Investigation of Improved Masking Noise for the Speech Privacy

Allen CHO[1]; Issa PANAHI[2]

The University of Texas at Dallas, USA

ABSTRACT

Many sound masking products for the purpose of achieving speech privacy have emerged on the market in recent years. Over the time, quality of the masking noise has been improved a lot and still is studied for a better method. This paper introduces a new way of generating masking noise which outperforms the commonly used methods. Using Adaptive Linear Predictive Coding (LPC) we can derive the model for various target sound and create a new masking signal which has exactly the same spectral envelope of the target sound and different phase. Two important aspects in evaluating the performance quality of the masking noise are masking capability and pleasantness. By testing these criteria in objective and subjective ways, we show superior performance of the proposed adaptive method in comparison with several existing techniques.

Keywords: Sound Masking, Speech Privacy, Psychoacoustics, Adaptive LPC

1. INTRODUCTION

As the importance of information has grown considerably due to the progress of technology, it has also become important to keep those knowledge confidential. Therefore, many ways to achieve privacy has been investigated in different approaches. But most of them is not appropriate for the places where is in need of privacy immediately and cost effective at the same time, for they need a few requirements which will take more time and money than given. In other cases, it is significant for those who work in highly sensitive condition to stay focused on their tasks. Sound masking is an up to date breakthrough to solve all these issues above. By adding unstructured sound to where the privacy is needed, we can cover the structured human speech which we want to mask and make it unintelligible. We can settle problems above once we eliminate the comprehensible human speech which will distract people from concentration or unwanted leakage of information. There has been long years of studies done for sound masking and already used in practical situations in several places such as hospitals, schools and companies. Using combined knowledge of psychoacoustic in human auditory perception and experiment conducted with normal hearing people, we came up with an improved way of producing masking signal effectively. Furthermore by creating this masking signal adaptively on a regular basis in time period, we can choose the masker to provide the best privacy in every condition.

2. PSYCHOACOUSTICS

2.1 Auditory Masking

Phenomenon occurring when there are two sounds existing to human at the same time and one is affected by the other sound(1). It can be classified into two terms. Auditory masking in the frequency domain is known as frequency masking or simultaneous masking while it is called temporal masking in the time domain. We are using frequency masking to mask since the sound masking system will be constantly operated more than temporarily. We also need to understand the characteristic of human auditory perception to maximize the efficiency of masking effect. Knowing that human speech contains intelligible content concentrated in frequency range between 1,000Hz and 3,000Hz. Revise the same frequency component of the masking signal would help concluding better result. As shown in (Fig. 1), lower frequency signal tends to show wider extent to be masked.

[1] choallen1223@gmail.com
[2] imp015000@utdallas.edu

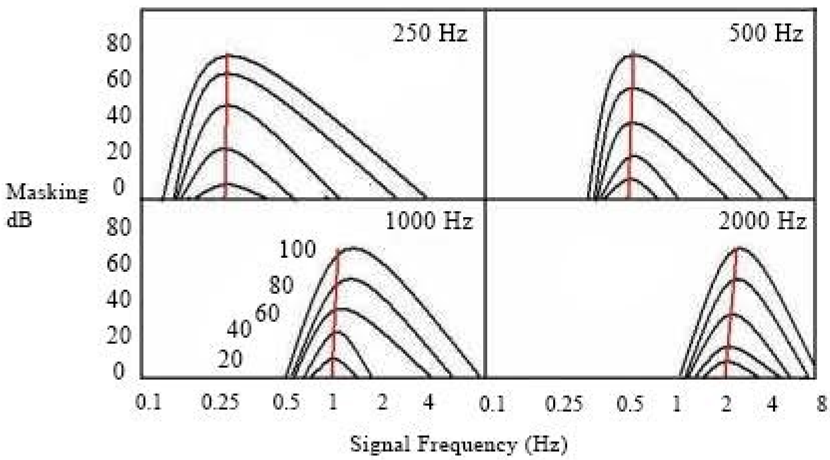


Fig. 1 Different masking patterns among various frequencies(2)

2.2 Weber-Fechner Law

Acknowledging Weber's law(3), certain sound can be more easily understood in silent environment than in the place with noise. In other words masking threshold to achieve desired masking level will change along with the time and environment. With changing frequency spectrum and power of the target sound, we can accomplish masking with the optimum signal which yields auditory masking while does not cause discomfort to the subjects.

2.3 Sensory Adaptation

Though we accomplish the unwanted speech unintelligible to the listeners, it is worse if the masking signal is more disturbing. To avoid the problems stated above, we need to refine the masking signal to make it sounds pleasant. Exposure to masking signal in a long time is not going to introduce another uneasiness to people based on the sensory adaptation(4). The sensitivity of human sensory receptors is changing relatively to the change of external stimuli. When the stimuli, masking signal, is applied continuously then people will adapted to the sound and neither detect the masking signal nor feel offense to it. Sudden change of stimuli in large amount could not trigger the sensory adaptation therefore we required to change the sound gradually. Sound masking system is equipped with the initial ramp-up function. This function will prevent the system from starting in loud volume suddenly so that the listener would not feel the system is on.

2.4 Zwicker's Model of Sensory Pleasantness

There is a way to measure how much comfort human accept quantitatively in hearing. Model of Sensory Pleasantness was first presented by Dr. Zwicker in 1990 from his book, Psychoacoustics(5). Equation of the model consisted of four different parameters each examining the property of sound.

Loudness

$$N = \int_0^{24\,Bark} N' dz \tag{1}$$

Sharpness

$$S = \frac{0.11 \int_0^{24\,Bark} N' g(z) z dz}{\int_0^{24\,Bark} N' dz} acum \tag{2}$$

Roughness

$$R = \frac{0.3 f_{mod}}{kHz} \int_0^{24\,Bark} \frac{\Delta L_E(z) dz}{dB/Bark} asper \tag{3}$$

Pleasantness

$$P = e^{-0.7R} e^{-1.08S} (1.24 - e^{-2.43T}) e^{-(0.023N)^2} \tag{4}$$

This model can be applied only onto stationary signals. Masking signal combined with Back Ground Music or other nonstationary masking signals can not be measured with the equation.

3. METHOD

3.1 Linear Predictive Coding

From the knowledge we have acquired in psychoacoustics, we can consider what factors we should focus on to create new masking signal. Research of sound masking up to this point has used white noise or pink noise for their masking noise. However in practical, those signals are inadequate to be used though it shows great level of masking quality. Our original goal was to achieve privacy for those who need it but when it comes to reality, comfort should be taken into consideration as well as with privacy since it is applied to human subjects. Even though it guarantees good privacy if it is hard to listen to long times, it is difficult to be used in real life. Our proposed method meets all the requirements stated above. Since its spectrum is almost as same as original target spectrum, masking efficiency will outdoes than previous colored noise such as white noise or pink noise. Discomfort in hearing can be also reduced since this method does not yield useless frequency which can cause disturbance to whom hears.

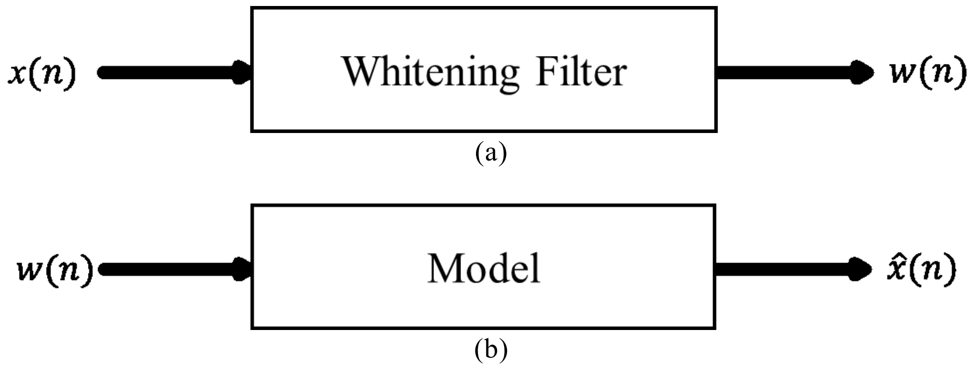


Fig. 2 Equation used for Linear Prediction Coding

Coefficients of a forward linear predictor(6) can be determined from the equation (5).

$$A(z) = \sum_{k=0}^{p} (a_k z^{-k}) \tag{5}$$

Whitening Filter $A(z)$ has order of p. From the LPC yields coefficients a_k of whitening filter which make the original signal $x(n)$ into white noise $w(n)$.

$$H(z) = \frac{1}{A(z)} \tag{6}$$

With the white noise that we can generate from the sound masking system, we can generate masking noise in real time by applying the inverse whitening filter Eq. (6). Strength of the outcome from this filter is that its power spectrum follows almost the same as the original signal $x(n)$. Furthermore, the phase has been modified during the filtering process so there remains no understandable speech component in it. Comparison of the power spectrum between two signals is shown in (Fig. 3).

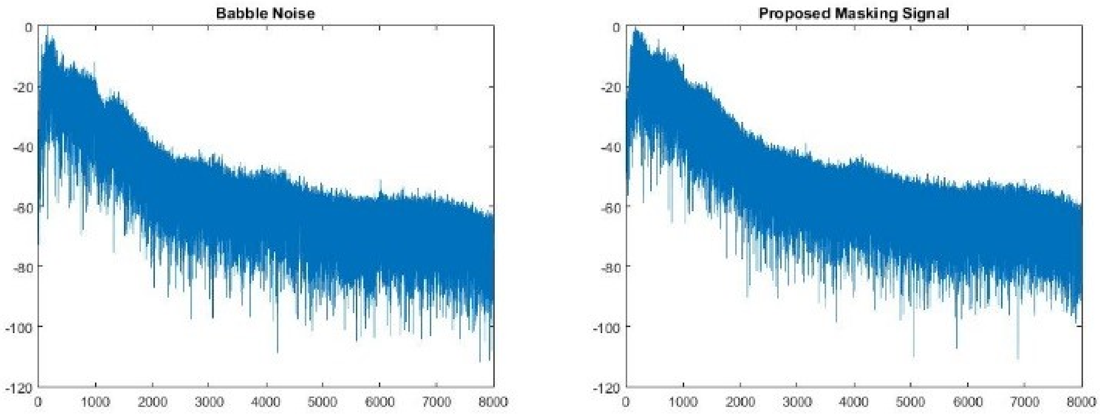


Fig. 3 Comparison between the original babble noise and proposed masking signal

3.2 A-Weighting Filter

Considering the process for human to recognize the speech, there exists limited frequency range which determines whether it is possible to comprehend or not. The critical frequency range differs from the property of speech. Vowel sound is delivered predominantly in the frequency of 250Hz to 500Hz and consonant is delivered in 2,000Hz to 4,000Hz. The sensitivity with regard to the frequency is represented on the equal-loudness contour(7) and Speech intelligibility index with regard to different frequency as shown in (Fig. 4) and (Fig. 5)(8). A-weighting filter(9) is made out of the data to enhance the specific frequencies which will have speech components. By adapting this filter to our proposed masking signal, we can emphasize the speech containing frequency range while suppressing other parts which will cause unwanted noise or disturbance to the listeners.

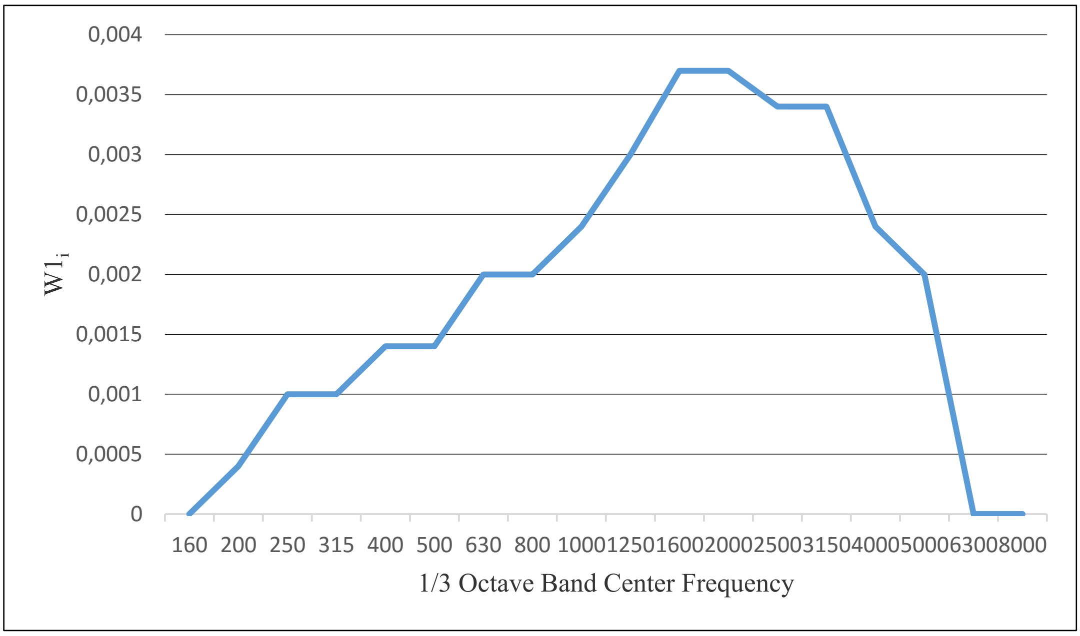


Fig. 4 Speech Intelligibility varies with frequency

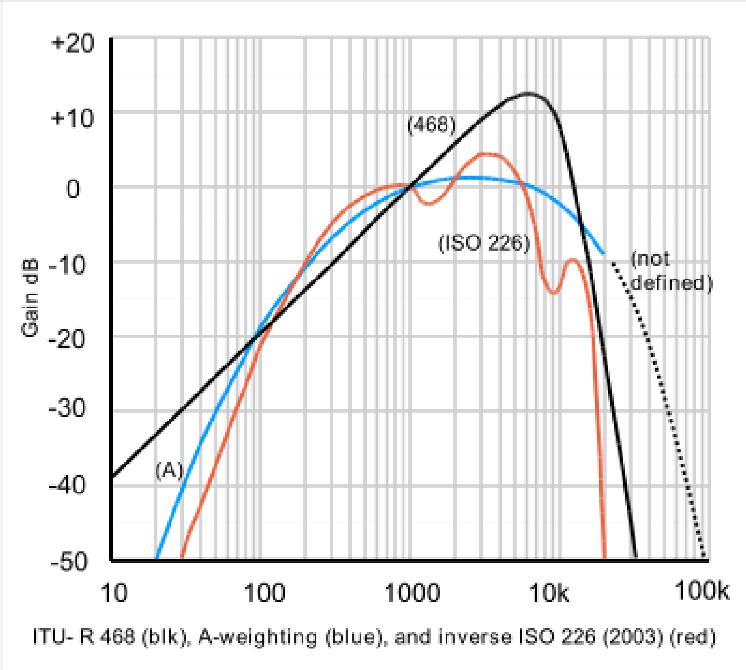


Fig. 5 Inverse of equal-loudness contour and A-weighting function

3.3 Moving Average Filter

The outcome of the original proposed masking noise has good masking proficiency. But the first product of the process includes little crackling sound in it which make listeners unpleasant. We can eliminate the problem with adding another filter called Moving Averaging Filter(6)(10).

$$Y(i) = \frac{1}{2N+1}\big(y(i+N) + y(i+N-1) + \cdots + y(i-N)\big) \qquad (7)$$

N : *Number of neighboring data points*

Basically this filter smooth the signal by replacing each data point with the average of its adjacent ones. Crackling sound is originated from the discontinuity of the signal and by having average with neighboring signals, we can remove disjoint in the continuous signal. From the reason that this filter works as low pass filter, large span may destroy intended frequency spectrum of masking noise. $N = 5$ is the ideal number to make the best of both pleasantness and masking capability.

3.4 Multiband Equalizer

Sensational preference differs from gender, age and other numerous characteristics. The proposed method could result in better response than the previous ones but it is impossible to make the best signal satisfying everyone who hears. By adapting equalizer onto the signal that we make, we can allow users to manipulate the sounds so that they can change with their own taste. In this process we also consider that human perception in hearing differs from each frequency. Dividing frequency into 20 different frequency not in linear scale but in logarithmic scale, non-uniformity among frequency can be adjusted. We use 1/3 octave bands to compose frequency bins. On each frequency bin we add gain function so we can alter the frequency range which we want to make change.

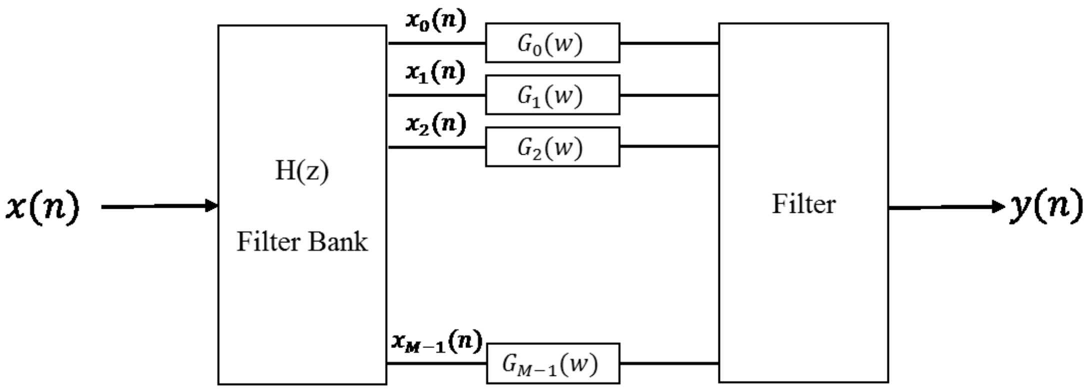


Fig. 6 Multiband filter equalizer

4. Test Setup

To verify the performance of proposed method, we conducted subjective test and objective test. We compare proposed method with white noise, pink noise and formerly used masking products. Signal A and Signal B in the results refers to state of the art respectively. Through speakers installed in the ceiling, masking noise will be played and other loudspeaker will play the human speech. Human speech can be understood in silence and subjects will score whether the masking noise is pleasant and speech under masking noise is intelligible. Scoring procedure is following Mean Opinion Score (MOS) measurement. Subjective test is done for 11 normal hearing people. To show the result in quantitative way we also use objective measuring process in two criteria, intelligibility and pleasantness. Coherence and the Speech Intelligibility Index (CSII) and Zwicker's sensory pleasantness model is used in the tests.

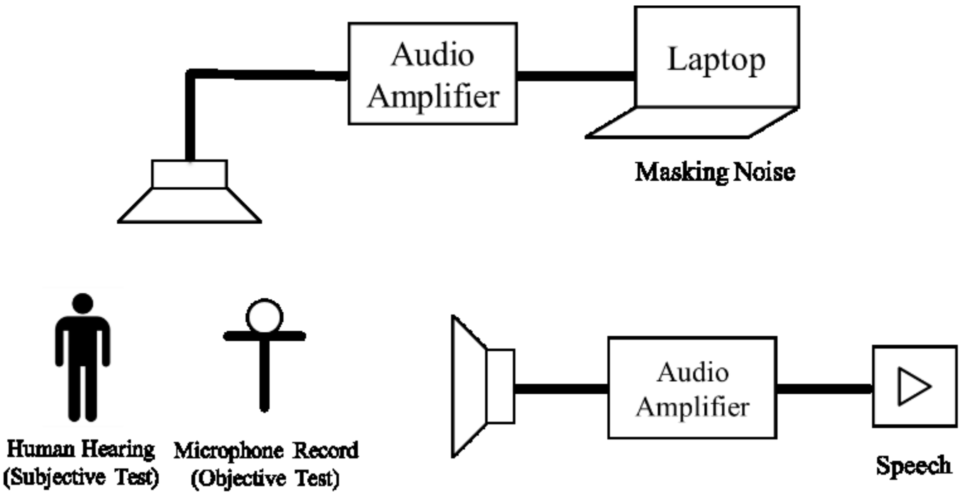
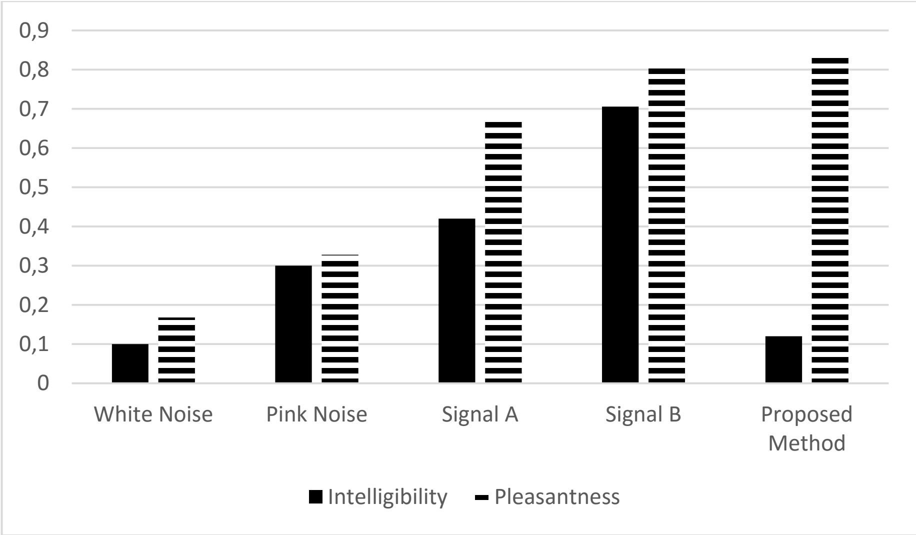


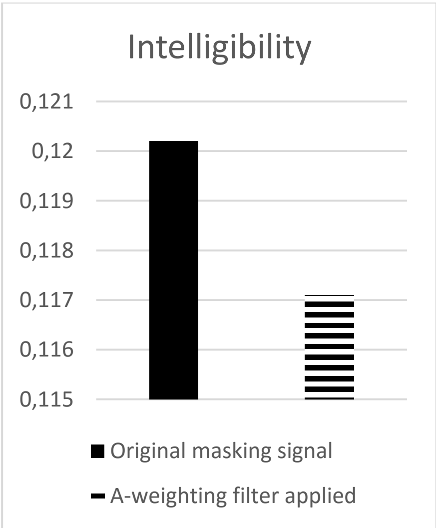
Fig. 7 Test set up for subjective test and objective test

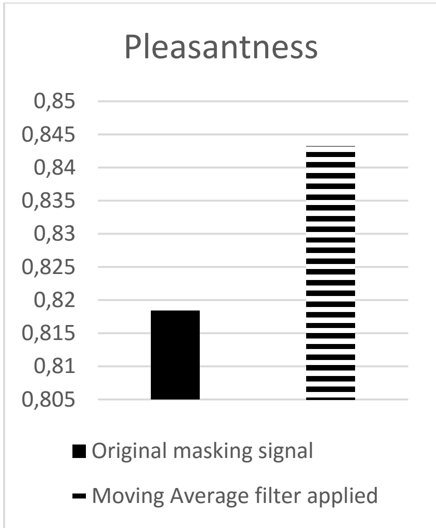
4.1 Objective Test Result

Since the purpose of masking is to lower the intelligibility of unwanted speech from outside, lower intelligibility indicates higher performance. In the opposite the sound should be feel comfortable for the subject, pleasantness should be higher.



(a)





(b) (c)

Fig. 8 Objective result for A-weighting filter and Moving Average Filter processed signals

4.2 Subjective Test Result

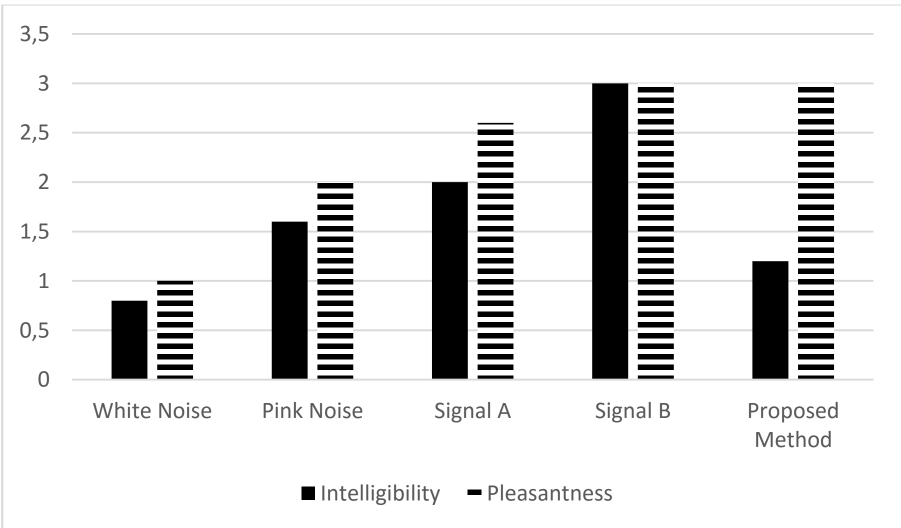
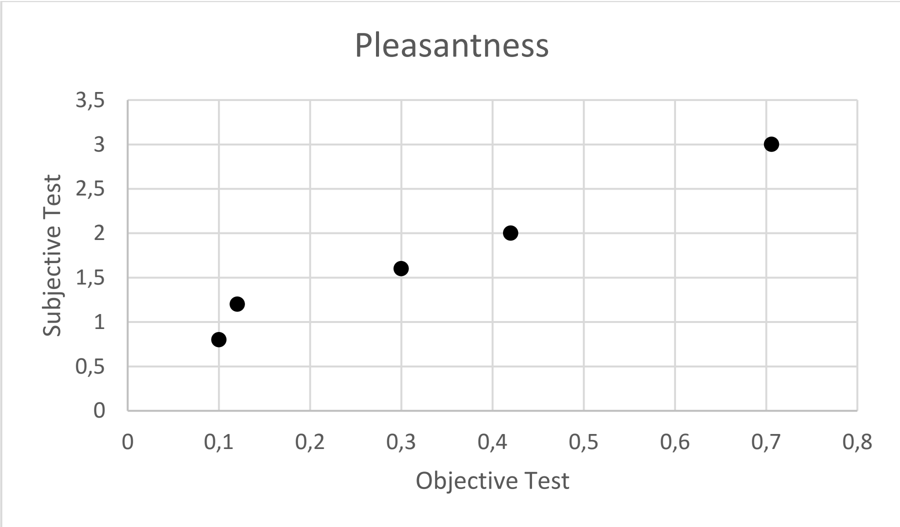
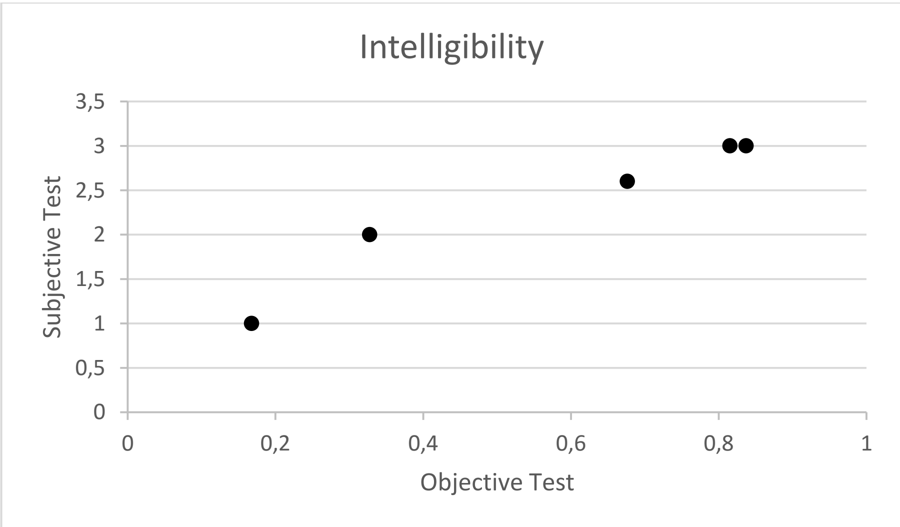


Fig. 9 Subjective Result



(a)



(b)

Fig. 10 Relation between the subjective and objective test results

5. Conclusion

A new algorithm generating signal to be used for masking unwanted speech and other noises was proposed in the paper which would contribute to achieve better privacy and less distraction. By adding two more filters after the process, we can promote better masking capability and comfort in hearing. The result compared to the previous methods shows proposed method is more helpful in practical usage.

REFERENCES

1. Moore, Brian C. J. An Introduction to the Psychology of Hearing. London: Academic, 1982.
2. Richard H. Ehmer, "Masking Patterns of Tones," The Journal of the acoustical society of America Vol 31. 1959
3. Stout, G. F. "The Weber-Fechner Law." A Manual of Psychology.: 199-209.
4. Webster, Michael. "Evolving Concepts of Sensory Adaptation." F1000 Bio Rep F1000 Biology Reports 4 (2012).
5. E. Zwicker, H. Fastl, “Psychoacoustics: Facts and Models,” Springer-Verlag Berlin Heidelberg, 3rd ed., 2007, pp. 220–246.
6. J.G. Proakis, D.G. Manolakis, Digital signal processing: principles, algorithms, and applications, Prentice Hall, Upper Saddle River, NJ, 1996.
7. Robinson, D. W., and R. S. Dadson. "A Re-determination of the Equal-loudness Relations for Pure Tones." Br. J. Appl. Phys. British Journal of Applied Physics 7.5 (1956): 166-81.
8. Richard L., St. Pierre, Jr. "The Impact of A-weighting Sound Pressure Level Measurements during the Evaluation of Noise Exposure." NOISE-CON, 2004
9. B. Bauer and E. Torick, "Researches in loudness measurement," in IEEE Transactions on Audio and Electroacoustics, vol. 14, no. 3, pp. 141-151, Sep 1966. doi: 10.1109/TAU.1966.1161864
10. Ahmad, M., and D. Sundararajan. "A Fast Algorithm for Two Dimensional Median Filtering." IEEE Trans. Circuits Syst. IEEE Transactions on Circuits and Systems 34.11 (1987): 1364-374. Web.