STI direct test method based on wavelet transform to extract envelope

Wenxin Wang¹, Ranran Zhu² and Hongliang Zhao¹*
¹Electrical and Automation Engineering, Shandong University of science and technology, Qingdao, Shandong, 266510, China
²Electronic information engineering, Shandong University of science and technology, Qingdao, Shandong, 266510, China
*Corresponding author’s e-mail: zhl6401@126.com

Abstract. Speech transmission index (STI for short) is an important index to evaluate the quality of speech transmission of the room, it can better reflect the degree of voice signal affected by room reverberation and noise in the transmission process. This paper presents an algorithm for directly measuring STI index, white noise is filtered by Paul Kellet filter to generate pink noise, the signal envelope is extracted by wavelet transform, which improves the extraction accuracy of signal envelope and makes the measurement of STI index more accurate.

1. Introduction
In 1947, Fletcher first proposed intelligibility index (AI), then speech intelligibility index (SII) and speech transmission index (STI) [1,2,3]. In more than 100 years, it developed a complete theory to measure speech intelligibility based on intelligibility index, of which the most important index is speech transmission index (STI). Because it can better reflect the influence of reverberation time and signal-to-noise ratio, and has certain anti system distortion ability, it is recommended as an objective evaluation parameter of speech intelligibility by IEC standard. The speech transmission index is defined as the degree to which the intensity envelope of the speech signal is affected by the transmission channel. It is obtained by the acoustic separation of the speech signal and is an objective observation based on the weighted contribution of multiple bands within the speech spectrum.

STI measurement methods are divided into direct method and indirect method. Now most of the measurement of STI is indirect method, but the indirect method has a lot of limitations, such as only for linear time-invariant system, and the parameters of measurement system need to be adjusted according to experience. At present, most of the software on the market such as ODEON, DIRAC, also use indirect method to measure STI, which is different from the direct method. Whether the signal is corrected by frequency response, whether the spectrum is correct, and whether the sound pressure level is correctly calibrated will have a greater impact on the results when indirect measurement is used to measure the synthetic signal of operating language spectral level. Therefore, indirect measurement requires a higher quality of the measuring personnel[4]. With the further development of the theory of language clarity, there are also methods to build indoor sound field models and predict STI index by simulating the acoustic propagation characteristics in the room with software. Most of these methods also use indirect testing methods: Sun Rui et al. proposed a method to analyze the acoustic signal characteristics from the time-frequency joint domain, extract relevant objective parameters according to the human perception characteristics of sound, and construct a sound quality
evaluation model[5]. However, Zhu Peisheng et al. pointed out that there are many factors that affect the results of sound quality simulation, and the acoustics effect between simulation model and real space is also complex, especially for rooms without actual reference calibration, it is difficult to obtain accurate STI prediction results[6].

Wavelet transform is a new method of transformation analysis. It is a local transformation in time and frequency domain, so it can effectively extract information from the signal. Through the operation functions such as scaling and translation, multi-scale thinning analysis of functions or signals is performed, which solves many difficult problems that Fourier transform cannot solve. Wavelet transform is now applied to various signal analysis fields. Barik Tarakeswar et al performed time-domain and time-frequency wavelet analysis of the collected thrust and torque signals[7]. Finally, they tried to extend the force and torque decomposition wavelet to the force-torque hybrid wavelet with the best characteristics of the regression model to improve the predictability of drilling quality. Khalid et al proposed a wavelet transform (WT) based non-HS OFDM technique (NHS-OFDM) for the MIMO-VLC system has been presented using non-imaging and imaging receivers to study the performance of visible light communication (VLC) systems numerically[8].

For STI calculation, signal envelope extraction is the most important step in STI data processing. Compared with the traditional Fourier transform, wavelet transform is an effective improvement on the basis of Fourier transform. This paper describes the STI direct test method systematically and in detail, and the envelope is extracted by wavelet transform in the demodulation process, so as to improve the accuracy of the envelope, and the envelope before and after improvement is compared and analyzed.

2. **STI direct measurement method**

The direct measurement method is divided into the following steps: generating test signal, signal transmission, extracting signal envelope and STI calculation. The specific measurement process is shown in Figure 1:

![Flow chart of direct measurement of STI index.](image)

**2.1. STI test signal generation**

The test signal is the sinusoidal modulation pink noise signal sent by the signal source. Pink noise can be generated by white noise, and the slope of each octave band is -3dB. At present, the generation methods of pink noise mainly include Paul Kellet's weighted sum filter method, rob ERT Bristow Johnson's zero pole filter method, Voss algorithm and transfer function approximation method, among which Paul Kellet method has the best performance. Therefore, in this paper, Paul Kellet filter is used to generate pink noise. The transfer function of the filter is shown in formula (1):
The pink noise generated by this filter has the advantages of good performance and small fitting error. Fig. 2 and Fig. 3 are the time domain and frequency domain images of pink noise generated by Paul Kellet filter:

\[
G(s) = \frac{0.05s^2 - 0.063s + 0.014}{s^2 - 1.8s + 0.803}
\]  

(1)

Figure 2. Time domain diagram of pink noise generated by Paul Kellet filter.

Figure 3. Frequency domain diagram of pink noise generated by Paul Kellet filter.
Figure 4. Wavelet transform and Hilbert transform envelope
Extraction of 500Hz octave band modulated 1Hz sinusoidal signal

Figure 5. Wavelet transform and Hilbert transform envelope

The generated pink noise signal is passed through seven octave band filters with center frequencies of 125Hz, 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz and 8000Hz to generate seven octave band signals corresponding to the spectrum of the language. Each octave band modulates 14 sinusoidal signals with frequencies from 0.63hz to 12.5hz respectively, corresponding to different degrees of signal fluctuation. Therefore, the complete STI test has a total of 98 signal points.

2.2. Extracting signal envelope by wavelet transform
Envelope can reflect the important information of acoustic signal. The slow fluctuation of envelope intensity corresponds to the pronunciation of words and sentences, while the fast fluctuation corresponds to a single phoneme in words. Usually, the envelope extraction uses Hilbert transform to construct analytical signal to obtain the absolute value. However, if the original waveform is asymmetric and fluctuates violently, the extracted envelope waveform is mostly accompanied by uneven burrs and can not wrap the original waveform well. Wavelet transform is to convolute the wavelet based with each window of the signal to obtain the wavelet coefficient. The wavelet coefficient contains important information of the signal. Generally, the noise coefficient is small, but the signal coefficient is large. Select an appropriate threshold to eliminate the noise, and the reconstructed signal amplitude is the signal envelope. Based on this principle, wavelet transform is often used in the field of signal denoising. In order to embody the advantages of wavelet transform
denoising, figure 4 and figure 5 show the envelope of the signal after Hilbert transform and wavelet transform with MATLAB.

Comparing the envelope extracted by hiltet transform with that extracted by wavelet transform, it will be found that the envelope extracted by wavelet transform has a higher fit with the original signal. Compared with Hilbert transform, it can better retain the original signal information, have better denoising effect and smoother waveform. The 6th order Symlets wavelet base is selected for transformation in here. Symlets wavelet has approximate symmetry and biorthogonal, so it can reduce the phase distortion during signal analysis and reconstruction to a certain extent, improve the signal-to-noise ratio, and is suitable for envelope extraction.

Set the test signal according transmission to:

\[ f(t) = I[1 + m \cos 2\pi f_m(t + \tau)] \]  

Where \( I \) is the output strength and \( f_m \) is the modulation frequency. Let \( \int_{-\infty}^{+\infty} \phi(t)dt = 0 \) be the 6-order Symlets wavelet generating function, and \( \int_{-\infty}^{+\infty} \varphi(t)dt = 0 \). Then the wavelet transform of function \( f(t) \in L^2(R) \) is:

\[
W_j(a,b) = (a)^{-1/2} \int_{-\infty}^{+\infty} f(t) \varphi \left( \frac{t-b}{a} \right) dt
\]  

Where, \( a \) is the amplitude factor reflecting the frequency domain characteristics, and \( b \) is the translation factor reflecting the time domain characteristics. The input signal is decomposed by four layers through wavelet transform to obtain high-frequency coefficients, and the noise standard deviation \( \sigma \) is estimated from the high-frequency coefficients of the first layer, that is, take the absolute value of the first level wavelet coefficient, then take the median value and divide it by the adjustment coefficient. The noise standard deviation \( \sigma \) is calculated as shown in formula (4):

\[
\sigma = \text{middle}(\tilde{W}_{1,k}) / 0.6745, 0 \leq k \leq 2^{-j-1} - 1
\]  

\( \tilde{W}_{1,k} \) represents the first level wavelet coefficient, and \( j \) is the order of the selected wavelet. The key to signal denoising lies in the selection of threshold and threshold function. Threshold calculation is the key to wavelet denoising and envelope extraction. Here, Birge -Massart penalty algorithm is used for threshold selection, and the penalty coefficient is 2. The threshold is obtained from the following steps:

\[
\lambda_m = |c(n_m)|, 1 \leq m \leq 4
\]  

\[ |c(m)| \] is a decreasing sequence composed of the absolute value of the m-th wavelet coefficient, where \( n_m = D/(k + 1 - m)\alpha \). the value range of \( D \) generally is \( S(1) \leq D \leq 2S(1) \). \( S(1) \) is the wavelet coefficient length after the first layer decomposition, where 2 is taken. Next, the wavelet coefficients are compared by threshold, and the soft threshold processing function is used:

\[
0, |w| < \lambda_m, 1 \leq m \leq 4
\]

\[
|w| \geq \lambda_m, 1 \leq m \leq 4
\]

Where \( w \) is the wavelet transform value of \( f(t) \). The last step is to extract the envelope of the denoised and reconstructed signal, and set the reconstructed signal as \( f'(t) \). \( F(t) \) is the convolution signal of \( f'(t) \), the signal envelope \( W \) after wavelet transformation is defined as formula (8):

\[
F(t) = f'(t) * \frac{1}{\pi} = \frac{1}{\pi} \int_{-\infty}^{+\infty} f'(\tau) d\tau
\]

\[
W = \{\text{Re}^2[F(t)] + \text{Im}^2[F(t)]\}^{1/2}
\]

2.3. STI calculation

STI is calculated based on modulation transfer function (MTF). The modulation transfer function of the transmission path can be determined by many methods. The main method is to quantify by
comparing the ratio of the modulation depth at the output and input of the test signal. The modulation depth of each octave band is obtained by associating the envelope strength with sinusoidal and cosine signals of a specific time length and a specific modulation frequency. Firstly, the modulation depth of the received signal $mdr_{k,f_m}$ output by each octave band is derived:

$$mdr_{k,f_m} = 2 \sqrt{\sum \left[ I_k(t) \cdot \sin(2\pi f_m t) \right]^2 + \sum \left[ I_k(t) \cdot \cos(2\pi f_m t) \right]^2} \sum I_k(t)$$

(9)

Here, $k$ is the octave band; $f_m$ is the modulation frequency; $I_k(t)$ is the envelope strength, it is the square of the envelope. After the modulation depth of the transmitted signal $mdt_{k,f_m}$ is obtained by the same method, the modulation transmission ratio can be calculated. All derived modulation transmission ratios $m(k,f_m)$ form the so-called MTF matrix. The modulation transmission ratio is:

$$m_{k,f_m} = \frac{mdr_{k,f_m}}{mdt_{k,f_m}}$$

(10)

In order to more accurately analyze the accuracy difference between envelope extraction based on wavelet transform and envelope extraction based on Hilbert transform, a section of pink noise modulated sinusoidal signal is simulated with MATLAB as the original signal, and the envelope is demodulated by wavelet transform and Hilbert transform respectively. The modulation depth of the input signal is obtained from the sinusoidal signal of the modulated original signal, then the modulation transmission ratio under different octave bands is calculated from the envelope strength. By comparing the modulation transmission ratio on each octave band, we can see the fit degree between the envelope obtained by wavelet transform and Hilbert transform and the original signal. Fig. 6 shows the modulation transmission ratio of wavelet transform and Hilbert transform in 7 octave bands:

![Figure 6. Modulation transmission ratio of wavelet transform and Hilbert transform in 7 octave bands](image)

It can be seen that the modulation transmission ratio obtained by wavelet transform is higher than that obtained by Hilbert transform, so the error between the envelope obtained by wavelet transform and the original signal is smaller and the fit degree is higher.

The next step is to calculate the effective signal-to-noise ratio $SNR_{eff}$. The most important step in converting the modulation transfer function into the language transmission index is to convert the modulation transfer matrix into the effective signal-to-noise ratio, that is, the effective signal-to-noise ratio is obtained by inverse transformation of the modulation transmission ratio affected by background noise and reverberation time:
The value of effective signal-to-noise ratio is between -15dB and +15dB, and a matrix about effective signal-to-noise ratio is obtained. Then, the effective signal-to-noise ratio is converted into the transmission index TI. The calculation method of each octave band and modulation frequency is as shown in formula (12):

$$SNR_{\text{eff, } k, f_n} = 10 \times \log \frac{m_{k, f_n}}{1 - m_{k, f_n}}$$

(11)

$TI_{k, f_n} = \frac{SNR_{\text{eff, } k, f_n} + 15}{30}$

(12)

Next, the derived transmission index matrix is averaged on the modulation frequency to obtain the modulation transmission index (MTI) of each octave band K. The use method is as shown in formula (13):

$$MTI_{k} = \frac{1}{n} \sum_{n=1}^{n} TI_{k, f_n}$$

(13)

The STI index of this room can be finally obtained by using the weighted average method for the modulation transmission index on the seven octave bands:

$$STI = \sum_{k=1}^{7} \alpha_k \times MTI_{k} - \sum_{k=1}^{6} \beta_k \times (MTI_{k} \times MTI_{k+1})^{1/2}$$

(14)

$\alpha_k$ is the weight coefficient of octave band $f_k$; $\beta_k$ is the redundancy factor between octave band $k$ and octave band $k + 1$. Table 1 shows the relationship between $\alpha_k$, $\beta_k$ and octave band $k$:

| $k$ | 1   | 2   | 3   | 4   | 5   | 6   | 7   |
|-----|-----|-----|-----|-----|-----|-----|-----|
| $\alpha_k$ | 0.085 | 0.127 | 0.230 | 0.233 | 0.309 | 0.224 | 0.173 |
| $\beta_k$   | 0.085 | 0.078 | 0.065 | 0.011 | 0.047 | 0.095 | -   |

3. Conclusion
This paper systematically summarizes the flow of STI direct test method specified in IEC 60268-16, and introduces in detail the calculation formula involved in the direct method. The direct method is more accurate in reflecting the language clarity. The fitting error of the pink noise generated by Paul Kellet filter is smaller and the performance is better; the envelope extracted by wavelet transform is smoother than that extracted by Hilbert transform, has higher fit with the original signal, and has a good denoising effect on the original signal. The application of wavelet transform can significantly improve the extracted envelope signal.

References
[1] French, N.R., Steinberg, J.C. (1947) Factors Governing the Intelligibility of Speech Sounds. Acoustical Society of America Journal, 19(1):90-119.
[2] Kryter, K.D. (1962) Proposed methods for calculation index. Acoustical Society of America Journal, 34(11):1689-1697.
[3] International Electrotechnical Commission. (2011) Sound system equipment—part 16: objective rating of speech intelligibility by speech transmission index, IEC 60268-16
[4] Zhu, P.S., Mo, F.S. (2014) Experimental comparison of direct and indirect measurement methods of language transmission index, 42(11):1759-1764.
[5] Sun, R., Zuo, Y.Y., Wu, C.G. (2021) Extracting objective parameters from time-frequency joint domain to establish sound quality prediction model. Noise and vibration control, 41(02):70-76.
[6] Zhu, P.S., Zhu, G.F., Zhu, T. (2014) ODEON comparison between simulation and actual measurement of language transmission index. Applied Acoustics, 33(05):419-425.

[7] Tarakeswar, B., Kamal, P. (2021) Prediction of drilled hole quality in bidirectional woven carbon fiber reinforced plastic using wavelet packets of force–torque signals. Journal of Reinforced Plastics and Composites, 40(21-22): 800-826.

[8] Khalid, A. (2021) Wavelet transform based non-hermitian symmetry OFDM technique for indoor MIMO-VLC system with an imaging receiver. Wireless Networks, 27(7): 4649-4663.