing the possibility of variability in difficulty of the different randomly chosen subsets. Third, each new 72-word subset required significant training before testing began. Experiment 2 established a new test methodology to address each of these limitations. First, the set size was varied systematically to choose the appropriate size for each processing condition. Second, subsets of the larger word set were chosen methodically such that each token had a similar number of likely confusions. Third, the need for training on each individual subset was removed by limiting the largest possible set size to 144 words, and training listeners in this

condition. Thus, listeners were already trained on all of the words that could make up a smaller subset. A preliminary experiment (2a) determined the set sizes needed for equalized performance in the stationary-noise conditions. Experiments 2b and 2c used the set sizes determined in experiment 2a to estimate the FMB for a range of fluctuating maskers at a fixed stationary-noise SNR.

B. Experiment 2a: Determining the set sizes for equal stationary-noise performance

1. Methods

This experiment used an adaptive tracking procedure to measure the SRT (i.e., the SNR required for a given level of performance) in a stationary-noise masking condition. The same three processing conditions from experiment 1 were tested (unprocessed, TFS-removed, and spectrally smeared). SRTs were estimated for the unprocessed conditions presented with a full set of 144 word response choices. SRTs were estimated for the processed conditions with response sets sizes ranging from 12 to 72 words. The idea was to determine the set size needed in the processed conditions to yield the same SRT as the 144-word unprocessed case.

The target words were monosyllabic consonant–vowel–consonant words selected from the rhyme test materials described by House *et al.* (1965). The rhyme test contains 50 lists of six words, with the words in each list differing only in their initial (25 lists) or final consonant (25 lists). The 144-word set for this experiment was constructed by selecting 24 of these 50 lists, with four lists for each of six vowel contexts (/æ/ as in bat, /e/ as in bait, /ɛ/ as in bet, /I/ as in bit, /i/ as in beet, and /ʌ/ as in but). Each word in the set was embedded in a carrier phrase ("You will mark ____ please"). Each token was spoken by three different male talkers (mean F0s: 109, 108, and 105 Hz; overall mean F0: 107 Hz) and recorded at a sampling frequency of 44.1 kHz, then down-sampled to 24.414 kHz following the application of an anti-aliasing low-pass filter. The stationary-noise masker used in this experiment was spectrally shaped to match the average long-term spectrum across the 144 tokens spoken by each of the three talkers. As in experiment 1, target speech was presented at a level of 60 dB SPL and the desired SNR was achieved by adjusting the level of the masker. The details regarding the masker duration, ramping, onset and offset relative to the target, and level were the same as in experiment 1. The signal-processing algorithms and sound-generation hardware were also as described in experiment 1, except that spectral smearing was applied to simulate filter broadening by a factor of 5 (instead of 4 as in experiment 1) to more closely equalize spectrally-smeared and TFS-removed performance.

Response buttons for the 144 words were arranged in 12 rows and 12 columns on the screen. The words were organized first by vowel context, with each subsequent pair of rows associated with one vowel. Within each vowel group spanning two rows, words were arranged in alphabetical order. Following each stimulus presentation, listeners were instructed to push the response button on the touch screen associated with the word presented, then press the "OK" button to confirm the response. Feedback was provided following each response by highlighting the button containing the correct response.

For the 144-word condition, the token presented on any given trial was selected at random from the 432 possible stimuli (144 words spoken by the three different talkers). All 144 word choices were printed in black font, indicating that all choices were available. For the conditions involving fewer than 144 words, smaller set sizes were created by choosing at random fewer than six words from each of the 24 rhyme-test lists. As in the 144-word condition, the target talker was selected at random on every trial. Word choices associated with this smaller set of words were printed in black font, indicating a valid choice, whereas the remainder of the 144 word choices was printed in gray font, indicating inactive buttons. For these smaller word sets, the number of words chosen from each list was either the same for all 24 lists (for set sizes divisible by 24) or a subset of the 24 lists was chosen at random to receive one additional word. For example, a set size of 36 words was generated by selecting two words each from 12 of the lists, and three words each from the other 12 lists. In the latter case, the lists containing one extra word in the valid response set were distributed as evenly as possible across the six vowel contexts. A new word subset was generated for each adaptive tracking run, but remained fixed for the duration of the run. The SRT was measured for a particular combination of set size and processing condition by adapting the SNR in a two-down, one-up procedure tracking the 70.7% correct point (Levitt, 1971). The starting SNR was −12 dB, changed by 4 dB until the first upper reversal, and changed by 2 dB for 12 more reversals. The SRT was taken to be the mean SNR at the last eight reversal points.

Each listener began with an initial training phase. The first part of the training phase was intended to familiarize the listener with the larger word list and response format—presenting each of the 144 word choices in quiet (the talker selected for each word presentation was selected at random). The second part of the training phase was intended to further familiarize the listener with the word set, reduced set sizes, and the signal-processing conditions. This phase consisted of 20 SRT measurement runs representing a variety of set sizes and processing conditions. The training phase took approximately 2 h to complete.

The test phase consisted of SRT measurements for each combination of the two processing conditions (TFS-removed or spectrally smeared) and five word-set sizes (12, 24, 36, 48, or 72 words), plus the unprocessed condition presented with the 144-word set, for a total of 11 conditions. Each of these conditions was presented four times for each listener, for a total of 44 SRT measurement runs. The first run for each condition was treated as additional training. The SRT estimate for each listener was taken as the mean SRT across the remaining three runs for each condition.

Seven NH listeners (age range 32-56 yr, four female) participated in the experiment. Six of these listeners had also participated in experiment 1. All listeners had pure-tone audiometric thresholds of 20 dB HL or better in both ears for octave frequencies between 250 and 4000 Hz, as well as at 6000 Hz.

2. Results

Mean SRTs are shown in Fig. 4. The error bars represent the standard error across listeners, calculated after subtracting the global-mean (across the 11 conditions) SRT for each individual listener. This is done to give an idea of the variability in each condition relative to the other conditions. For the processed conditions that were tested with different set sizes, SRTs increased with increasing set size, as expected. To estimate the set sizes for the two processed conditions needed to yield comparable performance to the unprocessed 144-word condition, the mean SRTs were linearly interpolated between successive set sizes (solid lines). The set sizes required to yield a SRT equal to that for the unprocessed conditions (SRT = −4.43 dB, horizontal dashed line) were found to be 31.0 words for the TFS-removed condition (gray vertical dashed line) and 32.5 words for the spectrally-smeared condition (black vertical dashed line).

3. Discussion

The purpose of this experiment (2a) was to determine the set sizes needed in the subsequent experiment (2b) to estimate the FMB for a variety of fluctuating maskers with the same baseline stationary-noise SRT for processed and unprocessed speech. The SRTs plotted in Fig. 4 are based on the full set of data collected from all seven listeners who participated in the experiment. However, data collection for experiment 2b was initiated before data collection was complete for all listeners in experiment 2a. Thus, the incomplete data from experiment 2a (i.e., the data that were available just before the initiation of data collection for experiment 2b) were analyzed at that time to estimate the set sizes required to equalize stationary-noise performance across processing conditions. At that time, preliminary estimates of the set sizes required to yield stationary-noise SRTs equal to the SRT for the unprocessed, 144-word condition were 31
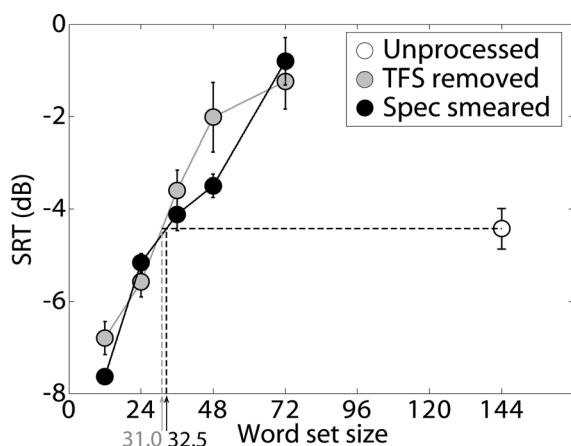

FIG. 4. The results of experiment 2a showing the stationary-noise SRT as a function of word-set size for the TFS-removed (gray symbols) and spectrally smeared (black symbols) conditions, for the 144-word set-size for the unprocessed condition. Error bars indicate ±1 standard error across listeners, calculated after normalizing the data for each listener by the mean SRT across the 11 conditions. Vertical dashed lines indicate the estimates of the word-set sizes needed in the processed conditions to yield the same stationary-noise SRT as the 144-word unprocessed condition.

words for the TFS-removed condition and 34 words for the spectrally smeared condition.

C. Experiment 2b: The FMB for a variety of fluctuating-masker conditions

This experiment used the word-set sizes estimated in the preliminary stages of experiment 2a that were intended to yield SRTs for the processed stationary-noise conditions equal to the SRT for the 144-word unprocessed stationary-noise condition. All of the TFS-removed conditions were tested with a 31-word set, the spectrally-smeared conditions were tested with a 34-word set, and the unprocessed conditions were tested with a 144-word set. For each processing type, a SRT was estimated for a stationary-noise masker condition, the same two fluctuating-masker conditions from experiment 1 (speech-modulated noise and an opposite-gender interfering talker) and two additional fluctuating-masker conditions (two same-gender interfering talkers and a 16-Hz interrupted noise).

1. Methods

The stimulus processing and SRT measurement procedure was as described for experiment 2a. The stationary-noise, speech-modulated noise, and female interfering-talker maskers were generated as in experiment 1, except that the long-term spectra for these two signals were also matched to the average long-term spectrum of the 144-word corpus for this experiment. The mean F0 for the female interfering talker was 199 Hz. Two new fluctuating maskers were also added. A condition with two male interfering talkers was included because this type of condition has been shown to produce a large amount of informational masking (IM), i.e., masking that cannot be accounted for in terms of peripheral energetic-masking effects (Brungart *et al*., 2001; Freyman *et al*. 2004). Instead, IM is thought to reflect central confusion about which portions of the audible signal are attributable to the target and which are attributable to the masker. A recording was made of a single male reading the same passage ("The Unfruitful Tree") as the female interfering-talker masker. The recording was then processed in the same way as the interfering female masker to remove silent gaps and match the long-term spectrum of the target speech corpus. In each trial, two segments of this long-duration recording were summed together to create the two-talker masker. The SNR for this masker condition was calculated based on the ratio of the target level and the summed two-talker masker level. The mean F0 for the male interfering-talker masker was 131 Hz, or about 3.5 semitones higher than the mean F0 of 107 Hz for the target stimuli produced by three other male talkers. The interrupted noise was generated by square-wave modulating the speech-spectrum shaped stationary noise at 16 Hz with a 50% duty cycle. The "on" portions of the modulated noise were ramped on and off (2 ms raised cosine).

SRTs were estimated for all combinations of three processing conditions (unprocessed, TFS-removed, and spectrally smeared) and five noise types (stationary, speech-modulated noise, female interfering talker, two male interfering talkers,

and interrupted noise). Each of these 15 combinations was presented four times in random order, for a total of 60 runs. The first run for each condition was considered to be a training run. The SRT for each listener and condition was taken to be the mean SRT measured in the last three runs.

Eleven NH listeners participated in this experiment (age range 29-66 yr, six female). All listeners had pure-tone audiometric thresholds of 20 dB HL or better in both ears for octave frequencies between 250 and 4000 Hz, as well as at 6000 Hz. Seven of these listeners had participated in experiment 2a. The four listeners that had not taken part in experiment 2a were presented with the same 2 h of training from experiment 2a to familiarize them with the word set, testing algorithm, and processing conditions.

2. Results

Mean stationary-noise SRTs for the three processing conditions are shown in Fig. 5(a). As in Fig. 2(a), lower SRT values indicate better performance, with a lower SNR required to achieve 70.7% correct performance. A RM-ANOVA computed on the SRT data shown in Fig. 5(a) indicated a significant effect of processing condition [$F(2,20) = 5.95$, $p < 0.01$]. Bonferroni-corrected two-tailed *post hoc* t-tests indicated that SRTs were significantly poorer for the spectrally smeared

condition than the unprocessed condition ($p < 0.03$) and that the difference between the TFS-removed and unprocessed SRTs just failed to reach significance ($p = 0.066$). Nevertheless, SRT differences between processing conditions were all 1.5 dB or less, even though the set-size manipulation did not precisely equalize stationary-noise performance across the processing conditions.

Mean FMB estimates for each fluctuating-masker and processing condition were calculated by subtracting the fluctuating-masker SRT from the stationary-noise SRT. Mean FMB estimates are shown in Fig. 5(b). Larger FMB values indicate a larger benefit received from the masker fluctuations. A RM-ANOVA with two within-subject factors (fluctuating-masker type and processing condition) was computed on the FMB data. A significant main effect of fluctuating-masker type [$F(31.9, 2.2) = 127$, $p < 0.0005$] reflected the substantial variation in FMB across fluctuating-masker conditions. The FMB was positive for the female-interferer, speech-modulated noise and interrupted-noise conditions, indicating that listeners received a benefit from the masker fluctuations relative to the stationary-noise condition. The FMB was slightly negative for the two-talker masker condition, indicating that listeners did not receive any benefit from masker fluctuations, and in fact were negatively impacted by the interfering talkers relative to the stationary-noise condition. This result suggests that the same-gender two-talker masker produced substantial IM, even with a difference in mean F0 between the target and interferers.

There was no significant main effect of processing type ($p = 0.87$), suggesting that the TFS removal and spectral smearing did not have a consistent effect on the FMB across masker conditions. There was a significant interaction between fluctuating-masker type and processing condition [$F(53.9, 5.39) = 3.11$, $p < 0.05$], suggesting that processing might have had different effects on the FMB for different masking types. However, *post hoc* t-tests showed no significant differences ($p > 0.05$) between the FMB for unprocessed and processed (TFS-removed or spectrally smeared) conditions for any of the fluctuating maskers, whether or not a Bonferroni correction was applied. Thus, these results showed no evidence that the FMB was affected by stimulus processing for any of the fluctuating maskers tested, once stationary-noise performance was approximately equalized across processing conditions using the set-size adjustment.

D. Experiment 2c: Set-size adjustment to more closely equalize stationary-noise performance

1. Methods

In experiment 2b, SRTs were about 1.5 dB larger for the processed conditions than for the unprocessed conditions [Fig. 5(a)]. Experiment 2c made small adjustments to the word-set sizes employed for the processed conditions to more closely equalize stationary-noise performance. Based on the slopes of the performance-versus-set size functions (Fig. 4) from experiment 2a, it was estimated that the set sizes would need to be decreased to 28 words (TFS-removed) and 29 words (spectrally smeared condition) to yield more similar performance to the 144-word unprocessed
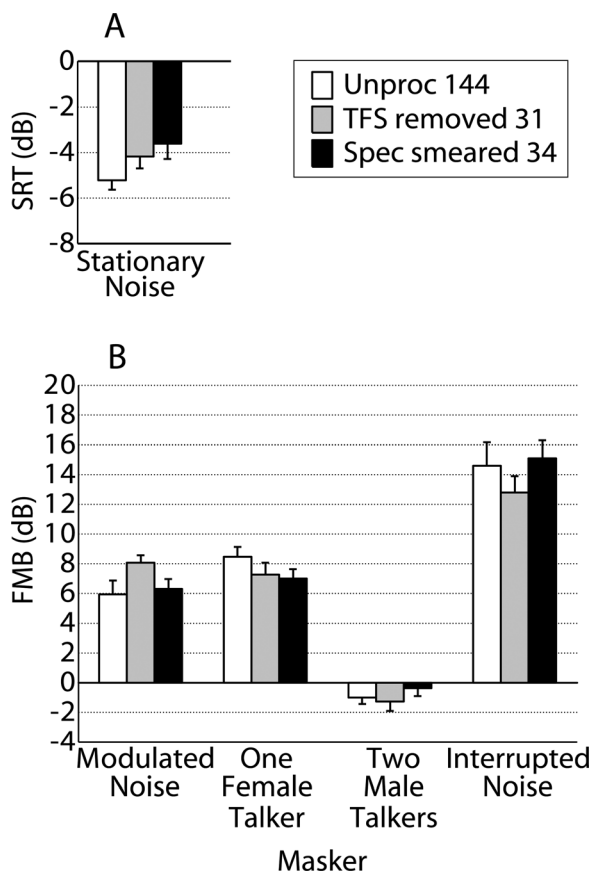
FIG. 5. (a) The stationary-noise SRT estimated using adaptive tracking and (b) the FMB for the four fluctuating maskers tested in experiment 2b. Error bars indicate ±1 standard error across listeners. The word-set sizes for the processed conditions were adjusted (to 31 and 34 words) based on the results of experiment 2a to yield similar stationary-noise performance to the 144-word unprocessed condition.

stationary-noise condition. SRT measurements were repeated using these set sizes for two processing conditions (TFS-removed and spectrally smeared) and three masker types (stationary-noise, female interfering talker and interrupted noise). Only these two fluctuating-masker conditions were repeated because these conditions showed some hint of a reduction in FMB (albeit nonsignificant) as a result of stimulus processing in experiment 2b. Four SRT measurements were made for each of these six combinations of processing condition and masker type. The first of these four measurements was discarded as a training run, and the SRT for each listener and condition was taken to be the mean of the last three SRT measurements. The same 11 listeners from experiment 2b participated.

2. Results

Figure 6 shows mean stationary-noise SRTs for the three processing conditions [Fig. 6(a)] and FMBs for the two fluctuating maskers [Fig. 6(b)] tested in experiment 2c. SRTs and FMBs for the unprocessed conditions were not retested in experiment 2c and are instead replotted from experiment 2b. Figure 6(a) shows that the adjusted word-set sizes nearly succeeded in equalizing stationary-noise performance across processing conditions. A RM-ANOVA computed for the stationary-noise SRT data shown in Fig. 6(a) showed a significant main effect of processing condition
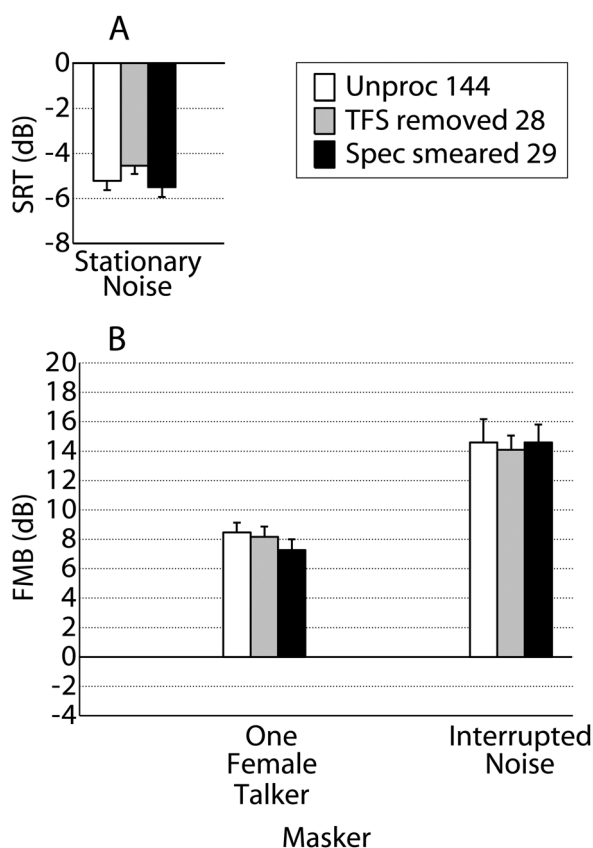

FIG. 6. (a) The stationary-noise SRT and (b) the FMB for the two fluctuating maskers retested in experiment 2c. Word-set sizes for the processed conditions were adjusted from the values used in experiment 2b to more closely equalize stationary-noise performance between unprocessed and processed conditions.

$[F(1.55,15.5) = 4.13, p < 0.05]$. However, *post hoc* Bonferroni t-tests showed no significant differences in stationary-noise SRT between the unprocessed and TFS-removed ($p = 0.30$) or spectrally smeared ($p = 1$) conditions, indicating that the set-size manipulation successfully offset the effect of stimulus processing on the SRT relative to the unprocessed condition. Only the *post hoc* comparison of the TFS-removed and spectrally smeared SRTs showed a significant difference ($p < 0.005$). Figure 6(b) shows that the FMB for the unprocessed and processed conditions were even more similar than in experiment 2b. A two-way (processing condition and masker type) RM-ANOVA computed for the FMB data indicated only a main effect of masker type $[F(10,1) = 40.1, p < 0.0005]$ with no significant main effect of processing condition ($p = 0.65$) or interaction between the two variables ($p = 0.59$). Thus, there was no evidence that stimulus processing affected the FMB in this experiment.

IV. DISCUSSION

The results of experiments 1 and 2 clearly indicate that stimulus processing to smear the stimulus spectrum or remove TFS information had a negative impact on speech-recognition performance overall, a result that is consistent with many previous studies. This is consistent with the idea that TFS and spectral cues relay important information about speech, and that when these cues are removed, speech discrimination becomes more difficult. However, when the response set sizes for the different stimulus processing conditions were adjusted to make them equally difficult for a stationary-noise masking condition, the negative impacts of stimulus processing on speech recognition did not appear to be any different for the fluctuating maskers than they were for the stationary-noise masker.

This result is important, because it suggests that the reduced FMBs that have been reported in previous studies with degraded spectral or TFS cues may be driven more by the overall decrease in difficulty introduced by these two kinds of processing than by any particular type of auditory processing that is specifically impaired by these manipulations. For example, in experiment 1, we found that both TFS processing and spectral smearing produced significant reductions in the FMB. However, they also both decreased the overall difficulty of the task with the stationary-noise masker. When the set size was increased from 72 words to 1000 words to produce an equivalent reduction in stationary-noise performance with the unprocessed speech, the FMB for the unprocessed speech was also reduced by roughly the same amount. In other words, increasing the set size from 72 to 1000 words had the same effect on the FMB as stimulus processing to remove TFS or smear the spectrum. This result strongly questions the idea that these particular signal-processing manipulations target specific auditory processing strategies that listeners use to extract information from gaps of a fluctuating masker to better receive speech information. A more parsimonious explanation for the apparent reduction in the FMB observed for fluctuating maskers in previous studies may be a SNR-based phenomenon based on a combination of (1) differences between stationary-noise and

fluctuating-masker conditions in the slopes of the performance-SNR functions and (2) the higher SNR values required to achieve a given performance level in stationary noise as a result of the generally poorer performance associated with stimulus processing.

This SNR-related confound is particularly apparent when the FMB is characterized in terms of the difference in SRT (i.e., the long-term SNR required to yield a given performance level) between stationary-noise and fluctuating-masker conditions. Because of the difference in the slopes of the performance-intensity functions for these two conditions, the FMB will necessarily be smaller for the higher (i.e., worse) stationary-noise SRTs associated with more difficult listening conditions. The effect is the same whether the listening conditions are made more challenging by distortion induced by stimulus processing or by an increase in the difficulty of the task with an increased number of responses to choose from.

Although it is not possible to rule out the possible existence of some fluctuating-masker types with FMBs that might depend on spectral resolution or TFS processing ability, the two experiments presented here have demonstrated that this is not the case for a wide variety of masker types, including speech-modulated noise, interrupted noise, opposite-gender speech, and two-talker same-gender speech. Thus, even in challenging masking conditions where TFS coding of F0 information would seem to be critical for providing cues for the perceptual segregation of simultaneous talkers (Bregman, 1990), such as those with multiple same-gender simultaneous talkers (e.g., Carhart et al., 1969; Bronkhorst and Plomp, 1992; Brungart et al., 2001; Freyman et al., 2004; Iyer et al., 2010), there is little evidence that TFS processing has a major impact on the ability to extract a target speech signal from a fluctuating masker. These results are therefore in general agreement with those of Oxenham and Simonson (2009), who found that high- and low-pass filtering yielded the same FMB for speech. They argued that the low-pass conditions contained salient pitch information associated with low-order, resolved harmonics, presumably relayed by TFS cues, whereas the high-pass conditions contained only the weak pitch information associated with high-order, unresolved harmonics, coded by phase locking to the temporal envelope. Thus, they concluded that salient, TFS-based, resolved-harmonic pitch cues were not necessary to receive a FMB, even in the case of target and interfering talkers with similar F0s.

Although all of the listeners used in these studies had normal hearing, the primary motivation for the studies was to obtain a better understanding of the factors that might impact the FMB for HI listeners. The main purpose for using simulations of reduced frequency selectivity and TFS-coding ability rather than actual HI listeners is to allow a test of the independent contributions that these distortions and audibility might make to the reduced FMB observed for HI listeners. Nevertheless, there are several reports in the literature of fluctuating-masker situations where HI or elderly listeners show a reduced benefit relative to NH listeners, even in the absence of a SNR confound. These include interfering-talker (Bernstein and Grant, 2009), interrupted-noise (Jin and Nelson, 2006), and sinusoidally amplitude-modulated

(SAM) noise conditions (Takahashi and Bacon, 1992; Grose et al., 2009). To the extent that the simulations used in experiments 1 and 2 accurately simulate suprathreshold processing deficits, and to the extent that the maskers used here are representative of those used in these previous studies, the current results suggest that the reduced FMB for HI listeners in these conditions is not likely to be due to a reduction in TFS processing ability or frequency selectivity. (Although SAM noise was not tested in this study, it seems unlikely that reduced FMB would be observed for SAM noise when it was not observed for the speech-modulated and interrupted-noise conditions that were tested in the current study.)

One possible cause for the reduced FMB for HI listeners observed in certain cases is the reduced audibility associated with elevated absolute thresholds. Reduced audibility will have relatively little impact on speech understanding in a stationary-noise condition because the audibility of the speech signal will be primarily determined by the noise level. In contrast, absolute threshold may play a larger role in a condition that involves a fluctuating masker, especially one with deep dips in the masker level or periods of quiet, such as the interfering-talker, SAM-noise, and interrupted-noise conditions. In these situations, the speech information unmasked by the masker valleys might be above threshold for the NH listener, but inaudible to the HI listener. Desloge et al. (2010) used hearing-loss simulations presented to NH listeners to mimic the loss of speech audibility experienced by individual HI listeners. They found that audibility effects could account for most of the reduced FMB associated with hearing loss in a 10-Hz interrupted-noise masking condition. The NH and HI listeners in the Bernstein and Grant (2009) study had a difference in audibility above about 2 kHz, which might account for the FMB difference between the two groups for the interfering-talker masking condition. Jin and Nelson (2006) attempted to use amplification to overcome audibility differences between NH and HI listeners to examine the role of suprathreshold factors in reducing the FMB in 8- or 16-Hz interrupted noise for HI listeners. They suggested that audibility was restored to the HI listeners, as both groups of listeners obtained the same performance (nearly 100% correct) in quiet. However, the ceiling effect on speech-identification performance prevents any conclusion as to whether audibility was restored for the HI listeners. In fact, articulation theory (ANSI, 1969) points out that 100% correct performance in a task with a limited response set can be obtained with an AI ranging from 0.5 (half of the speech information is audible) to 1.0 (all of the speech information is audible). Two other studies that attempted to equalize NH and HI audibility by using noise masking to elevate NH thresholds (Eisenberg et al., 1995; Bacon et al., 1998) had a small residual difference between stationary-noise SRTs of the two listener groups that might have confounded the FMB estimate. If audibility limitations, rather than suprathreshold processing deficits, are the main determining factor in reducing the FMB for HI listeners, this would bode well for the likelihood of success of traditional gain-based hearing-aid signal-processing algorithms for improving speech-reception performance in fluctuating-

masker conditions. In particular, a fast compression algorithm, where the gain that could vary as quickly as the changes occurring in the masker level, could provide audibility for low-level speech sounds that occur during dips in the fluctuating-masker level (Naylor and Johannesson, 2009).

There are, however, a few examples of studies that have found differences in the FMB for NH and HI listeners that are difficult to fully explain with differences in audibility or by differences in the stationary-noise SRT between the two groups. Phatak and Grant (2009) used amplification to offset the reduction in audibility for HI listeners. They then used a spectral analysis to confirm that the amplification successfully restored audibility for the vowel (but not the consonant) features. Their results show that, under these conditions, the FMB for vowel identification relative to a fixed stationary-noise SRT was similar for NH and HI listeners for 2-, 4-, and 12-Hz SAM maskers, but that the FMB for HI listeners was reduced for the 32-Hz SAM masker. Grose et al. (2009) found a reduced FMB for the older relative to younger NH listeners for 16- and 32-Hz SAM noise in an experiment where there was no difference in stationary-noise performance between the two groups, suggesting that suprathreshold distortions might be responsible for the reduced FMB benefit (although audibility effects of the FMB cannot be ruled out, as there were some audiometric threshold differences between the two nominally NH listener groups for frequencies of 2 kHz and above). The particular SAM conditions that yielded FMB differences in the studies of Phatak and Grant and Grose et al. were not tested in the current study. It therefore remains unknown whether reduced TFS-processing ability or frequency selectivity could be responsible for the reduced FMB for these relatively high-rate SAM maskers. However, it seems likely that reduced temporal resolution for older and HI listeners could explain these results, consistent with the findings of Dubno et al. (2003) and George et al. (2006) of a correlation between the FMB for high-rate (16–50 Hz) modulated maskers and measures of temporal resolution.

A key assumption made in this study was that the set-size manipulation affected only the transformation from available speech information to percentage-correct performance (ANSI, 1969), and not the nature of the speech information required to perform the task. In this study, we verified this assumption by demonstrating that changes in set size had roughly the same impact on the proportion of correct responses for the stationary-noise, interfering-talker and speech-modulated noise conditions tested in this experiment (Fig. 3). Nevertheless, this result might not necessarily hold for all masker types and response-set conditions. Buss et al. (2009) found that the differences in the FMB between an open-set and a three-alternative closed-set word-identification task were much larger for low-rate (2.5–10 Hz) than for high-rate (20–40 Hz) SAM maskers. This suggests that, contrary to the assumptions made in the current study and supported by the analysis shown in Fig. 3, the speech cues required for accurate word identification might in fact depend on the context of the task, at least for certain fluctuating maskers or word-set conditions that were not tested in the current experiment.

There were two main differences between the current study and that of Buss et al. that could have contributed to the different results obtained in the two studies. First, Buss et al. examined the effects of changes in masker rate, whereas the current study examined interfering-talker and speech-modulated noise maskers, both of which are dominated by relatively slow modulations below about 10 Hz (Steeneken and Houtgast, 1980). Second, the contrast between the open set and the very small closed set (three word choices) in the study by Buss et al. might have had a greater effect on the nature of the speech cues required for word identification than the more modest set-size manipulations (1000-word open set versus 72-word closed set) employed in experiment 1 of the current study. For example, accurate vowel identification might have often been sufficient for word identification with only three response alternatives, whereas accurate consonant identification would be required for word identification in the open-set case. Phatak and Grant (2009) showed that the masker modulation rate has a greater effect on the FMB for vowel than for consonant recognition. Further work is needed to determine the range of set-size and masking conditions for which this basic assumption holds.

In this study, we introduced an experimental procedure designed to measure word-recognition performance in a test with adjustable difficulty. The procedure uses pilot testing to establish the relationship between the SRT and word-set size for a given condition, and then uses this information to adjust the set sizes to equate SNR and stationary-noise speech-reception performance across a variety of processing conditions. In our experiment, this equalization was used to eliminate any potential influence that a baseline SNR confound might have on the FMB estimate. However, this technique might be useful in a variety of other situations where it is desirable to equalize performance across multiple experimental conditions without modifying the acoustic stimulus.

One obvious example where this might be helpful is in comparing the FMBs for NH and HI listeners, where HI listeners show a reduction in speech-recognition performance for stationary-noise conditions that must be offset to rule out a possible confound due to the SNR dependence of the FMB. An adaptive set-size adjustment procedure might also useful when measuring IM in competing-talker situations where the differences in the relative levels of the talkers can be used as a cue to differentiate between the different competing talkers in the stimulus (Brungart, 2001). A SNR confound related to this level cue might be responsible, at least in part, for results suggesting that HI listeners are less susceptible to IM than NH listeners (e.g., Arbogast et al., 2005), because their poorer overall performance might drive their SRTs to a SNR value where they have access to level-related segregation cues that are unavailable to NH listeners operating at threshold.

Another area where such a technique might be useful is in the evaluation of hearing-aid compression algorithms. Naylor and Johannesson (2009) showed that a fast compression algorithm became more beneficial (in terms of increasing the effective SNR for speech presented in a fluctuating noise) for more negative SNRs. For positive SNRs, the algorithm tended

to amplify the noise more than the signal, thereby reducing the effective SNR. A standard speech test that estimates the SRT such as the Hearing in Noise Test (HINT; Nilsson *et al.*, 1994) will not be able to control the SNR range. Such a procedure could just as easily yield a SRT at a positive SNR, where the compression algorithm is not operating efficiently, as it could at a negative SNR, where the algorithm is more likely to be beneficial. The other standard speech-testing alternative—measuring performance in percentage-correct terms at a fixed SNR—is likely to present problems when the performance level approaches 0 or 100% correct levels. A variable word-set size SRT procedure could help to avoid these problems by allowing an adaptive test with more control over the SNR range of the measurement.

V. SUMMARY AND CONCLUSIONS

Two methods for simulating the psychophysical deficits associated with sensorineural hearing loss were examined to determine their effect on the FMB for NH listeners. When the processed and unprocessed conditions were tested with the same closed-set response paradigm, both noise vocoding (to simulate a reduction in the ability to use TFS information) and spectral smearing (to simulate reduced frequency selectivity) produced substantial reductions in the measured FMB. However, when stationary-noise performance was equated by adjusting the response set sizes in the different processing conditions, the processing was found to not reduce the FMB. This result held for a variety of fluctuating-masker types, including speech-modulated noise, an opposite-gender interfering talker, two same-gender interfering talkers, and a 16-Hz interrupted noise. These results suggest that deficits in frequency selectivity or in the ability to use TFS information do not fundamentally limit a NH listener's ability to use dips of a fluctuating masker to extract information from a target speech signal. Thus, it is unlikely that these kinds of processing deficits can explain the reduced FMB often reported for HI listeners. Further testing is needed to determine if the reduced FMB for HI listeners is the result of a SNR-related confound similar to the one addressed in this study, or whether other factors such as reduced audibility or reduced temporal resolution might account for the difficulties HI listeners appear to experience when listening in a fluctuating-masker background.

ANSI (**1969**). *Methods for the Calculation of the Articulation Index, S3.5* (American National Standards Institute, New York).

ANSI (**1997**). *Methods for Calculation of the Speech Intelligibility Index, S3.5* (American National Standards Institute, New York).

Arbogast, T. L., Mason, C. R., and Kidd, G. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," J. Speech Lang. Hear. Res. **41**, 549–563.

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

Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.

Bernstein, J. G. W., and Grant, K. W. (**2009**). "Auditory and auditory-visual speech intelligibility in fluctuating maskers for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **125**, 3358–3372.

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–114.

Bregman, A. S. (**1990**). "Auditory organization in speech perception," in *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), pp. 529–594.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Buss, E., Hall, J. W., and Grose, J. H. (**2004**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear Hear. **25**, 242–250.

Buss, E., Whittle, L. N., Grose, J. H., and Hall, J.W. (**2009**). "Masking release for words in amplitude-modulated noise as a function of modulation rate and task," J. Acoust. Soc. Am. **126**, 269–280.

Carhart, R., Tillman, T., and Greetis, R. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

Carhart, R. C., and Tillman, T. W. (**1970**). "Interaction of competing speech signals with hearing losses," Arch. Otolaryng. **91**, 273–279.

Desloge, J. G., Reed, C. M., Braida, L. D., Perez, Z. D., and Delhorne, L. A. (**2010**). "Speech reception by listeners with real and simulated hearing impairment: Effects of continuous and interrupted noise," J. Acoust. Soc. Am. **128**, 342–359.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2003**). "Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with impaired hearing," J. Acoust. Soc. Am. **113**, 2084–2094.

Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Hear. Res. **38**, 222–233.

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.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2008**). "Spatial release from masking with noise-vocoded speech," J. Acoust. Soc. Am. **124**, 1627–1637.

George, E. L. J., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295–2311.

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. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gnansia, D., Jourdes, V., and Lorenzi, C. (**2008**). "Effect of masker modulation depth on speech masking release," Hear. Res. **239**, 60–68.

Gnansia, D., Pean, V., Meyer, B., and Lorenzi, C. (**2009**). "Effects of spectral smearing and temporal fine structure degradation on speech masking release," J. Acoust. Soc. Am. **125**, 4023–4033.

Grant, K. W., and Seitz, P. F. (**2000**). "The recognition of isolated words and words in sentences: Individual variability in the use of sentence context," J. Acoust. Soc. Am. **107**, 1000–1011.

Grose, J. H., Mamo, S. K., and Hall, J. W. (**2009**). "Age effect in temporal envelope procesing: Speech unmasking and auditory steady-state responses," Ear Hear. **30**, 568–575.

Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.

Hopkins, K., and Moore, B. C. J. (**2009**). "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise," J. Acoust. Soc. Am. **125**, 442–446.

Hopkins, K., and Moore, B. C. J. (**2010**). "The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **127**, 1595–1608.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D. (**1965**). "Articulation-testing methods: Consonantal differentiation with a closed-response set," J. Acoust. Soc. Am. **37**, 158–166.

IEEE (**1969**). *IEEE Recommended Practice for Speech Quality Measures* (Institute of Electrical and Electronic Engineers, New York).

Iyer, N., Brungart, D. S., and Simpson, B. D. (**2010**). "Effects of target-masker contextual similiarity on the multimasker penalty in a three-talker diotic listening task," J. Acoust. Soc. Am. **128**, 2998–3010.

Jin, S. H., and Nelson, P. B. (**2006**). "Speech perception in gated noise: The effects of temporal resolution," J. Acoust. Soc. Am. **119**, 3097–3108.

Lacher-Fougère, S. and Demany, L. (**2005**). "Consequences of cochlear damage for the detection of interaural phase differences," J. Acoust. Soc. Am. **118**, 2519–2526.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). "The intelligibility of speech as a function of the context of the test materials," J. Exp. Psych. **41**, 329–335.

Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (**2006**). "Frequency discrimination of complex tones by hearing-imparied subjects: Evidence for loss of ability to use temporal fine structure," Hear. Res. **222**, 16–27.

Naylor, G., and Johannesson, R. B. (**2009**). "Long-term signal-to-noise ratio at the input and output of amplitude-compression systems," J. Am. Acad. Audiol. **20**, 161–171.

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). "A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 2045–2064.

Nilsson, M., Soli, S., and Sullivan, J. A. (**1994**). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Oxenham, A. J., and Moore, B. C. J. (**1997**). "Modeling the effects of peripheral nonlinearity in normal and impaired hearing," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ), pp. 273–288.

Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference," J. Acoust. Soc. Am. **125**, 457–468.

Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.

Phatak, S. A., and Grant, K. W. (**2009**). "Consonant and vowel perception in modulated maskers: Effect of hearing impairment," Aging and Speech Communication: Third International and Interdisciplinary Research Conference, Bloomington, IN (A).

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2006**). "Extended speech intelligibility index for the prediction of the speech reception threshhold in fluctuating noise," J. Acoust. Soc. Am. **120**, 3988–3997.

Steeneken, H. J. M., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326.

Stickney, G. S., Assmann, P. F., Chang, J. E., and Zeng, F. G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.

Summers, V., and Molis, M. R. (**2004**). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47**, 245–256.

Takahashi, G. A., and Bacon, S. P. (**1992**). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," J. Speech Hear. Res. **35**, 1410–1421.

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.

Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Helfer, K. S. (**2007**). "Speech intelligibility in cochlear implant simulations: Effect of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2388.