OPTIMAL TRUNCATION TIME FOR MATCHED FILTER ARRAY PROCESSING

D. Rabinkin, R. Renomeron, J. Flanagan

CAIP Center, Rutgers University
[rabinkin,renomero,jlf]@caip.rutgers.edu

Dwight F. Macomber

SEAS, University of Pennsylvania,
macomber@ee.upenn.edu

ABSTRACT

Matched filter array processing (MFA) has been shown to improve signal-to-noise (SNR) quality for array speech capture in reverberant environments. However, under non-optimum conditions, MFA processing is computationally costly, and may produce little improvement or even subjective quality degradation as compared with simple time delay compensation (TDC). Appropriate truncation of the MFA filter bank is shown to reduce the computational burden without significantly reducing the capture SNR. This work attempts to find an optimal truncation time with respect to room size, wall absorption and the number of microphones used for the system. Simulations were conducted to evaluate MFA performance as a function of truncation length as these parameters were varied in situations typical of teleconferencing applications. It was demonstrated that judicious MFA truncation allows a reduction in computation load without sacrificing capture SNR.

1. INTRODUCTION

Microphone arrays provide a means to selectively capture distant sound sources. Matched filter array (MFA) processing has proven to be a useful technique as an extension of time delay compensation (TDC) when the array operates in a reverberant environment [5, 6]. The MFA processing technique performs coherent additions of direct and reflected signal energy in order to improve the signal-to-noise ratio (SNR) of the desired signal over background noise.

MFA processing consists of convolving the captured signal set from the array with the corresponding set of time-reversed focus-to-sensor responses. These can be obtained from the image method [1] or measured using a test signal [2, 3]. In large or highly reverberant rooms, the focus-to-sensor responses can be quite long — on the order of 1 second. As a result, the computational complexity of the algorithm, as well as system delay and subjective considerations, become factors in determining system feasibility. Truncating the focus-to-sensor responses used to construct the matched filters is an effective method for alleviating the effects of these problems.

The length of the focus-to-sensor responses is a function of both enclosure geometry and acoustic reflectivity. After the first arrivals, the sound intensity in a room decays approximately exponentially with time; hence the strongest reflections usually occur early in the focus-to-sensor response. As the later reflections are much weaker, they can be excluded from the matched filters without significantly affecting the performance of MFA processing. In addition, due to the nature of matched filter processing, the later reflections will cause an anticausal echo in the processed signal. Truncation of the MFA reduces this effect [6].

Work supported by NSF Grant No. MIP-9314625.

The effectiveness of reflection matching is dependent on the number of sensors used in the array and the placement of said sensors. Intuitively, having a larger number of sensors will improve the performance of the algorithm. However, the geometric arrangement of the sensors can also have a significant impact. In [5], it was shown that MFA's with low spatial cross-correlations between sensors required fewer microphones for a given SNR improvement. The SNR performance of a truncated MFA was investigated as a function of enclosure size, wall reflection coefficient, and the number of sensors in the microphone array.

2. THEORY

In an enclosure, sound propagation from source to microphone can be modeled as a transfer function. Thus, a sound signal $s(t)$ captured by a microphone in an enclosure can be expressed as:

$$m(t) = s(t) * h(t) \tag{1}$$

where $h(t)$ is the filter corresponding to the model transfer function, and "$*$" denotes convolution. Impulse responses of Room Transfer Functions (RTF's) may be simply modeled (in continuous-time) as the sum of a set of impulses. The first impulse represents the arrival of the direct wave. The progressively delayed and attenuated impulses which follow represent the multiple reflections arriving at the sensor location. They are commonly referred to as reverberation. Thus, the acoustic source-to-sensor pressure response is of the form

$$h(t) = \sum_{j=1}^{\infty} p_j \delta(t - \tau_j), \tag{2}$$

where the corresponding sets of p_j and τ_j depend on the acoustic properties of the signal path between the sound source and the sensor.

The MFA algorithm consists of filtering the input signals obtained from each microphone with the time reverse of the corresponding focus-to-sensor impulse response. For a sound source located at the array focus, the effect of the matched filter is to convolve the undistorted signal with the autocorrelation of the focus-to-sensor response:

$$y_i(t) = m_i(t) * h_{fi}(-t) = s(t) * h_{fi}(t) * h_{fi}(-t) \tag{3}$$

where $m_i(t)$ is the signal received at sensor i. A simple case for a single matched filter corresponding to an enclosure with two reflections is shown in Figure 1.

The output of the MFA is the sum of the outputs of each individual matched filter. For a single sound source at the focal point, this is:

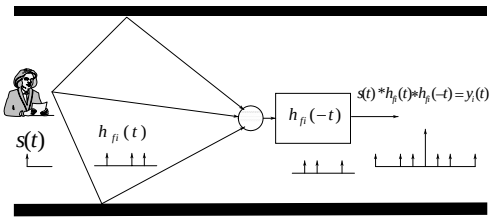

Figure 1: Effect of matched filtering in an enclosure.

$$y_{\mathrm{ON}}(t) = \sum_{i=1}^{N} \{s_{\mathrm{ON}}(t) * h_{fi}(t)\} * h_{fi}(-t)$$
$$= s_{\mathrm{ON}}(t) * \sum_{i=1}^{N} h_{fi}(t) * h_{fi}(-t) \qquad (4)$$

where $s_{\mathrm{ON}}(t)$ is the signal originating from the focal region, $h_{fi}(t)$ is the impulse response from the focal point to microphone i, and N is the number of sensors. When N is sufficiently large, the summation term on the right side of (4) will approximate a large amplitude impulse and will enhance the component of desired signal $s_{\mathrm{ON}}(t)$ appearing at the microphones.

A distinct advantage of the MFA over simple beamforming is its ability to suppress perceptual reverberation from the captured signal. In a reverberant environment, the performance of the delay-sum beamformer suffers as reflected sound waves appear along the "bore" of the beam. It is shown in [4] that the potential SNR for the MFA is independent of the number of reflections, as compared with beamforming, where SNR decreases monotonically with the number of reflections. The ratio of signal to reverberant energy for the MFA was shown to be

$$\mathrm{SNR}_{\mathrm{MFA}} = \frac{(NK)^2}{NK(K-1)} = \frac{NK}{K-1}, \qquad (5)$$

while the signal-to-reverberation ratio for the delay-sum beamformer is

$$\mathrm{SNR}_{\mathrm{BF}} = \frac{N^2}{N(K-1)} = \frac{N}{K-1}. \qquad (6)$$

In the above, K represents the number of reflections in the acoustical environment, and N represents the number of microphones in the arrays. Principles of operation of the two arrays are shown in Figure 2. Spherical spreading and the attenuation due to absorption and propagation were ignored in derivations of Equations 5 and 6 and Figure 2.

In practice, the source-to-sensor responses used for the MFA must be truncated. Otherwise, at an 8-kHz sampling rate, the 0.5 to 1.5-second reverberation times typical of medium-sized conference rooms would require processing each microphone's output with an FIR filter several thousand taps long. Processing delay would also introduce an unnatural lag in teleconferencing applications.

Another problem caused by using full-length matched filters is the precursor (anticausal echo) in the output of the MFA. The on-focus system impulse response of the MFA described in (4) has two long tails with a large impulse at the center, as shown schematically in Figure 2. The forward tail in the MFA impulse response generates a precursor of the desired signal at the output. These early arrivals are subjectively unpleasant and become more easily perceptible as they occur further in advance of the dominant arrival. With truncation, the length of early arrival lead time is significantly reduced, and the precursor is not perceptually prominent in the output signal.

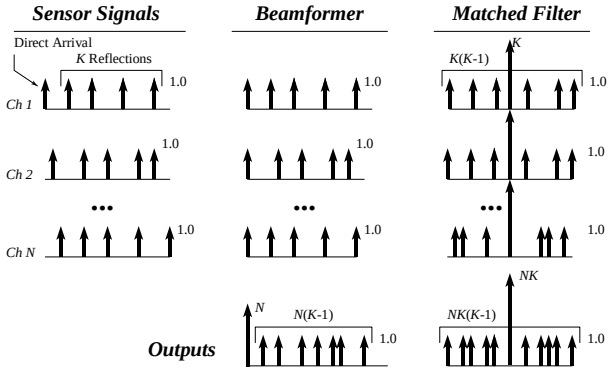

Figure 2: Alignment of captured signals using "delay-and-sum" beamforming and using MFA processing.

Use of matched filtering on each of the sensor signals in the MFA system provides spatial discrimination—signals arriving from the focus position will be enhanced relative to signals arriving from other locations. The processing combines the direct and reverberant arrivals coherently and suppresses the arrivals corresponding to off-focus signals, which are combined incoherently.

A conventional matched filter produces a peak in the amplitude of its output waveform at some time instant following the arrival of the input signal to which it is matched [7]. The matching process maximizes the filter's output energy (represented by the square of filter's output amplitude peak) relative to the output noise energy due to the presence of stationary additive noise at the matched filter's input. The output signal-to-noise ratio is considered optimum for this condition. The signal-to-noise criterion is different, however, for sound capture in a reverberant environment. In the following paragraphs, the SNR performance of MFA processing will be qualitatively evaluated for two general situations—full matched filtering for rooms with varying absorption, and matched filtering with truncated matched filters.

The transfer function of a single, fully matched MFA channel is essentially the autocorrelation of the RTF for that sensor. The single RTF can be decomposed into the direct arrival $h_d(t)$ and the successive arrivals of the reverberant tail $h_r(t)$:

$$h_i(t) = h_d(t) + h_r(t). \qquad (7)$$

The $i-$th channel output for a source signal $s(t)$ is then made up of the convolution of the signal with the sum of the respective correlations:

$$y_i(t) = s(t) * [\phi_{dd}(t) + \phi_{dr}(t) + \phi_{rd}(t) + \phi_{rr}(t)]. \qquad (8)$$

Output energy is the integral of $y^2(t)$. The total output energy can be seen to contain products of the auto- and cross-correlations of the direct and reverberant components of $h_i(t)$. Signal energy in the output is due primarily to $\phi_{dd}(0)$ and $\phi_{rr}(0)$, while the reverberant "noise" energy is due to $\phi_{dd}(t)$ and $\phi_{rr}(t)$, $t \neq 0$, along with the smaller cross-correlations $\phi_{dr}(t)$ and $\phi_{rd}(t)$.

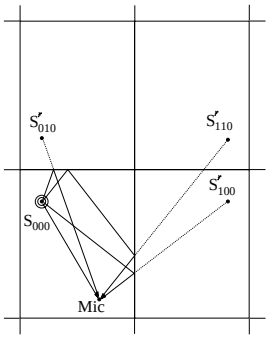

Figure 3: Illustration of the direct arrival and 3 reflections for sound propagation in a rectilinear enclosure

As the room absorption is reduced and reverberation time of the room increases, the amplitude of the reverberant tail $h_r(t)$ increases. This results in a significant increase in $\phi_{rr}(0)$. With squaring, the output signal energy due to the peak in the MFA response increases dramatically. The increase in the noise component is less than for the signal due to the incoherence in the reverberant arrivals. Hence, the SNR improvement offered by the MFA over the delay-sum beamformer increases as room reverberation times increase.

In the case of truncated MFA processing, $\phi_{dd}(0)$ is the only *auto*-correlation that contributes to energy of the desired signal in the MFA output—all other contributions come from *cross*-correlations. The large signal contribution due to $\phi_{rr}(0)$, the peak of the auto-correlation of $h_r(t)$, is lost. It is replaced by the cross-correlation between the full and the truncated $h_r(t)$. As the early, and strongest, portions of $h_r(t)$ are maintained in the truncated version, the desired signal component in the MFA output due to this cross-correlation are reduced only slightly from $\phi_{rr}(0)$.

3. SIMULATIONS

The effect of truncation on matched filtering was studied with simulated RTF's, which were computed using the Allen and Berkley image method for rectangular enclosures [1] as shown in Figure 3. It is assumed in the model that sound reflects from the enclosing walls at an incident angle equal and opposite to the arrival angle, and the amplitude of the reflection is attenuated by a reflection coefficient β, which is related to a wall's absorption coefficient α by the relation $\alpha = 1 - \beta^2$.

A multi-path transfer response between an acoustic source and a transducer is illustrated in Figure 3, where images appear to be arriving from "virtual" sources located in image rooms. The amplitude of sound arriving from these sources is attenuated by the usual $1/r$ propagation factor, as well as the product of the reflection coefficients $\{\beta\}$ associated with the set of reflection walls. The form of the resultant impulse response is given in (2). Discrete sampling of the transfer function implies lowpass filtering of the response by an anti-aliasing filter. Since both practical transducers and typical enclosures attenuate very low frequencies (below 50 Hz), there is also a "built-in" physical highpass filter associated with the transfer function.

The computation of the transfer functions for a given enclosure and microphone configuration is as follows. For each microphone in the array:

1. Compute ideal transfer function $h(t)$ to 9th-image order.
2. Compute a discrete-time response using a 8kHz lowpass filter.
3. Convolve with a 50Hz digital highpass filter.

The resultant impulse response set $\{h_i(n)\}$ is assumed to comprise the discrete-time acoustical system response.

To compute the truncated MFA response, each matched filter is convolved with a truncated version of itself, and then all the responses are added together to compute the overall transfer function. Truncation length is varied from length of time to first arrival[1] to the T_{60} time[2] in 30 increments.

No simply calculated measure for SNR that reflected the subjective improvement afforded by the MFA was known to the authors. Therefore, it was chosen to classify all output energy associated with the large peak in the *array*'s impulse response as *signal*. Room reverberation produces the tails in the auto-correlated RTF's seen at each sensor's matched-filter output. The energy in the MFA output due to the resultant sum of all these reverberant tails was considered *noise*.

SNR was computed as a function of the following parameter variations:

α	0.01, 0.1, 0.2, 0.4
# Microphones	10, 32, 100
Room size (meters) Small(S), Medium(M), Large(L)	$3 \times 4 \times 5$, $6 \times 8 \times 10$, $12 \times 16 \times 20$

The source and sensors were randomly placed around the corresponding enclosure using a uniform distribution.

4. RESULTS AND DISCUSSION

Figure 4 shows a typical response function $h_i(n)$ and a corresponding system SNR plotted against truncation length. The SNR is defined as system output SNR when all MFA filters are truncated to a length of time indicated on the abscissa of Figure 5. The SNR gain plateaus after 150 ms which is seen to be the prescribed optimal truncation time.

Figure 5 shows the superimposed SNR curves plotted against truncation times for all parameter permutations. The maximum SNR for all curves falls into 3 distinct bands, which are segregated solely on according to the number of sensors in the system. The figure indicates that the potential SNR of the array system is independent of the enclosure geometry and the reverberation time [3] of the room. These results are in agreement with (5).

The data from Figure 5 is summarized in Table 1. For each set of physical parameters the table entries contain a maximum system SNR value, and the required truncation length necessary to achieve an SNR within 0.5 dB of the maximum. Truncation length increases with room size, and decreases with the absorption constant. These results are in agreement with intuition.

An interesting observation from the results in Table 1 is that when all other parameters are fixed, there is a decrease in optimal truncation length with increase in the number of array sensors. This is explained by the greater density of reflections at the tails of $\{h_i(n)\}$. When many sensors are used there is a greater likelihood that these late reflections will produce coherent peaks in the tail of the MFA response. While these extraneous peaks are offset

[1] This corresponds to "straight" TDC.

[2] T_{60} defines the time by which the sound energy density in a room drops to 60 decibels below the initial sound energy density.

[3] For reverberant interference only.

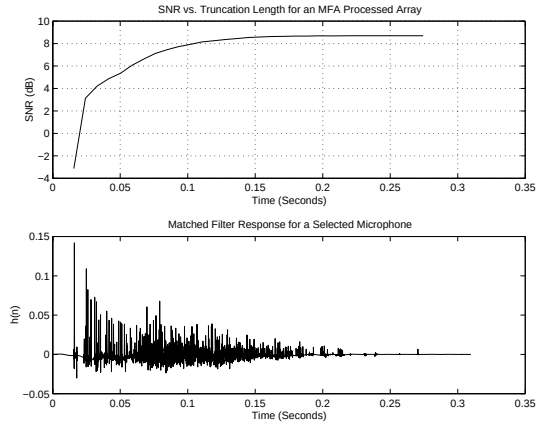

Figure 4: $h_i(n)$ for a selected microphone in a 10-element array, and the corresponding SNR vs. truncation length plot. In this instance $\alpha = 0.1$ and the room is $6 \times 8 \times 10$ meters in size.

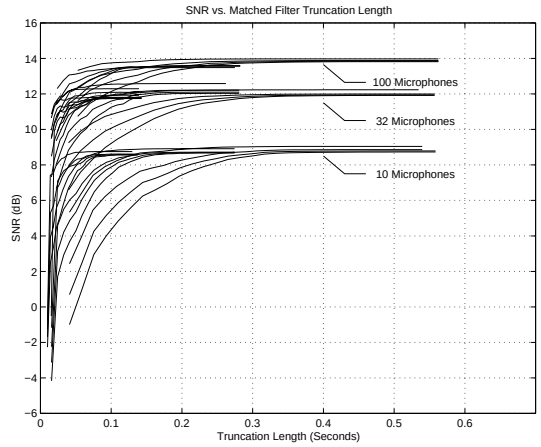

Figure 5: SNR vs truncation length for various array configurations

by greater growth in the main peak of the autocorrelation, the incremental benefit is less significant. Hence it is desirable to truncate the filter bank earlier for systems with a greater number of channels.

A related observation is that the growth in SNR with the number of sensors is less than predicted by (5). According to (5), there should be a 5.2dB growth for every tripling of sensor count. Table 1 indicates an approximate growth of 3.3dB with an increase from 10 to 32 sensors, and a growth of 1.5dB with the increase from 32 to 100 sensors. This is again attributed to the partial coherence of individual channel autocorrelations with a corresponding growth in tail energy. It is predicted that SNR growth will be incrementally very small with sensor count increases when a very large number of sensors is used.

5. CONCLUSION

The MFA technique is shown to be effective in eliminating reverberant noise for array sound capture. Simulation results demonstrate that system SNR for reverberant noise can be made independent of enclosure characteristics. Simulation results are in good agreement with theoretical considerations, and predicted performance trends.

		Absorption α			
# mics	Room Size	0.01	0.1	0.2	0.4
10	S	8.56 61	8.59 57	8.61 44	8.75 26
	M	8.68 128	8.70 111	8.72 102	8.91 59
	L	8.73 248	8.79 231	8.86 207	9.04 124
32	S	11.79 61	11.93 44	12.09 40	12.29 24
	M	12.03 117	12.10 100	12.21 74	12.59 30
	L	11.92 212	11.94 195	12.00 160	12.23 73
100	S	13.21 54	13.23 37	13.23 32	13.22 20
	M	13.59 93	13.59 84	13.56 75	13.48 32
	L	13.82 189	13.84 155	13.88 138	13.98 53

Table 1: Maximum MFA SNR (dB) - top number in block; Optimum truncation length (ms) - bottom number in block.

6. REFERENCES

[1] J. B. Allen and D. A. Berkley. Image method for efficiently simulating small room acoustics. *J. Acoust. Soc. Am.*, 65(4):943–950, April 1979.

[2] N. Aoshima. Computer-generated pulse signal applied for sound measurement. *J. Acous. Soc. Am*, 69(5):1484–1488, May 1981.

[3] J. Borish and J. Angell. An efficient algorithm for measuring the impulse response using pseudorandom noise. *J. Audio Eng. Soc.*, 31:478–487, 1993.

[4] J.L. Flanagan, A.C. Surendran, and E.E. Jan. Spatially selective sound capture for speech and audio processing. *Speech Communication*, 13:207–222, 1993.

[5] E. E. Jan. *Parallel Processing of Large Scale Microphone Arrays for Sound Capture*. PhD thesis, Rutgers University, New Brunswick, NJ, May 1995.

[6] R. J. Renomeron. Spatially selective sound capture for teleconferencing systems. Master's thesis, Rutgers University, New Brunswick, NJ, October 1997.

[7] George L. Turin. An introduction to matched filters. *IRE Transactions on Information Theory*, IT-6(3):311–329, June 1960.