Benefit of temporal fine structure to speech perception in noise measured with controlled temporal envelopes

Joanne M. Eaves[a)]
Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom

A. Quentin Summerfield
Department of Psychology and Hull-York Medical School, University of York, Heslington, York YO10 5DD, United Kingdom

Pádraig T. Kitterick
Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom

(Received 12 August 2010; revised 26 April 2011; accepted 27 April 2011)

Previous studies have assessed the importance of temporal fine structure (TFS) for speech perception in noise by comparing the performance of normal-hearing listeners in two conditions. In one condition, the stimuli have useful information in both their temporal envelopes and their TFS. In the other condition, stimuli are vocoded and contain useful information only in their temporal envelopes. However, these studies have confounded differences in TFS with differences in the temporal envelope. The present study manipulated the analytic signal of stimuli to preserve the temporal envelope between conditions with different TFS. The inclusion of informative TFS improved speech-reception thresholds for sentences presented in steady and modulated noise, demonstrating that there are significant benefits of including informative TFS even when the temporal envelope is controlled. It is likely that the results of previous studies largely reflect the benefits of TFS, rather than uncontrolled effects of changes in the temporal envelope. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3592237]

PACS number(s): 43.71.Gv, 43.71.Rt [CJP] Pages: 501–507

I. INTRODUCTION

Perceiving speech in a background of noise is challenging for many listeners. One factor which may be important is the ability to process the temporal fine structure (TFS) of sound. When a broadband signal such as speech is passed through a narrowband filter, the output of the filter can be thought of as consisting of two components: (i) the TFS and (ii) the temporal envelope. The TFS is the component of the output waveform whose amplitude oscillates at a frequency close to the center frequency of the filter. The temporal envelope of the output waveform is the slower variation in amplitude over time which is carried by the TFS (Moore, 2008).

The importance of TFS is shown indirectly by the difficulties experienced in noise by users of cochlear implants, which do not convey TFS precisely (Nie *et al.*, 2005), and by listeners with moderate-to-severe cochlear hearing loss who have been found to be poorer at encoding TFS than normal-hearing listeners (Hopkins and Moore, 2007; Lorenzi *et al.*, 2009). The importance of TFS has been shown directly in studies with listeners with normal hearing. Those studies have compared performance under two conditions. In one condition (ENV), the stimuli have useful information only in their temporal envelopes. In the other condition (ENV&TFS), the stimuli have useful information in both their temporal envelopes and their TFS. Listeners perceive speech more accurately in the ENV&TFS condition (Qin and Oxenham, 2003; Gnansia *et al.*, 2008, 2009; Lorenzi *et al.*, 2009; Hopkins and Moore, 2009).

In these studies, ENV stimuli were generated in four steps: (1) The broadband signal (e.g., a speech-plus-noise stimulus) was filtered into a series of narrow-band channels. (2) The temporal envelope was extracted in each channel. (3) The original TFS was replaced either by a sinusoid at the center frequency (CF) of the channel or by a narrow band of noise centered on the CF. The amplitude of the tone or noise was modulated by the envelope extracted in Step 2. (4) The resulting modulated signals were summed together across channels. In comparison, ENV&TFS stimuli were either generated by summing the channel signals formed in Step 1 or were identical to the original broadband signals.

Although different authors have implemented these steps in different ways (Table I), studies have in common the fact that the temporal envelope was processed differently in creating ENV stimuli compared with ENV&TFS stimuli. Differences arose in two studies (Qin and Oxenham, 2003; Gnansia *et al.*, 2009) because the unprocessed stimuli were used in the ENV&TFS condition, but rectification and low-pass filtering were used to extract the envelope in creating stimuli for the ENV condition. An additional difference arose in a third study (Gnansia *et al.*, 2008) which filtered the modulated sinewaves generated at Step 3 with the original analysis filters before summation at Step 4. The fourth study (Lorenzi *et al.*, 2009) created ENV&TFS stimuli by summing the channel signals from Step 1, but low-pass filtered the envelope in creating stimuli for the ENV condition. The fifth study (Hopkins and Moore, 2009) also created ENV&TFS stimuli by summing the channel signals from Step 1, but filtered the modulated sinewaves generated at Step 3 with the original analysis filters before summation at Step 4.

[a)]Author to whom correspondence should be addressed. Electronic mail: jo_eaves@yahoo.com

TABLE I. Processing steps used to create ENV and ENV&TFS stimuli in five studies. The steps are described in the Introduction. (ERB_N: equivalent rectangular bandwidth of an auditory filter in a young healthy adult with normal hearing. CF: center frequency. HWR: half-wave rectification. FWR: full-wave rectification. LPF: low-pass filter.)

Study	Condition	Step 1: Channel filtering	Step 2: Envelope extraction	Step 3: Replacement of TFS	Step 4: Reconstruction of signal
Qin and Oxenham (2003)	ENV	24 channels, with bandwidths of 1 ERB_N, and CFs ranging from 80 to 6000 Hz	HWR followed by LPF at the minimum of 300 Hz and $(ERB_N)/2$	Band-limited noise	Summation of band-limited noises
	ENV&TFS	Unprocessed			
Gnansia *et al*. (2008)	ENV	32 channels, with bandwidths of 1 ERB_N, and CFs ranging from 80 to 8583 Hz	FWR followed by LPF at 64 Hz	Sine wave	Band-pass filtering of each sine wave, followed by summation
	ENV&TFS	Unprocessed			
Gnansia *et al*. (2009)	ENV	32 channels, with bandwidths of 1 ERB_N, and CFs ranging from 80 to 8583 Hz	FWR followed by LPF at 64 Hz	Sine wave	Summation of sine waves
	ENV&TFS	Unprocessed			
Lorenzi *et al*. (2009)	ENV	16 channels, with bandwidths of 1 ERB_N at low CF*s*, and of 2 ERB_N at higher CFs, and CFs ranging from 80 to 8020 Hz	Hilbert Transform, followed by downsampling, LPF at 64 Hz, and upsampling	Sine wave	Summation of sine waves
	ENV&TFS	As above			Summation of channel signals from Step 1
Hopkins and Moore (2009)	ENV	32 channels, with bandwidths of 1 ERB_N, and CFs ranging from 100 to10,000 Hz	Hilbert Transform	Sine wave	Band-pass filtering of each sine wave, followed by summation
	ENV&TFS	As above			Summation of channel signals from Step 1

Thus, differences in the temporal envelope between ENV and ENV&TFS stimuli arose in either or both of Steps 2 and 4. Low-pass filtering in Step 2 explicitly removes higher modulation frequencies from the temporal envelope. Band-pass filtering in Step 4 has a similar effect. Consider that modulating a sinusoid introduces frequency components (side-bands) at frequencies of $f+m$ and $f-m$ Hz, where f and m are the carrier and modulation frequencies, respectively. Band-pass filtering attenuates the side-bands which might otherwise enter adjacent channels, but also distorts the temporal envelope because the amplitudes of the side-bands determine the envelope. The distortion is greater the higher the modulation frequency, because those side-bands are more widely spread from the carrier frequency. Therefore, band-pass filtering before summation effectively low-pass filters the temporal envelope.

Thus, in each study, the processing not only excluded informative TFS from ENV stimuli, but also modified the temporal envelope compared with ENV&TFS stimuli. Therefore, the comparison of the presence and absence of informative TFS was confounded with a difference in the temporal envelope. In the present study, we examined the benefit of including informative TFS on the intelligibility of speech masked by noise, while imposing the same temporal envelope both on ENV&TFS stimuli and on ENV stimuli.

We also addressed two further issues. The addition of informative TFS improves speech-reception thresholds more when the background noise is modulated sinusoidally in amplitude than when it is steady (Hopkins and Moore, 2009; Gnansia *et al.*, 2008), although this effect has not always been found (Qin and Oxenham, 2003).[1] Accordingly, we measured the benefit of TFS in the presence of both steady and modulated noise. We also addressed the question of whether intelligibility is affected by side-bands falling into adjacent channels. In one condition, the band-pass filtered (BPF) channel signals were band-pass filtered a second time before being summed at Step 4 ("twice-BPF"). In another condition, the channel signals were summed without being band-pass filtered a second time ("once-BPF"). In all conditions, we measured the benefit of TFS without confounding differences in the temporal envelope.

II. METHODS

A. Participants

Forty adults aged 18 – 42 yr (mean = 21.3 yr) were paid to participate. Sixteen participated in the main experiment. Sixteen, four, and four participated in the first, second, and third control experiments. All were native English speakers and had pure-tone sensitivity better than 20 dB hearing level at octave frequencies from 250 to 8000 Hz, inclusive. Each subject took part in one experiment only and none had previous experience of the stimuli or test procedures.

B. Stimuli

1. Speech materials

Two sets of speech materials were used: 270 IHR sentences (Macleod and Summerfield, 1990) and 360 IEEE sentences (Institute for Electrical and Electronic Engineers, 1969). Both sets had been spoken clearly and recorded digitally at a sampling rate of 20 000 samples/s with 16-bit amplitude quantization. Both sets were up-sampled to 44 100 samples/s before being processed. IHR sentences are lexically, syntactically, and semantically simple and of neutral predictability. They were spoken by an adult male talker of Standard British English. Each sentence contains three key words and the average duration of the sentences is 1.53 s. An example, with the key words in italics, is "*They moved* the *furniture*." IEEE sentences are more complex and less predictable. Each sentence contains five key words and the average duration is 2.66 s. They were spoken by a different adult male speaker of Standard British English. An example is "*Hop over* the *fence* and *plunge in*." The average voice fundamental frequency extracted using PRAAT (Boersma and Weenink, 2010) was 145 Hz (range 127–176 Hz) for the IHR sentences and 162 Hz (range 142–189 Hz) for the IEEE sentences.

2. Noises

Steady noises were synthesized by summing random-phase constant-amplitude sine waves at frequencies from 1 to 22 050 Hz, at 0.2 Hz intervals. The average long-term spectrum of each type of sentence was calculated with half-overlapping Hanning-windowed 4096-point Fast Fourier Transforms applied to every sentence in each set. These spectra were imposed on the noises that were used to mask each type of sentence. Noises had 100-ms raised-cosine onset and offset ramps. The duration of the noises was such that they started before, and finished after, each sentence. Modulated noises were generated by modifying the steady noises with Eq. 1 (Hopkins and Moore, 2009) where $F(t)$ and $N(t)$ are the waveforms of the modulated and steady noises, respectively,

$$F(t) = N(t) \times 10^{[\cos(2\pi 8t)-1]\times S}, \quad (1)$$

when S is 0.75, Eq. (1) produces noises that are modulated at a rate of 8 Hz on a dB scale with a 30-dB peak-to-valley difference. When S is 1.5, Eq. (1) produces noises that are modulated at a rate of 8 Hz on a dB scale with a 60-dB peak-to-valley difference.

3. Processing

Each sentence was combined with noise at 37 signal-to-noise ratios (SNRs) ranging from −36 dB to +36 dB in 2-dB steps. At 0 dB SNR, the total RMS powers of the speech and steady noise samples were equal on average. Negative SNRs were created by reducing the level of the speech. Positive SNRs were created by reducing the level of the noise. This method ensured that the stimuli would not be uncomfortably loud at any SNR.

The resulting speech-plus-noise stimuli were filtered using an array of 32 finite impulse response (FIR) filters described by Hopkins and Moore (2009). The filters were designed to be approximately 1-ERB_N wide (Glasberg and Moore, 1990) and to have moderately steep transition bands and minimal spectral ripple in their passbands. The frequency at the 6-dB-down point on the low-frequency side of the lowest filter was 100 Hz. The frequency at the 6-dB-down point on the high-frequency side of the highest filter was 10 000 Hz. This filterbank mimics the frequency analysis performed by a healthy adult peripheral auditory system. The signals at the output of each channel were realigned in time to compensate for the delays introduced by filtering.

Each filtered speech-plus-noise stimulus was processed in each of four ways. (1) ENV&TFS(once-BPF) stimuli were formed by summing the time-aligned signals at the outputs of the 32 filters. (2) To create ENV(once-BPF) stimuli, the temporal envelope of the signal in each channel was extracted using the Hilbert transform (Hilbert, 1912). The envelope was then used to modulate a sinusoid at the center frequency of the channel, and the resulting modulated sinusoids were summed across the 32 channels. The phases of the 32 sinusoids were chosen randomly. (3) ENV(twice-BPF) stimuli were created in the same way as ENV(once-BPF) stimuli except that, before being summed, the modulated sinusoids were band-pass filtered again with the original band-pass filters. (4) To create ENV&TFS(twice-BPF) stimuli, the envelope in each channel in the ENV(twice-BPF) condition was superimposed on the TFS in that channel from the ENV&TFS(once-BPF) condition. To do this, the analytic signal of the ENV&TFS(once-BPF) signal in that channel, derived using the Hilbert transform, was scaled at each time-point by the ratio difference between the ENV(once-BPF) and ENV(twice-BPF) envelopes. A Matlab (The Mathworks, 2007) script which illustrates the processing methods can be accessed from http://tinyurl.com/jasayork082010. In total, 630 sentences were processed at 37 SNRs with two types of noise, two processing strategies, and two filtering strategies, generating 186 480 stimuli.

Figure 1 plots the envelope of a segment of speech that was processed in each of the four ways described above. The effect on the temporal envelope of the stimuli of band-pass filtering once or twice can be seen by comparing the upper (once-BPF) and lower (twice-BPF) panels. The success of the processing to impose the same temporal envelope on ENV and ENV&TFS stimuli can be seen by comparing the

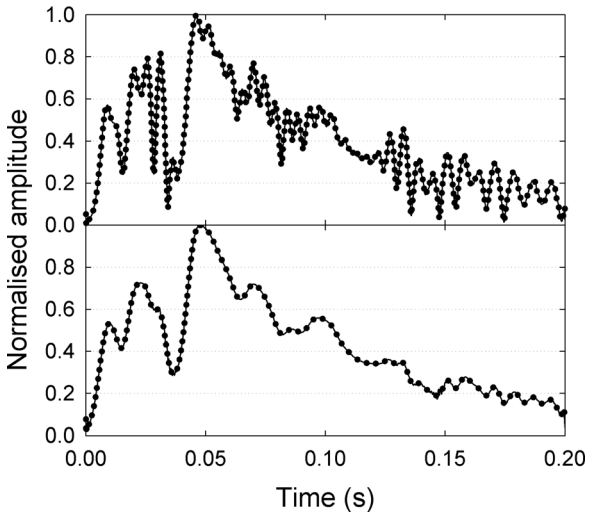

FIG. 1. Temporal envelopes, derived using the Hilbert transform, from stimuli processed in four ways. Upper panel: ENV&TFS(once-BPF) (continuous line) and ENV(once-BPF) (circles). Lower panel: ENV&TFS(twice-BPF) (continuous line) and ENV(twice-BPF) (circles). The alignment of the circles with the corresponding continuous lines demonstrates that the steps taken to control temporal envelopes were successful. The figure was generated by passing a 200-ms segment of the word "bull" from the sentence "The bull chased the lady" spoken in quiet through the FIR filter centered on 1033 Hz with low and high cut-off frequencies of 965 and 1100 Hz, respectively.

continuous lines (ENV&TFS) and the filled circles (ENV) in each panel.[2]

C. Procedures

The main experiment had a $2 \times 2 \times 2 \times 2$ factorial design with two sentence types (IHR and IEEE), two noise types (steady and modulated), two processing strategies (ENV and ENV&TFS), and two filtering strategies (once-BPF and twice-BPF). Stimuli were presented at a level (defined below) of 60 dB(A) sound pressure level (SPL). Sixteen subjects took part in this experiment. Three control experiments were run to check whether the audibility of the sentences or other floor effects limited the size of differences between conditions. In the first control experiment, 16 subjects listened to IHR sentences presented in quiet as well as in noise. Six conditions were created by combining two processing strategies (ENV, ENV&TFS) with three noise types (steady, 30-dB-modulated, and in quiet), at a presentation level of 60 dB(A) SPL. In the second control experiment, four subjects listened to IEEE sentences at presentation levels of 60 and 65 dB(A) SPL. Eight conditions were created by combining the two presentation levels with two processing strategies (ENV, ENV&TFS) and two noise types (steady, 30-dB-modulated). The third control experiment compared effects of modulated noises with peak-to-valley differences of 30 and 60 dB. Four subjects listened to IEEE sentences in six conditions, created by combining two processing strategies (ENV, ENV&TFS) and three noise types (steady, 30-dB-modulated, 60-dB-modulated) at a presentation level of 60 dB(A) SPL. The stimuli used in the control experiments had been passed through the band-pass filters once.

Stimuli were delivered binaurally through headphones (Sennheiser HD580). In different experiments, the average level of the sentences in noise at 0 dB SNR was either 60 or 65 dB(A) SPL at each ear. Levels were measured with a Brüel & Kjær artificial ear (type 4153) using the flat-plate adaptor, sound level meter (type 2260 Investigator), and microphone (type 4189).

A speech-reception threshold (SRT), defined as the SNR at which the accuracy of identifying key words was 50% correct, was estimated for each participant in each condition. Lists of 30 sentences were presented. One sentence was presented on each trial and the participant's task was to repeat back as much of the sentence as possible. Responses were scored as correct if three out of three key words were reported correctly in IHR sentences, or 3, 4, or 5 out of 5 key words were reported correctly in IEEE sentences. The SNR was controlled adaptively using the ascending method of limits (e.g., Plomp and Mimpen, 1979). The first sentence in each list was presented repeatedly, starting at an SNR below the participant's SRT. The SNR was increased in 4-dB steps until the response to the sentence was correct. The SNR was then reduced by 2 dB. The other sentences in the list were then presented, and the SNR was decreased/increased by 2 dB following each correct/incorrect response, respectively.

The total number of key words reported correctly and the total number of key words presented were calculated for each SNR. These data were converted into Probit units (Finney, 1971) and the slope and intercept parameters of the transformed data were estimated using linear regression. The probit data were converted back into the proportion of key words correct at each SNR so that the SNR at which listeners identified key words with an accuracy of 50% correct could be estimated. In this way, performance with the two types of sentence could be compared at the same point on the psychometric function.[3]

Across participants, sentence lists were presented an equal number of times in each condition, and the order of the conditions was counterbalanced using a Williams-square design (Williams, 1949). Participants responded to 15 practice trials in each condition, presented in accordance with the adaptive routine, before each SRT was estimated.[4]

III. RESULTS

A. Effects of sentence type, noise type, processing strategy, and number of band-pass filtering operations

In the main experiment, SRTs were lower with IHR sentences than IEEE sentences by 0.6 dB [95% confidence interval (c.i.) 0.3 to 0.8 dB], with modulated noise than steady noise by 8.2 dB (95% c.i. 7.9 to 8.5 dB), with ENV&TFS than ENV by 2.4 dB (95% c.i. 2.2 to 2.7 dB), and with once-BPF than twice-BPF stimuli by 0.2 dB (95% c.i. -0.1 to 0.4 dB). A $2 \times 2 \times 2 \times 2$ analysis of variance confirmed significant effects of sentence type [$F(1, 15) = 15.50$, $p = 0.001$], noise type [$F(1, 15) = 1292.70$, $p < 0.001$], and processing strategy [$F(1, 15) = 337.95$, $p < 0.001$], but not the number of band-pass filtering operations [$F(1, 15) = 2.22$, *n.s.*]. There was one significant interaction. It arose between noise

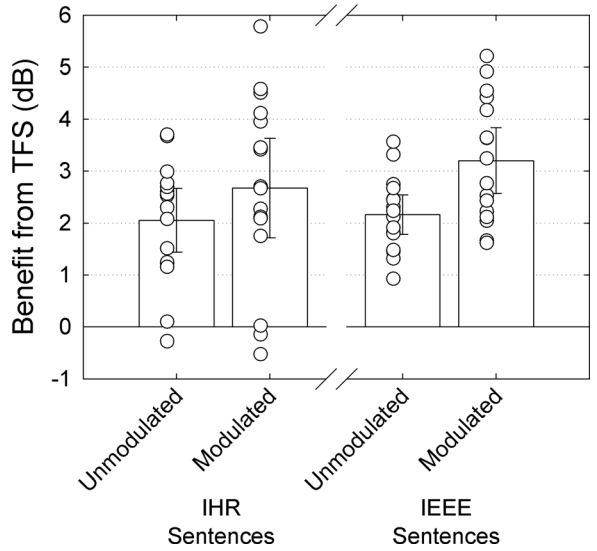

FIG. 2. Benefit from TFS with each type of sentence in steady and modulated noise. Benefit was calculated by subtracting the SRT in each ENV&TFS condition from the SRT in the corresponding ENV condition. Bars plot the mean benefit for each condition. Error bars plot 95% confidence intervals of the means. Open circles plot data for individual participants.

type and processing strategy [$F(1, 15) = 15.02$, $p = 0.001$] because the effect of including informative TFS was larger with modulated noise (3.0 dB) than with steady noise (1.9 dB) (Fig. 2).

B. Control experiments

Three aspects of the results of the control experiments demonstrate that SRTs in noise were not limited by the absolute audibility of the sentences or by other floor effects. First, when sentences were presented in quiet, the level of the speech at threshold was 18.0 dB lower (95% c.i. 16.7 to 19.2 dB) than the level of the speech at the lowest SRT in noise (ENV&TFS in a modulated noise). Second, mean SRTs in noise did not differ between conditions where stimuli were presented at 65 dB SPL (− 10.6 dB) and at 60 dB SPL (− 10.9 dB) (mean difference 0.3 dB, 95% c.i. − 0.3 to 0.9 dB). Additionally, in conditions with modulated noise there was no difference in the size of the benefit of TFS which was 2.3 dB (95% c.i. 0.6 to 4.1 dB) when stimuli were presented at 60 dB SPL and 2.2 dB (95% c.i 0.5 to 3.9 dB) when stimuli were presented at 65 dB SPL (mean difference 0.2 dB, 95% c.i. −3.1 to 2.7 dB). Third, while mean SRTs were significantly lower by 5.0 dB (95% c.i. 4.0 to 6.0 dB) with noises that had a 60−dB modulation depth than noises that had a 30-dB modulation depth, the benefit of TFS did not differ. The benefit was 3.0 dB (95% c.i. 1.1 to 4.8 dB) with a 30-dB-modulated noise and 3.5 dB (95% c.i. 2.3 to 4.7 dB) with a 60-dB-modulated noise (mean difference 0.6 dB, 95% c.i. − 2.4 to 3.6 dB).

Table II summarizes the results of the main experiment and the three control experiments by listing the average improvement in SRT for each condition, relative to the condition that, on average, produced the highest SRT (once-BPF IEEE sentences in the ENV condition, presented in steady noise at 60 dB SPL). The numerical similarity of the entries within each of the four main quadrants of the table reinforces the point that the factors which influenced SRTs were noise type, processing strategy, and modulation depth of the noise. Sentence type, presentation level, and number of band-pass filtering operations did not have a systematic influence on SRTs.

IV. DISCUSSION

Previous studies that have assessed the role of TFS in speech perception in noise, by comparing performance with ENV&TFS stimuli to performance with ENV stimuli, have not controlled differences in the temporal envelope between the two types of stimuli. The present study demonstrates that there are significant benefits of including informative TFS even when the temporal envelope is controlled. Moreover, the size of the benefit does not differ significantly between conditions in which the channel signals are band-pass filtered once or twice before summation.

TABLE II. Average improvement in SRT in dB for each condition, using different combinations of processing strategy (ENV&TFS, ENV), sentence type (IHR, IEEE), number of band-pass filtering operations (once, twice), noise type (steady, 30-dB-, and 60-dB-modulated), and presentation level (60 dB SPL, 65 dB SPL). Averages were calculated by subtracting the mean SRT in each condition from the mean SRT in the condition with the highest mean SRT (ENV, IEEE, Once, Steady, 60 dB). Participants in both the main experiment and the three control experiments contributed to the averages.

			Steady		30-dB-modulated		60-dB-modulated
			Presentation level				
			60 dB (SPL)	65 dB (SPL)	60 dB (SPL)	65 dB (SPL)	60 dB (SPL)
ENV	IHR	Once-BPF	0.2 ($N = 32$)	N/A	8.1 ($N = 32$)	N/A	N/A
		Twice-BPF	0.7 ($N = 16$)	N/A	7.9 ($N = 16$)	N/A	N/A
	IEEE	Once-BPF	0.0 ($N = 24$)	0.2 ($N = 4$)	8.0 ($N = 24$)	9.5 ($N = 4$)	13.0 ($N = 4$)
		Twice-BPF	0.2 ($N = 16$)	N/A	7.4 ($N = 16$)	N/A	N/A
ENV&TFS	IHR	Once-BPF	2.1 ($N = 32$)	N/A	11.1 ($N = 32$)	N/A	N/A
		Twice-BPF	2.3 ($N = 16$)	N/A	11.2 ($N = 16$)	N/A	N/A
	IEEE	Once-BPF	2.0 ($N = 24$)	2.1 ($N = 4$)	11.0 ($N = 24$)	11.6 ($N = 4$)	16.6 ($N = 4$)
		Twice-BPF	1.9 ($N = 16$)	N/A	10.2 ($N = 16$)	N/A	N/A

A. Band-pass filtering once or twice

In the main experiment, band-pass filtering twice compared to once raised SRTs, on average, by only 0.2 dB and by only 0.7 dB in the condition showing the largest effect (IHR sentences in 30-dB-modulated noise containing envelope-only information). Nor did the effect of including informative TFS depend on whether stimuli were band-pass filtered once (mean benefit 2.5 dB) or twice (mean benefit 2.3 dB). Thus, previous estimates of the benefit of TFS (Qin and Oxenham, 2003; Gnansia *et al*., 2008, 2009; Lorenzi *et al.,* 2009; Hopkins and Moore, 2009) are likely mainly to have reflected the consequences of enhancing TFS, rather than the consequences of enhancing the temporal envelope.[5]

The likely reason why there was only a small difference between SRTs measured with stimuli which had been band-pass filtered once or twice is that twice filtering selectively attenuates higher modulation frequencies (upper vs lower panel of Fig. 1), which make only a small contribution to speech perception in noise. For example, Drullman *et al.* (1994) demonstrated that attenuating modulation frequencies of 16 Hz and above raised SRTs for sentences in speech-spectrum shaped steady noise by only about 1 dB, whereas attenuating modulation frequencies of 4 Hz and below raised SRTs by more than 6 dB. We note, however, that Stone *et al*. (2008) demonstrated that the intelligibility of vocoded speech can drop by about 20 percentage points by removing modulation frequencies above 45 Hz when the target speech is presented against a background of a competing talker. Thus, a larger difference between once band-pass filtering and twice band-pass filtering might be found when speech is masked by speech compared to when speech is masked by noise.

The size of the difference in the temporal envelope between once-BPF and twice-BPF stimuli depends on the method used to extract the envelope. The Hilbert transform exposes more differences (e.g., in Fig. 1) than are found if the envelope is extracted by rectification and low-pass filtering at half the channel bandwidth, because the low-pass filter removes the higher modulation frequencies that would otherwise be removed by the second band-pass filter. A more general issue that affects all studies which manipulate the TFS while seeking to control the temporal envelope is that any definition of the envelope in the signal domain may differ from its representation in the auditory system. Thus, changes in the TFS which preserve the Hilbert envelope may nonetheless change the auditory envelope. As a result, differences in performance between conditions with informative and uninformative TFS may be influenced by uncontrolled changes in the auditory envelope. These effects are likely to be small, but their size is yet to be determined.

B. Comparison with Hopkins and Moore (2009)

The inclusion of informative TFS improved SRTs significantly more in modulated noise than in steady noise (Fig. 2), as shown by Hopkins and Moore (2009).[6] However, the benefit of TFS in modulated noise in the present experiment (3.4 dB) was only about half the benefit reported by Hopkins and Moore (6.0 dB). This difference may have arisen from any or all of five differences between their procedures and ours. First, the experiments used different recordings of the IEEE sentences, spoken by different talkers. However, the similarity of the results obtained in the present experiment with IHR and IEEE sentences suggests that differences in talker and in the linguistic complexity of clearly spoken sentence materials may have only a small effect on the benefit from including informative TFS. Second, in the current study, negative SNRs were created by attenuating the speech stimuli, whereas in Hopkins and Moore the noise level was increased. This difference, however, might have been expected to enlarge the benefits of TFS in the present experiment, because increasingly good performance was not penalized by an increasingly high level of noise. Third, the subjects differed. The present results, however, show strong consistency between experiments, suggesting that differences among groups of young normally hearing participants on the present tasks are small. Fourth, Hopkins and Moore compared once-BPF ENV&TFS stimuli with twice-BPF ENV stimuli, whereas we made comparisons while controlling the number of filtering operations. In the present data, the comparison of once-BPF ENV&TFS IEEE sentences with twice-BPF ENV IEEE sentences in a modulated noise yielded an effect of TFS of 3.4 dB. This effect was numerically (though not statistically) larger than the effect of 2.8 dB when twice-BPF ENV&TFS IEEE sentences were compared with twice-BPF ENV IEEE sentences. Thus, differences in the number of filtering operations may have contributed to the difference between the benefit of TFS reported by Hopkins and Moore (2009) and the benefit measured in the present study, but do not explain all of the difference.

A fifth possible reason for the difference in the measured benefit of TFS between the present experiments and Hopkins and Moore (2009) is that they exposed their participants to more tone-vocoded sentences before measuring SRTs than we did. The effects of TFS on SRTs could be larger when participants are more familiar with processed stimuli and their SRTs are lower. To examine this issue, we conducted a supplementary analysis of the data from each pair of ENV and ENV&TFS conditions in Table II for which 16 or more participants provided data. We tested the idea that better-performing participants show larger benefits from TFS by calculating the product-moment correlation between the benefit of TFS (i.e., the difference in SRTs between the ENV and ENV&TFS conditions) and the average of the SRTs in the ENV and ENV&TFS conditions, thus avoiding the problem of mathematical coupling (Oldham, 1962). Only one of the eight correlations was significant. However, it showed the opposite pattern from the one predicted; i.e., a larger benefit was associated with poorer performance. Thus, there is no evidence in the present data that a lack of familiarity with the stimuli limited the benefits of TFS.

In summary, it is not clear why we found smaller benefits of including informative TFS than did Hopkins and Moore (2009). Further work is required to identify all of the factors which influence the benefit of TFS to speech perception in noise.

V. CONCLUSIONS

Including informative TFS significantly increases the intelligibility of sentences in noise even when the temporal envelope of the sentences is controlled. Moreover, the benefit of TFS is largely unaffected by whether the temporal envelope in each channel is band-pass filtered once or twice. Thus, the advantage of replacing uninformative TFS with informative TFS measured in previous studies is most likely due to the enhanced representation of TFS rather than the increased definition of the temporal envelope, or the avoidance of unwanted incursions of side-bands into adjacent channels. Nonetheless, researchers wishing to estimate the benefits of TFS while avoiding confounding the processing strategy with the number of filtering operations could process stimuli in the ways described in Sec. II B 3 of this article and at http://tinyurl.com/jasayork082010.

ACKNOWLEDGMENTS

J.M.E. was supported by the Goodricke Appeal Fund for Deaf Children and Young People in North Yorkshire. P.T.K. was supported by Advanced Bionics SARL. We thank Kathryn Hopkins for helpful discussions, and Brian Moore and an anonymous reviewer for suggestions which improved the clarity of the manuscript.

[1]Compare the difference between steady and modulated noise in the unprocessed and 24-channel conditions in Fig. 3 of Qin and Oxenham (2003).

[2]Altering the temporal envelope in this way can generate artifacts when the sampled amplitude of the original envelope is close to zero. No such artifacts arise in Fig. 1, but artifacts are illustrated in the output of the Matlab script.

[3]SRTs were also estimated by averaging the SNRs at which the last 26 sentences in each list were presented. SRTs estimated in this way were 0.9 dB lower on average than SRTs estimated by Probit analysis, but showed the same pattern of statistical significance.

[4]Due to the limited number of IHR sentences, participants were familiarized with Bamford-Kowal-Bench (BKB) sentences (Bench *et al*., 1979) in conditions involving IHR sentences. BKB sentences are similar in complexity to IHR sentences and were spoken by the same talker.

[5]Gnansia *et al*. (2009) noted that extracting the temporal envelope in their ENV condition by full-wave rectification followed by low-pass filtering at 64 Hz distorted the envelope compared with their ENV&TFS condition which was unprocessed. They reported a supplementary experiment in which the 64-Hz low-pass filter was replaced by one set to half the channel bandwidth for channels whose bandwidths exceeded 128 Hz, thereby preserving envelope modulations at higher rates. This change had no overall effect on masking release for vowel-consonant-vowel syllables between steady and amplitude-modulated noise.

[6]A full explanation for why these differences in SRT arise would require an analysis of complete psychometric functions using, for example, the methods described by Bernstein and Grant (2009).

Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol. **13**, 108–112.

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

Boersma, P., and Weenink, D. (**2010**). "Praat: doing phonetics by computer [Computer program]. Version 5.1.38," URL http://www.praat.org/, retrieved on 1 May 2010.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Finney, D. J. (**1971**). *Probit Analysis* (Cambridge University Press, Cambridge), pp. 1–252.

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.

Hilbert, D. (**1912**). *Grundzüge einer Allgemeinen Theorie der linearen Integralgleichungen (Basics of a general theory of linear integral equations)* (Teubner, Leizpig), pp. 1–310.

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.

Institute For Electrical And Electronic Engineers (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Lorenzi, C., Debruille, L. Garnier, S. Fleuriot, P., and Moore, B. C. J. (**2009**). "Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal," J. Acoust. Soc. Am. **125**, 27–30.

Macleod, A., and Summerfield, Q. (**1990**). "A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," Br. J. Audiol. **24**, 29–43.

Moore, B. C. J. (**2008**). "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," J. Assoc. Res. Otolaryngol. **9**, 399–406.

Nie, K., Stickney, G., and Zeng, F. G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.

Oldham, P. D. (**1962**). "A note on the analysis of repeated measurements of the same subjects," J. Chronic Dis. **15**, 969–977.

Plomp, R., and Mimpen, A. (**1979**). "Improving the reliability of testing speech reception threshold for sentences," Audiol. **18**, 43–52.

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.

Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (**2008**). "Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region," J. Acoust. Soc. Am. **124**, 2272–2282.

The Mathworks (**2007**). *MatLab Rv7.1*, The Mathworks, Natick MA.

Williams, E. J. (**1949**). "Experimental designs balanced for the estimation of residual effects of treatments," Aust. J. Sci. Res. Ser. A – Phys. Sci. **2**, 149–168.