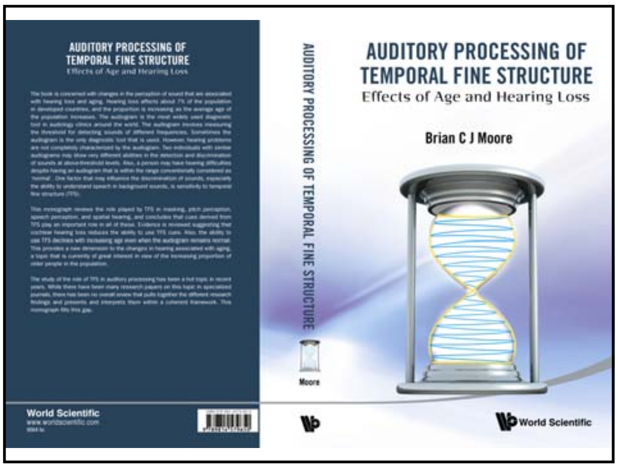
The role of temporal fine structure in the perception of speech in background sounds by people with normal and impaired hearing

Brian C.J. Moore

Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

UNIVERSITY OF CAMBRIDGE



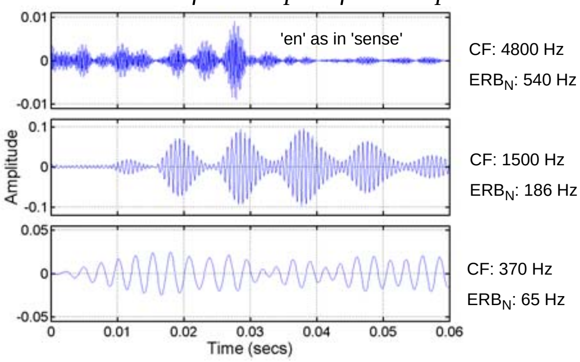
Overview

- What is meant by temporal fine structure (TFS)
- Psychoacoustic tests of sensitivity to TFS
  - The TFS1 test
  - The TFS-LF test
- Approaches to studying the role of TFS in speech perception
  - Vocoder processing
  - Correlational
- Effects of hearing loss and age on the use of TFS for speech perception
- Conclusions and take-home messages

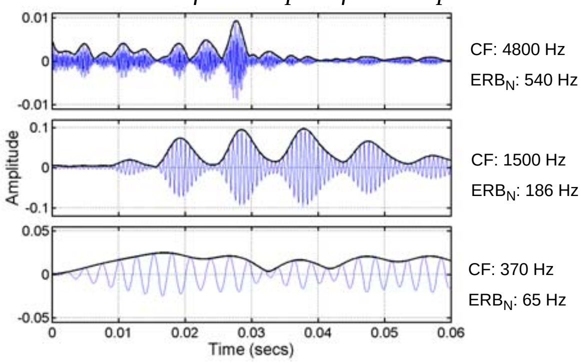
Auditory representation of TFS and E: Normal hearing

- Each place on the basilar membrane (BM) behaves like a bandpass filter
- The response at each place can be considered as composed of
  - Temporal fine structure (TFS): carried by patterns of phase locking in the auditory nerve to individual stimulus cycles
  - Envelope (E): carried by fluctuations in firing rate over time and/or phase locking to envelope

E and TFS for bandpass filtered speech



E and TFS for bandpass filtered speech



Terminology

Three kinds of ENV and TFS:

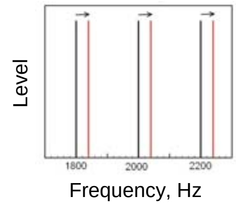
- Physical ENV and TFS of the input signal
  - ENV_p and TFS_p
  - These are not well defined when the signal is broadband
  - Will sometimes be used to refer to signals resulting from filtering into channels (each channel signal is narrowband)
- The ENV and TFS at a given place on the BM
  - ENV_{BM} and TFS_{BM} (can be estimated using gammatone filters)
- The neural representation of ENV and TFS
  - ENV_n and TFS_n
- TFS_p and TFS_{BM} exist over a wide frequency range
- TFS_n weakens at high frequencies
  - The upper limit in humans is unknown

A psychoacoustic measure of TFS sensitivity based on pitch perception

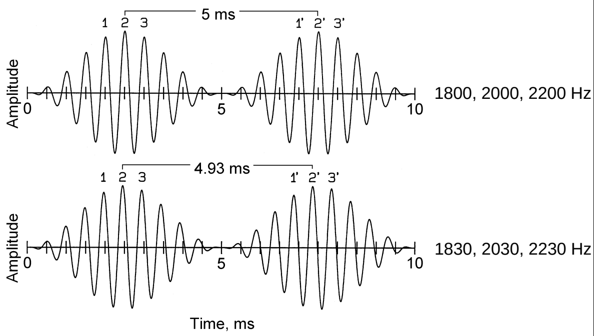
"Frequency-shifted" tones can be created from harmonic (H) tones by shifting each component upwards by the same amount in Hz. Such inharmonic (I) tones have the **same** envelope repetition rate as the original harmonic tone, but **different** TFS.

Schouten, J. F. (**1940**). "The perception of pitch," Philips Tech. Rev. **5**, 286-294.
de Boer, E. (**1956**). "Pitch of inharmonic signals," Nature **178**, 535-536.
Patterson, R. D. (**1973**). "The effects of relative phase and the number of components on residue pitch," J. Acoust. Soc. Am. **53**, 1565-1572.

A shift in pitch is usually heard



Waveforms of unshifted and shifted tones



Reducing excitation-pattern and ENV cues

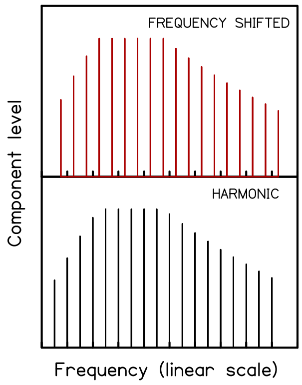
Moore, G. A., and Moore, B. C. J. (**2003**). "Perception of the low pitch of frequency-shifted complexes," J. Acoust. Soc. Am. **113**, 977-985.

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.

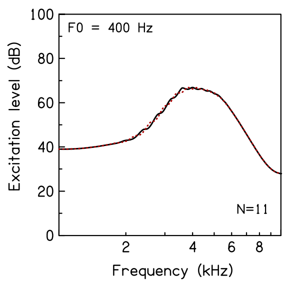
Moore, B. C. J., and Sek, A. (**2009**). "Development of a fast method for determining sensitivity to temporal fine structure," Int. J. Audiol. **48**, 161-171.

- Complex tones with many components
- Passed through a fixed bandpass filter centred on high (unresolved) components
- Passband slopes relatively shallow (30 dB/oct)
  - avoids marked changes in level when a component moves in or out of the passband
- Broadband noise added to mask components on edges of passband and to mask combination tones
- To avoid possible ENV_{BM} and ENV_n cues, the component phases are selected randomly from trial to trial

Spectra (without background noise)



Excitation patterns:
Harmonic (black) and shifted (red) complexes

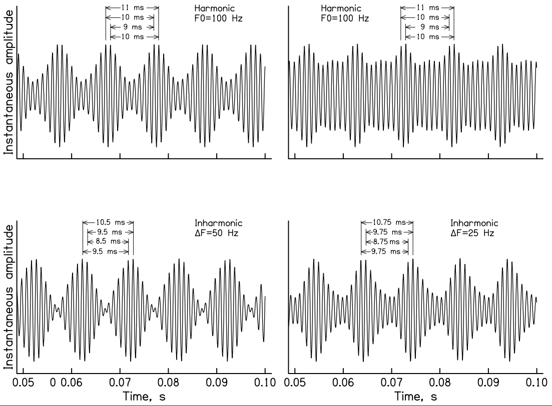


Nominal F0 = 400 Hz

Filter centred on 11th harmonic

With background noise

Simulation of waveforms on basilar membrane (CF = 1000 Hz)



The TFS1 test

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

Sek, A., and Moore, B. C. J. (**2012**). "Implementation of two tests for measuring sensitivity to temporal fine structure," Int. J. Audiol. **51**, 58-63.

- Two-interval forced choice
- Each interval has four 200-ms tone bursts with 100-ms intervals between bursts
- 300-ms silence between intervals
- One interval has H H H H
- Other interval has H I H I
- Task – pick the interval in which the tones alternate in pitch
- Frequency shift (ΔF) varied adaptively to determine "threshold"
- Training effects with this procedure are very small
  - Moore and Sek (IJA, 2009)
  - King et al. (JASA, 2013)

Effect of hearing loss on Performance of the TFS1 test

- People with cochlear hearing loss usually perform poorly on the TFS1 test
- But … excitation patterns do differ slightly for the H and I tones
- Hearing loss is usually associated with broader-than-normal auditory filters
- Could the poor performance of hearing-impaired subjects reflect a reduced ability to use excitation-pattern cues?

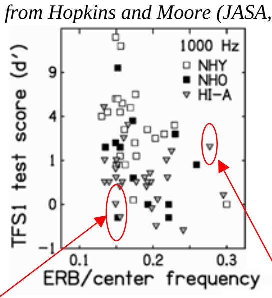
Testing the possible role of excitation-pattern cues (1)

- Auditory filters broaden with increasing level
- If excitation-pattern cues are used, performance should worsen with increasing level
- It doesn't (except at very high levels: Marmel et al., ARO, 2012, abstract 632)

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

Moore, B. C. J., and Sek, A. (**2011**). "Effect of level on the discrimination of harmonic and frequency-shifted complex tones at high frequencies," J. Acoust. Soc. Am. **129**, 3206-3212.

Testing the possible role of excitation-pattern cues (2)
Relationship of TFS1 scores to auditory filter sharpness:
Data from Hopkins and Moore (JASA, 2011)



Chance performance despite normal ERB

Good performance despite large ERB

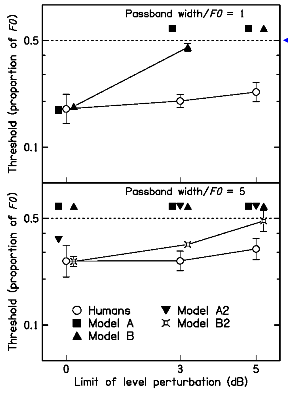
Results suggest that performance on the TFS1 test is not based primarily on excitation-pattern cues

Testing the possible role of excitation-pattern cues (3)

Jackson and Moore (JASA, submitted)

- Randomly perturbed the level of each of the components in each of the H and I tones over ranges ± 3 dB and ± 5 dB
  - Disrupts the pattern of ripples in the excitation patterns
- Models based on the use of excitation-pattern cues predicted that the level perturbation would markedly impair performance
- The performance of human subjects was only slightly (non-significantly) affected by the level perturbation

Results of Jackson and Moore



Highest possible threshold

F0 = 200 Hz

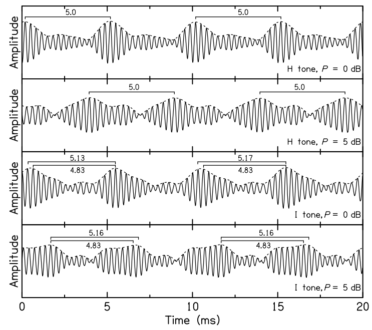
Lowest component in passband = 13th

Conclusion:

Performance of the TFS1 task is not excitation-pattern cues

Effects of level perturbation on TFS_{BM} cues

Simulations of the outputs of auditory filters showed that TFS_{BM} cues (the time intervals between peaks in TFS_{BM} close to adjacent envelope maxima) were only slightly affected by the level perturbation



F0 = 200 Hz, shift = 100 Hz

H tone, no level rove

H tone, 5-dB level rove

I tone, no level rove

I tone, 5-dB level rove

Conclusions on the TFS1 test

- The outcome of the test does not depend (solely) on the use of excitation patterns cues (when the bandpass filter is centred on high harmonics)
- The outcome of the test almost certainly depends on the use of TFS cues

Binaural sensitivity to TFS: the TFS-LF test

- The phase of low-frequency tones can be compared at the two ears and used to localise sounds
- Depends on comparing TFS_n at the two ears
- Hopkins, K., and Moore, B. C. J. (**2010**). "Development of a fast method for measuring sensitivity to temporal fine structure information at low frequencies," Int. J. Audiol. **49**, 940-946:
  - two-alternative forced-choice task
  - each interval contains four tones with frequency *f*
  - in one interval all tones are diotic
  - in the other tones one and three are diotic while tones two and four have an interaural phase shift
  - 0 0 0 0 *vs* 0 ϕ 0 ϕ or 0 ϕ 0 ϕ *vs* 0 0 0 0
  - envelopes always synchronous across ears – TFS_n is needed to perform the task

The role of TFS in speech perception

Two general approaches:

- Various forms of vocoder processing
  - Attempt to reduce TFS cues while preserving ENV cues
  - Attempt to reduce ENV cues while preserving TFS cues
  - Assess effects on speech perception
- Correlational
  - Assess the performance of normal-hearing, hearing-impaired, young, and older subjects on speech-perception tasks
  - Compare to performance on TFS1, TFS-LF (and other) tests

Vocoder processing

- Speech in quiet or in a background sound is filtered into *N* frequency bands or channels
- ENV_p and TFS_p are estimated for each channel
  - ENV_p estimated by rectification and lowpass filtering or via the Hilbert transform
  - TFS_p estimated by dividing the channel signal by ENV_p
- The signal in each channel is manipulated so as to alter either ENV_p or TFS_p
- Each manipulated channel signal is filtered to restrict its spectrum to the passband of the channel
- The filtered channel signals are combined

Vocoder processing intended to disrupt TFS cues

- TFS_p in each channel is replaced by noise (noise vocoder) or a tone at the centre frequency of each channel (tone vocoder)
- The modified TFS_p is modulated by the unmodified ENV_p for that channel
- Speech processed in this way is described as "ENV-speech"
- Often described as "removing" TFS while preserving ENV
- In fact:
  - Any audio signal has TFS
  - Such vocoder processing replaces the original TFS_p by less informative TFS_p
  - The less informative TFS_p still conveys information about the spectro-temporal characteristics of the signal
  - ENV_{BM} and ENV_n for the processed signal are different from ENV_{BM} and ENV_n for the original signal

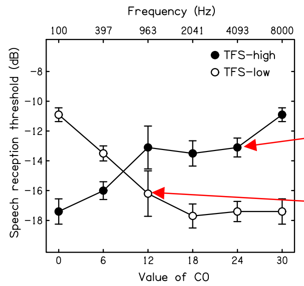
Studies using ENV-speech

- Speech in quiet can be intelligible when *N* is four or more, and intelligibility increases with increasing *N*
  - Drullman, 1995; Shannon et al., 1995; Loizou et al., 1999; Lorenzi et al., 2006
- The intelligibility of ENV-speech decreases markedly when the speech is presented in a background sound
  - Nelson et al., 2003; Qin and Oxenham, 2003; Stone and Moore, 2003
- Possible explanations:
  - Original TFS cues may be important for the segregation of speech from background sounds
  - The poor intelligibility may be a consequence of "modulation masking"
  - Envelope fluctuations in the background sound impair the ability the extract ENV_n information about the target speech (Stone et al., JASA, 2011; 2012)
  - Modulation masking may be especially important when TFS cues are degraded

ENV-Speech continued: The role of TFS in different frequency regions

- Hopkins *et al*. (JASA, 2008) and Hopkins and Moore (JASA, 2010) measured speech reception thresholds (SRTs) for a target talker in a background talker as a function of the frequency range over which original TFS information was available
- The signal was split into 32 1-ERB_N wide channels
- Above or below a cut-off channel, CO, channels were tone or noise vocoded, to remove the original TFS information
- Remaining channels were not processed
- As the number of channels with original TFS information was increased, SRTs decreased (improved)

Results of Hopkins and Moore (2010): tone vocoder



Original TFS information added starting from high frequencies

Original TFS information added starting from low frequencies

Original TFS is important over a wide frequency range

The change of SRT with changing CO was greater for normal-hearing (NH) subjects than for hearing-impaired subjects (not shown)

→ TFS is used more effectively by NH than HI subjects

Effect of type of speech material

- Lunner *et al*. (Ear and Hearing, 2012) repeated the experiment of Hopkins *et al*. (2008) using different types of speech materials and a tone vocoder
- TFS information was added starting from low frequencies
- Danish HINT sentences:
  - similar to the materials used by Hopkins *et al*. (2008)
  - somewhat unpredictable structure
  - drawn from an open set
  - results similar to those of Hopkins *et al*.
- Dantale 2 sentences:
  - highly predictable structure
  - drawn from a closed set
  - the decrease in SRT with increasing CO was similar for young normal-hearing (YNH) and older hearing-impaired (OHI) groups
  - SRTs for YNH subjects were very low even with fully vocoded signal (CO = 0).

SRTs measured by Lunner et al.

Corpus	Participant type	ENV-speech *CO* = 0	Intact speech *CO* = 32	Benefit of original TFS
HINT	OHI	3.1	-0.1	3.2
	YNH	-3.0	-10.3	7.3
Dantale 2	OHI	-3.7	-7.7	4.0
	YNH	-14.9	-18.0	3.1

Larger benefit for YNH than for OHI

Similar benefit for YNH and for OHI

Interpretation

- Speech has a sparse representation in the auditory system
  - the energy is high in only a few spectro-temporal regions, with low energy elsewhere (Darwin, 2009)
- For a mixture of two talkers, there is little overlap between the spectro-temporal regions dominated by one talker and the regions dominated by the other talker
- The identification the target speech is limited mainly by informational masking
- TFS information may reduce informational masking, by providing cues that aid the perceptual segregation of the target and the background
- With highly predictable speech material (Dantale 2) the original TFS information in the signal may not be required

Vocoder processing intended to disrupt ENV cues

- The TFS_p signal (which has a “flat” envelope) for each channel is filtered to restrict its spectrum to the passband of that channel
- The filtered TFS_p signals are combined
- Speech processed in this way is described as TFS-speech
- Often described as “removing” ENV cues while preserving TFS cues
- In fact:
  - ENV cues are reconstructed by filtering of the channel signals
  - Even if the second filtering stage is not applied, filtering in the auditory system results in envelope reconstruction
  - TFS_{BM} and TFS_n for the processed signal are different from TFS_{BM} and TFS_n for the original signal

Illustration of envelope reconstruction

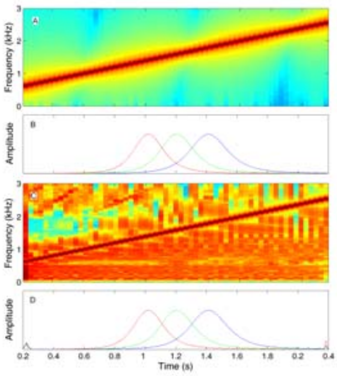


Figure created by Tom Baer

Spectrogram of frequency glide from 600 to 2600 Hz

ENV_{BM} at the outputs of auditory filters with CFs of 1407, 1593, and 1800 Hz

Spectrogram after processing to remove ENV cues while preserving TFS cues

ENV_{BM} for processed signal at the outputs of auditory filters with CFs of 1407, 1593, and 1800 Hz

Studies using TFS-speech

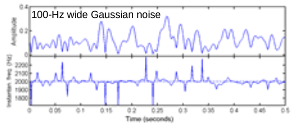
- Sheft et al. (JASA, 2008) created TFS-speech using 16 channels that were approximately 2 ERB_N wide or 32 channels that were approximately 1 ERB_N wide.
- They created three modulators representing the fluctuations in instantaneous frequency of the TFS for each channel
  - FMu contained the unmodified pattern of frequency modulation (FM)
  - FMs: the amount of FM was scaled (reduced) so that deviations in instantaneous frequency were restricted to the passband of the channel
  - The modulator for a given channel was applied to a sinusoidal carrier at the centre frequency of that channel
- A third condition, FMr, was obtained by using the FMu modulator for each channel, but randomizing the starting phase of the carrier, separately for each channel carrier

Studies using TFS-speech (Sheft et al. cont)

- A model of peripheral auditory processing was used to estimate the amount of envelope reconstruction:
  - FMs and FMr modulators led to similar amounts of envelope reconstruction
- But …. intelligibility of processed VCV syllables was lower for the FMs than for the FMr modulator
- Intelligibility scores were poorly correlated with estimates of the amount of envelope reconstruction
- Intelligibility scores were more highly correlated with estimates of the fidelity of preservation of TFS_{BM} and TFS_n cues
- Conclusion: TFS cues contribute to intelligibility and this is not solely a consequence of envelope reconstruction

Studies using TFS-speech (3)

- The instantaneous frequency of TFS_p for a given channel contains wild excursions when ENV_p has a low amplitude
  - partly a consequence of amplification of low-level noise in the signal

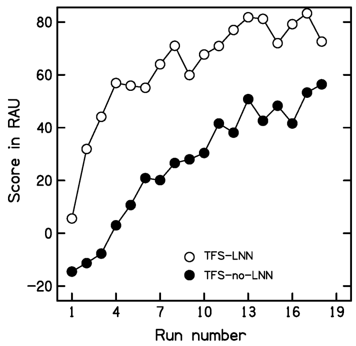


ENV_p

TFS_p

- Hopkins *et al*. (JASA, 2010) added low-noise noise (LNN) to TFS_p for each channel
  - The LNN for each channel had the same long-term average spectrum as the target speech within that channel
  - The LNN reduced the excursions in instantaneous frequency of TFS_p for each channel
  - The LNN made the TFS-speech sound less noisy
  - Simulations and additional experiments suggested that the LNN did not increase the extent of envelope recovery

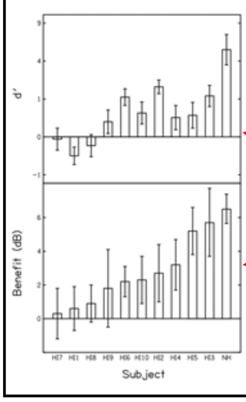
Results of Hopkins et al (2010)



Conclusions: adding LLN reduces artifacts in TFS_p and makes TFS_n more like that for the original signal

TFS_n probably contributes to speech intelligibility

Correlational studies (1):Benefit from original TFS information is correlated with sensitivity to TFS



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

Scores on TFS1 task (mean for two frequency regions)

Overall benefit from original TFS: SRT for all channels vocoded minus SRT for all channels unprocessed

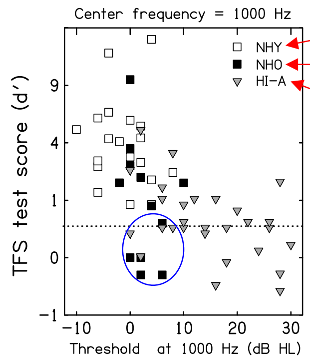
Correlation (HI only) = 0.67, $p < 0.05$

Correlational studies (2): Effect of age and hearing loss on the use of TFS

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.

- Measured:
  - performance on the TFS1 task
  - audiometric threshold at the test frequency
  - sharpness of the auditory filter at the test frequency
  - sensitivity to inter-aural phase (TFS-LF test)
- Three groups of subjects
  - Young, normal hearing (YHN)
  - Old, normal hearing (ONH) (up to 6 kHz)
  - Hearing impaired (HI), wide age range

Hopkins and Moore (2011): TFS1 results



Young normal hearing

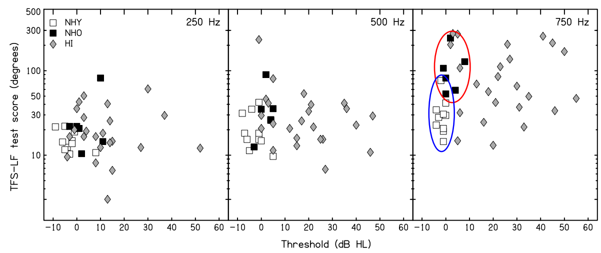
Old normal hearing

Hearing impaired (29-82 years) (mild loss)

Scores above the dashed line are above chance

Some NHO and HI subjects with mild losses cannot do the task, even when audiometric thresholds at the test frequency are 10 dB HL or less

Hopkins and Moore (2011): TFS-LF results



For 750 Hz, older NH (■) perform more poorly than young NH (□)

Other data confirm that TFS-LF scores worsen with increasing age even when audiometric thresholds at the test frequency are "normal"

Moore, B. C. J., Glasberg, B. R., Stoev, M., Füllgrabe, C., and Hopkins, K. (**2012**). "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. (**2012**). "The effects of age on temporal fine structure sensitivity in monaural and binaural conditions," Int. J. Audiol. **51**, 715-721.

Hopkins and Moore (2011): Correlations

- SRTs were measured for speech in a steady background noise, and noise with spectral and temporal dips
- When the effect of mean audiometric threshold was partialled out, SRTs for speech in the modulated noise were correlated with scores on the TFS1 test, but not with scores on the TFS-LF test or with the measures of frequency selectivity
- The results suggest that a reduction in sensitivity to TFS can partly account for the speech perception difficulties experienced by hearing-impaired and by older subjects

Correlational studies (3): Spatial hearing

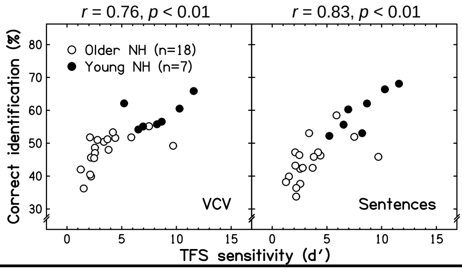
- Neher et al. (JASA 2012)
  - 17 older hearing-impaired subjects
  - SRTs were measured for a female speech target presented from directly in front (0° azimuth), in the presence of two female speech maskers presented from ±50° azimuth
  - subjects wore hearing aids that preserved interaural level cues and ensured audibility for frequencies up to 6 kHz
  - TFS-LF test for frequencies of 250, 500 and 750 Hz; geometric mean gives TFS-LFav
- Results
  - significant correlation between age and TFS-LFav (r = 0.75)
  - TFS-LFav values not correlated with low-frequency audiometric thresholds
  - SRTs were significantly correlated with age (r = 0.71), TFS-LF$_{av}$ (r = –0.63), and with measures of cognitive abilities
  - correlations between SRTs and TFS-LF$_{av}$ scores and between SRTs and measures of cognitive ability became non-significant when the effect of age was partialled out
  - performance on the various measures may be influenced by a common, age-related mechanism

Correlational studies (4): Füllgrabe, Moore and Stone (unpublished)

- Measured sensitivity to TFS using the TFS1 test at 1 and 2 kHz and the TFS-LF test at 0.5 and 0.75 kHz
- Two groups:
  - young (mean age = 23 yrs)
  - older (mean age = 67 yrs)
  - matched for verbal IQ
  - all audiograms bilaterally normal (≤20 dB HL) for frequencies up to 6 kHz
  - mean audiograms closely matched across groups
- TFS performance significantly poorer for the older group than for the young group
- Auditory filters do not broaden with increasing age when the audiogram is normal (Peters and Moore, 1992)
  - the worse performance of the older group cannot be attributed to reduced frequency selectivity

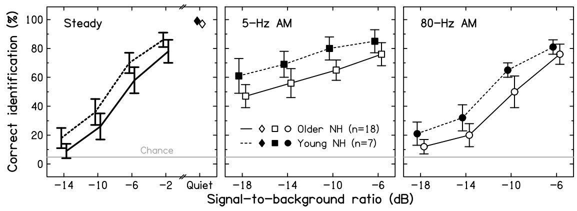
Füllgrabe, Moore and Stone (2)

- Measured the intelligibility of consonants in steady noise and noise that was sinusoidally amplitude modulated at 5 or 80 Hz
- Also measured intelligibility for sentences in a single talker background
- Same SBRs for the two groups
- Scores (averaged across all noise types and all SBRs) were correlated with scores on the TFS tasks, averaged across all centre frequencies and tasks



Füllgrabe, Moore and Stone (3)

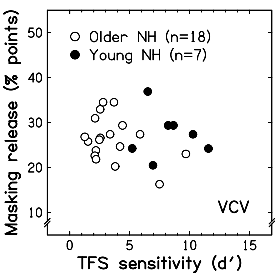
- The older subjects had poorer identification of speech in both the steady and modulated noises



- Masking release was *not* reduced for the older subjects, despite their deficit in TFS processing

Füllgrabe, Moore and Stone (4)

- Masking release for each subject was quantified as the difference in identification scores for consonants in steady noise and in modulated noise, averaged across corresponding SBRs.



No significant correlation between TFS scores and masking release

Conclusions

(1) Monaural TFS sensitivity (TFS1 test) is adversely affected by both hearing loss and age

(2) Binaural TFS sensitivity (TFS-LF test) is mainly affected by age (but haven't studied large low-frequency loss)

(3) The ability to understand speech in background sounds is correlated with TFS sensitivity

(4) TFS sensitivity does not seem to be critical for dip listening
  - Dip listening can occur when the original TFS information is severely degraded by vocoder processing
  - Masking release for speech is not significantly correlated with psychoacoustic measures of sensitivity to TFS

TFS information may be mainly important for aiding perceptual segregation of the target and background

Acknowledgments

Supported by the MRC (UK), the Oticon Foundation, and Action on Hearing Loss (formerly RNID). Thanks to:

Christian Füllgrabe

Brian Glasberg

Kathryn Hopkins

Thomas Lunner

Aleksander Sek

Michael Stone