MULTI-CHANNEL DECONVOLUTION USING PADÉ APPROXIMATION

Hong Wang
PictureTel Corporation M/S 635, 222 Rosewood Dr., Danvers, MA 01923
Tel:(508)623-4468; Fax: (508)749-2804; email:wangh@pictel.com

ABSTRACT

A new approach of estimating acoustic transfer functions from multi-channel reverberant signals is proposed. The ratio of two transfer functions is approximated by a polynomial by estimating one reverberant signal from another using least mean square estimation. Each transfer function is then estimated using Padé approximation. Two criteria are used to search for the most appropriate pair of transfer functions among all pairs generated by Padé approximation. The dereverberated signal is derived by multi-channel inverse filtering using the multiple input/output inverse theorem[1](MINT). Simulation using Gaussian noise convolved with minimum phase transfer functions gave a reconstruction SNR of 66dB. The approach is also applied to sub-band inverse filtering of reverberant speech recorded in a real room.

1. Introduction

Room reverberation is one of the most disturbing factors in audio/video conferencing 'since people are usually a distance away from the microphone. Many speech enhancement approaches based on signal processing have been proposed, however, there is still no satisfactory solution. One of the most promising ways to realize nearly perfect dereverberation of speech is to do inverse filtering using multiple microphones which pick up the signals from different acoustic paths [1, 2]. In this case, the impulse responses of each acoustic path must be known. Since the impulse response changes greatly with change of the shape of the enclosure, position of the sound source, temperature and humidity etc.[3], it is very hard to do calibration in a real application. The purpose of this work is to estimate the impulse responses from the multi-channel reverberant signals, and realize dereverberation by inverse filtering.

A cepstrum subtraction approach[4] has been proposed for blind deconvolution using two microphone signals. In this approach, the complex cepstrum of the speech signals recorded by two microphones are first calculated. The maximum and minimum phase cepstrum coefficients of the two impulse responses are calculated from the subtraction of the cepstrum of the two reverberant speech signals. The dereverberated speech is reconstructed using its cepstrum which is derived by the subtraction of the cepstrum of the reverberant speech and the impulse response. Since the speech signals always have zeros very close to the unit circle, the calculation of complex cepstrum is usually difficult. The reconstruction of the maximum and minimum phase part of the impulse responses is an iteration process which is very computationally intensive.

In this paper, we propose a new approach of estimating the impulse responses from the reverberant signals. We assume that there are no common zeros in the two acoustic paths considered. First, the ratio of the z-transforms of two impulse responses is approximated by a polynomial using least mean square estimation. For given orders of numerator and denominator, the two transfer functions can be uniquely decided by Padé approximation. Since the order of the transfer functions are unknown, two criteria are introduced to search among the many pairs of transfer functions generated by Padé approximation for all combinations of orders of the two transfer functions. The recovered signal is obtained by using two channel diophantine inverse filtering using the two impulse responses.

Simulation using Gaussian noise convolved with two minimum phase impulse responses showed a reconstruction SNR of 66dB. The algorithm is also applied to multi-channel reverberant speech recorded in a real room. The above criteria were used to search for the most appropriate pair of impulse responses from a pre-selected set. The sub-band signals are then recovered by diophantine inverse filtering. The complex amplitude of each sub-band is equalized before reconstructing the full band speech. In this case calibration is used in the pre-selection of impulse responses and amplitude equalization. The SNR of recovered speech is 18dB, which is 12dB higher than the amplitude equalization approach.

2. Principle

Consider the case of using two microphones to pick up sound in a room. Let $x(n)$ represent the sound source signal, $y_1(n)$, $y_2(n)$ the signal received at the two microphones, and $h_1(n)$, $h_2(n)$ the impulse responses of the two acoustic paths. The following relations exist

$$\begin{aligned} y_1(n) &= x(n) * h_1(n) \\ y_2(n) &= x(n) * h_2(n) \end{aligned} \quad (1)$$

The Z transforms of the above equations are

$$\begin{aligned} Y_1(z) &= X(z)H_1(z) \\ Y_2(z) &= X(z)H_2(z) \end{aligned} \quad (2)$$

If we assume that $X(z)$ does not have zeros on the unit circle, $h_1(n)$ and $h_2(n)$ are minimum phase, and there are no common zeros in $H_1(z)$ and $H_2(z)$, the ratio of the two transfer functions $H_{12}(z)$ can be estimated from the two reverberant signals by least mean square estimation.

$$\epsilon = \sum_{n=0}^{\infty} |y_2(n) - \sum_{i=0}^{N} h_{12}(i) y_1(n-i)|^2 = minimum \quad (3)$$

Where the Z transform of $h_{12}(n)$ is the ratio of the two transfer functions.

$$H_{12}(z) = \frac{H_2(z)}{H_1(z)} \quad (4)$$

0-7803-2431-5/95 $4.00 © 1995 IEEE

If the transfer function is not minimum phase, delay should be introduced in $y_2(n)$ of Eq.(3). A polynomial can be approximated by the ratio of two polynomials using Padé approximation. Given the order of the two unknown polynomials, the solution is unique[5]. Let $H'_1(m,z)$ and $H'_2(l,z)$ be the approximation of $H_1(z)$ and $H_2(z)$ at the order of m and l, respectively. The coefficients of $H'_1(m,z)$ and $H'_2(l,z)$ are determined by solving the following simultaneous linear equations with $m+l+1$ variables.

$$H_{12}(z)H'_1(m,z) - H'_2(l,z) = O(z^{m+l+1}). \qquad (5)$$

The following two criteria are used to search for the most appropriate pair of impulse responses from all possible pairs generated by Padé approximation for all combinations of m and l.

The first criterion is used to evaluate the performance of the approximation of the polynomial $H_{12}(z)$ by $H'_1(m,z)$ and $H'_2(l,z)$.

$$snr_1(m,l) = 10\log_{10}\frac{\sum_{n=0}^{N}|h_{12}(n)|^2}{\sum_{n=0}^{N}|h_{12}(n) - h'_{12}(m,l,n)|^2}, \qquad (6)$$

where $h'_{12}(m,l,n)$ represents the coefficients of the division of the estimated polynomials, $H'_2(l,z)/H'_1(m,z)$.

The second criterion is based on multi-channel inverse filtering theory. The diophantine inverse filters $G_1(m,z)$ and $G_2(l,z)$ are derived by solving the following linear equations:

$$H'_1(m,z)G_1(m,z) + H'_2(l,z)G_2(l,z) = 1 \qquad (7)$$

The recovered signal is

$$x'(m,l,n) = y_1(n) * g_1(m,n) + y_2(n) * g_2(l,n), \qquad (8)$$

where $g_1(m,n)$ and $g_2(l,n)$ are the inverse Z transform of $G_1(m,z)$ and $G_2(l,z)$, respectively.

This criterion evaluates how well the estimated impulse responses $h'_1(m,n)$, $h'_2(l,n)$, and recovered speech signal $x'(m,l,n)$ can approximate the real reverberant signals, where $h'_1(m,n)$ and $h'_2(l,n)$ are the inverse Z transform of $H'_1(m,z)$ and $H'_2(l,z)$.

$$snr_2(m,l) = 10\log_{10}\frac{\sum_{n=0}^{\infty}|y_1(n)|^2}{\sum_{n=0}^{\infty}|y_1(n) - x'(m,l,n) * h'_1(m,n)|^2} + 10\log_{10}\frac{\sum_{n=0}^{\infty}|y_2(n)|^2}{\sum_{n=0}^{\infty}|y_2(n) - x'(m,l,n) * h'_2(l,n)|^2} \qquad (9)$$

The optimum orders (M,L) of the impulse responses are decided according to the above two criteria.

$$(M,L) = \underset{m,l}{\text{argmax}}(snr_1(m,l) + snr_2(m,l)) \qquad (10)$$

The impulse responses $h'_1(M,n)$ and $h'_2(L,n)$ are the most appropriate pair.

3. Simulation Results

A Gaussian noise is used as the source signal. The reverberant signals are obtained by the convolution of the Gaussian noise with two minimum phase filters. The length of the Gaussian noise is 8000 samples, and the order of the impulse responses is 5.

Orders of m and l were searched from 3 to 10, and the criterion $(snr_1(m,l) + snr_2(m,l))$ gave the maximum value at $m = l = 5$. The SNR of the recovered signal is 66dB. Table 1 shows the result of the impulse response estimations.

Table 1: Estimation results of the impulse responses using Gaussian noise

n	$h_1(n)$	$h'_1(5,n)$	$h_2(n)$	$h'_2(5,n)$
0	1.00	1.0000	1.00	1.0000
1	0.80	0.7999	0.70	0.6999
2	0.60	0.5999	0.80	0.8000
3	0.70	0.7000	0.30	0.3000
4	0.40	0.4000	0.60	0.6000
5	0.10	0.1000	0.05	0.0500

4. Application to Sub-band Speech Dereverberation

The impulse response of a real room at speech bandwidth is always very long (more than 1000 taps). The search for the correct order (M,L) is too computationally intensive to be practically realized. By dividing the speech signals into a large number of sub-bands, the impulse response of each sub-band will be considerably simpler. Also we found in our previous work[2] that when the reverberant speech signals are divided into many sub-bands, for example 256 sub-bands with a sampling rate of 8kHz, many of the sub-bands have minimum phase impulse responses.

In this experiment, a speech sample recorded by seven microphones is used. The reverberation time of the room is 0.43s at 500Hz. The sampling frequency is 8kHz. The number of sub-bands is 256 with a decimation rate of 128.

The experiment follows the steps below: 1). Divide the reverberant signals into 256 sub-bands and down sample by 128. 2). Calculate $h_{ij}(n)$ for each sub-band and for each combination of microphone pairs, where i,j=1,2,..7. 3). Calculate impulse responses $h'_i(m,n)$, $h'_j(l,n)$ using Padé approximation for each combination of m and l, where $m,l =$ 3,4,...,20. 4). Pre-select two candidates of impulse response for each microphone, resulting in fourteen candidates for seven microphones. This is accomplished by calculating the SNR of the estimated impulse response compared to the real impulse response. The real impulse response is calculated by least mean square estimation using the sub-band signal of the original non-reverberant speech and the reverberant speech. Note that this step is not blind, i.e., it uses the information from the original signal. 5). Use the proposed criteria to choose two impulse responses from the fourteen impulse responses derived from step 4. 6). Do diophantine inverse filtering using the selected pair of impulse responses by step 5. 7). Equalize the complex amplitude of each subband using the original non-reverberant signal. This step is not blind. 8). Reconstruct the full band dereverberated speech.

The SNR of the dereverberated speech is 18.2dB. However, the SNR of dereverberated speech using only step 1, 7 and 8, i.e. by sub-band complex amplitude equalization only, is 5.9dB. The SNR of the reverberant speech is -0.97dB.

Figure 1 shows the sound spectrogram of the original speech, the reverberated speech, the dereverberated speech by complex amplitude equalization, and the inverse filtering approach proposed in this paper. Compared to the complex amplitude equalization only approach, the inverse filtering approach gives much more reverberation reduction, and produces much less pre-echo in the reconstructed full band speech.

5. Discussions

In step 4 of Section 4, the original reference signal is used to pre-select two impulse responses. This was done because the criteria proposed in this paper do not always give the correct evaluation of the impulse responses for all cases. In some cases, the criteria may choose a totally different function as the impulse response. However, for the pre-selected set of impulse responses, the criteria work very well. The pre-selected impulse response pairs (7*24=168 pairs in all) were checked. It was found that most of the pairs gave worse recovery of sub-band reverberant speech than the pair selected by the proposed criteria. That means, although the criteria are not perfect, it helps to eliminate the improper impulse responses. Another very important conclusion from this experiment is that it is possible to get the sub-band impulse responses of an acoustic enclosure only from the reverberant signals. Although the problem of searching for the most appropriate ones still remains.

Since this approach works perfectly with Gaussian noise convolved with minimum phase filters, and it works well for some sub-bands of real speech, the reasons why the criteria fails for some cases are considered as follows: 1). All sub-band impulse responses were treated as causal, and this might not be appropriate. 2). When the sub-band signal has zeros on the unit circle, the estimation of the ratio might not be accurate.

Another unsolved problem of this approach is the equalization of the complex amplitude of each sub-band. In order to show the importance of inverse filtering in each sub-band, the full band speech was reconstructed from the sub-band inverse filtered signals using complex amplitude equalization. This was compared with the recovered speech with amplitude equalization only approach. From Figure 1 we can see that amplitude equalization only approach does not solve the problem of reverberation, and the inverse filtering in each sub-band is necessary. However how to estimate the equalization parameter for each sub-band is still unknown.

6. Conclusions

We proposed a new method of estimating the impulse responses from multi-channel reverberant signals. Simulation of blind deconvolution using Gaussian noise showed 66dB SNR for the reconstruction signal. A sub-band dereverberated speech signal using this approach obtained an SNR of 18.2dB, which is a 12.3dB increase compared to the one without inverse filtering. Although the dereverberation process for real speech is not fully blind, the result showed the possibility of estimating sub-band acoustic impulse responses only from the reverberant signals. In order to solve the blind dereverberation problem, further studies on the choice of criteria and complex amplitude equalization are necessary.

ACKNOWLEDGMENTS

The author thanks PictureTel research staff for their discussions, comments and support.

7. References

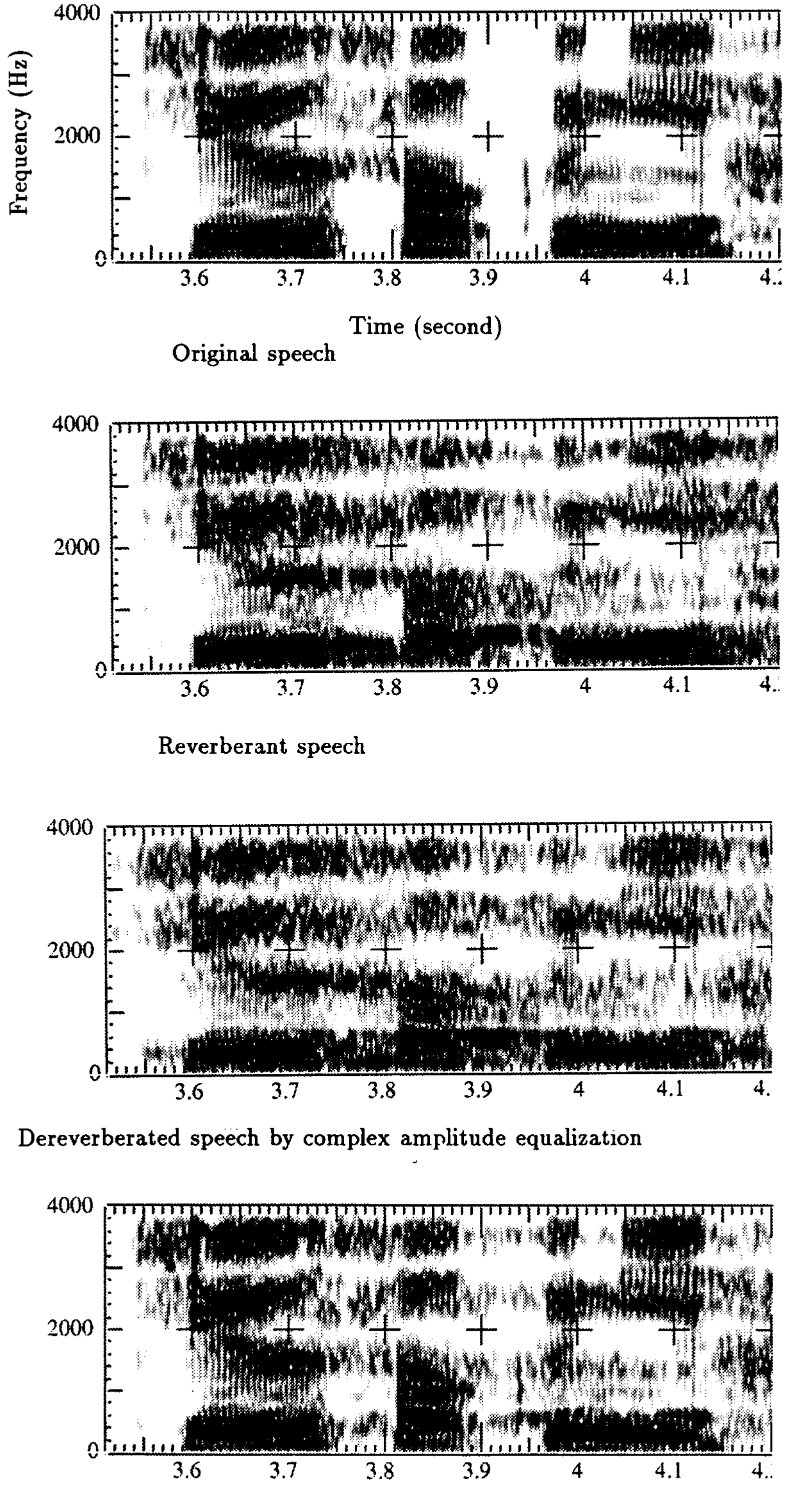
[1] M. Miyoshi and Y. Kaneda. "Inverse Filtering of Room Acoustic". *IEEE Trans. ASSP*, 36(2):145–152, 1988.

[2] H.Wang and F.Itakura. "An approach of dereverberation using multi-microphone sub-band envelope estimation". In *proceedings of the ICASSP IEEE*, pages 953–956, 1991.

[3] T. Hikichi and F. Itakura. "Compensation of the Time Variation of the Acoustic Transfer Function Using Linear Time Warping of the Impulse Response". In *Proceedings of Acoust. Soc. Japan*, 1994.

[4] A. P. Petropulu and S. Subramaniam. "Cepstrum Based Deconvolution for Speech Dereverberation". In *proceedings of the ICASSP IEEE*, pages I–9–12, 1994.

[5] G. A. Baker Jr. *Essentials of Pade Approximation*. Academic Press, 1975.



Original speech

Reverberant speech

Dereverberated speech by complex amplitude equalization

Dereverberated speech by inverse filtering and complex amplitude equalization

Figure 1: Speech sound spectrogram