The effect of tone-vocoding on spatial release from masking for old, hearing-impaired listeners[a)]

Andrew King[b)], Kathryn Hopkins and Christopher J. Plack
Manchester Centre for Audiology and Deafness, University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom

Niels Henrik Pontoppidan, Lars Bramsløw, Renskje K. Hietkamp, Marianna Vatti, Atefeh Hafez,
Eriksholm Research Centre, Oticon A/S, Rørtangvej 20, Snekkersten, Denmark

Running title: Temporal fine structure and spatial hearing

a) Portions of this work were presented at the 4th International Symposium on Auditory and Audiological Research, Nyborg, Denmark, August 2013.

b) Author to whom correspondence should be addressed.
Current address: Laboratoire des systèmes perceptifs, Département d’études cognitives, École normale supérieure, PSL Research University, CNRS , 75005 Paris, France
Email: andrew.king@ens.fr

Old, hearing-impaired listeners generally benefit little from lateral separation of multiple talkers when listening to one of them. This study aimed to determine how spatial release from masking (SRM) in such listeners is affected when the interaural time differences (ITDs) in the temporal fine structure (TFS) are manipulated by tone-vocoding (TVC) at the ears by a master hearing aid system. Word recall was compared, with and without TVC, when target and masker sentences from a closed set were played simultaneously from the front loudspeaker (co-located) and when the maskers were played 45° to the left and right of the listener (separated). For 20 hearing-impaired listeners aged 64 to 86, SRM was 3.7 dB smaller with TVC than without TVC. This difference in SRM correlated with mean audiometric thresholds below 1.5 kHz, even when monaural TFS sensitivity (discrimination of frequency-shifts in identically filtered complexes) was partialed out, suggesting that low-frequency audiometric thresholds may be a good indicator of candidacy for hearing aids that preserve ITDs. The TVC difference in SRM was not correlated with age, pure-tone ITD thresholds, nor fundamental frequency difference limens, and only with monaural TFS sensitivity before control for low-frequency audiometric thresholds.

PACS number: 4366Pn, 4366Qp, 4371Rt, 4371Ky, 4371Lz

I. INTRODUCTION

Old, hearing-impaired (HI) people struggle to understand speech when they are in a noisy environment (Duquesnoy, 1983; Working Group on Speech Understanding and Aging, 1988; Gatehouse and Noble, 2004; Divenyi *et al.*, 2005; Helfer and Freyman, 2008). This may be due, in part, to a degraded ability to separate sounds that come from different directions. Interaural time differences (ITDs) and interaural level differences (ILDs) occur when a sound source is located outside the median plane; the sound reaches one ear before the other and the head causes an acoustic shadow for wavelengths below the head size (Rayleigh, 1907; Kuhn, 1977). It is possible to use these differences to separate sounds that arrive from different directions (Middlebrooks and Green, 1991; Best *et al.*, 2004; Bremen and Middlebrooks, 2013). In the laboratory, speech intelligibility is better when the target and masker sounds are presented from different directions around the listener, rather than all from the same direction (Bronkhorst, 2000). When the maskers are speech signals, the spatial release from masking (SRM) is typically 10 to 15 dB for normal-hearing (NH) listeners (Behrens *et al.*, 2008; Marrone *et al.*, 2008b). HI people (particularly if they are elderly) vary widely in the SRM they gain, from close-to-normal to none at all (Marrone *et al.*, 2008a; Gallun *et al.*, 2013; Jakien *et al.*, 2017), even with compensation for reduced audibility (Marrone *et al.*, 2008c; Neher *et al.*, 2009). This suggests that deficits in the coding of the spatial separation of sounds may occur with some forms of hearing impairment. A better understanding of why some people benefit from spatial cues, and others do not, would help develop and prescribe hearing aids that preserve or sacrifice spatial cues depending on a listener’s ability to use these cues.

ITDs, which are typically less than 0.7 ms in humans, may be encoded for sound below 1.5 kHz by comparison across ears of the synchronized firing of the auditory nerve fibers to the phase of basilar membrane displacement (Rose *et al.*, 1967). This phase locking codes the time intervals between corresponding peaks in the band-pass filtered output and represents the temporal fine structure (TFS) of sounds, which would not be available from the place of excitation on the basilar membrane alone (Cariani and Delgutte, 1996). TFS appears to benefit pitch perception and lateralization (Smith *et al.*, 2002; Drennan *et al.*, 2007). Some studies suggest TFS aids speech perception in fluctuating background noise (*e.g.*, Hopkins *et al.*, 2008), whilst others do not (e.g., Oxenham and Simonson, 2009; Apoux *et al.*, 2013).

The benefit of TFS to speech perception can be measured by manipulating the TFS with signal processing such as vocoding (Dudley, 1939). Vocoding involves filtering a waveform into a series of frequency band-pass channels, extracting the envelopes from each band and multiplying them by new carrier signals. These modified channels are then summed. In quiet, vocoded speech can be understood even with the use of a few frequency channels (*e.g.*, Shannon *et al.*, 1995; Loizou *et al.*, 1999; Smith *et al.*, 2002). However, in the presence of noise, vocoded speech requires higher signal-to-noise ratios (SNRs) than non-vocoded speech to be intelligible (Qin and Oxenham, 2003; Stone and Moore, 2003).

With sinusoidal vocoder carriers (tone-vocoding, or TVC), changing the phase of the sinusoids between the ears allows one to test how ITDs carried in the TFS contribute to SRM. Andersen *et al.* (2010) compared SRM with and without TVC. They used headphones and head-related impulse responses (HRIRs) from a head and torso simulator (Algazi *et al.*, 2001) to simulate lateral separation of speech signals. Andersen *et al.* (2010) found that young, NH listeners' speech reception thresholds (SRTs) were lowest (best) when no TVC was applied; SRTs were elevated (poorer) by 5.9 dB when the TVC was applied to the signals before the HRIRs (preserving the ITDs but not the original monaural TFS), suggesting that vocoding decreases performance even if spatial cues are preserved. However, SRTs were elevated a further 2.4 dB when the TVC was applied in phase across ears after the HRIRs, effectively removing the differences in ITDs between the speech signals. This extra elevation suggests that binaural TFS cues, even carried by TVC signals, can help speech intelligibility. More recent evidence suggests a greater role of binaural TFS cues and a smaller role of the original speech TFS: Swaminathan *et al*. (2016) studied the effect on SRM of noise vocoding before convolving speech with HRIRs. They retained or removed ITD-based cues of spatial separation in the TFS by vocoding with either the same or uncorrelated noises at each ear, respectively. Swaminathan *et al*. (2016) found that correlated-noise vocoding produced similar SRTs and SRM to non-vocoded stimuli, but vocoding with interaurally uncorrelated noises produced poorer SRTs and SRM. The effect of noise vocoding on SRM was predominantly driven by the interaural correlation of the noise carriers below 1.5 kHz. On the other hand, Garadat *et al.* (2009) conducted a similar study to Andersen *et al.* (2010), but found that greater SRM was achieved with the TVC than without, and that there was no difference in performance between applying vocoding before or after HRIRs. However, Garadat *et al.* (2009) only used a single masker sentence lateralized to one side, allowing the listeners to take advantage of the improved SNR at the other ear.

The reduced effect of vocoding for HI listeners, compared to NH listeners, suggests that hearing loss may reduce the ability to use TFS for speech perception with modulating maskers, such as competing talkers (Hopkins *et al.*, 2008; Hopkins and Moore, 2010b). Hopkins *et al*. (2008) found that NH listeners performed worse with TVC speech than with the original speech, whereas HI listeners performed about the same as the NH listeners did with the TVC speech, regardless of the speech processing. This suggests that HI listeners do not benefit from the TFS in the original speech as much as NH listeners.

Both hearing loss and old age are associated with impaired TFS sensitivity, as seen by poorer discrimination of harmonic from inharmonic complexes filtered with identical pass-bands (*e.g.*, Hopkins and Moore, 2007; 2011) and pure-tone or TFS ITD detection (*e.g.*, Lacher-Fougère and Demany, 2005; King *et al.*, 2014). Even with hearing loss below 2 kHz absent or controlled for, old listeners have poorer TFS sensitivity (*e.g.*, Ross *et al.*, 2007;

Moore *et al.*, 2012b) and a reduced advantage of binaural hearing (Warren *et al.*, 1978; Pichora-Fuller and Schneider, 1992) than young listeners. Aging effects may underlie poor speech perception in noisy environments and result in less SRM (Gelfand *et al.*, 1988; Marrone *et al.*, 2008a; Marrone *et al.*, 2008c). It is possible that age exacerbates the effect of hearing loss on the ability to use TFS in speech perception.

The first aim of the current study was to determine whether or not older, HI listeners are able to use TFS to achieve SRM, by comparing SRM for speech with and without TVC processing. This may help determine whether or not hearing devices for older HI listeners need to preserve TFS ITDs or not. The second aim was to identify a good measure to predict which HI listeners can benefit from TFS in separating speech, to improve hearing aid prescription and fitting. For this, individual differences in HI listeners' SRTs and SRM, with and without TVC, were compared to their age, hearing loss and performance on three psychoacoustic tasks: one of monaural TFS sensitivity, one of binaural TFS sensitivity, and one of monaural temporal envelope sensitivity.

The monaural measure of TFS sensitivity was the TFS1 task (Moore and Sęk, 2009a), which has previously been shown to correlate with monaural SRTs in modulated speech-shaped noise, even after controlling for listeners' audiometric thresholds (Hopkins and Moore, 2011). It involves discrimination of harmonic from inharmonic complexes filtered with identical pass-bands. As the components in the inharmonic complex are shifted in frequency equally, the modulation rate (envelope) is the same for both harmonic and inharmonic complexes. This task is thought to rely exclusively on TFS sensitivity if the components within the pass-band are of a sufficiently high harmonic number to remain unresolved by the cochlea. However, small excitation pattern differences might still allow discrimination (Micheyl *et al.*, 2010).

To measure TFS sensitivity at low frequencies, Hopkins and Moore (2011) used the binaural task of ITD detection. Monaural SRTs were not correlated with ITD detection after controlling for audiometric threshold. However, ITD detection may relate to speech tests that emphasize binaural advantages to speech perception, such as the one in the current study, more than to monaural SRTs. Strouse *et al.* (1998) found no correlation between detection of ITDs in click trains and the binaural masking level difference (BMLD; the masking release produced by presenting a signal in anti-phase with a diotic masker) in speech for young and old NH listeners. However, detection of ITDs in click trains does not test TFS sensitivity, unlike detection of ITDs in pure tones (with synchronous onsets and offsets). Strelcyk and Dau (2009) measured ITD-based lateralization and BMLDs for tones in noise, and found both to be correlated with SRTs in laterally separated, speech-shaped noise. Lower SRTs in laterally separated speech maskers are also associated with lower ITD detection thresholds (Neher *et al.*, 2012) and the ability to detect ITDs at higher carrier frequencies (Neher *et al.*, 2011). Therefore, in the current study, pure-tone ITD detection was selected as the binaural task of TFS sensitivity.

The measure of temporal envelope sensitivity used in the current study was the fundamental frequency difference limen (F0DL). An F0DL is a listener's threshold for discriminating two harmonic complexes with different F0s and hence different modulation rates. If the harmonics are resolved, envelope cues may be weak and spectral cues may be used, but if the harmonics are unresolved, the use of TFS cues may be limited and listeners may rely on envelope cues (Oxenham *et al.*, 2009). This task was selected as the differing F0s of speech from different talkers may be used as a cue to segregate speech streams (*e.g.*, Brokx and Nooteboom, 1982). Sensitivity to this cue may impact on the amount of masking interfering talkers produce and possibly interact with SRM (Best *et al.*, 2012).

The hypotheses for the relations between the speech perception and psychoacoustic tasks were as follows:

1) That better TFS1 scores and ITD detection would correlate with lower SRTs in the conditions without TVC, and with greater differences between SRTs with and without TVC;

2) That ITD detection, in particular, would correlate with SRTs with spatially separated talkers without TVC, SRM without TVC and the difference in SRM due to TVC;

3) That F0DLs would correlate with SRTs in conditions with TVC, and SRM with TVC.

Additionally, if any relations between TFS sensitivity and SRTs or SRM were found, partial correlations were planned to determine if the thresholds were explained by age, hearing loss or F0DLs (the latter may reveal variance in TFS1 scores and ITD detection not specific to TFS, such as general temporal processing or psychoacoustic task performance).

II. METHODS

A. Listeners

Twenty listeners aged 64–86 (mean=72 years) and all had bilateral, gently-sloping sensorineural hearing loss (see Fig. 1). Mean audiometric pure-tone thresholds across listeners and frequencies (0.125, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, and 8 kHz) were 47 dB hearing level (HL) for the right ears and 46 dB HL for the left ears. Thresholds ranged from 10 to 40 dB HL at 0.125 kHz and 45 to 85 dB HL at 8 kHz; the average spread was 35 dB HL. Listener age was not correlated with either low-frequency (mean from 0.125 to 1.5 kHz) audiometric thresholds (PTA_{LF}; r=0.01), or high-frequency (mean from 2 to 8 kHz) audiometric thresholds (r=0.19). Listeners were screened for conductive or mixed hearing losses. Asymmetry across ears was < 10 dB for the frequency-average (0.125 to 8 kHz) audiogram. The largest asymmetries at a single frequency were 20 dB (3 listeners). All listeners spoke Danish as their first language and were recruited from a database maintained at the Eriksholm Research Centre, where all the experiments took place. The same database was used by Neher *et al.*, (2011; 2012) approximately five years earlier, but only one listener in the current study also participated in the studies described by Neher *et al.* (2011; 2012).

The sample size of 20 was selected to achieve a power of at least 0.8 for the interaction between the effects of TVC and spatial configuration. This was determined from an expected medium-to-large effect size for the interaction, based on an approximate Cohen's *d* of 0.7 for the effect of TVC on speech perception in competing speech for HI listeners (Hopkins *et al.*, 2008) and an approximate Cohen's *d* of 1.1 for the effect of SRM for HI listeners (Neher *et al.*, 2009). However, it is possible that the correlations between performance on the psychoacoustic tasks and SRM were not sufficiently powered. Observed power for the correlations, based on an effect size of the mean coefficient of determination (r^2) from all the correlations of 0.17 and an α of 0.05, was 0.55 for 17 listeners and 0.62 for 20 listeners. Power calculations were performed with GPower 3.1.3 (Kiel, Germany).

B. Speech test

1. Stimuli and Setup

A loudspeaker setup and a Master Hearing Aid system for simulating hearing-aid signal processing (MHA; HörTech, 2008; described in Grimm *et al.*, 2006) were used rather than headphones and dummy-head HRIRs (as used by Andersen *et al.*, 2010). This was expected to be more ecologically valid for behind-the-ear (BTE) hearing aid users, because although the MHA provides gain corrections for the lost outer ear gain, its microphones are located in BTE hearing aid shells, which disrupt the natural directional cues provided by the pinnae. Also, using a loudspeaker setup allows head movements without the sound sources moving with the head (as headphones do).

A special version of the Danish Dantale II corpus (Wagener *et al.*, 2003) designed for spatial speech-on-speech testing was used (Behrens *et al.*, 2008). Recorded words were spoken by three Danish females. Words were selected from a closed set. Sentences were five words long and always followed the same structure and order: a person's name, verb, number, adjective, and object. For example, "*Henning købte tre smukke ringe"* ("Henning bought three beautiful rings"). The target sentence was always played from the front loudspeaker (0° relative to the listener) in an anechoic room. Two masker sentences (M_1 and M_2) were played at the same time as the target, at various target-to-masker ratios (TMRs). The three sentences—each spoken by a different female talker—began with a "Ready" prompt. The first word (the person's name) in the target sentence was displayed to the listener via a computer screen hanging above the front loudspeaker to cue which sentence to listen to and recall. In each trial, listeners had to recall as many words from the target sentence as possible. Two spatial configurations were used: a Co-located configuration, where M_1 and M_2 were played from the same front loudspeaker as the target, and a Separated configuration, where loudspeakers positioned at azimuths of −45° and +45° relative to the listener played signals M_1 and M_2, respectively. Symmetrically separated maskers were used to minimize the benefits of increased TMR at one ear that occur with asymmetrical maskers (Marrone *et al.*, 2008a). The maximum root mean square (RMS) sound pressure level at the center of the listener's head position was 70 dB sound pressure level (SPL). For positive TMRs, M_1 and M_2 were attenuated whilst the target level remained fixed. For negative TMRs, the target was attenuated whilst M_1 and M_2 remained fixed in level.

2. Master Hearing Aid signal processing

Listeners wore behind-the-ear hearing aid microphones which recorded the sounds and sent the signals to a control computer running the hearing-aid simulation on the MHA research platform. The MHA split the signals into 512 linear sub-bands using a fast Fourier transform and corrected for the microphones' frequency responses. The complex values in these sub-bands were summed into 32 logarithmically spaced sub-bands with rectangular bandwidths equivalent to NH listeners' auditory bandwidths at moderate sound levels (ERB_N, Glasberg and Moore, 1990). The 32 non-linear sub-bands had a combined pass-band from 0.1 to 10 kHz. Since the complex value in each sub-band had an imaginary part that was a 90° phase-shifted version of the real part of any given time sample, the envelopes were extracted as the absolute value of each sub-band complex value (equivalent to the Hilbert envelope). For each sub-band, the cut-off frequency of the extracted envelope was equal to half the sub-band bandwidth. Two MHA conditions were used, one with TVC and one without TVC. In the TVC condition the envelopes were multiplied by a pure tone, in phase across ears, at the band center frequency, thus making the ITD in the TFS zero. The modulated tones were combined and amplified with multi-band linear gain following CAMEQ specifications (Moore and Glasberg, 1998) to correct for hearing loss and also for outer-ear gain and the hearing aid receiver (output) frequency response. Finally, the signal was high-pass filtered with a 100-Hz cutoff. In the condition without TVC, the processing followed the same procedure, except that each extracted envelope was multiplied by the phase angle of the complex signal of the corresponding sub-band, rather than pure tones, to restore the original TFS. The MHA signal processing (either with, or without, TVC) produced an overall delay of approximately 40 ms. The output was presented to the listener from the hearing aid receiver via Etymotic Research foam plugs for each listener.

Fig. 2 shows the effects of TVC on the ITD of a speech signal from 45° right of a head and torso simulator wearing the MHA devices. The lag in the cross-correlation function at which the unsigned correlation coefficient was greatest was taken as the ITD. It is clearly non-zero in both the broadband signal and envelope without TVC (top row), but with TVC (bottom row) the ITD in the broad-band signal becomes zero, whilst the envelope ITD and the ILD are still non-zero (reduced from 540 to 340 μs and from 4 to 3.4 dB, respectively). The very broad peak in the envelope cross-correlation suggests that the envelope ITD may not be a precise cue for lateralizing or localizing sounds, if the auditory system extracts comparable information to this analysis.

3. Training

To minimize the confounding effect of learning the task or the nature of the stimuli, listeners were familiarized with the stimuli in a training session at least one week prior to the test session. The training session consisted of eight blocks, alternating between stimuli without TVC (odd-numbered steps) and stimuli with TVC (even-numbered steps). Blocks one and two each consisted of 12 trials of a target sentence from 0° in quiet. Blocks three and four each consisted of 12 trials of a target sentence from 0° and one masker sentence from either +45° or −45° (TMR=5 dB). The last four blocks each consisted of 12 trials at each of two TMRs with two maskers in the conditions used in the test session. The maskers were in the Separated configuration for blocks five and six, without and with TVC respectively, and in the Co-located configuration for blocks seven and eight, without and with TVC respectively. Table I gives the TMRs for each of these four blocks with the percent correct that each TMR was expected to produce and was observed to produce. Feedback was given.

4. Procedure

The test session began with a ‘warm-up’ block of 12 trials in the Separated condition without TVC, then 12 with TVC (TMR=8.7 dB for both). After this, four test blocks of 50 trials were performed with short breaks in between. Test-condition order was pseudo-randomized with spatial configuration nested inside MHA condition, so both spatial configurations were completed for one MHA condition before beginning the second MHA condition. Randomization resulted in eight and twelve participants completing the conditions with TVC before and after the conditions without TVC, respectively. Although full counter-balancing of the ordering would have been more appropriate, Student’s *t*-tests between those who completed conditions with TVC first and those who completed conditions without TVC first showed no significant differences in performance in the four test conditions [Colocated, without TVC: $t(18)=0.4$, $p>0.05$; Colocated, with TVC: $t(18)=0.9$, $p>0.05$; Separated, without TVC: $t(18)=0.5$, $p>0.05$; Separated, with TVC: $t(18)=0.4$, $p>0.05$].

For each test condition, a listener's psychometric function was estimated from the proportion of correctly recalled words in the 50 trials using the following logistic function:

$$\Psi = \left\{1 + e^{4s_{50}(L_{50}-TMR)}\right\}^{-1} \qquad (1)$$

where s_{50} and L_{50} denote the slope and TMR (respectively) at 50% correct word recall estimated by negative logarithmic maximum-likelihood from the proportion of correctly recalled words per TMR. The first 24 trials were set at pre-defined TMRs (given in Table II) estimated to produce 30, 40, 50, 60, 70, 80, and 90% correct word recall, based on a psychometric function with L_{50} set at group mean performance in training and a shallow slope (s_{50}=0.02) to produce a wide range of TMRs to minimize floor and ceiling effects. At equivalent percent-correct estimations, the pre-defined TMRs in the test sessions (Table II) differed from the TMRs used in the last four blocks of training (Table I). This occurred because of the shallow function slope to calculate the pre-defined TMRs. Three trials were presented at each TMR after an initial three at the TMR expected to produce 80% correct.

A further 26 trials were presented at TMRs estimating 40, 60, 70, and 90% correct on an interim psychometric function fitted to the listener's performance on the 24 pre-defined trials. Six trials were presented at each personalized TMR after two trials at the TMR expected to produce 70% correct. A final psychometric function was calculated from the results of both the pre-defined and personalized TMR trials (excluding the initial three pre-defined and two personalized). Inclusion of responses to both pre-defined and personalized TMR trials allowed as many data points as possible to be used in calculating the final psychometric function. The TMR that would give 50% correct was taken as threshold ($TMR_{50\%}$) for analysis. No feedback was given in the test session.

SRM was defined as the $TMR_{50\%}$ in the Separated condition subtracted from the $TMR_{50\%}$ in the Co-located condition.

C. Psychoacoustic tasks

For the psychoacoustic tasks, listeners were tested in a sound insulating listening booth. All stimuli were created via MATLAB (Natick, MA), an RME Hammerfall II digital-to-analog converter, and a custom-made amplifier. Stimuli were presented over a pair of Sennheiser HDA200 circumaural headphones at 30 dB sensation level (SL) based on the listener's pure-tone audiogram.

1. Stimuli

The stimuli in the monaural tests (TFS1 and F0DL) were presented to the left ear. All stimuli were presented in the presence of a threshold-equalizing noise (TEN; Moore *et al.*, 2000) to mask combination tones and components of the complex tones falling outside of the pass-band of the filtering described below. The TEN level at 1 kHz was 15 dB/ERB_N RMS below the overall RMS level of the test stimulus. This corresponded to an effective SNR of 25 dB. In the monaural conditions, the TEN was played in the same ear as the test stimulus. For ITD detection, uncorrelated samples of TEN were played to the two ears. Uncorrelated noise was chosen to avoid any competing spatial cues in the noise, as an interaural correlation could produce an ITD coherent across frequency.

a. TFS1 tasks Reference stimuli were harmonic complexes with components spaced by a modulation rate (f_m). Each component began in a random phase. They were band-pass filtered around a center frequency (f_c) of 1.2 kHz. The bandwidth was dictated by f_m, passing five components with a 30 dB/octave roll off. Two different f_m's were tested: 100 and 200 Hz, giving two different ranges of harmonics. When f_m=100 Hz, the 10^{th} to 14^{th} harmonics were passed by the filter, with the 12^{th} harmonic at f_c ($TFS1_{H12}$, middle panel of

Fig. 3). When f_m=200 Hz, the fourth to eighth harmonics were passed by the filter, with the sixth harmonic at f_c (TFS1$_{H6}$, top panel of Fig. 3). Listeners discriminated between these harmonic complexes and inharmonic versions where all frequency components were shifted by δ Hz. Both shifted and reference stimuli were identically band-pass filtered based on the reference f_c and f_m. δ started at 50 Hz for both TFS1$_{H12}$ and TFS1$_{H6}$ and was limited to a maximum of f_m/2. Above f_m/2 the shifted stimulus becomes increasingly similar to the reference stimulus as δ approaches f_m. If a listener cannot discriminate an f_m/2 shift, it is impossible to measure a threshold TFS1 score. For NH listeners, thresholds can be obtained consistently for f_c as high as the 14th harmonic (Moore and Sęk, 2009b), but HI listeners appear to have a much lower maximum harmonic number (Hopkins and Moore, 2007). TFS1$_{H6}$ was included in case TFS1$_{H12}$ thresholds could not be obtained. However, TFS1$_{H6}$ thresholds may be partly based on discrimination of resolved harmonics using tonotopic cues, rather than purely TFS from unresolved harmonics.

b. F0DLs For the measurement of F0DLs, both reference and shifted stimuli were harmonic; δ was multiplied by the harmonic number of each component (f / f_m) to produce the shifted stimulus. Like TFS1$_{H12}$, the F0DL reference f_c was 1.2 kHz, and f_m was 100 Hz. Again, both reference and shifted stimuli were band-pass filtered with a 30 dB/octave roll off, passing five components. However, for F0DLs the filter shifted with the stimulus; so whilst the reference filter pass-band f_c was 1.2 kHz, the shifted filter pass-band was centered at f_c + $\delta(f_c / f_m)$, or 1.2+ δ (1.2/0.1) kHz. The bottom panel of Fig. 3 shows example F0DL stimuli. The starting δ was 5 Hz and no maximum limit was imposed.

c. ITD detection ITDs were presented using 500 Hz pure tones (ITD500). Onset and offset ramps were synchronous across ears so there was no ITD in the envelope. The left ear stimulus was $\sin(2\pi\cdot 500t)$ where t is the time-sample vector, the right ear stimulus was $\sin(2\pi\cdot 500t + \delta)$. The starting δ was π radians (1 ms) and this was also the maximum δ limit. A different f_c for ITD detection and the TFS1 tasks was used in order to minimize the number of listeners who could not perform these tasks at above chance performance. Whilst other studies have measured both monaural and binaural TFS tasks in the same frequency region (e.g., 850 Hz; Moore *et al.*, 2012b) in young NH listeners, it is difficult to obtain ITD thresholds above 500 Hz for older listeners (Grose and Mamo, 2010; Moore *et al.*, 2012b), and it is difficult to obtain TFS1 thresholds below an f_c of 1 kHz (Moore and Sek 2009; Moore *et al*., 2012b). Also, previous studies have found only moderate correlations between TFS1 and ITD detection and have suggested that performance is at least partly driven by different mechanisms (Hopkins and Moore, 2011; Moore *et al*., 2012b).

2. Procedure

A two-interval, two-alternative, forced-choice task was used where one interval contained four 200 ms bursts of the reference stimuli (RRRR) and the other interval contained four 200 ms bursts, alternating between reference and shifted stimuli (RSRS). Each burst was separated by 100 ms, and ramped on and off by the rising and falling halves (respectively) of a 20 ms Hanning window. The two intervals were separated by 400 ms of silence. This paradigm is described further elsewhere (Moore and Sęk, 2009a; Hopkins and Moore, 2010a). For all four tasks, a geometric, two-down, one-up adaptive procedure tracked 71% correct (Levitt, 1971) over eight reversals. Step sizes of δ were a factor of 1.5^3 until the first reversal, 1.5^2 until the second reversal and 1.5 thereafter. The geometric mean of the last six reversals was used as the threshold estimate from a given track. Each listener attempted at least two (and a maximum of three) tracks for each task and the geometric mean of the thresholds from the completed tracks was taken as their threshold. If the listener failed to discriminate the maximally shifted stimulus from the reference stimulus three times within a track, 40 extra trials were presented with the maximum shift. In these cases, percent correct was calculated for all the trials in which the maximum shift was presented. The adaptive track was run again if the listener scored better than 63% correct (above this performance can be assumed, with at least 95% confidence, not to be due to chance). Otherwise, another 40 trials at maximum shift were presented and the percent correct was recalculated from these trials. If the listener still scored worse than 63% correct, testing for that condition stopped and no threshold was obtained; otherwise, a final adaptive track was run. F0DL thresholds were obtainable from all listeners; for ITD500, no threshold could be obtained for three listeners; for $TFS1_{H6}$, no threshold could be obtained for two listeners (not the same listeners as for ITD500, however); for TFS_{H12}, no threshold could be obtained for 19 listeners (one listener managed one successful track out of three). Because $TFS1_{H12}$ thresholds could only be measured for one listener, this task was discarded from analysis.

For F0DLs, $TFS1_{H6}$ and ITD500, the test-retest reliability was quantified for the listeners who completed two or three adaptive tracks per task by calculating the intraclass correlation coefficient (ICC; repeated measures case, based on single scores rather than the mean; see Shrout and Fleiss, 1979). For F0DLs the ICC was 0.48 for two tracks and 0.40 for the three tracks, for $TFS1_{H6}$ the ICC was 0.76 for two tracks and 0.75 for three tracks, and for ITD500 the ICC was 0.38 for two tracks and 0.46 for three tracks.

Before beginning the experiment, the listeners were given a brief training period. In this period they heard example trials with the maximum shift in the shifted stimuli for $TFS1_{H12}$, $TFS1_{H6}$ and ITD500, and a 10 Hz shift for F0DL. They heard eight trials without TEN and eight with TEN per condition.

III. RESULTS AND DISCUSSION

A. Speech test

The $TMR_{50\%}$ mean and individual values are plotted in Fig. 4. With TVC, $TMR_{50\%}$ was similar in the Co-located (mean=3.7 dB) and Separated configurations (mean=3.4 dB); individual SRM ranged between −1.8 and 2.7 dB. Without TVC, the mean $TMR_{50\%}$ was 2.8

dB in the Co-located condition, and −1.2 dB in the Separated condition (where $TMR_{50\%}$ varied from −6.5 to 5.5 dB across listeners), leading to SRM ranging from −0.9 to 8.4 dB.

Performance was analyzed with a repeated-measures analysis of variance (ANOVA) with two within-subjects factors: spatial configuration (Co-located *vs*. Separated) and processing (without TVC *vs*. with TVC). Both factors produced significant main effects. $TMR_{50\%}$ was lower (better) in the Separated conditions than the Co-located conditions [$F(1,19)=37.8$, $p<0.001$] and $TMR_{50\%}$ was lower without TVC than with TVC [$F(1,19)=87.0$, $p<0.001$]. The interaction was also significant [$F(1,19)=38.6$, $p<0.001$], confirming that the SRM without TVC was greater than SRM with TVC. Because the TVC removed the ITDs in the TFS in each sub-band by generating the sine carriers in phase across the ears, it appears that old HI listeners gained significantly by using ITDs in the TFS of sounds from different azimuths.

The similarity between $TMR_{50\%}$ in both spatial configurations with TVC suggests the envelope ITDs and ILDs are not sufficient for speech unmasking, assuming that TVC does not disrupt these cues substantially. However, vocoding does not guarantee independent manipulation of TFS and envelope. It can affect the spectro-temporal envelope (Kates, 2011) and models suggest neural representations of TFS and envelope are comparably degraded by vocoding with noise carriers (Shamma and Lorenzi, 2013). Whilst noise carriers may affect envelopes more than tone carriers (Kates, 2011), envelope information at a neural level (including interaural envelope cues) may not be faithfully preserved after TVC. Therefore, the effect of TVC on the envelope cues on SRM cannot be discounted. Alternatively, the small SRM observed with TVC may be because old listeners are less sensitive than younger listeners to envelope ITDs (King *et al.*, 2014). Comparing performance in the Separated condition with TVC in the current study and in Andersen *et al.* (2010) suggests that $TMR_{50\%}$ is about 12 dB lower for young NH listeners than for old HI listeners when only envelope ITDs and ILDs are available.

Without TVC, the mean SRM and the variation in SRM was very similar to that found for HI listeners by Marrone *et al.* (2008a) and Neher *et al.* (2009; 2011). HI listeners' mean SRM is approximately 8-14 dB less than NH listeners' SRM and exhibits larger individual differences (Behrens *et al*., 2008; Marrone *et al.*, 2008a; Neher *et al.*, 2009). In the present study, the difference between $TMR_{50\%}$ in the Separated configuration with TVC and without TVC was 4.6 dB, which is roughly half that found for young NH listeners in a simulated spatial setup using headphones and HRIRs (8.3 dB; Andersen *et al.*, 2010). Whilst old HI listeners are less sensitive to TFS-ITDs than young NH listeners (*e.g.*, Lacher-Fougère and Demany, 2005; Hopkins and Moore, 2011), it should not be assumed that they are unable to use ITDs in the TFS of speech. Differences in the nature and severity of hearing loss may explain the variation in old HI listeners' ability to use TFS-ITDs in SRM.

To determine whether individual differences in the effects and interaction of TVC and spatial configuration were driven by age or hearing loss, each listener's age and PTA_{LF} were used as covariates in an extension of the ANOVA model described above. Higher (worse) PTA_{LF} was related to higher $TMR_{50\%}$, across all conditions [$F(1,17)=8.8$, $p<0.01$], but age was not related to a change in $TMR_{50\%}$ over all conditions [$F(1,17)=0.7$, $p>0.05$]. Age did not interact with the effects of MHA processing [$F(1,17)=0.1$, $p>0.05$] or spatial configuration [$F(1,17)=1.2$, $p>0.05$]. Higher PTA_{LF} was associated with a smaller difference in performance between the vocoded and non-vocoded conditions [$F(1,17)=17.3$, $p<0.01$]. PTA_{LF} did not interact significantly with the effect of spatial configuration on $TMR_{50\%}$ [$F(1,17)=3.6$, $p>0.05$]. The three-way interaction between PTA_{LF}, MHA processing and spatial configuration was significant [$F(1,17)=23.2$, $p<0.001$]; SRM without TVC, compared to with TVC, was smaller with increasing PTA_{LF}. This is shown in Fig. 5, where $TMR_{50\%}$ is plotted as a function of PTA_{LF} with least-squares linear fits for each test condition separately. The variation in performance in the Separated condition without TVC is partially explained by an increase in $TMR_{50\%}$ with increasing PTA_{LF}. This suggests that low-frequency audiometric hearing loss is related to how well a listener can understand speech from a conversational partner when others are talking around them. Low-frequency audiometric thresholds may be a convenient way to determine which type of hearing aid processing may be best for an individual with hearing loss. If performance in the Separated condition without TVC is mediated by a listener's sensitivity to TFS ITDs, then an individual with a low PTA_{LF} may benefit from hearing aids that preserve these ITDs. If their PTA_{LF} is high, they may benefit from processing strategies that increase the SNR at the expense of sacrificing TFS ITDs, such as directional microphone sensitivity (Van de Bogaert *et al.*, 2005).

There was a small, but significant, difference in $TMR_{50\%}$ between the Co-located conditions with TVC and without TVC [mean difference=0.9 dB, $t(19)=4.8$, $p<0.001$]. Worse performance with TVC suggests that the TVC disrupted information that is used to separate the speech signals, even without spatial cues. This effect was smaller than the 5 dB effect of TVC on monaural speech intelligibility found by Hopkins *et al.* (2008) with old HI listeners. This may possibly be due to differences in the speech corpora. Lunner *et al.* (2012) compared the effect of TVC on speech intelligibility in competing speech with three different speech corpora, including the closed-set Dantale II used in the current study and the open-set corpus used by Hopkins *et al.* (2008). Lunner *et al.* (2012) found that there was a larger effect of TVC in the open-set corpora than the closed-set corpus. However, this was only evident with young NH listeners, not with old HI listeners.

B. Psychoacoustic tasks

The geometric mean scores for the three psychoacoustic tasks were 210 μs, 11.9 Hz and 5.9 Hz for ITD500, $TFS1_{H6}$, and F0DLs respectively. The standard deviation of the scores were factors of 1.79, 2.11, and 1.56 for ITD500, $TFS1_{H6}$, and F0DLs respectively. The F0DLs had a bimodal distribution, with modes at 4 and 9 Hz.

A correlation matrix between age, PTA_{LF}, the psychoacoustic task scores, and SRM without TVC minus SRM with TVC ($SRM_{\Delta TVC}$) is given in Table III. $TFS1_{H6}$ and ITD500 thresholds were logarithmically transformed to distribute them normally before calculating Pearson's product–moment correlation coefficients (r). F0DLs were correlated against the other variables using Spearman's ranked correlation (ρ). A sequentially rejective Bonferroni correction (Holm, 1979) was applied to compensate for multiple comparisons. This was chosen over the traditional Bonferroni correction because it is less conservative, without requiring additional assumptions. Additionally, the SRTs in each speech-test condition and the differences between them reflecting SRM and the effect of TVC were correlated against age, PTA_{LF} and the psychoacoustic task scores (Table IV). However, due to the large number of correlations in Table IV and the low power for these correlations, no correction for multiple comparisons was applied, resulting in an inflated chance of a false positive. The correlations in Table IV should be considered purely exploratory and any significant correlations should be replicated with a larger sample before inferring any conclusions. All five partial correlations mentioned in this section were corrected for multiple comparisons separately from the correlations in Table III, again using the Holm-Bonferroni procedure.

It was hypothesized that two measures of TFS sensitivity, TFS1 and ITD500, would correlate with SRTs and SRM without TVC and the difference between SRTs and SRM with and without TVC (*i.e.*, the effect of TVC). $TFS1_{H6}$ correlated significantly with $SRM_{\Delta TVC}$ before, but not after, correcting for multiple comparisons (Table III). $TFS1_{H6}$ correlated significantly with SRTs in the Separated condition without TVC, the effect of TVC in the Separated condition, and with SRM without TVC (Table IV). However, $TFS1_{H6}$ was not significantly correlated with SRTs in the Co-located condition without TVC, or the effect of TVC in the Co-located condition. ITD500 did not correlate significantly with either SRTs in the Separated condition without TVC, with SRM without TVC, or with $SRM_{\Delta TVC}$. $SRM_{\Delta TVC}$

is plotted against listeners' ITD500 and TFS1H6 scores in the right and middle panels of Fig. 6.

It is surprising that the measure of TFS-ITD processing, ITD500, did not correlate significantly with performance in these conditions, as some studies show relations between TFS-ITD processing and SRTs (Neher *et al.*, 2011; Neher *et al.*, 2012). However, it is consistent with a recent study which found that SRM produced by ITDs alone in target and interferer speech streams was not related to 250-Hz pure tone ITD discrimination thresholds in quiet, which, surprisingly, was better for listeners with moderate hearing loss than those with mild hearing loss (Lőcsei *et al.*, 2016). In the current study ITD500 did not correlate significantly with PTA_{LF} or with audiometric threshold at 500 Hz ($r=-0.05$, $p>0.05$).Whilst King *et al.* (2014) found a moderate correlation between TFS-ITD detection and absolute threshold at the same carrier frequency, most others have not (Lacher-Fougère and Demany, 2005; Hopkins and Moore, 2011; Strelcyk and Dau, 2009; Moore *et al.*, 2012a).

Whereas Neher *et al.* (2011; 2012) and Hopkins and Moore (2011) tested ITD detection in quiet, the current study tested ITD detection in noise. The current study included binaurally uncorrelated TEN at an effective SNR of 25 dB because ITD500 was compared to speech reception in a background of competing-talkers rather than speech reception in quiet. Three out of 20 listeners could not perform the ITD500 task. This is not a substantially larger proportion than when HI listeners perform ITD discrimination in quiet, which is typically 5 to 10% (e.g., Moore *et al*., 2012b), but can be as high as 40% (Whitmer *et al*., 2014). Nonetheless, it is possible that the TEN disrupted the ITD in the target tone, obscuring measurement of the listeners' sensitivity to ITDs in the tone, or confounded the mechanism that links ITD sensitivity to SRM.

Strelcyk and Dau (2009) found a significant correlation between SRT in lateralized noise and pure-tone ITD discrimination in noise presented 10 dB below the level which just masked the tone. However, they only found this for diotic noise and at a high level above absolute threshold, not for uncorrelated noise and a lower SL similar to the ITD500 stimuli in the current study. Whilst dichotic maskers aided masking release in speech reception, they presented an additional challenge in lateralization (footnote 7 in Strelcyk and Dau, 2009). Also, Strelcyk and Dau (2009) found that the differences in ITD discrimination between NH and HI listeners were most pronounced with a high level tone in quiet. Therefore, ITD detection of low-SL pure tones in the presence of uncorrelated TEN (as in the current study) might use different mechanisms than using ITDs in SRM.

If the $TFS1_{H6}$ scores are assumed to not have produced false positive results, it is perhaps surprising that $TFS1_{H6}$ scores correlated significantly with $SRM_{\Delta TVC}$ (before correction for multiple comparisons), whilst ITD500 scores did not. One possible explanation is that $TFS1_{H6}$ is a more reliable measure than ITD500. Test-retest reliability for the ITD500 task was weak to moderate (ICC of 0.38 to 0.46), compared to the high reliability of the $TFS1_{H6}$ task (ICC of 0.75 to 0.76). This lack of reliability suggests that measurement error may have affected the correlations between the ITD500 task and the other measures. However, taking multiple comparisons into account, neither TFS measure correlated significantly with $SRM_{\Delta TVC}$. A replication with greater statistical power is required to determine if such measures are or are not related to SRM and the effect of TVC on SRM.

$TFS1_{H6}$ scores correlated with age and PTA_{LF} (Fig. 7) before correction for multiple comparisons. Partial correlations were tested to see if $TFS1_{H6}$ score still correlated with $SRM_{\Delta TVC}$ after the variance due to PTA_{LF} or age was accounted for. Partialing out age actually increased the strength of the correlation between $TFS1_{H6}$ and $SRM_{\Delta TVC}$ ($r=-0.70$, $p<0.01$). However, after controlling for PTA_{LF}, $TFS1_{H6}$ and $SRM_{\Delta TVC}$ were not correlated ($r=-0.34$, $p>0.05$). Conversely, there was still a moderate correlation between PTA_{LF} and $SRM_{\Delta TVC}$, when $TFS1_{H6}$ score was accounted for ($r=-0.50$, $p<0.05$). However, this was not significant after correction for multiple comparisons. This suggests that hearing loss may be related to how TVC affects SRM independently from the individual differences explained by $TFS1_{H6}$ score.

The effect of TVC in the Co-located configuration was expected to correlate with $TFS1_{H6}$ scores due to the vocoding disrupting monaural TFS cues useful for speech perception. However, this was not the case. This may be because either $TFS1_{H6}$ did not measure TFS sensitivity exclusively (harmonics 4 to 8 were passed by the filter applied to the stimuli, which may be sufficiently low as to be resolved, even for HI listeners), or TVC did not exclusively disrupt TFS cues (Shamma and Lorenzi, 2013), or for both reasons. It is not clear what form of auditory processing SRM requires (and that is affected by TVC) that is reflected by $TFS1_{H6}$ performance. Perhaps the sharpness of frequency selectivity (which could arguably affect $TFS1_{H6}$ performance) affects SRM, but this was not tested in the current study.

The partial correlation between $TFS1_{H6}$ and $SRM_{\Delta TVC}$, controlling for F0DLs, was performed to see if performance on a task unlikely to provide resolved harmonics or TFS cues could explain the relation in terms of general temporal processing or pitch discrimination. The correlation was still moderate ($r=-0.54$, $p<0.05$), but not significant after correction for multiple partial correlations. F0DLs moderately correlated with all SRTs (see Table IV), particularly in the Separated configuration (Fig. 8), but this is without correcting for multiple comparisons. However, F0DLs did not correlate significantly with SRM with TVC. Nevertheless, this suggests that the ability to discriminate pitch shifts may be related to a listener's ability to segregate one talker's speech from that of others, possibly due to differences in the F0s of the various talkers' speech. F0DLs were not related to $SRM_{\Delta TVC}$ or to $TFS1_{H6}$ scores, suggesting that F0DLs do not relate well to TFS processing that contributes to spatial unmasking. A lack of a relationship to TFS processing is perhaps not surprising if

F0DLs are driven by spectral cues, or envelope or distortion product cues from unresolved components (Oxenham *et al.*, 2009).

Consistent with the ANOVA with PTA_{LF} included as a covariate in section III.A, $SRM_{\Delta TVC}$ was strongly correlated with PTA_{LF}; $SRM_{\Delta TVC}$ decreases as PTA_{LF} increases. This can be seen in the left panel of Fig. 6. PTA_{LF} also correlated significantly with SRTs in the Separated condition without TVC and with SRM without TVC (see table IV). Peissig and Kollmeier (1997) found that audiometric threshold was a poor predictor of SRM with either noise or speech maskers, but they only tested eight HI listeners, whereas Neher *et al.* (2011) found, with 23 HI listeners, that low frequency audiometric threshold was moderately correlated with SRTs in conditions similar to the Separated condition without TVC in the current study. Neher *et al.* (2011) did not measure SRTs in a co-located condition, so SRM could not be calculated. Jakien *et al.* (2017) found that audiometric threshold predicted SRM.

Higher PTA_{LF} thresholds may have been associated with less SRM due to insufficient audibility. However, foam plugs provide a closed seal which minimizes leakage, and the CAMEQ prescription to compensate for hearing loss has been shown to restore audibility of speech stimuli (between 0.5 and 5 kHz) for listeners with audiometric thresholds greater than those in the current study (Hopkins *et al.*, 2008; Hopkins and Moore, 2011). Although it is likely that audibility of the signals was restored, it was not measured here. With linear amplification, it is possible that some listeners did not receive the same audibility that compressive amplification could provide, rendering soft sounds inaudible. In the current study SRM and $SRM_{\Delta TVC}$ were analyzed as difference measures; poor audibility of the speech may be expected to affect the four speech test conditions similarly. However, Glyde *et al.* (2015) showed that increasing high-frequency audibility beyond prescribed values did not affect SRT in co-located conditions, but did improve SRTs in separated conditions, thus increasing SRM. Best *et al.* (2016) and Jakien *et al.* (2017) reported that presenting speech stimuli at an SL above each listener's SRT in quiet resulted in less SRM for listeners with more hearing loss. However, compensating for the loss of audibility within frequency bands from 0.25 to 8 kHz resulted in similar SRM for listeners with different amounts of hearing loss, even though those with more hearing loss had higher SRTs in both co-located and separated conditions. It may be that the audibility of the high-frequency (particularly above 5 kHz) content of the stimuli in the current study was insufficient.

Age has been associated with poor TFS sensitivity (*e.g.*, Grose and Mamo, 2010; Moore *et al.*, 2012b) and reduced SRM (*e.g.*, Gelfand *et al.*, 1988; Marrone *et al.*, 2008a, Gallun *et al.*, 2013). Neher *et al.* (2012) suggested that a common, age-related mechanism could affect performance on speech perception, psychoacoustic and cognitive tasks alike. However, in the current study age was not correlated with $SRM_{\Delta TVC}$ or ITD500, and the correlation between $SRM_{\Delta TVC}$ and PTA_{LF} remained significant after age was partialed out ($r=-0.76$, $p<0.001$). Neher *et al.* (2011) argued that a narrow age range (60–78 years) limited the effect size of cognitive measures previously found to relate to age and SRTs with spatially separated speech (Neher *et al.*, 2009). Indeed, with a large age range, Gallun *et al*. (2013) found that age and hearing loss (SRTs in quiet) were independent contributors to reduced SRM—with age alone explaining reduced SRM in simulated auditory space using headphones and HRIRs; however a follow-up study with an overlapping sample but a larger range of hearing losses found hearing loss was the greater contributor to reduced SRM (Jakien *et al.*, 2017). It is possible that the age effect was limited in the current study, where the age range was also narrow (22 years). However, worse $TFS1_{H6}$ scores were correlated with increasing listener age. This indicates some age-related deficit in monaural TFS processing.

IV. CONCLUSIONS

The current study aimed to determine the contribution of TFS ITDs to SRM by manipulation of TFS ITDs via vocoding, and to find a good predictor of individual differences in the benefit from TFS ITDs to SRM. When presented with closed-set sentences spoken by three female talkers, older, HI listeners benefited from spatial separation, although more variably and less than younger, NH listeners in previous studies (Marrone *et al.*, 2008a; Marrone *et al.*, 2008c; Neher *et al.*, 2011). Applying TVC to the signals recorded at the ears (in phase, to remove differences in the ITDs) significantly reduced the benefit of spatial separation. This suggests that some older HI listeners are capable of using binaural TFS cues to aid speech perception in multi-talker environments. This has implications for the potential trade-offs of disrupting binaural temporal acoustic information with signal processing strategies employed in modern, digital hearing aids.

The current study assessed how well several measures that are thought to rely on auditory temporal processing might predict the effect of TVC on SRM. The audiogram below 1.5 kHz and monaural TFS sensitivity were both correlated with the effect of TVC on SRM (before correction for multiple comparisons), whilst pure-tone ITD detection was not. This result is surprising, as both pure-tone ITD detection and SRM rely on binaural processing of TFS information, but it may be due to the modest test-retest reliability of the ITD detection task, or the inclusion of background noise. Further research with a larger sample is needed to clarify whether SRM is truly unrelated to binaural TFS sensitivity as the ITD detection results here suggest and to verify that monaural TFS sensitivity is indeed related to SRM, rather than a false positive result. Low-frequency audiometric thresholds may be a convenient metric to determine who may benefit from hearing aids that preserve binaural cues in the TFS.

ACKNOWLEDGMENTS

The authors thank two anonymous reviewers for constructive comments on an earlier version of the manuscript. The authors also thank the volunteer test participants for their time. This work was supported by the United Kingdom Medical Research Council (grant ref: G1001609) and Oticon A/S.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF database," in *2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 99-102.

Andersen, M. R., Kristensen, M. S., Neher, T., and Lunner, T. (**2010**). "Effect of binaural tone vocoding on recognising target speech presented against spatially separated speech maskers," in *International Hearing Aid Research Conference 2010* (Lake Tahoe, California, USA).

Apoux, F., Yoho, S. E., Youngdahl, C. L., and Healy, E. W. (**2013**). "Role and relative contribution of temporal envelope and fine structure cues in sentence recognition by normal-hearing listeners, " J. Acoust. Soc. Am. **134**, 2205-2212.

Behrens, T., Neher, T., and Johannesson, R. B. (**2008**). "Evaluation of speech corpus for assessment of spatial release from masking.," in *Auditory signal processing in hearing-impaired listeners*, edited by T. Dau, J. M. Buchholz, J. Harte, and T. U. Christiansen (Centertryk A/S, Copenhagen, Denmark), pp. 449-457.

Best, V., Marrone, N., Mason, C. R., and Kidd, G. Jr (**2012**). "The influence of non- spatial factors on measures of spatial release from masking," J. Acoust. Soc. Am. **131**, 3103-3110.

Best, V., Mason, C. R., Swaminathan, J., Kidd, G. Jr., Jakien K. M., Kampel S. D., Gallun F. J., Buchholz J. M., and Glyde H. (**2016**). "On the contribution of target audibility to performance in spatialized speech mixtures," Adv. Exp. Med. Biol., **894**, 83-91.

Best, V., van Schaik, A., and Carlile, S. (**2004**). "Separation of concurrent broadband sound sources by human listeners," J. Acoust. Soc. Am. **115**, 324-336.

Bremen, P., and Middlebrooks, J. C. (**2013**). "Weighting of spatial and spectro-temporal cues for auditory scene analysis by human listeners," Plos One **8**, e59815.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). “Intonation and the perceptual separation of simultaneous voices,” J. Phonetics **10**, 23-36.

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acta Acust. Utd. Acust. **86**, 117-128.

Cariani, P. A., and Delgutte, B. (**1996**). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," J. Neurophysiol. **76**, 1698-1716.

Divenyi, P. L., Stark, P. B., and Haupt, K. M. (**2005**). "Decline of speech understanding and auditory thresholds in the elderly," J. Acoust. Soc. Am. **118**, 1089-1100.

Drennan, W. R., Won, J. H., Dasika, V. K., and Rubinstein, J. T. (**2007**). "Effects of Temporal Fine Structure on the Lateralization of Speech and on Speech Understanding in Noise," J. Assoc. Res. Otolaryngol. **8**, 373–383.

Dudley, H. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169-177.

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

Gallun, F. J., Diedesch, A. C., Kampel, S. D., Jakien, K. M. (**2013**) "Independent impacts of age and hearing loss on spatial release in a complex auditory environment," Front. Neurosci. **7** (252), 1-11.

Garadat, S. N., Litovsky, R. Y., Yu, G., and Zeng, F.-G. (**2009**). "Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech," J. Acoust. Soc. Am. **126**, 2522-2535.

Gatehouse, R. W., and Noble, W. (**2004**). "The speech, spatial and qualities of hearing scale (ssq)," Int. J. Audiol. **43**, 85-99.

Gelfand, S. A., Ross, L., and Miller, S. (**1988**). "Sentence reception in noise from one versus two sources: Effects of aging and hearing loss," J. Acoust. Soc. Am. **83**, 248-256.

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

Glyde, H., Buchholz, J. M., Nielsen, L., Best, V., Dillon, H., Cameron, S., and Hickson, L. (**2015**). "Effect of audibility on spatial release from speech-on-speech masking" J. Acoust. Soc. Am. **138**, 3311-3319.

Grimm, G., Herzke, T., Berg, D., and Hohmann, V. (**2006**). "The master hearing aid: A pc-based platform for algorithm development and evaluation," Acta Acust. Utd. Acust. **92**, 618-628.

Grose, J. H., and Mamo, S. K. (**2010**). "Processing of temporal fine structure as a function of age," Ear Hear. **31**, 755-760.

Helfer, K. S., and Freyman, R. L. (**2008**). "Aging and speech-on-speech masking," Ear Hear. **29**, 87.

Holm, S. (**1979**). "A simple sequentially rejective multiple test procedure," Scand. J. Stat. **6**, 65-70.

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. (**2010a**). "Development of a fast method of measuring sensitivity to temporal fine structure information at low frequencies.," Int. J. Audiol. **49**, 940-946.

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

Hopkins, K., and Moore, B. C. J. (**2011**). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," J. Acoust. Soc. Am. **130**, 334-349.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "The 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.

Jakien, K. M., Kampel, S. D., Gordon, S. Y., and Gallun, F. J. (**2017**). "The benefits of increased sensation level and bandwidth for spatial release from masking," Ear Hear. **38**, 13-21.

Kates, J. M. (**2011**). "Spectro-temporal envelope changes caused by temporal fine structure modification," J. Acoust. Soc. Am. **129**, 3981-3990.

King, A., Hopkins, K., and Plack, C. J. (**2014**). "The effects of age and hearing loss on interaural phase difference discriminationa)," J. Acoust. Soc. Am. **135**, 342-351.

Kuhn, G. F. (**1977**). "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157-167.

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.

Lőcsei, G., Pedersen J. H., Laugesen, S., Santurette, S., Dau, T., MacDonald, E. N. (**2016**). "Temporal fine-structure coding and lateralized speech perception in normal-hearing and hearing-impaired listeners, " Trends Hear. **20**, 1-15.

Loizou, P. C., Dorman, M., and Tu, Z. (**1999**). "On the number of channels needed to understand speech," J. Acoust. Soc. Am. **106**, 2097-2103.

Lunner, T., Hietkamp, R. K., Andersen, M. R., Hopkins, K., and Moore, B. C. J. (**2012**). "Effect of speech material on the benefit of temporal fine structure information in speech for young normal-hearing and older hearing-impaired participants," Ear Hear. **33**, 377-388.

Marrone, N., Mason, C. R., and Gerald Kidd, J. (**2008a**). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," J. Acoust. Soc. Am. **124**, 3064-3075.

Marrone, N., Mason, C. R., and Kidd, G. (**2008b**). "Tuning in the spatial dimension: Evidence from a masked speech identification task," J. Acoust. Soc. Am. **124**, 1146-1158.

Marrone, N., Mason, C. R., and Kidd, G., Jr. (**2008c**). "Evaluating the benefit of hearing aids in solving the cocktail party problem.," Trends Amplif. **12**, 300-315.

Micheyl, C., Dai, H., and Oxenham, A. J. (**2010**). "On the possible influence of spectral and temporal-envelope cues in tests of sensitivity to temporal fine structure," J. Acoust. Soc. Am. **127**, 1809-1809.

Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Ann. Rev. Psychol. **42**, 135-159.

Moore, B. C. J., and Glasberg, B. R. (**1998**). "Use of a loudness model for hearing-aid fitting. I. Linear hearing aids," Br. J. Audiol. **32**, 317-335.

Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcántara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205-224.

Moore, B. C. J., and Sęk, A. (**2009a**). "Development of a fast method for determining sensitivity to temporal fine structure," Int. J. Audiol. **48**, 161-171.

Moore, B. C. J., and Sęk, A. (**2009b**). "Sensitivity of the human auditory system to temporal fine structure at high frequencies," J. Acoust. Soc. Am. **125**, 3186-3193.

Moore, B. C. J., Glasberg, B. R., Stoev, M., Füllgrabe, C., and Hopkins, K. (**2012a**). “The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies,” J. Acoust. Soc. Am. **131**, 1003–1006.

Moore, B. C. J., Vickers, D. A., and Mehta, A. (**2012b**). "The effects of age on temporal fine structure sensitivity in monaural and binaural conditions," Int. J. Audiol. **51**, 715-721.

Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A. S., and Schaik, A. (**2009**). "Benefit from spatial separation of multiple talkers in bilateral hearing-aid users: Effects of hearing loss, age, and cognition," Int. J. Audiol. **48**, 758-774.

Neher, T., Laugesen, S., Søgaard Jensen, N., and Kragelund, L. (**2011**). "Can basic auditory and cognitive measures predict hearing-impaired listeners’ localization and spatial speech recognition abilities?," J. Acoust. Soc. Am. **130**, 1542-1558.

Neher, T., Lunner, T., Hopkins, K., and Moore, B. C. J. (**2012**). "Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners (l)," J. Acoust. Soc. Am. **131**, 2561-2564.

Oxenham, A. J., Micheyl, C., and Keebler, M. V. (**2009**). "Can temporal fine structure represent the fundamental frequency of unresolved harmonics?," J. Acoust. Soc. Am. **125**, 2189-2199.

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.

Peissig, J., and Kollmeier, J. (**1997**). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660-1670.

Pichora-Fuller, M. K., and Schneider, B. A. (**1992**). "The effect of interaural delay of the masker on masking-level differences in young and old subjects," J. Acoust. Soc. Am. **91**, 2129-2135.

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.

Rayleigh, L. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214-232.

Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," J. Neurophysiol. **30**, 769-793.

Ross, B., Fujioka, T., Tremblay, K. L., and Picton, T. W. (**2007**). "Aging in binaural hearing begins in mid-life: Evidence from cortical auditory-evoked responses to changes in interaural phase," J. Neurosci. **27**, 11172-11178.

Shamma, S., and Lorenzi, C. (**2013**). "On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system," J. Acoust. Soc. Am. **133**, 2818-2833.

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

Shrout, P. E., and Fleiss, J. L. (**1979**). "Intraclass correlations: Uses in assessing rater reliability," Psychol. Bull. **86** (2), 420-428.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature **416**, 87-90.

Stone, M. A., and Moore, B. C. J. (**2003**). "Tolerable hearing-aid delays. Iii. Effects on speech production and perception of across-frequency variation in delay," Ear Hear. **24**, 175-183.

Strelcyk, O., and Dau, T. (**2009**). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," J. Acoust. Soc. Am. **125**, 3328-3345.

Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385-2399.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Roverud, E., and Kidd, G. Jr (**2016**). "Role of binaural temporal fine structure and envelope cues in cocktail-party listening," J. Neurosci. **36**, 8250-8257.

Van de Bogaert, T., Wouters, J., Klasen, T. J., and Moonen, M. (**2005**). "Distortion of interaural time cues by directional noise reduction systems in modern digital hearing aids," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.* (New Paltz, New York, USA), pp. 57-60.

Wagener, K., Josvassen, J. L., and Ardenkjær, R. (**2003**). "Design, optimization and evaluation of a danish sentence test in noise: Diseño, optimización y evaluación de la prueba danesa de frases en ruido," Int. J. Audiol. **42**, 10-17.

Warren, L. R., Wagener, J. W., and Herman, G. E. (**1978**). "Binaural analysis in the aging auditory system," J. Gerontol. **33**, 731-736.

Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (**2014**). "The perception of apparent auditory source width in hearing-impaired adults," J. Acoust. Soc. Am. **135**, 3548-3559.

Working Group on Speech Understanding and Aging (**1988**). "Speech understanding and aging," J. Acoust. Soc. Am. **83**, 859-895.

Table I. The TMRs in the training phase, the percent correct of word recall the TMR was expected to produce, and the mean (across listeners) percent correct that was observed for the final four training blocks.

Condition	TMR (dB)	Estimated percent correct	Observed percent correct
Separated without TVC	8	70%	74%
	2	40%	41%
Separated with TVC	9	70%	76%
	3	40%	45%
Co-located without TVC	10	80%	87%
	4	40%	49%
Co-located with TVC	11.5	80%	90%
	6	40%	56%

Table II. A table of the pre-defined TMRs (dB) used for the first 24 trials of each block by spatial configuration (rows) and by the percent correct word recall the TMR was estimated to produce (columns).

	Estimated percent correct word recall						
Configuration	30%	40%	50%	60%	70%	80%	90%
Co-located, test	−1.3	1.9	5.1	8.3	11.5	14.7	17.9
Separated, test	−7.3	−4.1	−0.9	2.3	5.5	8.7	11.9

Table III. A correlation matrix between listeners' PTA_{LF}, age, $TFS1_{H6}$ score, F0DL, ITD500 and the difference between SRM with and without TVC ($SRM_{\Delta TVC}$). Correlations with F0DL were Spearman's rank correlations, the rest were Pearson's product–moment correlations. The number of listeners is in parentheses for each correlation.

	PTALF	Age	$TFS1_{H6}$	F0DL	ITD500	$SRM_{\Delta TVC}$
PTALF	1 (20)	0.01 (20)	0.64[b] (18)	0.50[a] (20)	0.11 (17)	-0.76[c] (20)
Age	0.01 (20)	1 (20)	0.68[b] (18)	0.23 (20)	0.29 (17)	-0.10 (20)
$TFS1_{H6}$	0.64[b] (18)	0.68[b] (18)	1 (18)	0.43 (18)	0.20 (15)	-0.64[b] (18)
F0DL	0.50[a] (20)	0.23 (20)	0.43 (18)	1 (20)	-0.09 (17)	-0.44 (20)
ITD500	0.11 (17)	0.29 (17)	0.20 (15)	-0.09 (17)	1 (17)	-0.10 (17)
$SRM_{\Delta TVC}$	-0.76[c] (20)	-0.10 (20)	-0.64[b] (18)	-0.44 (20)	-0.10 (17)	1 (20)

[a]Correlations significant at α of 0.05
[b]Correlations significant at α of 0.01
[c]Correlations significant at α of 0.001. Significant after Holm–Bonferroni correction.

Table IV. A table of correlations of listeners' PTA_{LF}, age, $TFS1_{H6}$ score, F0DL and ITD500 against SRTs for the Co-located with TVC (Col_{TVC}), Co-located without TVC (Col_{noTVC}), Separated with TVC (Sep_{TVC}) and Separated without TVC (Sep_{noTVC}) conditions and various differences between these conditions. The nomenclature is similar to that of Table III; however, no correction for multiple comparisons was performed and these correlations should be considered purely exploratory and requiring further research with larger samples.

	PTALF	Age	$TFS1_{H6}$	F0DL	ITD500
Sep_{noTVC}	0.69[c] (20)	0.18 (20)	0.61[b] (18)	0.64[b] (20)	0.14 (17)
Col_{noTVC}	0.52[a] (20)	0.07 (20)	0.30 (18)	0.54[a] (20)	0.03 (17)
Sep_{TVC}	0.17 (20)	0.19 (20)	0.26 (18)	0.64[b] (20)	0.05 (17)
Col_{TVC}	0.56[b] (20)	0.10 (20)	0.46 (18)	0.53[a] (20)	0.01 (17)
$Col_{noTVC} - Sep_{noTVC}$	-0.63[b] (20)	-0.19 (20)	-0.62[b] (18)	-0.60[b] (20)	-0.15 (17)
$Col_{TVC} - Sep_{TVC}$	0.30 (20)	-0.19 (20)	0.08 (18)	-0.40 (20)	-0.06 (17)
$Col_{TVC} - Col_{noTVC}$	0.08 (20)	0.05 (20)	0.19 (18)	0.04 (20)	-0.02 (17)
$Sep_{TVC} - Sep_{noTVC}$	-0.77[c] (20)	-0.09 (20)	-0.60[b] (18)	-0.38 (20)	-0.12 (17)

[a]Correlations significant at α of 0.05
[b]Correlations significant at α of 0.01
[c]Correlations significant at α of 0.001

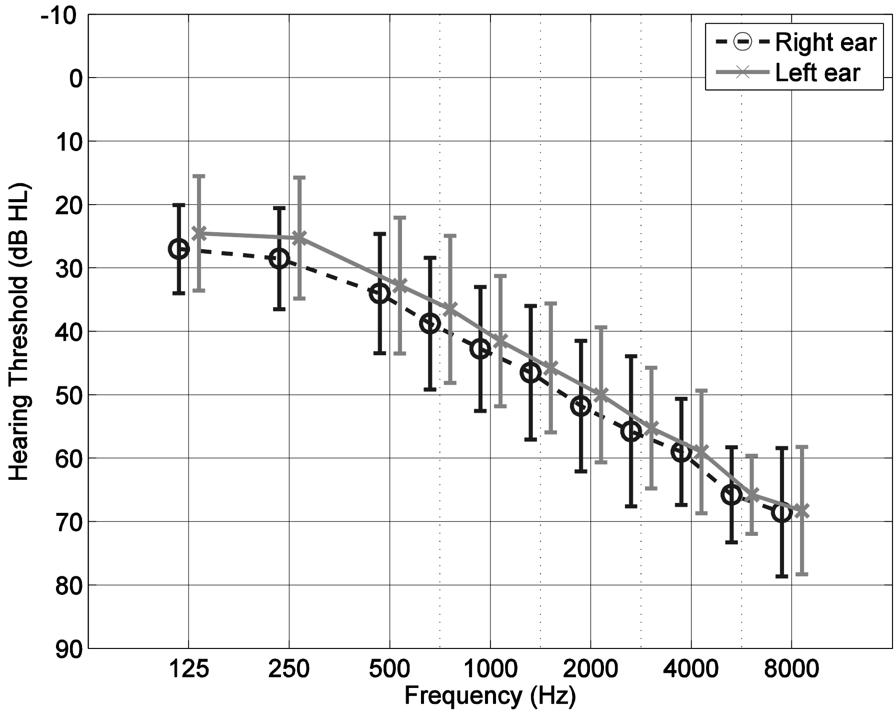

FIG 1. The mean audiograms of the 20 listeners (± 1 standard deviation) for left (gray crosses) and right (black circles) ears separately.

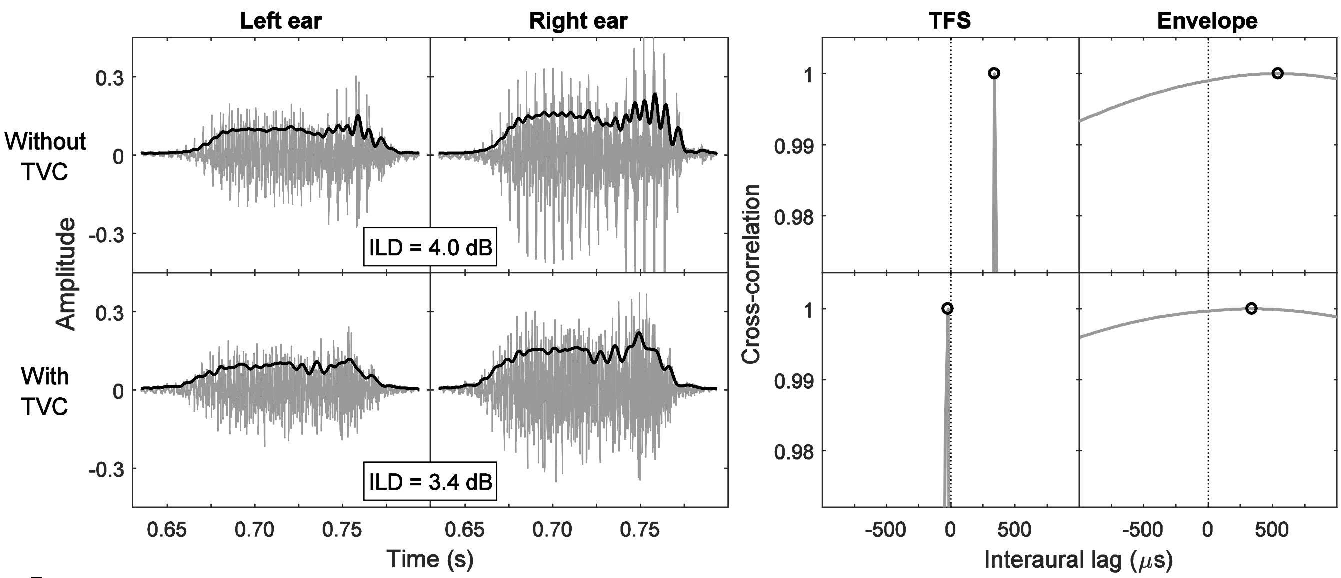

FIG 2. Example excerpts of sentences played from 45° to the right of the listener's position, recorded on a head and torso simulator (B&K). The top row of panels shows the sentence recorded without tone-vocoding and the bottom row of panels show the same sentence recorded with tone-vocoding. The time domain waveform (in gray) of a portion of speech (the word "*valgte*") and its envelope (half-wave rectified and low pass filtered at 160 Hz; overlaid in black) at the left ear (far-left panels) and the right ear (middle-left panels) with the ILDs inset, the interaural cross-correlation of these two signals (middle-right panels) and of the two signals' envelopes (far-right panels). Maximum cross-correlations (ITDs) indicated by black circles.

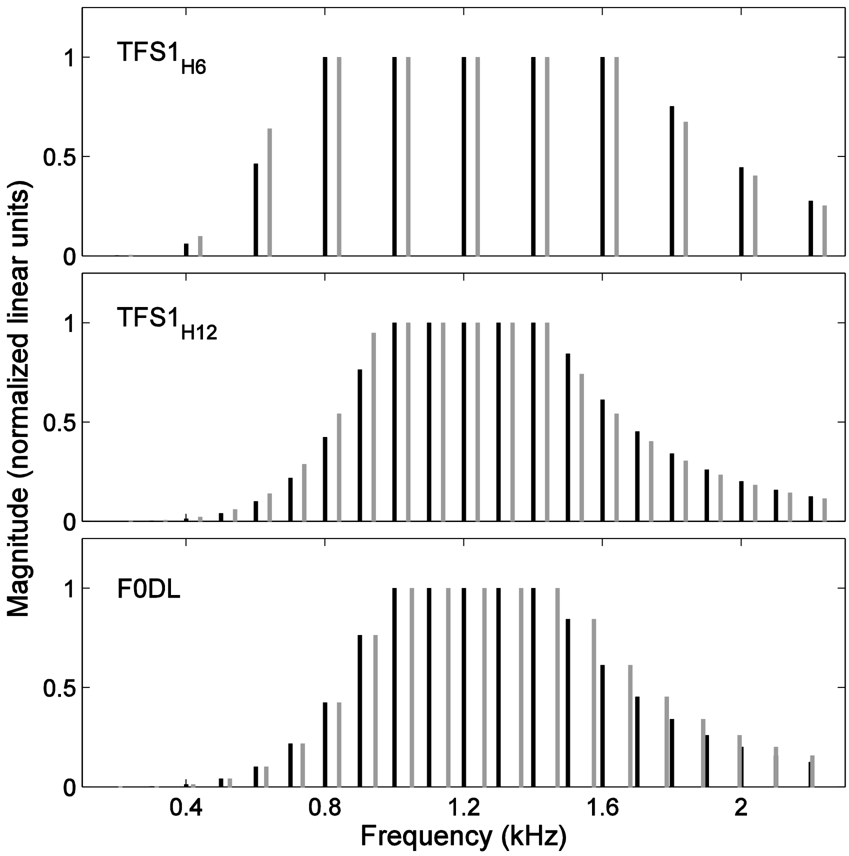

FIG 3. Schematic diagrams of the frequency spectra of the pitch discrimination stimuli on a linear frequency scale. The panels show, from top to bottom, $TFS1_{H6}$, $TFS1_{H12}$ and F0DL. Black lines show the reference stimuli and the gray lines show the shifted stimuli. The top two panels show a shift in all components of +40 Hz, and the bottom panel shows the shift in components due to a modulation rate shift of +5 Hz.

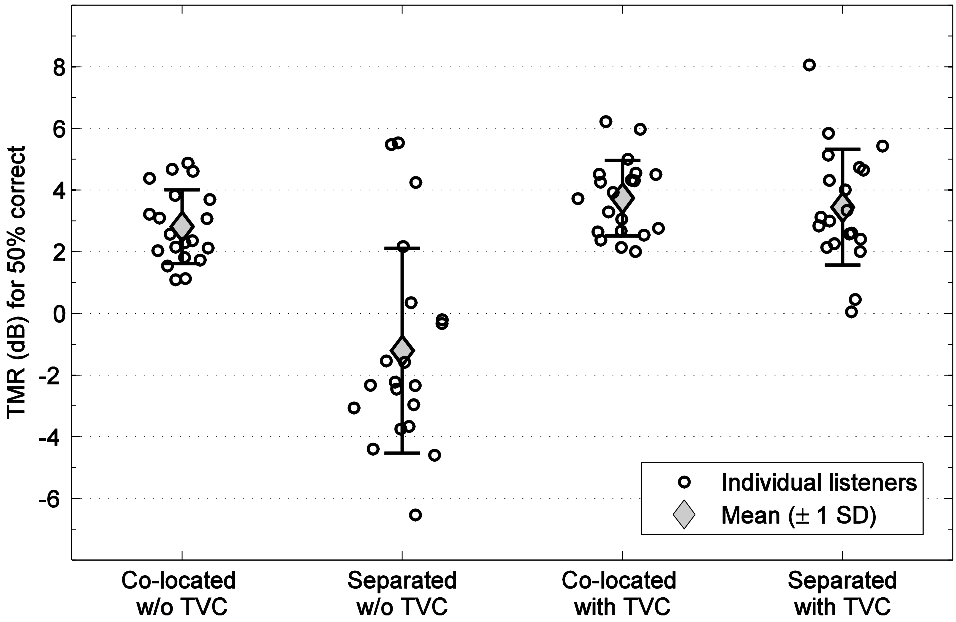

FIG 4. Estimated TMRs for 50% correct word recall ($TMR_{50\%}$), plotted by speech test condition. Black circles indicate $TMR_{50\%}$ for individuals, gray diamonds indicate mean ± 1 standard deviation.

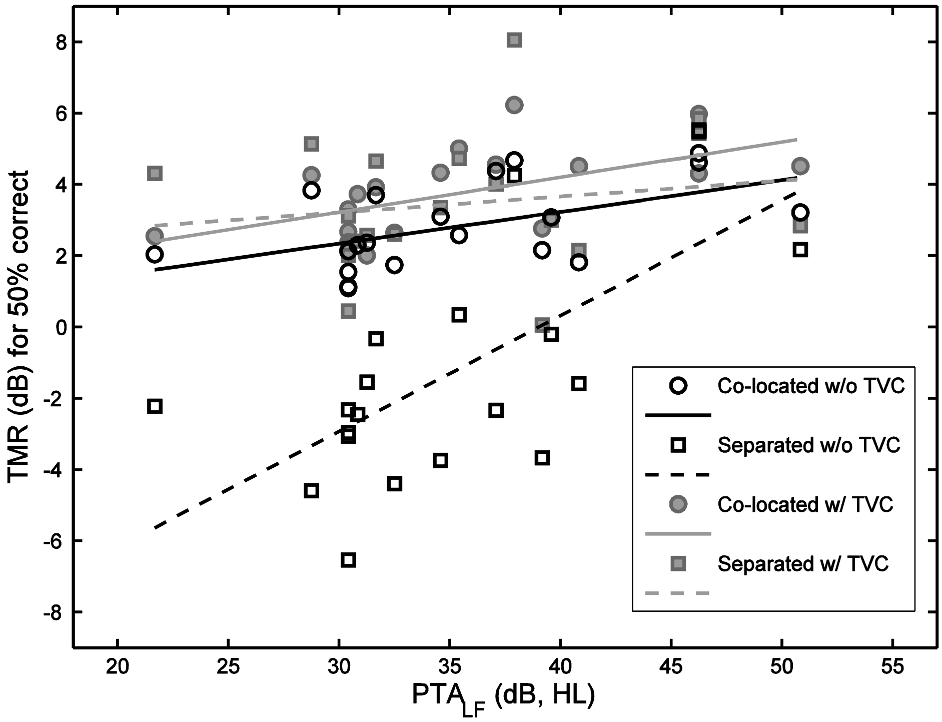

FIG 5. $TMR_{50\%}$ plotted for each test condition as a function of the listeners' low-frequency-average audiometric threshold (PTA_{LF}). Lines indicate least-squares best linear fit for each condition.

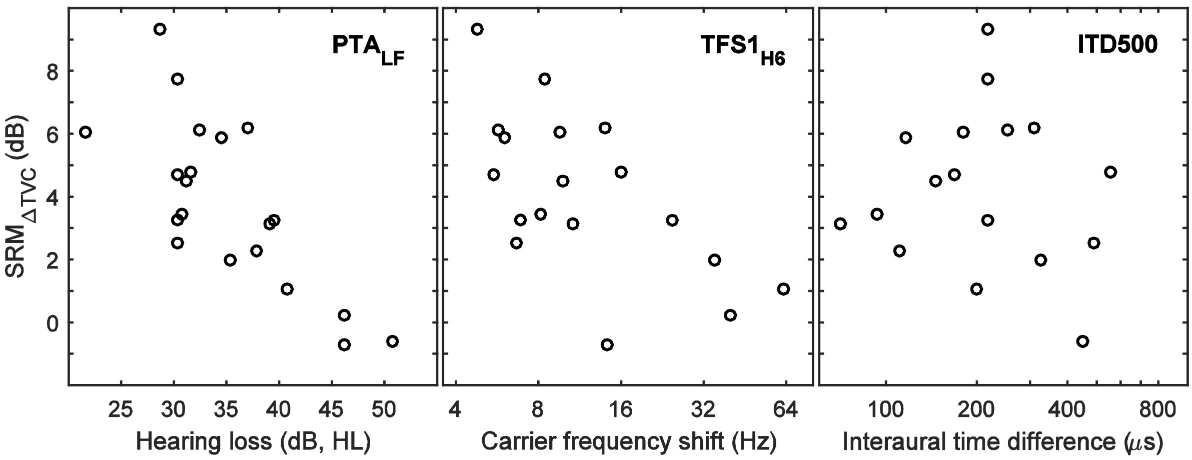

FIG 6. Individual differences in the effect of TVC on SRM ($SRM_{\Delta TVC}$) plotted as a function of PTA_{LF} (left), $TFS1_{H6}$ score (middle), and ITD500 (right).

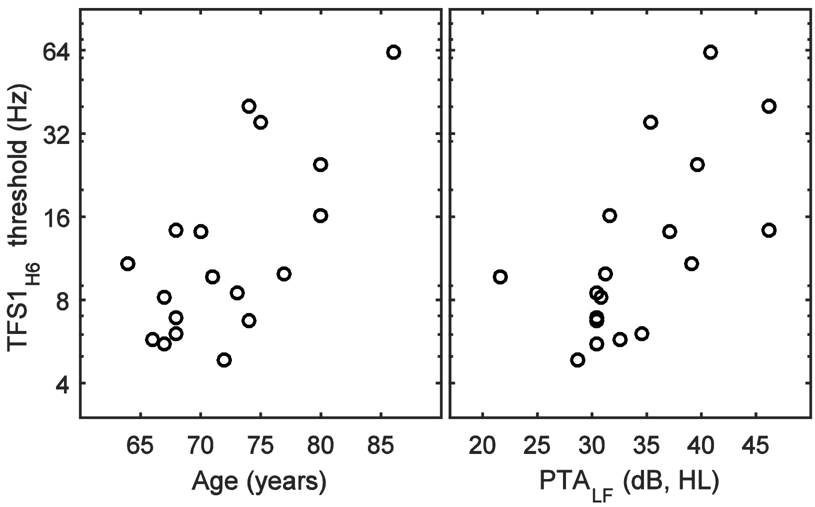

FIG 7. Individual listeners' $TFS1_{H6}$ thresholds plotted as a function of age (left), and PTA_{LF} (right).

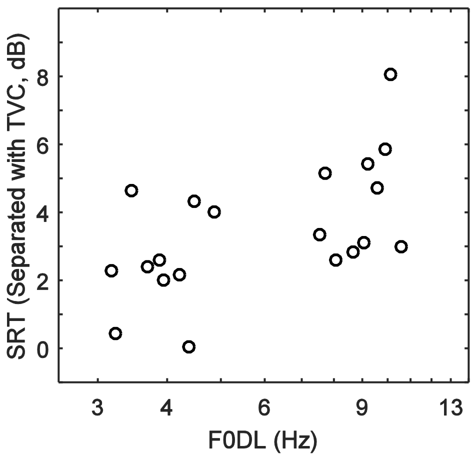

FIG 8. $TMR_{50\%}$ SRTs in the Separated condition with TVC plotted as a function of listeners' F0DLs.