IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

An approach to Intelligibility Improvement and Noise Reduction in Speech Processing Applications

TULLURI SATEESH , ECE, CMR Technical Campus, Hyderabad

D. SREEKANTH, ECE, CMR Technical Campus, Hyderabad

ABSTRACT

The perceptual effects of noise reduction differ between hearing aids. The results agree well with those of normal-hearing listeners in a previous study. None of the noise-reduction algorithms improved speech intelligibility, but all reduced the annoyance of noise. The noise reduction that scored best with respect to noise annoyance and preference had the worst intelligibility scores. The trade-off between intelligibility and listening comfort shows that preference measurements might be useful in addition to intelligibility measurements in the selection of noise reduction. The speech intelligibility of indoor public address systems is degraded by reverberation and background noise. This paper proposes a preprocessing method that combines speech enhancement and inverse filtering to improve the speech intelligibility in such environments. An energy redistribution speech enhancement method was modified for use in reverberation conditions, and an auditory-model-based fast inverse filter was designed to achieve better dereverberation performance. An experiment was performed in various noisy, reverberant environments, and the test results verified the stability and effectiveness of the proposed method.

1. INTRODUCTION

The evaluation of noise-reduction algorithms has recently shifted from a focus on SI toward a focus on the objective assessment of cognitive measures, particularly listening effort. As yet there is no standard definition of listening effort. The British Society of Audiology recently proposed the following definition in a white paper: ‘‘the mental exertion required to attend to, and understand, an auditory message’’ (McGarrigle et al., 2014, p. 434). Currently, there is also no standard method of measuring listening effort. In fact, for several known methods, it is not known whether they in fact correspond to listening effort (McGarrigle et al., 2014). Listening effort has been indirectly measured using response times (RTs), for instance, in a dual-task paradigm where the secondary task is nonauditory (Desjardins & Doherty, 2014; Downs, 1982; Neher et al., 2014; Pals, Sarampalis, & Baskent, 2013; Sarampalis et al., 2009). Sarampalis et al. (2009), for example, used a dual-task paradigm to evaluate the effect of noise reduction and found significantly better recall of words and faster reaction times as a benefit of noise reduction at the lower SNRs. They hypothesized that noise-reduction algorithms could reduce the noise in a way that is comparable with the ability of the auditory and cognitive systems in the brain to ignore the noise (Sarampalis et al., 2009). Noise reduction could support this function of the brain, not by improving the SI, but rather by relieving the cognitive load, thereby resulting in a perceived improvement in listening comfort and a decrease in listening effort (Brons, Houben, & Dreschler, 2013; Huckvale & Frasi, 2010; Lunner et al., 2009; Marzinzik, 2000; Sarampalis et al., 2009). In contrast, noise-reduction algorithms might also introduce signal distortions, leading to a reduction in perceived listening comfort or listening effort (Lunner et al., 2009; Ng, Rudner, Lunner, Pedersen, & Ro¨ nnberg, 2013).

In a recent study, we compared noise reduction from different hearing aids to gain some insight in the effects of noise reduction (the black box) on the speech signal (Brons, Houben, & Dreschler, 2013). In short, acoustical analyses showed that noise-reduction implementations differ among hearing aids, and perceptual measurements showed that these differences are perceptually relevant for normal-hearing listeners. Noise-reduction implementations differed perceptually from each other in the degree to which they influenced the noise annoyance and speech naturalness perceived by normal-hearing listeners, resulting in differences in preference. Finally, small differences in speech intelligibility and listening effort were found among noise-reduction systems but not between noise reduction on and off. In this follow-up study, we investigated whether these findings also hold true for hearing-impaired listeners. It might be that hearing-impaired listeners are less sensitive to differences between processing conditions because of suprathreshold deficits such as reduced frequency selectivity or impaired modulation detection (Marzinzik, 2000). On the other hand, because of their decreased ability to understand speech in noise, it might be more important for hearing-impaired listeners to avoid distortions of the speech signal.

This paper proposes a new preprocessing method for improving speech intelligibility by a combination of the PDMSE method and the FIF method. The PDMSE method was modified for reverberant environments, and a new Gammatone (GT)-filter-based FIF method was designed to achieve better equalization and dereverberation performance. Compared with the A-EQ, W-EQ, and FIF equalization methods, the GT-filterbased FIF method can further decrease the distortion of the transmission channel. Compared with individual FIF and PDMSE methods, the improved combination method has better stability and higher speech quality. Furthermore, compared with the multizone and ASII methods, the combination method can significantly improve the speech intelligibility in different noisy and reverberant environments. To validate the method, an experiment was performed in real environments with various noise and reverberation conditions. The speech transmission index, spectrogram, log-spectral distortion measure, short-time objective intelligibility measure, and modified rhyme test were used to compare the performance. The objective and subjective evaluation results illustrate that the method can effectively improve the speech intelligibility of I-PA systems in noisy and reverberant environments.

1.1 PROBLEM STATEMENT

Based on this algorithm, warped domain equalization (W-EQ) method was proposed to improve the listening experience. This method uses the bark scale, which is related to auditory perception and low-frequency response equalization and produces a better listening experience than other equalization methods. However, the bark scale is not an auditory model and cannot simulate the frequency response characteristics of the basilar membrane in the cochlea. Moreover, these equalization methods do not account for the influence of background noise on speech intelligibility.

Increasing the playback level is one clear solution to improve the speech intelligibility in the event of background noise. However, it is impossible to increase the output level indefinitely due to the limited power output of loudspeakers and the pain-threshold pressure limitation of the ear. In addition, in the case of I-PA systems, the listener is located in a noisy environment, and the noise reaches the ears without any possibility of intercepting it beforehand. Therefore, a preprocessing speech enhancement method without increasing the output power would be more suitable for use with I-PA systems.

An energy redistribution voiced/unvoiced (ERVU) method was proposed to improve intelligibility without increasing the output power. The method redistributes more speech energy to the transient regions to reinforce speech signals. A perceptual distortion measure (PDM)-based speech enhancement (PDMSE) method was proposed based on the ERVU method and the PDM algorithm. Compared with the ERVU method, the PDMSE method can further improve speech quality without decreasing intelligibility. However, these methods do not consider the influence of reverberation on speech intelligibility.

1.2 OBJECTIVES

The goal of the current study was to answer the following research question: Does hearing aid noise reduction influence speech intelligibility, listening effort, noise annoyance, speech naturalness, and preference for listeners with a moderate sensor neural hearing loss, compared with (a) no noise reduction and (b) noise reduction from other linearly fitted hearing aids.

2. LITERATURE REVIEW

Maj van den Tillaart, have proposed Single-microphone noise reduction leads to subjective benefit, but not to objective improvements in speech intelligibility. We investigated whether response times (RTs) provide an objective measure of the benefit of noise reduction and whether the effect of noise reduction is reflected in rated listening effort. Twelve normal-hearing participants listened to digit triplets that were either unprocessed or processed with one of two noise-reduction algorithms: an ideal binary mask (IBM) and a more realistic minimum mean square error estimator (MMSE). For each of these three processing conditions, we measured (a) speech intelligibility, (b) RTs on two different tasks (identification of the last digit and arithmetic summation of the first and last digit), and (c) subjective listening effort ratings. All measurements were performed at four signal-to-noise ratios (SNRs): _5, 0, þ5, and þ1 dB. Speech intelligibility was high (>97% correct) for all conditions. A significant decrease in response time, relative to the unprocessed condition, was found for both IBM and MMSE for the arithmetic but not the identification task. Listening effort ratings were significantly lower for IBM than for MMSE and unprocessed speech in noise. We conclude that RT for an arithmetic task can provide an objective measure of the benefit of noise reduction.

Inge Brons, study evaluates the perceptual effects of single-microphone noise reduction in hearing aids. Twenty subjects with moderate sensor neural hearing loss listened to speech in babble noise processed via noise reduction from three different linearly fitted hearing aids. Subjects performed (a) speech-intelligibility tests, (b) listening-effort ratings, and (c) paired comparison ratings on noise annoyance, speech naturalness, and overall preference. The perceptual effects of noise reduction differ between hearing aids. The results agree well with those of normal-hearing listeners in a previous study. None of the noise-reduction algorithms improved speech intelligibility, but all reduced the annoyance of noise.

Nasir Saleem, Many forms of human communication exist; for instance, text and nonverbal based. Speech is, however, the most powerful and dexterous form for the humans. Speech signals enable humans to communicate and this usefulness of the speech signals has led to a variety of speech processing applications. Successful use of these applications is, however, significantly aggravated in presence of the background noise distortions. These noise signals overlap and mask the target speech signals. To deal with these overlapping background noise distortions, a speech enhancement algorithm at front end is crucial in order to make noisy speech intelligible and pleasant. Speech enhancement has become a very important research and engineering problem for the last couple of decades. In this paper, we present an all-inclusive survey on unsupervised single-channel speech enhancement (U-SCSE) algorithms.

Shiksha Pandita, Speech enhancement has become one of the most important tools of the modern generation and is widely used in various fields for various purposes. The past decade has seen dramatic progress in speech recognition technology, to the extent that systems and high-performance algorithms have become accessible. Speech enhancement depends on signal

processing. Speech enhancement techniques are widely used to enhance the quality and intelligibility of the speech signal in the noisy environment. Conventional noise reduction methods introduce more residual noise and speech distortion. The existing algorithms fail when there are abrupt changes in the noise level. To overcome the shortcomings of the conventional methods, improved noise tracking algorithm is proposed in this paper for speech enhancement. The noise signal is estimated for the existing and the proposed methods. Results are simulated using LabView. This report shows how to recognize and enhance the speech using filters in lab view.

Kalamani M,have proposed Speech enhancement techniques are widely used to enhance the quality and intelligibility of the speech signal in the noisy environment. Conventional noise reduction methods introduce more residual noise and speech distortion. The existing algorithms fail when there are abrupt changes in the noise level. To overcome the shortcomings of the conventional methods, improved noise tracking algorithm is proposed in this paper for speech enhancement. The noise signal is estimated for the existing and the proposed methods. Results are simulated using LabVIEW. From the evaluated results it is observed that the proposed noise tracking algorithm improve the average noise power from 90% to 97% of the original noise signal power for various noise conditions.

3. PROPOSED METHOD

The overall scheme of the proposed method is shown in Fig. 1. Initially, the input signal s (n) is captured, and a time-frequency (TF) decomposition and GT filter are applied to obtain the short-term clean speech frame sm,i. sm,i is then sent to the voice activity detection (VAD) module and to the preprocessing and synthesis module.

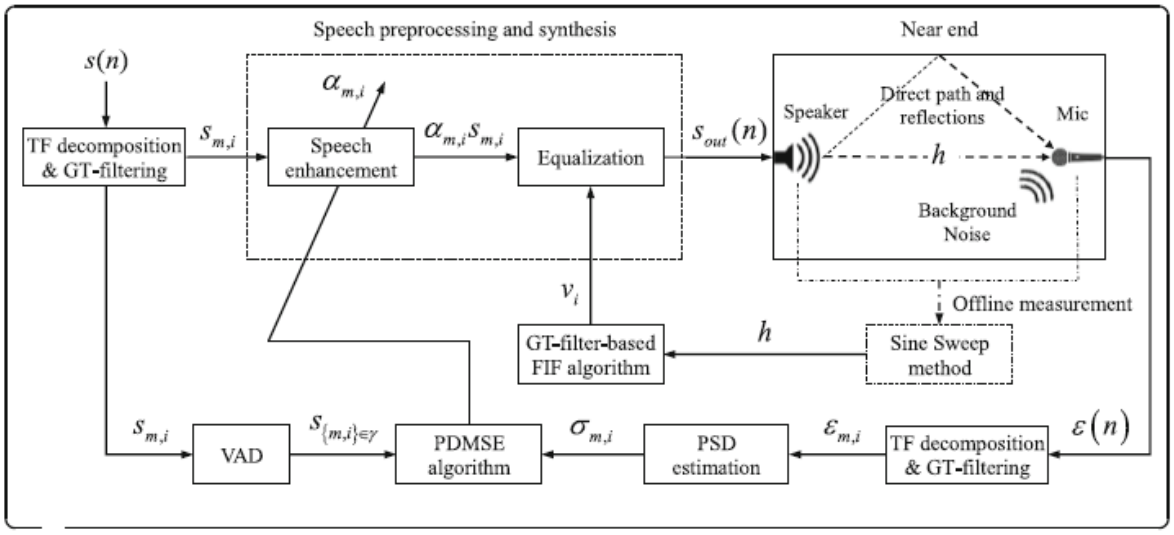

Fig. 1 Overall scheme of the proposed approach

The VAD module is applied to obtain the positions of the active voice in speech signals and prepare the detection information for the PDMSE algorithm. In the block for speech preprocessing and synthesis, a modified PDMSE method is used in the speech enhancement stage to increase the energy of transient speech. Next, a GT-filter-based FIF method is used in the equalization stage to pre-compensate the distortion of the transmission channel. The final preprocessing and synthesis signal sout (n) is used as an input for the loudspeaker to broadcast. The distortion signal ε (n) is then recorded by a microphone, and TF decomposition and GT filtering are once again performed to obtain the short-term distortion frame εm,i.

The power spectral density (PSD) estimation module is next applied to estimate the energy of background noise. Finally, the gain function α is calculated by the PDMSE algorithm, and the inverse sub-filters vi are obtained by the GT-filter-based FIF algorithm. Both parameters are used to adjust the preprocessing speech signal to obtain the best speech intelligibility. Furthermore, based on the method by Meng et al., a sine sweep signal with a length of 10 s is used as an excitation signal to obtain the RIR in advance to calculate the inverse filter.

3.1 Improved preprocessing speech enhancement

In the PDMSE algorithm, the PDM model plays an important role because it is more sensitive to transients than the spectral-only model. Furthermore, it can detect tiny differences between the input signal and the measured signal within a short time frame (20–40 ms). The PDM is a kind of TF decomposition method based on the spectro-temporal auditory model. The distortion measure D(s,ε) can be described simply by summing all the individual short-term distortion frames εm,i:

$$D(s,\varepsilon) = \sum_{m,i} d\left(s_{m,i}, \varepsilon_{m,i}\right),$$

3.2 Improved fast inverse filtering

The FIF method is used to achieve an “inverse filter” (an equalizer) of the RIR. Taking into account the sensitivity of the human ear to different frequencies, a FIF method based on GT filters was designed to achieve suitable dereverberation and equalization performance for human auditory characteristics. In contrast to the 1/ 3 octave and the bark scale, the GT filter is a kind of auditory filter that can simulate the characteristics of the basilar membrane. The central frequencies of the GT filter banks are distributed in a quasi-logarithmic form and are evenly distributed in the frequency range of the speech signal based on the equivalent rectangular bandwidth (ERB). The ERB is a measure used in psychoacoustics and approximates the bandwidths of the filters in human hearing. The GT filter banks can be represented as follows in the form of an impulse response in the time domain:

$$g(t) = ct^{n-1}e^{-2\pi bt}\cos(2\pi f_0 t + \phi), t > 0,$$

4. NOISE REDUCTION TECHNIQUES

The noise is classify into following category like, adaptive, additive, additive random, airport, background, car, Cross-Noise, exhibition hall, factory, multi-talker babble, musical, Natural, non-stationary babble, office, quantile-based, restaurant, street, suburban train, ambient, random, train-station, white Gaussian etc. Noise is mainly dividing into four categories: Additive noise, Interference, Reverberation and Echo. These four types of noise has led to the developments of four broad classes of acoustic signal processing techniques include, Noise reduction/Speech enhancement, Source separation, speech dereverberation and Echo cancellation/Suppression. The scope of this paper limited to noise reduction techniques only. Noise reduction techniques depending on the domain of analyses like Time, Frequency or Time- Frequency/Time-Scale.

4.1 Noise Reduction Algorithms

The Noise reduction methods are classified into four classes of algorithms: Spectral Subtractive, Subspace, Statistical-model based and Wiener-type. Some popular Noise reduction algorithms are, The log minimum mean square error logMMSE (Ephraim & Malah 1985), The traditional Wiener (Scalart & Filho 1996), The spectral subtraction based on reduced-delay convolution (Gustafsson 2001), The exception of the logMMSE-SPU (Cohen & Berdugo 2002), The logMMSE with speech-presence uncertainty (Cohen & Berdugo 2002), The multiband spectral-subtractive (Kamath & Loizou 2002), The generalized subspace approach (Hu & Loizou 2003), The perceptuallybased subspace approach (Jabloun & Champagne 2003), The Wiener filtering based on wavelet-thresholded multitaper spectra (Hu & Loizou 2004), Least-Mean-Square (LMS), Adaptive noise cancellation (ANC) [3], Normalized(N) LMS, Modified(M)- NLMS, Error nonlinearity (EN)-LMS, Normalized data nonlinearity (NDN)-LMS adaptation etc.

4.2 Fusion Techniques for Noise Reduction

4.2.1 The Fusion of Independent Component Analysis (ICA) and Wiener Filter

The fusion uses following steps: i. ICA is applied to a large ensemble of clean speech training frames to reveal their underlying statistically independent basis ii. The distribution of the ICA transformed data is also estimated in the training part. It is required for computing the covariance matrix of the ICA transformed speech data used in the Wiener filter iii. Then a Wiener filter is applied to estimate the clean speech from the received noisy speech iv. The Wiener filter minimizes the meansquare error between the estimated signal and the clean speech signal in ICA domain v. An inverse transformation from ICA domain back to time domain reconstructs the enhanced signal. vi. The evaluation is performed with respect to four objective quality measure criteria. The properties of the two techniques will yield higher noise suppression capability and lower distortion by combining them.

4.2.2 Recursive Least Squares (RLS) Algorithm: Fusion of DTW and HMM

Recursive Least Squares (RLS) algorithm is used to improve the presence of speech in a background noise. Fusion pattern recognition is used such as with Dynamic Time Warping (DTW) and Hidden Markov Model (HMM). There are a few types of fusion in speech recognition amongst them are HMM and Artificial Neural Network (ANN) and HMM and Bayesian Network (BN. The fusion technique can be used to fuse the pattern recognition outputs of DTW and HMM.

5 EXPERIMENTAL STEPS FOR IMPLEMENTING RLS ALGORITHM

- Recording speech, WAV file was recorded from different speakers
- RLS : The RLS was used in preprocessing for noise cancellation
- End point detecting: two basic parameters are used: Zero Crossing Rate (ZCR) and short time energy.
- Framing, Normalization, Filtering
- MFCC : Mel Frequency Cepstral Coefficient (MFCC) is chosen as the feature extraction method.
- Weighting signal, Time normalization, Vector Quantization (VQ) and labeling.
- Then HMM is used to calculate the reference patterns and DTW is used to normalize the training data with the reference patterns
- Fusion HMM and DTW**:**

o DTW measures the distance between recorded speech and a template.

o Distance of the signals is computed at each instant along the warping function.

o HMM trains cluster and iteratively moves between clusters based on their likelihoods given by the various models.

As a result, this algorithm performs almost perfect segmentation for recoded voice, recoding is done at noisy places, segmentation problem happens because in some cases the algorithm produces different values caused by background noise. This causes the cut off for silence to be raised as it may not be quite zero due to noise being interpreted as speech. On the other hand for clean speech both zero crossing rate and short term energy should be zero for silent regions.

6. EXPECTED RESULTS

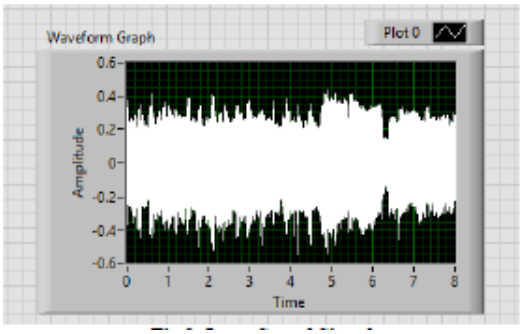

Fig.2. Input Sound Signal

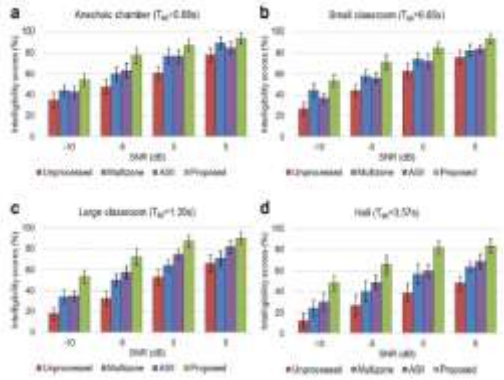

Fig. 3 Results of listening test under different RT and SNRconditions. a anechoic chamber (T60 = 0.08s), b small classroom (T60 = 0.65s), c large classroom (T60 = 1.39s), d hall T60 = 3.57s).

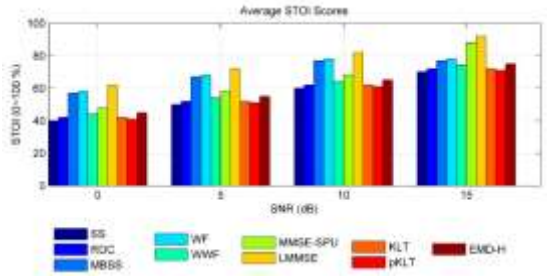

Fig.4. Average Speech Intelligibility prediction for U-SCSE algorithms in terms of STOI.

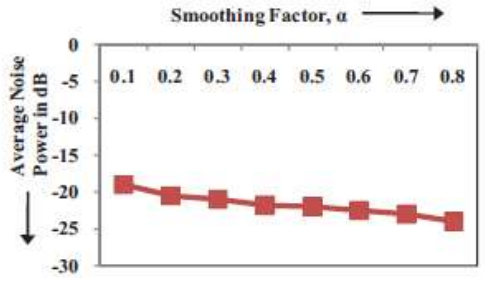

Figure 2. Estimated Average noise power by MCRA method for Smoothing Factor, α is varied from 0.1 to 0.8

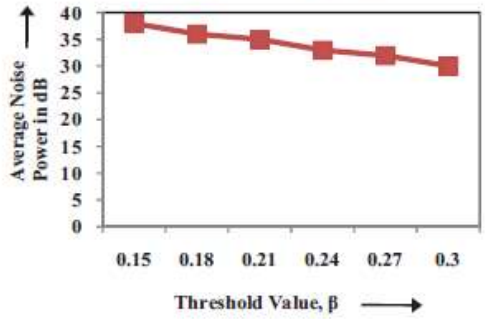

Figure 5. Estimated Average noise power by INT method for Threshold value, α is varied from 0.15 to 0.3 with optimal α is 0.1

CONCLUSION

The optimal filters can be designed either in the time or in a transform domain. The advantage of working in a transform space is that, if the transform is selected properly, the speech and noise signals may be better separated in that space, thereby enabling better filter estimation and noise reduction performance. The suppress noise from the speech signals without speech distortion it is an art of the noise removal approach. Noise reduction from three hearing aids tested was able to reduce the annoyance of babble noise perceived by listeners with moderate sensorineural hearing loss. The noise reduction that reduced noise annoyance the most and that was most preferred caused poorer intelligibility scores, confirming a trade-off between listening comfort and intelligibility.

REFERENCES

1. Maj van den Tillaart, The Influence of Noise Reduction on Speech Intelligibility, Response Times to Speech, and Perceived Listening Effort in Normal-Hearing Listeners, 2017.
2. Inge Brons, Effects of Noise Reduction on Speech Intelligibility, Perceived Listening Effort, and Personal Preference in Hearing-Impaired Listeners, 2014.
3. Huan-Yu Dong, Speech intelligibility improvement in noisy reverberant environments based on speech enhancement and inverse filtering, 2018.
4. Nasir Saleem, On Improvement of Speech Intelligibility and Quality: A Survey of Unsupervised Single Channel Speech Enhancement Algorithms,2019.
5. Kalamani M, Improved Noise Tracking Algorithms For Speech Enhancement Using LabVIEW, 2013.
6. Elliott, SJ, & Nelson, PA. (1989). Multiple-point equalization in a room using adaptive digital filters. Journal of the Audio Engineering Society, 37(11), 899–907.
7. Mourjopoulos, JN. (1994). Digital equalization of room acoustics. Journal of the Audio Engineering Society, 42(11), 884–900.
8. Tokuno, H, Kirkeby, O, Nelson, PA, et al. (1997). Inverse filter of sound reproduction systems using regularization. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 80(5), 809–820.
9. Kirkeby, O, Nelson, PA, Hamada, H, et al. (1998). Fast deconvolution of multichannel systems using regularization. IEEE Transactions on Speech and Audio Processing, 6(2), 189–194.
10. Kirkeby, O, & Nelson, PA. (1999). Digital filter design for inversion problems in sound reproduction. Journal of the Audio Engineering Society, 47(7/8), 583–595.
11. Radlovic, BD, & Kennedy, RA. (2000). Nonminimum-phase equalization and its subjective importance in room acoustics. IEEE Transactions on Speech and Audio Processing, 8(6), 728–737.
12. Cecchi, S, Romoli, L, Carini, A, et al. (2014). A multichannel and multiple position adaptive room response equalizer in warped domain: real-time implementation and performance evaluation. Applied Acoustics, 82, 28–37.
13. Mourjopoulos, J, Clarkson, P, Hammond, J (1982). A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals, IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP) (pp. 1858–1861).
14. Fuster, L, de Diego, M, Ferrer, M, et al. (2012). A biased multichannel adaptive algorithm for room equalization, In Proceedings of the 20th European Signal Processing Conference (EUSIPCO) (pp. 1344–1348).
15. B Sauert, P Vary, Improving speech intelligibility in noisy environments by near end listening enhancement. ITG-Fachbericht-Sprachkommunikation. (2006)