CONCLUSIONS

This study demonstrated that a SPIF can predict speech intelligibility for a range of hearing impairments. These results are promising, indicating that using the AN model, predict speech intelligibility results, even for aided listeners with SNHL. The NAL-RP and DSL 4.0 linear hearing-aid fitting algorithms were compared using simulated performance intensity functions. The results showed that, while for both a *flat moderate* and *flat severe* SNHL the simulated results matched those for real listeners, there was little to differentiate the results for the fitting algorithms. From a speech intelligibility perspective, the simulations predicted that both algorithms provide similar intelligibility gains which reinforces the empirical findings of Ching *et al*.

REFERENCES

Bondy, J., S. Becker, *et al*. (**2004**). "A novel signal-processing strategy for hearing-aid design: neurocompensation" Signal Processing **84** (7), 1239-1253.

Boothroyd, A. (**1968**). "Developments in Speech Audiometry" Sound **2** (1): 3 - 10.

Boothroyd, A. (**2006**). Computer-Aided Speech Perception Assessment (CASPA) 5.0 Software Manual. San Diego, CA. San Diego, CA.

Boothroyd, A. (**2008**). "The Performance/Intensity Function: An Underused Resource" Ear and Hearing **29** (4), 479-491.

Bruce, I. C., F. Dinath, *et al*. (**2007**). Insights into optimal phonemic compression from a computational model of the auditory periphery. *Auditory Signal Processing in Hearing-Impaired Listeners, Int. Symposium on Audiological and Auditory Research (ISAAR)*. eds T. Dau, J. Buchholz, J. M. Harte and T. U. Christiansen. Danavox Jubilee Foundation Denmark: 73-81.

Ching, T. Y. C., H. Dillon, *et al*. (**2010**). "Evaluation of the NAL-NL1 and the DSL v.4.1 prescriptions for children: Paired-comparison intelligibility judgments and functional performance ratings" International Journal of Audiology **49** (S1).

Dillon, H. (**2001**). *Hearing Aids*. Thieme Medical Publishers, New York.

Hines, A. and N. Harte (**2010**). "Speech Intelligibility from Image Processing" Speech Communication **52** (9), 736-752.

Hines, A. and N. Harte (**2011**). "Speech Intelligibility prediction using a Neurogram Similarity Index Measure" Speech Communication, [doi:10.1016/j.specom. 2011.09.004].

Kates, J. M. and K. H. Arehart (**2010**). "The Hearing-Aid Speech Quality Index (HASQI)" J. Audio Eng. Soc **58** (5), 363--381.

Zilany, M. S. A., I. C. Bruce, *et al*. (**2009**). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics" J. Acoust. Soc. Am. **126** (5), 2390-2412.

Predictive measures of the intelligibility of speech processed by noise reduction algorithms

KAROLINA SMEDS[1], FLORIAN WOLTERS[1,2], ARNE LEIJON[3], ANDERS NILSSON[1,3], SARA BÅSJÖ[1], AND SOFIA HERTZMAN[1]

[1]*Widex A/S, ORCA Europe, Stockholm, Sweden; karolina.smeds@orca-eu.info*
[2]*Univeristy of Applied Sciences, Oldenburg, Germany*
[3]*KTH, Stockholm, Sweden*

A number of predictive measures were evaluated in terms of their ability to predict the effect on speech intelligibility of different types of noise reduction (NR). Twenty listeners with hearing impairment and ten listeners with normal hearing participated in a blinded laboratory study. An adaptive speech test was used. The speech test produce results in terms of physical signal-to-noise ratios that correspond to equal speech recognition performance with and without the NR algorithms, which facilitates a direct statistical test of how well the predictive measures agree with the experimental results. Three NR algorithms and a reference condition were compared. The experimental results were used to evaluate a number of predictive measures, including a standard Speech Intelligibility Index (SII) method, two time-variable SII methods, and one coherence-based SII method. Further, one measure based on the correlation between band envelope magnitudes of clean and processed noisy speech was evaluated. The measures that make short-time analyses of both speech and noise did best in the comparison.

BACKGROUND

Noise reduction (NR) is commonly used in modern hearing aids (HAs). Previous measurements (Smeds *et al*., 2009) have shown that hearing aid NR algorithms function in very different ways. It would be of great value if predictive measures could be used to indicate the effect of various NR algorithms prior to laboratory or field testing with listeners. The now reported work was part of a larger study, where both speech intelligibility and sound quality of NR processed speech were evaluated. The sound-quality work has been reported by Smeds *et al*. (2010).

GENERAL METHOD

Twenty listeners with hearing impairment (HI) and ten listeners with normal hearing (NH) participated in an adaptive speech test. The listeners with impaired hearing were provided with individualized gain using tightly fitted linear hearing aids. Three NR algorithms and a reference condition were compared using pre-processed sound files. The experimental results were used to evaluate five predictive measures of speech intelligibility.

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

Participants

Twenty listeners with symmetrical, sensorineural, mild-to-moderate hearing loss (Fig. 1, left panel), eleven women and nine men, were recruited from a research database at ORCA Europe. Their ages ranged from 62 to 82 years (mean 71.5 years). Ten listeners with normal hearing (Fig. 1, right panel), six women and four men, were recruited by advertising at the Stockholm University. Their ages ranged from 19 to 28 years (mean 23 years).

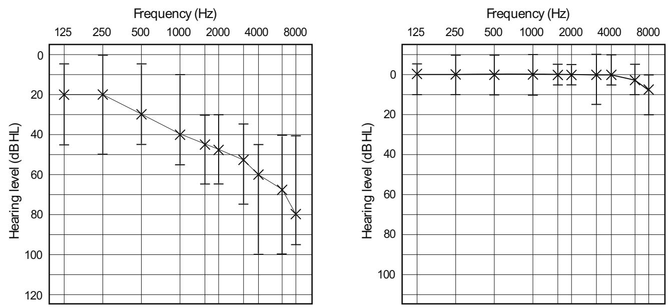

Fig. 1: Median thresholds (crosses) and range of hearing losses (bars) for 20 HI (left panel) and 10 NH listeners (right panel).

Hearing aids

High-quality hearing aids (Inteo 9, Widex A/S), linearly programmed according to the NAL-R prescription (Byrne and Dillon, 1986) reduced by 6 dB across the frequency range, were fitted bilaterally to the HI participants. All advanced signal processing was switched off. The hearing aids were used with tight earmoulds. The hearing aid fittings were verified using real-ear insertion gain measurements, and the linearity was confirmed using coupler-gain measurements.

Noise Reduction Algorithms

Three software-implemented NR algorithms were used. These were selected to create sound files that really differed after NR processing, evaluated by normal-hearing listeners in an informal listening test.

WEDM (a Bayesian noise estimator based on the Weighted Euclidean Distortion Measure) and Wiener (Wiener filtering based on a priori SNR estimation) are described in a textbook by Loizou (2007) and the Matlab codes provided in the textbook were used. The PSSLP (Perceptually tuned Spectral Subtraction algorithm with Low-Pass filtered spectral filter coefficients, (Luts *et al*., 2010)), was fine tuned for hearing aid use.

Hagerman speech test

A Swedish adaptive sentence test using 5-word sentences with a fixed syntax spoken by a female talker was used (Hagerman, 1982). The test result was the SNR at 80% correctly repeated keywords.

An unintelligible artificial babble noise was derived by superimposing the International Speech Test Signal (ISTS) (Holube *et al*., 2010) eight times with randomly varying starting points and the levels pair-wise decreased by 2, 4, and 6 dB relative to the first pair. The babble file was then filtered to the long-term average spectrum of the speech sentences. The sentences were mixed with the artificial babble in SNRs from -12 dB to +15 dB with a step size of 1 dB. These mixed speech and babble files were then processed by the three NR algorithms. The reported results are the nominal SNRs used, i.e., the SNR at the input to the NR algorithms. The speech level was fixed at 70.5 dB SPL while the noise level was varied. Data were collected twice at two visits.

The listening test was performed in a sound-proof booth (3.2×3.05×2.0 m). The participants listened binaurally under sound-field conditions using one loudspeaker (Jamo D400) placed one meter in front of the listener. The frequency response from the digital signal to the listening position was flat within ±4 dB. The measured frequency response was included in all theoretical calculations. The sound files were stored on a PC and played back with an external 24-bit RME Fireface 880 sound card.

PREDICTIVE MEASURES

The results from the speech test were compared to five theoretical measures of speech intelligibility. The speech test produces results in terms of physical signal-to-noise ratios that correspond to equal speech recognition performance with and without the NR algorithms. This facilitates a direct statistical test of how well the predictive measures agree with the experimental results. A good predictive measure will give the same calculated value for all four conditions. In the following, Friedman's two-way analysis of variance by ranks test was used (p=0.05). Another advantage with this method is that no assumptions about transfer functions from calculated scores to predicted speech recognition scores are necessary.

The individual speech test results were entered in all the calculations together with individual hearing thresholds and insertion gain values. Speech and noise files corresponding to the individual speech test results were prepared using a method of separating speech and noise (Hagerman and Olofsson, 2004). Average results from the two visits for the best ear are presented.

1. SPEECH INTELLIGIBILITY INDEX, SII

Method

The speech intelligibility index, SII, quantifies audibility of speech based on long-term average estimates of the speech and noise spectra and the hearing threshold levels (ANSI-S3.5, 1997). Audibility was determined in 1/3-octave bands using the band importance function for the Speech-In-Noise test. A desensitization factor suggested by Pavlovic *et al*. (1986) was used to incorporate supra-threshold deficits associated with sensorineural hearing loss.

Results

For both the participants with impaired and normal hearing (Fig. 2), the results for the four listening conditions differ, i.e., the SII is not a good predictor of the effect the NR algorithms have on speech recognition.

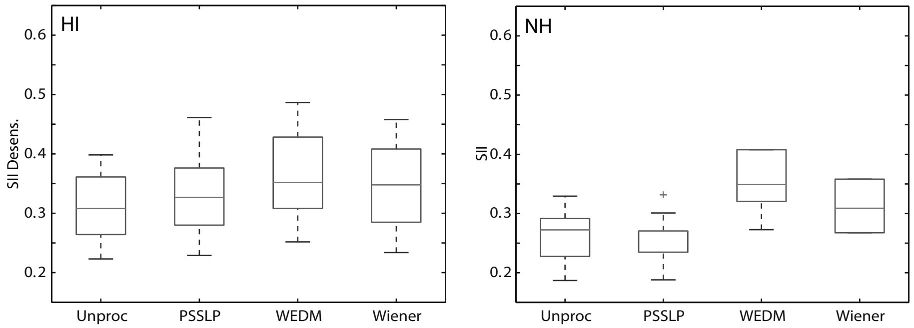

Fig. 2: SII for HI (left) and NH listeners (right). Each box shows inter-quartile values and the median is represented by the line in the box. Outliers (+) are defined as values outside 1.5 times the box length, and the whiskers extend to the highest and lowest values when the outliers are excluded.

2. EXTENDED SII, ESII

Method

Rhebergen and Versfeld (2005) have presented an extension to the SII with the purpose of predicting speech recognition in fluctuating noise. ESII is determined using the long-time average speech spectrum, but the short-time (9-35 ms) noise spectrum. The method takes forward masking into account. The short-time SII values are averaged to give one ESII value. This extension to the SII has shown promising results for fluctuating speech noise, interrupted noise and multi-talker babble noise (Rhebergen and Versfeld, 2005).

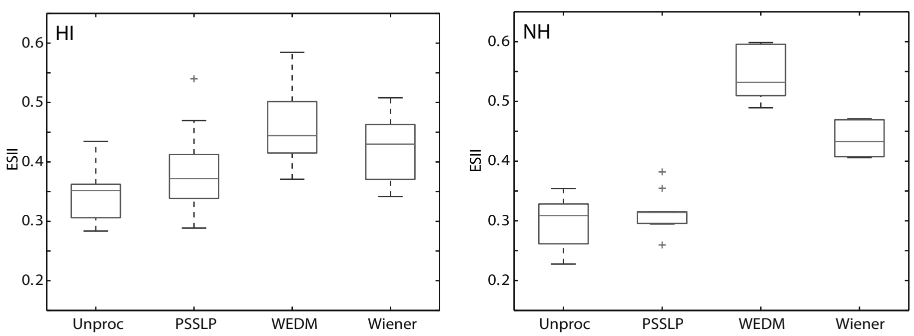

Fig. 3: ESII results for HI (left panel) and NH listeners (right panel). Explanation in Fig. 2.

Results

For both the participants with impaired and normal hearing (Fig. 3), the results for the four listening conditions differ, i.e., the ESII is not a good predictor of the effect the NR algorithms have on speech recognition.

3. SHORT-TIME SII, STSII

Method

As part of the current study, another short-time SII version was implemented. STSII calculates the SII using short-time (25 ms) speech and noise spectra. It uses the Pavlovic *et al*. (1986) desensitization factor, but it does not take forward masking into account. The final result is an average of the short-time SII values.

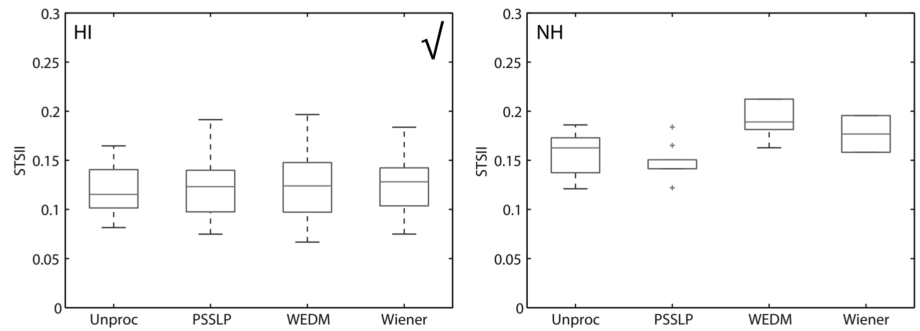

Fig. 4: STSII results for HI (left) and NH listeners (right). Explanation in Fig. 2. The tick mark indicates that the STSII seems to predict the performance of the listeners with impaired hearing well.

Results

For the participants with impaired hearing (Fig. 4, left panel), the results for the four conditions do *not* differ. However, for the participants with normal hearing (Fig. 4, right panel), the results for the listening conditions differ.

4. THREE-LEVEL COHERENCE SII, CSII

Method

Kates and Arehart (2005) have also presented an extension to the standard SII calculations. Short-time speech segments are divided into three level regions based on their RMS values. A signal-to-distortion ratio (SDR) is calculated from the coherence between the clean speech and the processed noisy speech. The SDR is calculated for each critical band and the CSII is calculated for each level region separately. The final measure is a weighted sum of the contributions from the three level regions. The weights presented by Kates and Arehart were used. The method has shown promising results for noisy speech subjected to peak-clipping and center-clipping distortion (Kates and Arehart, 2005).

Results

For the participants with impaired hearing (Fig. 5, left panel), the results for the four conditions differ, but the difference is close to non-significant (Friedman, p=0.044). For the participants with normal hearing (Fig. 5, right panel), the results for the listening conditions do *not* differ.

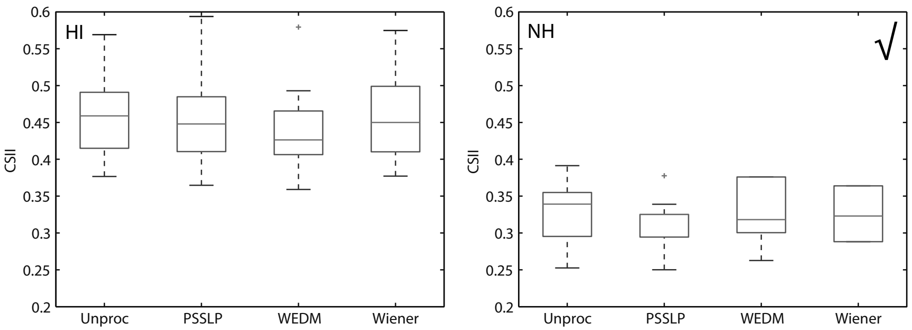

Fig. 5: CSII results for HI (left) and NH listeners (right). Explanation in Fig. 2. The tick mark indicates that the CSII seems to predict the performance of the listeners with normal hearing well.

5. SHORT-TIME OBJECTIVE INTELLIGIBILITY MEASURE, STOI

Method

Taal *et al*. (2010) have developed a measure of the amount of correlation between band envelope magnitudes of clean speech and processed noisy speech. A short-time (13 ms) spectral analysis is made. For each frequency band, the linear correlation coefficient is calculated within running overlapping time segments of about 400 ms, after scaling and clipping. The correlation coefficients are then averaged across time and frequency bands. The method has shown good agreement with speech recognition results obtained for normal-hearing listeners tested with noisy speech processed using noise suppression in the form of an ideal binary mask (Taal *et al*., 2010).

The original STOI calculation does not take hearing loss into account and the overall signal amplitudes do not influence the result at all. To apply this method for HI participants, two modifications were made: (1) The short-time spectra were adjusted to represent the sound-field pressure values actually presented to the listeners. (2) The absolute hearing threshold was simulated by adding an internal masking noise floor with a spectrum corresponding to the individual hearing thresholds. These equivalent noise spectrum levels were taken from the SII standard.

Results

For the participants with impaired hearing (Fig. 6, left panel), the results for the four conditions do *not* differ. For the participants with normal hearing (Fig. 6, right panel), the results for the listening conditions differ.

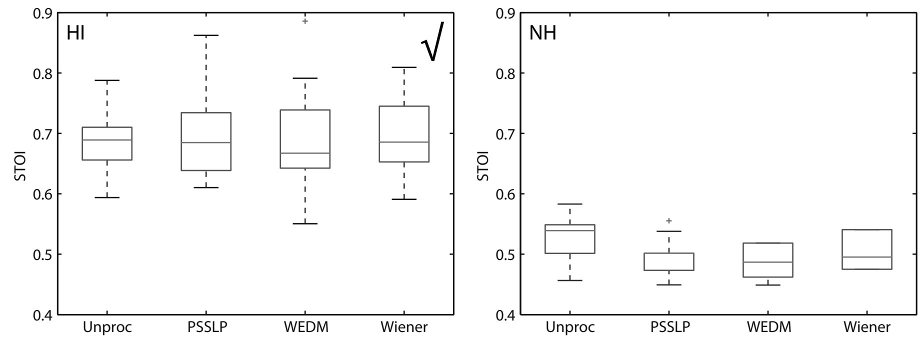

Fig. 6: STOI results for HI (left) and NH listeners (right). Explanation in Fig. 2. The tick mark indicates that the STOI seems to predict the performance of the listeners with impaired hearing well.

DISCUSSION

None of the theoretical measures was able to predict the speech test results for both groups of listeners. The standard SII and the ESII methods could not predict the results for any of the groups. The STSII measure seemed to predict the results for the listeners with impaired hearing, but not the listeners with normal hearing.

The WEDM method increased the long-time SNR the most, which lead to high calculated results for predictive measures that use long-time average speech and noise spectra, as the standard SII. However, the short-time SNRs, which the ear and brain are processing, are un-changed with this type of NR processing. This mismatch is seen in the results as an over-estimation of the WEDM results (Figs. 2 and 3). The ESII just takes the temporal characteristics of the noise, and not the speech, into account, which does not seem to be enough.

WEDM was also the processing scheme that produced the largest amount of distortion, which affected speech test results negatively. This effect could not be picked up by the simpler SII-measures. Measures that are based on correlation between the clean speech signal and the processed noisy speech, such as the CSII and the STOI, were able to predict the speech test results to some extent. Surprisingly, the STOI measure, which was not intended for use with listeners with impaired hearing, gave better results for this group than for the normal-hearing listeners, when using our suggested modifications.

CONCLUSIONS

None of the theoretical measures was able to predict the speech test results for both groups of listeners. Short-time analysis of the SNR and methods based on correlation of the clean speech and the processed noisy speech seems most promising.

REFERENCES

ANSI-S3.5 (**1997**). *American national standard methods for the calculation of the speech intelligibility index*. New York: Am National Standards Institute.

Byrne, D., and Dillon, H. (**1986**). "The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid" Ear Hear., **7**, 257-265.

Hagerman, B. (**1982**). "Sentences for testing speech intelligibility in noise" Scand. Audiol., **11**, 79-87.

Hagerman, B., and Olofsson, Å. (**2004**). "A method to measure the effect of noise reduction algorithms using simultaneous speech and noise" Acta Acustica, united with Acustica, **90**, 356-361.

Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (**2010**). "Development and analysis of an International Speech Test Signal (ISTS)" Int. J. Audiol., **49**, 891-903.

Kates, J. M., and Arehart, K. H. (**2005**). "Coherence and the speech intelligibility index" J. Acoust. Soc. Am., **117**, 2224-2237.

Loizou, P.C. (**2007**). *Speech enhancement theory and practice*. Boca Raton, FL, USA: Taylor & Francis Group.

Luts H., Eneman K., Wouters J., Schulte M., Vormann M., Büchler M., *et al*. (**2010**). "Multicenter evaluation of signal enhancement algorithms for hearing aids" J. Acoust. Soc. Am., **127**, 1491-1505.

Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). "An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals" J. Acoust. Soc. Am., **80**, 50-57.

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

Smeds, K., Bergman, N., Hertzman, S. and Nyman, T. (**2009**). "Noise reduction in modern hearing aids – Long-term average gain measurements using speech" in Proceedings of *The International Symposium on Auditory and Audiological Research (ISAAR). Binaural processing and spatial hearing.* Helsingør, Denmark. Edited by J. M. Buchholz, T. Dau, J. C. Dalsgaard and T. Poulsen, pp. 445-452.

Smeds, K., Wolters, F., Nilsson, A., Båsjö, S., Hertzman, S. and Leijon A. (**2010**). "Objective measures to quantify the perceptual effects of noise reduction in hearing aids" in Proceedings of *The AES 38th International Conference - Sound quality evaluation*. Audio Engineering Society, Piteå, Sweden. Edited by J. Vanderkooy, pp. 101-107.

Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J. (**2010**). "A short-time objective intelligibility measure for time-frequency weighted noisy speech" in Proceedings of *ICASSP 2010,* pp. 4214-4217.

On the relationship between multi-channel envelope and temporal fine structure

Peter L. Søndergaard[1], Rémi Decorsière[1] and Torsten Dau[1]

[1]*Centre for Applied Hearing Research, Technical University of Denmark, DK-2800 Lyngby Denmark*

The envelope of a signal is broadly defined as the slow changes in time of the signal, where as the temporal fine structure (TFS) are the fast changes in time, i.e. the carrier wave(s) of the signal. The focus of this paper is on envelope and TFS in multi-channel systems. We discuss the difference between a linear and a non-linear model of information-extraction from the envelope, and show that using a non-linear method for information-extraction, it is possible to obtain almost all information about the originating signal. This is shown mathematically and numerically for different kinds of systems providing an increasingly better approximation to the auditory system. A corollary from these results is that it is not possible to generate a test signal containing contradictory information in its multi-channel envelope and TFS.

INTRODUCTION

The *envelope* of a signal is broadly defined as the slow changes in time of the signal, whereas the temporal fine structure (TFS) are the fast changes in time, i.e. the carrier wave of the signal. A typical method for splitting a signal into envelope and temporal fine structure is by the use of the Hilbert transform, as first proposed in Gabor (1946). In the cochlea, it is generally assumed that the action of the inner hair cells performs an envelope extraction process for high frequencies. For low frequencies, they instead extract the temporal fine structure.

The Hilbert transform method works well if the signal is narrow-band or a chirp. In this case, there is no doubt as to which part of the signal should be regarded as part of the envelope and which part should be regarded as part of the TFS. For complex signals however, the splitting of a signal into a single envelope and a single TFS is not a good model. Consider for instance the superposition of two pure tones with well separated center frequencies: in this case the Hilbert transform method will return a modulated envelope and a TFS with a center frequency being the average of the center frequencies of the two tones. This splitting does not fit our perception of such a tone.

The most common method to analyze complex sounds is to split them into sub-bands using a filter bank with band-pass filters, and then find the narrow-band envelope and TFS for each sub-band channel. This is what is commonly done in most auditory models. If enough overlapping filters are used, this leads to the classic definition of the

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.