**Perception of Temporal Fine Structure in Individuals with Normal Hearing Sensitivity: A Comparison of Different Measures**

**Background**

The perception of speech involves the interpretation of complex acoustical patterns and perceiving them as linguistic units. It is a complex task because a single acoustic pattern may not always represent the same speech segments. Instead, the patterns may vary depending upon the preceding or following segments as well as the auditory environment.Speech perception becomes even more difficult in the presence of background noise (Moore, 2003)

The interpretation of speech depends on how well the auditory system decodes the acoustical cues present in it. These acoustical cues that are important for the interpretation of speech segments can be divided into temporal and spectral cues(Moon & Hong, 2014). Temporal fine structure (TFS) and temporal envelope (ENV) are considered as the major temporal cues, and their encoding is considered crucial for speech understanding, especially in the presence of background noise(Lorenzi, Debruille, Garnier, Fleuriot, & Moore, 2009). The TFS,sometimes called the carrier, is characterised by rapid oscillationsin the signal with a rate close to the centre frequency of the frequency band of the signal. The ENV, on the other hand, corresponds to the slow, amplitude modulation of the carrier (or the TFS)over time(Lorenzi et al., 2009; Moon & Hong, 2014).Both TFS and ENV are coded in the auditory nervous system in terms of time related changes in the neural firing. More specifically, TFS is better represented in terms of phase locking (the synchronisation of nerve firing with a particular phase of the stimulus), and ENV is described as amplitude variations in the nerve firings (Buss, Iii, & Grose, 2004; Moon & Hong, 2014).

It is well documented that the ENV cues are sufficient to have good speech perception in the quiet, but it isinadequatein the presence of background noise. This may be because the ENV cues alone fail to provide perceptual segregation in a complex listening environment (Moore, 2008). TFS information is reported to be useful whenlistening to speech in the presence of noise- fluctuating (Hopkins & Moore, 2009) as well as steady-state noise(Moore, Glasberg, Flanagan, & Adams, 2006).Individuals with normal hearing benefit from 'dip listening' in noise, especially fluctuating noise(gathering snippets of the signal when it is audible over the noise). This ability is considerably impaired in individuals with cochlear hearing loss,and it has been attributed to their inability to efficiently use TFS information due to changes in the cochlear mechanics(Henry & Heinz, 2013; Hopkins & Moore, 2007; Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006). Assessment of speech perception abilities in noise, therefore,indicates one'sabilities to utilise TFS information.

However, clear information about thedifferent aspectsof TFS perception is needed to understand the actual contribution of TFS cues to the perception of speech. Accordingly, the studies that explore the sensitivity to TFS information and the relative contribution of TFS and ENV components to speech perception are essential.

Sensitivity to TFS information (of complex tones) may be understood using tests like the TFS1 and TFS-LF(Hopkins & Moore, 2010a; Sȩk & Moore, 2012). These tests adaptively vary the TFS information, whileleaving the envelope unaltered.The TFS1 test varies thefrequency of the TFS component delivered to a single ear, and the TFS-LF test varies the phase of the TFS between the right and the leftears. The ability to detect the smallest change in TFS is used to understand a participant's sensitivity to TFS. For example, Hopkins and Moore (2011) showed that the TFS sensitivity is weak in the elderlywith normal hearing sensitivity (63-66 years), compared to young participantswith normal hearing sensitivity (20-35 years), even when their frequency discrimination abilities were comparable(Hopkins & Moore, 2011). In a similar study, using the same tests Moore, Vickers and Mehta (2012) showed that age and sensitivity to TFS cues were correlated(Moore, Vickers, & Mehta, 2012).

Various methods are used to study the relative role of TFS and ENV cues in speech perception. One such method is 'Vocoding'. It is the extraction of TFS from a speech signal to preserve ENV cues alone or vice versa. In this method, a signal is split into different frequency bands and envelope and TFS are extracted from each band using processes like the Hilbert transform. If the envelope information from the Hilbert analysed signal is to be retained, the extracted envelope is low pass filtered, and a sine wave with a frequency equal to the centre frequency of each band is then amplitude modulated with it. The output from all bands is then combined,and the final product is a signal with only envelope information. (any reference)?

In order to make a TFS only signal, after the Hilbert transformation, the envelope component is discarded. The TFS in each band is multiplied by a constant equal to the root-mean-square (RMS) power of the bandpass filtered signal. The 'power-weighted' TFS signals are then summed over all the frequency bands. These stimuli contain TFS information only and are termed as'TFS-speech'(Lorenzi et al., 2006). Perception of TFS-speech has good intelligibility when the TFS extraction is done from a wide frequency band, and it deteriorates when the number of frequency bands is increased(Zachary M. Smith, Delgutte, & Oxenhamt, 2002). However, studies have shown that the ENV cues are reconstructed at the output of the auditory filters and that the perception of this this'recovered envelope'aids in speech comprehension from TFS-speech(Ghitza, 2001).

To understand this phenomenon further, speech stimuli were made by recovering the envelope from the TFS speech (called 'RENV speech')(Gilbert & Lorenzi, 2006; Léger, Desloge, Braida, & Swaminathan, 2015). The recovered envelope is used to amplitude modulate a sine wave with a frequency equal to the centre frequency of its extracted frequency band. Studies found that these stimuli were also intelligible. However, the intelligibility deteriorated as the number of frequency bands used to create the TFS-speech increased(Lorenzi et al., 2006; Sheft, Ardoint, & Lorenzi, 2008). Nevertheless, findings of these studies indicated that cochlea could indeed re-create intelligible envelopes from TFS-speech.

In contrast, other studies that tried to eliminate theinfluence of RENV on speech perception have shown that TFS information does contribute to speech perception by its own; it does not function as a mere vehicle to carry ENV cues to the cochlea(Hopkins & Moore, 2010b; Hopkins, Moore, & Stone, 2010; Sheft et al., 2008). For example,Sheft,Ardoint and Lorenzi, (2008)assessed the contribution of TFS information to consonant identification using different speech processing methods. They observed the consonant identification patterns under different conditions where the ability for ENV reconstruction was restricted. They reported that TFS cues contributed moreto the perception of place cues, compared to manner cues.

From the above literature discussion, it is seenthat a number of methods exist to study the sensitivity to TFS cuesand its relative contribution to speech perception. Most of these studies were carried out on smaller units of speech like consonants and syllables.Further, an exploration involving perception of TFS cues using different methods and perception of speech in the presence of noisemight develop a new understanding between the relationship between these measures. Therefore, the aims of this study were 1) to explore the relationship between different measures of sensitivity to TFS (TFS speech and Recovered envelope speech using sentences, TFS1 and TFS LF tests), and 2) to explore how the results from the different measures of TFS are related to performance on SPIN testing using sentence stimuli.

**Methods**

**Sentence comprehension tests:**

***Stimuli***

Kannada sentence lists developed by Geetha, Kumar, Manjula, and Pavan (2014) were used to prepare stimuli for the study.There were 24 lists developed, and each list had 10 sentences, each with four keywords, resulting in 40 keywords per list. These sentences were processed in 3 different ways- Initially, removing the envelope and retaining only the TFS from the sentence; secondly, by reconstructing the envelope from the extracted TFS; thirdly, the sentences were mixed with noise to create stimuli for speech perception in noise (SPIN) testing. The procedure used for processing of the stimulus in the first three methodswas similar to Swaminathan et al. (2014).

***TFS-speech:***The sentences were first bandpass filtered into 8 and 16bands (8nb and 16nb) of equal bandwidth on a log frequency scale spanning 80 to 8020 Hz. The output from each band underwent Hilbert analysis and the TFS component within each band was extracted as the cosine of the phase of the Hilbert analytic signal. The TFS component was scaled to match the long-term average energy of the original signal in each bandpass. The resulting amplitude normalised TFS components wereadded to get the TFS-speech stimulus. (TFS speech using 2 and 4 bands were not usedin the study to avoid ceiling effect, since the participants in a pilot testing obtained complete scores in those conditions)

***RENV speech:*** From the TFS stimuli created with 2(2nb) and 4 (4nb) frequency bands (using method similar to the generation of TFS-speech mentioned above), the envelopes were extracted, and RENV speech was created. Each of the TFS-extracted sentenceswasfirst bandpass filtered into 40 frequency bands using 12th order digital Butterworth filter. The frequency bands were of equal bandwidth on a log frequency scale between 80 to 8020 Hz, simulating a cochlear filter bank. The signal was filtered in forward and backward directions. The envelope component within each band was extracted as the magnitude of the Hilbert analytic signal and low-pass filtered at 300 Hz using a sixth-order Butterworth filter. From each frequency band, the recovered envelope was used to vocode a pure tonecarrier with a central frequency of the corresponding frequency band and was band pass filtered. The resultant components were added to get RENV speech(for creating RENV speech, two and four bands of TFS speech was used since the intelligibility of RENV speech reduced drastically after four bands).

***Stimuli generation for speech identification in noise (SPIN) task:***As mentioned above, the same sentence lists (unprocessed) were used to create stimuli for this test condition. Speech shaped noise equivalent to the spectrum of each selected list was produced and mixed at 0 dB signal to noise (SNR) using custom code with MATLAB 2014 (Mathworks Inc., Natick, MA, USA).

**TFS perception of complex tones:**

The participants'sensitivity to changes in the TFSof complex tones was assessed using two tests developed by Moore and colleagues. One test assessed sensitivity to TFS within the one earfor high frequency components(Moore & Sek, 2009), whereas the other test assessed sensitivity to TFS across two ears for low frequency stimuli(Hopkins & Moore, 2010a).

**Participants:**

Twenty individuals in the age range of 18 to 25 years(mean age: 20.4 years) participated in the study. They had audiometric thresholds within 15 dB between 250 and 8000 Hz and no history of hearing or comprehension difficulties as reported. All the participants were native speakers of Kannada (a language spoken in the south Indian state of Karnataka) with proficiency in comprehending speech and script in the language. An informed consent form was obtained from all the participants, the method abided the ethical guidelines for bio behavioural research involving human subjects(Venkatesan, 2009)and was approved by the ethical committee for research at the institute.

**Procedure:**

The participants were seated comfortably in a chair, and the testing was carried out in a sound treated room.Sentence comprehension tests, as well as the TFS tests, were carried out in random orders, to rule out order effect. All the tests were carried out using aLenovo Laptop (running on Windows 10 OS, Intel(R) i3-2370M CPU) and the stimuli were presented through calibrated headphones (Sennheiser HDA200).

***Sentence comprehension tests:***The stimuli for the speech identification tests were conducted for the four different speech processing conditions- (TFSnb8, TFSnb16, RENVnb2 and RENVnb4) and speech perception in noise(SPIN)). The stimuliin these conditions were presented to the participants through the software Paradigm (version 2.5.0.68). The sentence list was randomly selected for each condition from the 24 lists,and no list was repeatedly presented to a participant. Under each condition, one list, with 40 keywords was presented. The stimuli were presented at 70 dB SPL. Stimuli for speech comprehension with processed speech and SPIN were presented, and the participants had to repeat the sentences heard verbatim. The responses were voice recorded for the scoring of keywords. Each correctly identified keyword was given a score of 1, and wrongly identified keywordwas assigned a score of 0. Therefore, the maximum achievable score was 40 after the presentation of a sentence list in each stimulus condition.

***TFS perception of complex tones:***The procedure followed for the test TFS1 was based on Moore and Sek (2009) and for TFS-LF was based on Hopkins and Moore (2010). Both tests used a two-interval two-alternative forced-choice method. The two intervals were separated by 500 ms, and each interval contained four tones of 400ms duration. The tones were consecutively presented with 100 ms gap between them. All the four tones in one of the intervals had identical TFS (the standard) and in the other interval (the target),the second and fourth tones had different TFS than the standard. The participants'task was to identify the target interval (which is perceived as an interval with tones that vary in pitch). Feedback was given after each trial. Thestarting and variable parameters of the test stimuli were set, and the software used a 2-down, 1-up adaptive procedure to arrive at the 71% correct point on the psychometric function (Levitt, 1971). Eight reversals were carried out with varying TFS parameters, and the values from the last six reversals were usedto calculate the threshold. If the SD of the last six reversals was more than 0.2, new testing was carried out. If during the adaptive procedure, the value of the variable parameter exceeded the maximum more than two times, the method of constant stimuli was used (40 trials) with the value of the parameter fixed at maximum. The thresholds were estimated once, after familiarisation of the task and stimulusby the participants. The tests were carried out using HP Laptop (running on Windows 10 OS, Intel(R) i5-6200U CPU), and the stimuli were presented through calibrated headphones(Hopkins & Moore, 2010a).

*TFS1 test:* This test assessed the monaural sensitivity to TFS in complex tones(Moore & Sek, 2009).All the tones used in the test had the same fundamental frequency (F0). Tone complexes in the standard interval had harmonic complex tones as the TFS. In the target interval, the second and the fourth tones had TFS where all the harmonics were shifted by a particular amount (initially set to 0.5F0 and manipulated during the adaptive procedure) resulting in inharmonic complex tones. The envelope repetition rate of the tone complexes remained the same, and only TFS was changed. The tones were filtered through a pass band centered on 9F0 (as harmonics beyond the 8thare not resolved on the basilar membrane)( Moore & Ohgushi, 1993; Plomp & Mimpen, 1968). The F0 used was 200 Hz, corresponding to a centre frequency of 1800 Hz.The stimuli were presented at 50 dB SPL to all participants as the presentation level was shown to have no significant impact between 30 and 50 dB SPL in normal hearing individuals(Moore & Sek, 2009).Threshold equalising noise (TEN)(Moore, 2010)was presented 15 dB below the presentation level of the harmonic and inharmonic tones to mask combination tones andcomponents falling on the skirts of the bandpass filter(Moore, Glasberg, Stoev, Fullgrabe, & Hopkins, 2012).

*TFS-LF test:*This test assessed the binaural sensitivity to TFS in complex tones (Hopkins & Moore, 2010a). The participants had to listen carefullyto the two intervals of stimuli and indicate which interval contained tone complex with a phase shift between the ears (perceived to shift between the ears) compared to an interval with tone complexes with identical phases at the two ears (perceived at the centre of the head). The phase shift between the ears (delta φ) was the manipulated variable, and the initial value was set at 1800. This was carried out using500 Hz stimulus considering good sensitivity to TFS at this frequency, compared to higher frequencies using this test. The stimuli were presented simultaneously to both the ears and were presented at 50 dB SPL in each ear.

**Results**

The participants’ individual scores on each test are presented as scatterplots in figures 1 and 2. The mean, median and standard deviation of raw scores of speech comprehension tasks (RENV, TFS conditions, and SPIN) and the thresholds from tests assessing TFS perception of complex tones (TFS tests) are shown in Table 1. The Shapiro-Wilk test was done on raw scores to check the normality of distribution in each condition. Non-parametric statistics were used for further analysis, as the data was not normally distributed across conditions. The correlation between scores on conditions of RENV, TFS speech perception, thresholds from TFS tests and these scores with SPIN scores were checked using Spearman's correlation coefficient. The scores were significantly correlated between the following conditions measuring TFS sensitivity: RENVnb2 with RENVnb4 (r = 0.449, *p* = 0.047) and TFSnb8 (r = .766, *p* = 0.000); TFS1 right ear with TFS1 left ear scores (r = 0.488, *p* = 0.029). There was no significant correlation (p< 0.05) between any of the RENV speech, TFS speech conditions, TFS tests’ scores with SPIN scores at 0 dB SNR.

Table 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | **Condition** | **Mean** | **Median** | **SD** |
| **RENV speech** | nb 2 | 10.30 | 10.50 | 7.72 |
| nb 4 | 1.15 | 0.00 | 1.72 |
| **TFS speech** | nb 8 | 16.70 | 19.00 | 10.01 |
| nb 16 | 0.15 | 0.00 | 0.49 |
| **TFS-1 test** | 1R | 16.24 | 13.25 | 9.21 |
| 1R SD | 0.15 | 0.17 | 0.05 |
| 1L | 17.24 | 14.55 | 6.70 |
| 1L SD | 0.14 | 0.14 | 0.05 |
| **TFS-LF test** | LF | 24.59 | 23.75 | 12.78 |
| LF SD | 0.12 | 0.11 | 0.04 |
| **SPIN** | 0 dB SNR | 38.50 | 38.50 | 1.87 |

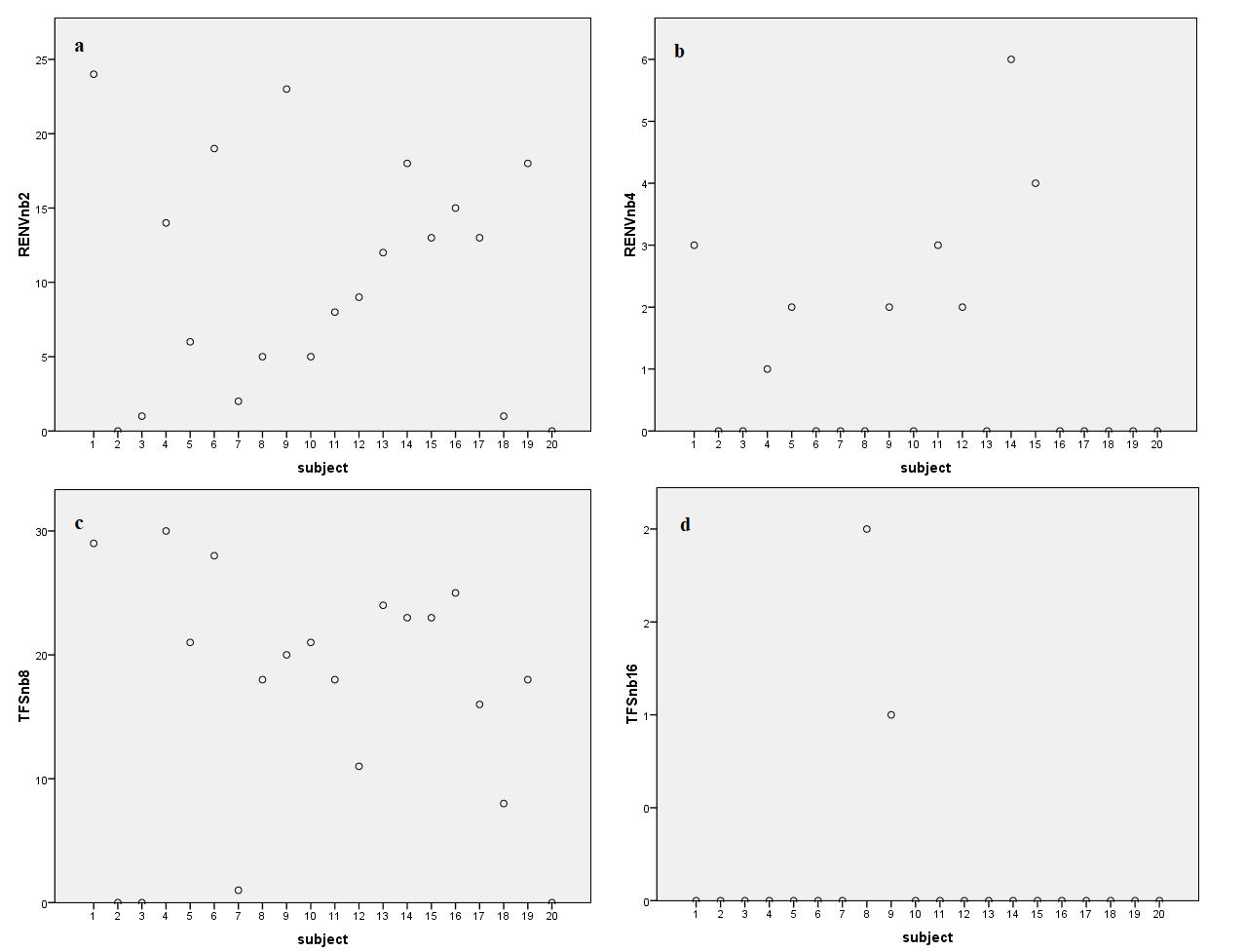


Figure 1

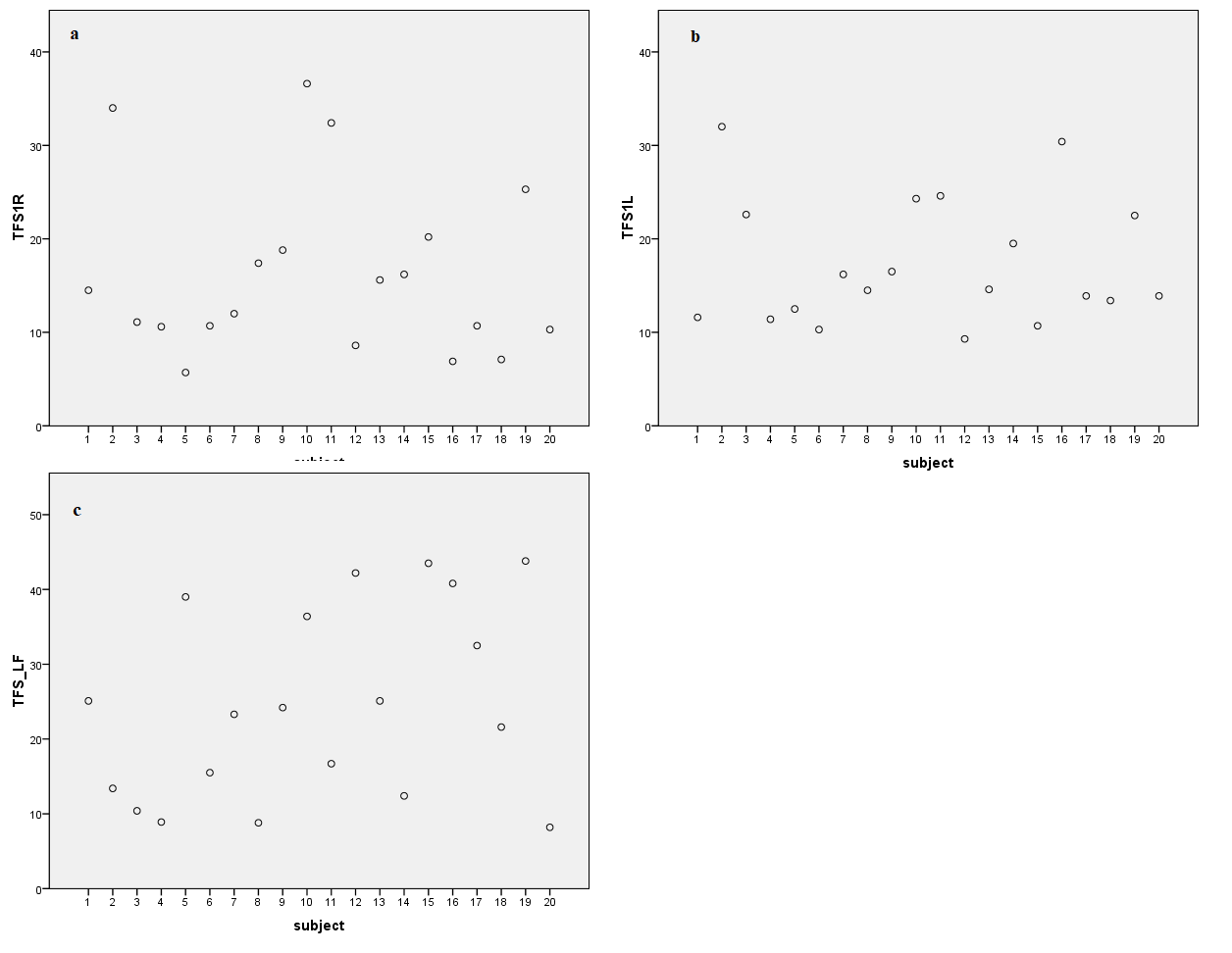


Figure 2

**Discussion**

The goal of the present study was to examine the correlation of measures of TFS perception and sensitivity to TFS cues to the participants' ability to perceive speech in the presence of noise. The findings of the study showed no significant correlation of the different measures considered in the study, namely, the perception of TFS-speech, the perception of RENV-speech, and the perception TFS in complex tones with SPIN scores in the participants of the study. The results of the study are in agreement with some of the literature(Neher, Laugesen, Søgaard Jensen, & Kragelund, 2011; Strelcyk & Dau, 2009)and in disagreement with the others(Füllgrabe, Moore, & Stone, 2015; Peters, Moore, & Baer, 1998).This discrepancy may be majorly attributed to the differences in the methods used in the studies. These differences may be related to parameters like the background noise used for SPIN, the population tested, the stimulus used for tests, etc.

Peters et al. (1998) measured Speech recognition thresholds (SRT) in steady and fluctuating background noise for individuals with normal hearing and young and older individuals with hearing loss. They compared the SRTs with TFS1 and TFS-LF tests and found a good correlation between SRTs in the modulated noise and scores on TFS1 test in older individuals and younger and older individuals with hearing loss. However, the correlation between the measuresis not considered for young participants with normal hearing.The test measure used here was the SRT and not the speech identification score (SIS), and the noise too differed from the present study.

But studies have been conducted where sentences were used to measure SIS (as a measure of speech perception), and they have useddifferent measures of TFS sensitivity (like the TFS1 and TFS-LF tests) to understand the relationship between TFS sensitivity and speech perception. Fullgrabe et al. (2015) observed a good correlation between the TFS1 and TFS-LF test scores and SPIN scores of their young normal hearing participants. However, they had used modulated noise (whereas the present study used non-fluctuating noise) for SPIN testing of sentences, and the testing was done in a sound field condition. On the other hand, with a similar testing paradigm, Neher et al. (2011) found no correlation between speech perception and scores on TFS-LF test in a group of 8 normal hearing participantsduring sound field testing. They had used sentences and spatially separated fluctuating background noise for stimuli. Strelcyk and Dau (2009) found no correlation between the measures of TFS perception and speech perception scores in the presence of modulated noise. But they found a significant correlation between measures of TFS perception and speech perception in the presence of two-talker babble. Therefore, we see that the findings of the studies vary depending upon the stimulus used, speech perception measures considered and the noise used. The present study used a non-fluctuating noise for assessing SPIN since speech identification can happen in places where the SNR is good while listening to a fluctuating noise. This means that in such occasions, speech is not effectively masked(Cooke, 2006). Therefore, a speech spectrum noise was used and the scores obtained using the same can be considered obtained from a true masker.

The study also found correlations between some, and not all the measures of TFS sensitivity used. The different tests used in the study measured sensitivity to TFS information, but possibly the different aspects of sensitivity to TFS. Perception of TFS information in TFSnb8 and TFSnb16 conditions involved perception of the extracted TFS (or the resultant recovery of the envelope at the level of the listener’s cochlea) from the sentence stimuli. The RENVnb2 and RENVnb4 conditions tested the listener’s ability to perceive the simulation of extracted envelope from the TFS speech. Significant correlation found between the two RENV conditions was possibly because the two tests measured the same construct underlying TFS perception. Good correlation was also seen between scores from RENVnb2 and TFSnb8 conditions. Even though RENV speech stimuli were derived from TFS speech, the number of bands used for extraction of TFS from the original stimuli were different, to avoid floor effects. It is possible that the recovery of the envelope from the TFS speech at the cochlear level is correlated with simulations of recovery of the envelope. At the extreme conditions, however, the perception of speech in either condition deteriorated in the participants. It was also seen that the deterioration varied among the participants and this variability could be the reason that correlations were not observed between these conditions.

The TFS perception of complex tones were not correlated with the measures of speech perception, but TFS1 test scores correlated between the right and left ears. All the participants of the study were young normal hearing individuals. Possibly their ability to perceive changes in high frequency TFS information correlated between the ears possibly due to this. However, these measures did not correlate with the results of TFS-LF test, indicating that the ability to perceive low frequency and high frequency TFS information were not comparable. Other studies that have reported correlation between TFS tests and speech perception in noise have compared these measures in different age groups (Füllgrabe, Moore, & Stone, 2015; Peters & Moore, 1992). A comparison of the same measures as administered in the present study and differences in the procedures used in the studies (like the noise and the speech test used) could have contributed to the differences in the findings. As stated, several studies have found correlations between measures of TFS sensitivity and speech perception in the elderly(Füllgrabe et al., 2015; Hopkins & Moore, 2010a; Peters et al., 1998)or in individuals with hearing loss(Füllgrabe et al., 2015; Hopkins & Moore, 2010a; Peters et al., 1998).The present study focussed on young individuals with normal hearing alone. Further studying the same measures in other populations susceptible to poorer processing of TFS cues may reveal more information regarding the relationship between different measures of TFS sensitivity and SPIN.

**Conclusions**

In the present study, an attempt was made to see if a young person's ability to understand speech in a commonly encountered adverse listening environment is related to a measure of sensitivity to TFS cues. The results of the study and the ensuing discussion show that in a normal hearing young adultspeech perception in the presence of continuous noise is not related to their sensitivity to different measures of TFS perception. However, the same may not be true for speech perception in the presence of fluctuating noise, or when the tests are administered on a different population.

**Figure Legends**

Figure 1: Scatterplots of individual speech perception scores from participants on RENVnb2 (a), RENVnb4 (b), TFSnb8 (c) and TFSnb16 (d) conditions.

Figure 2: Scatterplots of individual data from participants on TFS1 test (Right ear – a and Left ear -b) and TFS-LF test (c).

**Tables**

*Table 1: Mean, median, and standard deviation (SD) for scores obtained in RENV speech, TFS speech conditions, TFS tests and speech identification in noise.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | **Condition** | **Mean** | **Median** | **SD** |
| **RENV speech** | nb 2 | 10.30 | 10.50 | 7.72 |
| nb 4 | 1.15 | 0.00 | 1.72 |
| **TFS speech** | nb 8 | 16.70 | 19.00 | 10.01 |
| nb 16 | 0.15 | 0.00 | 0.49 |
| **TFS-1 test** | 1R | 16.24 | 13.25 | 9.21 |
| 1R SD | 0.15 | 0.17 | 0.05 |
| 1L | 17.24 | 14.55 | 6.70 |
| 1L SD | 0.14 | 0.14 | 0.05 |
| **TFS-LF test** | LF | 24.59 | 23.75 | 12.78 |
| LF SD | 0.12 | 0.11 | 0.04 |
| **SPIN** | 0 dB SNR | 38.50 | 38.50 | 1.87 |