Proceedings of the Institute of Acoustics

SOUND-SPACE EQUALIZATION AND INTELLIGIBLE SPEECH REPRODUCTION IN A REVERBERANT SPACE

Yoshinori Takahashi, Waseda University, Tokyo, Japan
Hiroaki Nomura, Kure College of Technology, Hiroshima, Japan
Mikio Tohyama, Waseda University, Tokyo, Japan
University of York, York, UK

1 INTRODUCTION

Reproducing intelligible speech or rendering a virtual sound image in a reverberant space is an important spatial-audio technology[1]. The key issue is inverse filtering or magnitude and phase equalization of the transfer function (TF) in a reverberant space. Inverse filtering, however, is formidable and can be instable in a reverberant space, since the TF can often be of the non-minimum phase[2]. The locations of the TF-zeros can be closely located to the unit circle in the complex frequency domain and those are highly sensitive to the locations of source and receiving positions even in a local area. If the inverse filter is utilized for speech reproduction at a target point, then receiving signals could be instable at other points even closely located to the target point. This paper investigates stable inverse filtering in a local area for the space-variant TF based on cepstrum decomposition of poles and zeros.

Tohyama et al. proposed inverse filtering of the minimum-phase component for minimum-phase source waveform recovery[3] and virtual sound image rendering[4]. Inverse filtering of the minimum-phase component is in principle stable and works well in the time domain. However the inverse filters are still sensitive to the minimum-phase zeros close to the unit circle in the complex frequency domain[5]. Moreover a source signal waveform, such as a speech, is not always of the minimum-phase. Therefore, the recovery of an entire waveform from a reverberant speech cannot be expected by minimum-phase inverse filtering only.

Speech intelligibility is highly sensitive to the narrow-band temporal envelope of speech. Thus, the modulation transfer function (MTF) was introduced as a measure in room acoustics for assessing the effect of an enclosure on speech intelligibility[6]. A temporal envelope of speech can be characterized by the phase spectrum. Therefore inverse filtering for the all-pass component of the TF might be important for recovering intelligible speech in a reverberant space rather than minimum-phase inverse filtering. Radlovic and Kennedy[7] presented that the inverse filtering of the all-pass component can be realized by using a matched filter with a time-delay. This makes it possible to perform the inverse filtering of a reverberant speech, which has both the minimum-phase and all-pass components. However speech reproduction through inverse filtering might be still dangerous, since reproduced speech quality at the receiving positions can be deteriorated even in a local area around the target. This is because the zeros of the TF are space variant.

This paper describes smoothing effects of the TF on inverse filtering for space-variant TF in a local area. Instability due to the TF zeros could be dissolved by adding a small positive number into the magnitude response of the TF, when the TF magnitude becomes nearly zero[5]. However it is difficult to smooth the all-pass phase at the same time, if the TF is of non-minimum phase and the zeros are located close to the unit circle. The authors will investigate a method for getting a smoothed TF with the minimum-phase and all-pass components independently by relocating the poles and zeros using the cepstrum decomposition. The relocating effect on the inverse filtering in a local area can be estimated by narrow-band temporal envelope recovery of reproduced speech signals in a reverberant space[8]. Significance of the phase equalization for the envelope recovery will be also demonstrated rather than the magnitude equalization[9].

2 TRANSFER FUNCTION DECOMPOSITION AND RELOCATING THE POLES AND ZEROS FOR STABLE INVERSE FILTERLING

2.1 Transfer Function Decomposition and Inverse Filtering

The inverse filtering for the TF is possible after the TF is decomposed of the minimum-phase and all-pass components. Figure 1 is an example of the TF decomposition and Figure2 illustrates the inverse filters for the components. Here inverse filtering for the all-pass component is possible by matched filtering with a delay[7].

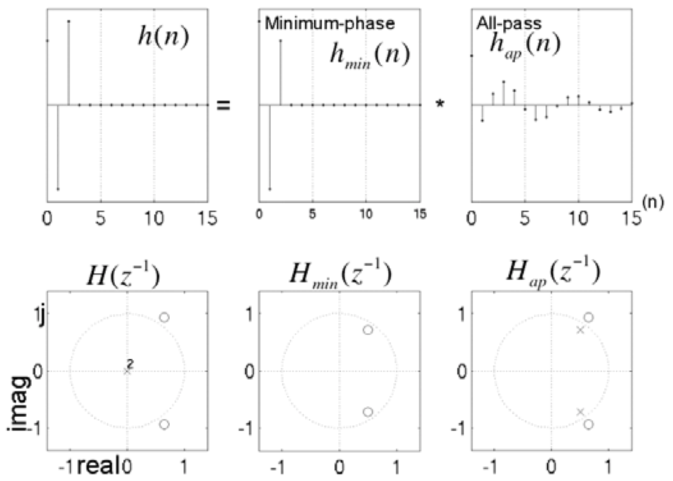

Figure 1. Decomposition of an impulse response into the minimum-phase and all-pass components.

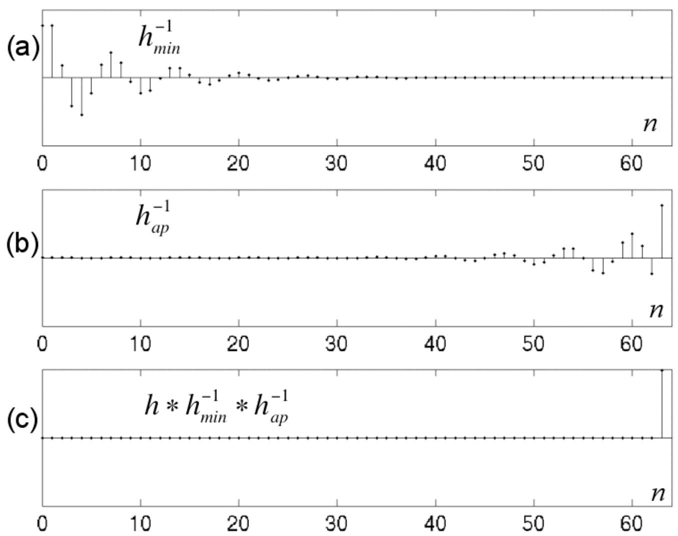

Figure 2. Impulse responses for the inverse filters ((a) Minimum phase component, (b) All-pass component), and output of de-convolution (c).

2.2 Relocating the Poles of the Minimum-phase Inverse

Inverse filtering is not always stable due to the TF zeros closely located to the unit circle. Figure 3 shows examples of the zero locations. If the minimum-phase (inverse) TF has the zeros (poles) as illustrated in Fig.3a(b), then the magnitude response of the (inverse) TF has the deep dips(steep peeks) as shown in Fig.4, although the product of the original and inverse TF becomes flat as a result. The dips of the original TF is highly sensitive to the source and receiving positions in a reverberant space, and thus the steep peeks in the magnitude response of the inverse are problematic for realizing stable inverse filtering.

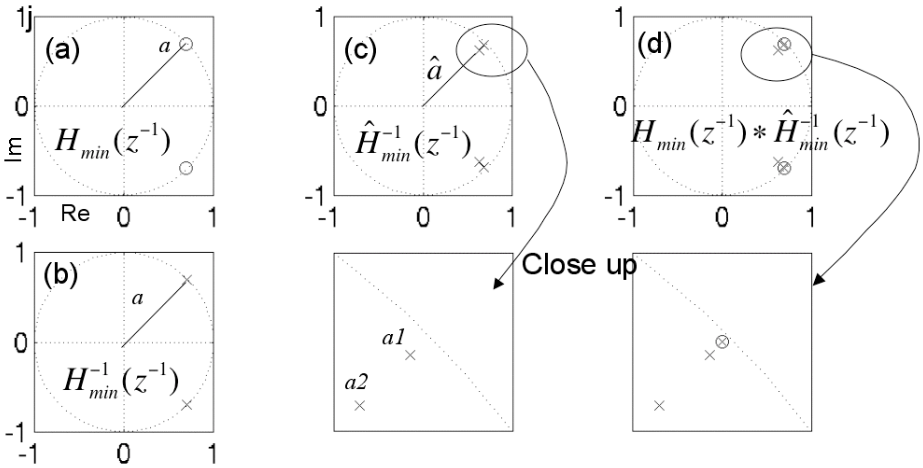

Figure 3. An example of zeros of minimum-phase transfer function (a), poles for the inverse filter (b), pole-relocated inverse filters (c), and poles and zeros for their product(d).

The steep peeks can be relaxed by relocating the poles of the inverse TF as shown in Fig.3(c). Examples of the magnitude response of the pole-relocated inverse are presented in Figure4 where steep peeks are loosened, although the notch-filter like characteristics are produced instead of a flat response after inverse filtering.

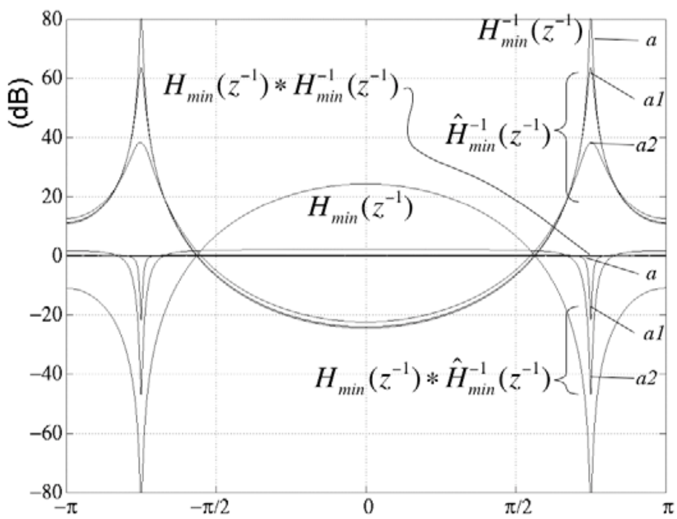

Figure 4. Magnitude frequency responses of Fig. 3(a-d).

2.3 Pole/Zero Relocating of the All-Pass TF

Figure 5(a) shows pairs of the poles and zeros for the all-pass component coupled with the minimum-phase shown in Fig.3(a). We can see abrupt phase changes in Fig.6 for the all-pass and its reversed phase, although those are cancelled each other. These rapid phase changes are also highly sensitive to the spatial locations of the source and receiving positions when the non-minimum phase zeros are located very close to the unit circle. The pairs of poles and zeros for the all-pass inverse can be relocated as well as the poles for the minimum-phase as shown in Fig.5(b). Figure 6 illustrates the pole/zero relocation effect on the all-pass phase inverse. We can see that abrupt phase change becomes slow by relocating the pairs of poles and zeros, although the sum of the original and the reversed phase is not cancelled.

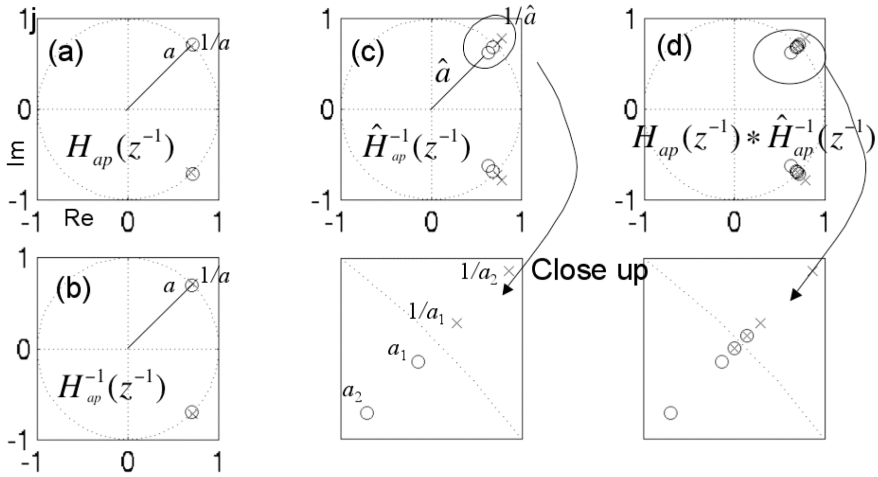

Figure 5. An example of poles and zeros of all-pass transfer function(a), poles and zeros for the inverse filter (b), for the pole/zero relocated inverse filters (c), and poles and zeros for their product(d).

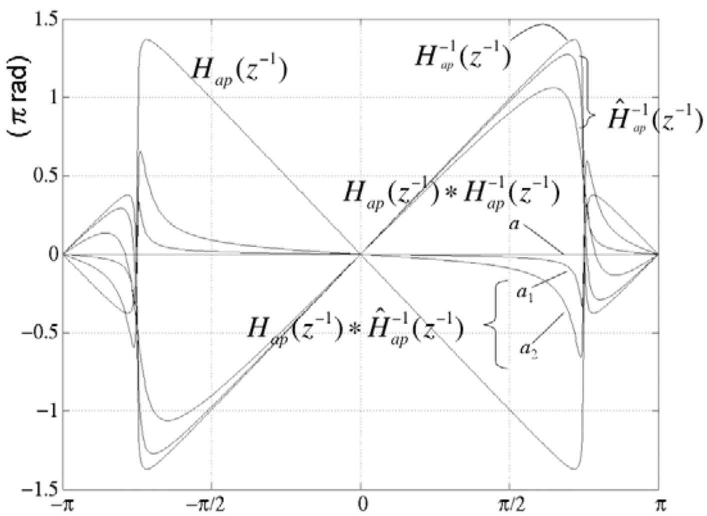

Figure 6. Phase frequency responses of Fig. 5(a-d).

2.4 Cepstrum Decomposition and Exponential Windowing

2.4.1 Exponential Windowing of the Minimum-phase Impulse Response

Relocating the poles of the minimum-phase inverse as described in 2.1 can be performed by the inverse filter for the exponentially-weighted minimum-phase impulse response. Exponential windowing moves the minimum-phase zeros farther from the unit circle.

2.4.2 Relocating the Poles and Zeros of the All-Pass TF

Let us take an example of all-pass impulse response record and its phase cepstrum as shown in Fig.7(a) and (b). Figure 7(c) is a newly extracted time-response from the causal part of the cepstrum. This time-response could be interpreted as the impulse response due to the poles of the all-pass. Figures 7(d) and (e) show the exponentially-windowed all-pass pole impulse response and its cepstrum. We can get the phase cepstrum as shown in Fig.7(f) by adding the non-causal part into the causal cepstrum given by Fig.7(e). The impulse response and its inverse can be constructed from the all-pass phase cepstrum shown in Fig. 7(f). The inverse corresponds to the inverse of pole/zero relocated all-pass. The processing scheme including both the minimum-phase and all-pass components is summarized in Fig. 8. Figure 9 is an example of inverse filtering after relocating the poles and zeros It is possible to relocated the poles and zeros of the minimum-phase and all-pass components independently.

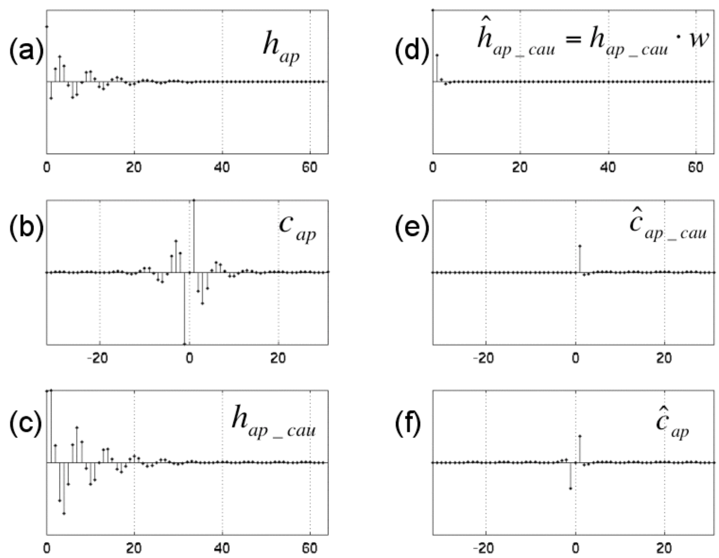

Figure 7. Pole/zero relocating of the all-pass TF using cepstrum decomposition of poles and zeros.

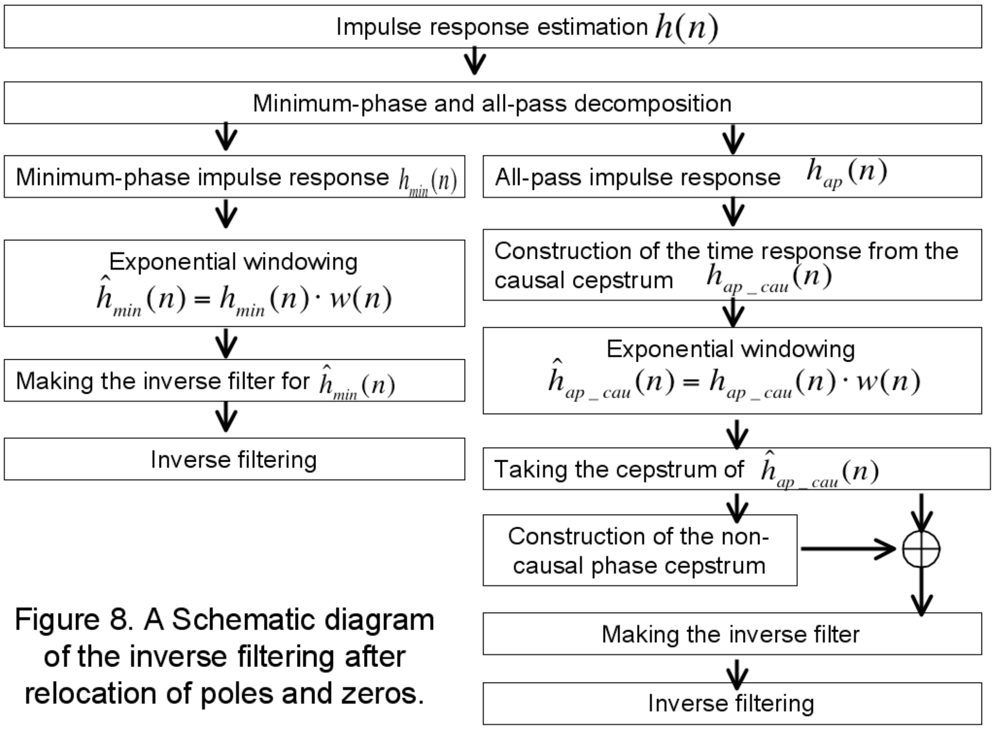

Figure 8. A Schematic diagram of the inverse filtering after relocation of poles and zeros.

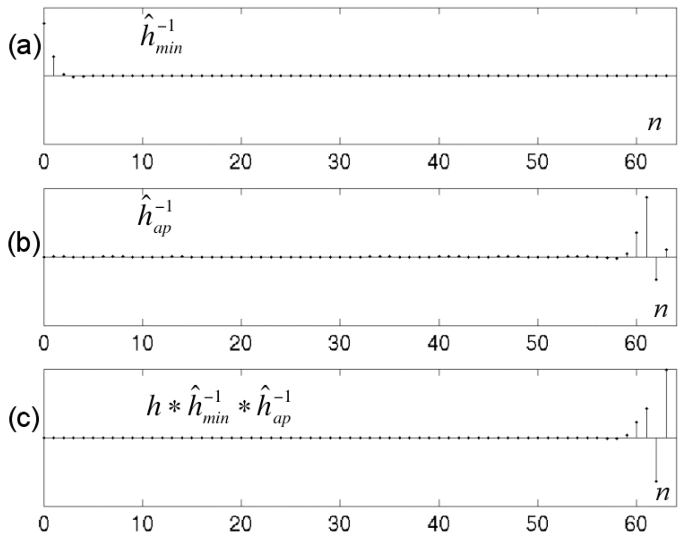

Figure 9. Impulse responses for the inverse filters after relocating poles and zeros((a) Minimum phase component, (b) All-pass phase component), and output of de-convolution (c).

3 EXAMPLES OF INVERSE FILTERING IN A REVERBERANT SPACE

Figure 10 shows a schematic of sound reproduction and equalization by inverse filtering in a reverberant space. Suppose that we have the inverse filter for the target position. Figures 11(a)-(d) illustrate samples of input and reverberant speech signals at the target and other positions in a local area, and similarly Figs.11(e)-(g) shows the receiving signals reproduced after the inverse filtering. We can see that almost perfect inverse filtering is possible at the target with a time delay; however, reproduced speech signals might be deteriorated at other positions by the target-oriented inverse filtering. This is because the zeros of the TF can be distributed closely to the unit circle and highly sensitive to the receiving positions.

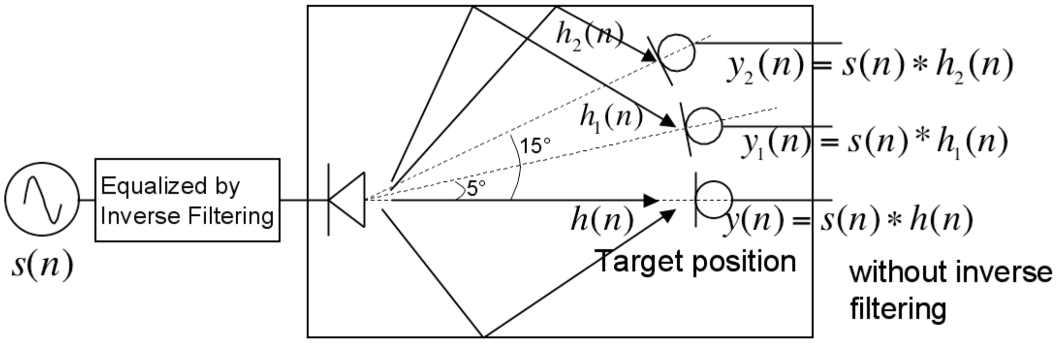

Figure 10. A schematic of the inverse filtering for sound reproduction.

Figures 12(a)-(c) present the inverse filtering results after relocating the poles and zeros of the TF at the target position following the procedure shown in Fig. 8. We can see the equalization at the target might be possible by the modest-inverse filtering which is not harmful at other positions. Effect of inverse filtering on reproduced speech could be estimated by temporal narrow-band (1/4 oct. band) envelopes[6]. Recovery of the narrow-band envelopes can be estimated by the correlation coefficients of the envelopes between input and reproduced signals with or without inverse filtering.

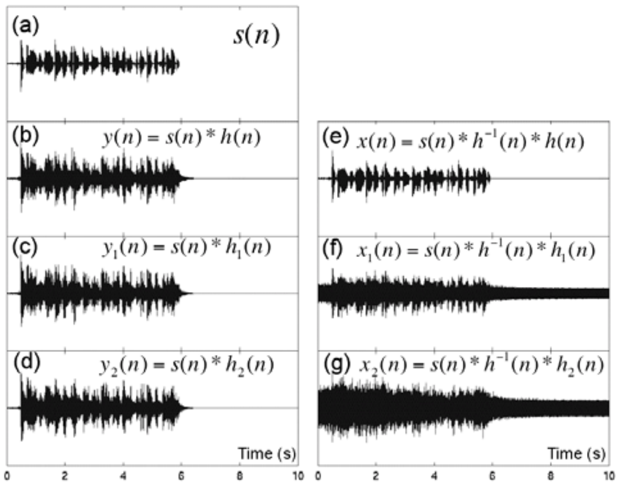

Figure 11. Examples for inverse filtering of *h(n)* for the target position
(a) Original dry speech (b-d) reverberant speech without the inverse filter
(e-g) reverberant speech with the inverse filter.

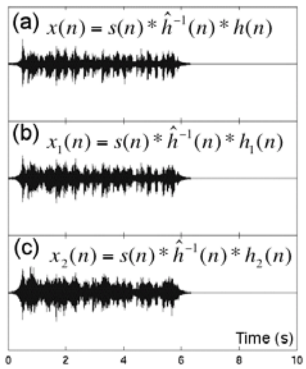

Figure 12. Inverse filtering results after relocating the poles and zeros of the target TF.

Figures 13(a)-(c) show the correlation coefficients between the input(dry) and reproduced signals every 1/4 oct.-band. The correlation coefficients without inverse filtering are plotted in Fig.13(a), and Fig.13(b) demonstrates the effect of the target inverse filtering on the correlation coefficients. Drastic changes can be seen at the receiving positions other than the target, although almost perfect inverse filtering can be performed at the target. We can find in Fig.13(c) that the correlations improve by the inverse filtering after poles and zeros relocations. Figure 14 demonstrates the magnitude and all-pass phase responses of the target position with and without relocating the poles and zeros.

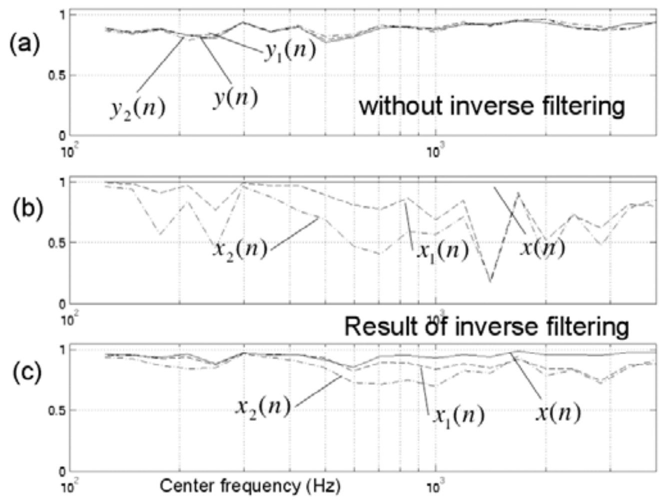

Figure13. Correlations of narrow-band temporal envelopes between the original and reverberant speech(a), reproduced speech through the target inverse filter (b), after relocation of the poles and zeros of the target TF(c).

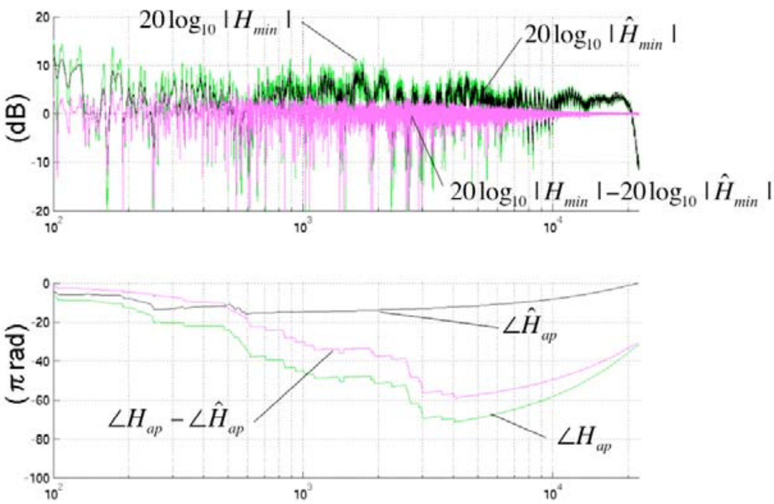

Figure 14. Magnitude and all-pass phase responses of the original and after relocating the poles and zeros.

4 MAGNITUDE AND PHASE EQUALIZATION IN A REVERBERANT SPACE

It is possible to relocate the poles and zeros for the minimum-phase and all-pass components, independently. Table 1 shows the combination of inverse filters for minimum-phase and all-pass components. Figure 15 shows normalized increase of the envelope-correlation coefficients by inverse filtering. The correlation increase is normalized by the correlation for without inverse filtering as shown in Fig.13(a).

We can see the pole/zero relocating effect is remarkable in minimum-phase inverse rather than the all-pass phase reverse. Here the minimum-phase inverse contains not only the magnitude inverse, but minimum-phase phase-inverse. However the severe sensitivity to the spatial locations might be due to the magnitude characteristics. The envelope correlation can be partly recovered by minimum-phase inverse, but all-pass phase inverse is still necessary at some frequency bands. Consequently we can surmise that pole-relocated minimum-phase with all-pass inverse at the target might be optimum in this example.

Table 1. Combination of Inverse Filters

	Without Inv. Filter	h_{min}^{-1} Target	$\hat{h}_{min}^{-1}$ Pole Relocated
Without Inv. Filter	(1)	(2)	(3)
h_{ap}^{-1} Pole/Zero Relocated	(4)	(5)	(6)
$\hat{h}_{ap}^{-1}$ Pole/Zero Relocated	(7)	(8)	(9)

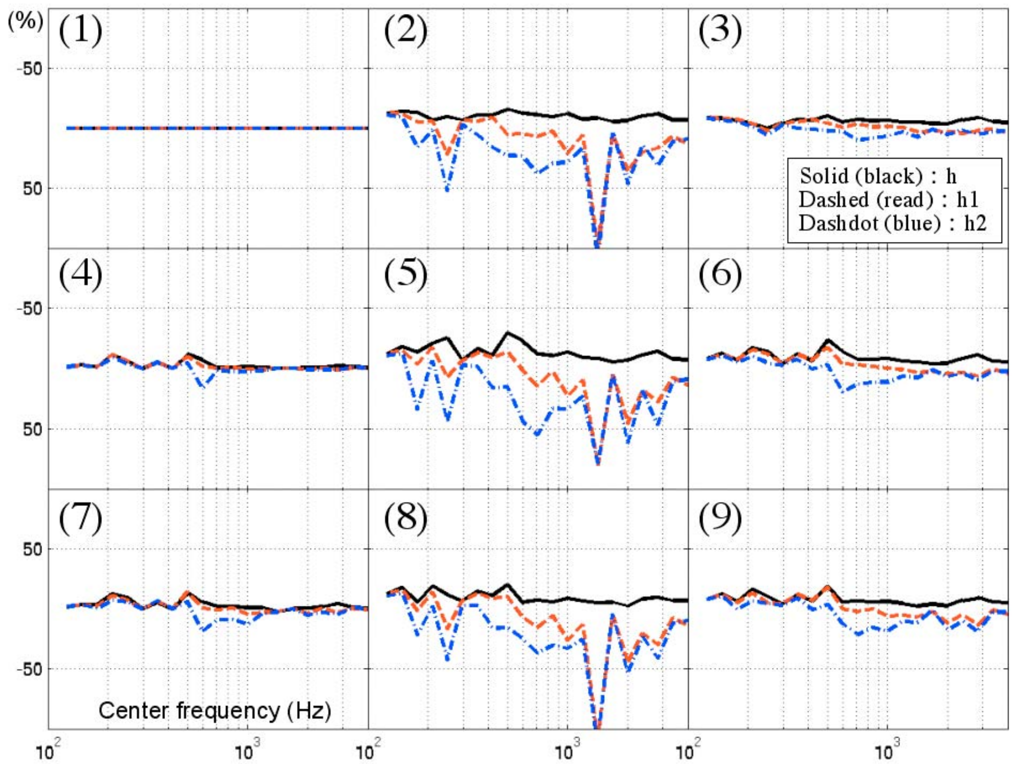

Figure15. Envelope Correlation Improvement defined by Correlation Increase by inverse filter normalized by the correlation without inverse filtering.

5 SUMMARY

This paper has described a method for inverse filtering of the TF including the all-pass component for intelligible speech reproduction in a reverberant space. We have confirmed that exponential windowing for the minimum-phase impulse response is effective for making inverse filtering to be robust against variations in the TF, since the poles of inverse TF move inside the unit circle in the complex frequency plane. Exponential windowing processing for the all-pass components is also formulated in this article according to pole/zero decomposition using cepstrum. However pole/zero relocating effect on the robustness of all-pass inverse is rather subtle, thus minimum-phase inverse with exponential windowing and all-pass inverse at a target position might be optimum for inverse filtering in a reverberant space.

6 REFERENCES

1. L. Savioja, et al., "Creating Interactive Virtual Acoustic Environments", J. Audio Eng. Soc. 47, pp. 675-705, 1999
2. M. Tohyama, R. H. Lyon and T. Koike, "Reverberant Phase in a Room and Zeros in the Complex Frequency Plane", J. Acoust. Soc. Am. 89, pp.1701-1707, 1991
3. M. Tohyama, R. Lyon, and T. Koike, "Pulse Waveform Recovery in a Reverberant Condition," J. Acous. Soc. Am. 91 (5) pp. 2805-2812 (1992)
4. M. Tohyama, "Room Transfer Functions and Sound Field Control," Active 95, pp.15-20 (1995)
5. O. Kirkeby and P. A. Nelson, "Digital Filter Design for Inversion Problems in Sound Reproduction," J. Audio Eng. Soc. 47. pp.583-595 (1999)
6. T. Houtgast, H. J. M. Steeneken and R. Plomp, "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics", Acustica vol. 46, pp. 60-72, 1980
7. B. D. Radlovic and Rodney A. Kennedy, "Nonminimum-Phase Equalization and Its Subjective Importance in Room Acoustics", IEEE SAP 8(6), pp. 728-737, 2000
8. Drullman, R., "Temporal Envelope and Fine Structure Cues for Speech Intelligibility," J. Acoust. Soc. Am. 97(1), pp. 585-592, 1995
9. H. Nomura, K. Yoshioda, M. Kazama, and M. Tohyama, "Speech Information Recovery by Magnitude or Phase Spectrum," Int. Cong. S&V. pp.4247-4251(2003)