Broadband Microphone Arrays for Speech Acquisition

Darren B. Ward

Acoustics and Speech Research Dept.
Bell Labs, Lucent Technologies
Murray Hill, NJ 07974, USA

Robert C. Williamson

Dept. of Engineering, FEIT
Australian National University
Canberra, ACT 0200

Rodney A. Kennedy

Telecommunications Engineering, RSISE
Australian National University
Canberra, ACT 0200

January 15, 1998

Abstract

A microphone array provides an effective means of hands-free speech acquisition, finding application in large-room video-conferencing, desktop teleconferencing, and hands-free telephones. Unlike a single omni-directional microphone, the array provides high spatial directivity, allowing it to attenuate room reverberation and other unwanted noise. In this paper, we give an overview of the techniques available for speech acquisition using microphone arrays. In particular we will focus on methods of frequency-invariant beamforming, in which the array maintains the same spatial response over a wide frequency range.

Corresponding Author:
Darren Ward
Room 2C-545
Bell Labs
600 Mountain Ave
Murray Hill, NJ 07974
USA
Phone: +1-908-582-5601
Fax: +1-908-582-7308
Email: dbward@research.bell-labs.com

1. INTRODUCTION

Hands-free audio communication finds use in many applications, including desktop teleconferencing and hands-free telephones. In these situations the desire is for a high-quality audio input, but without the requirement for the user to either hold or wear a microphone; this is especially important for hands-free telephones in cars. An effective technique for hands-free speech acquisition is microphone arrays. This technology was first applied to large-room teleconferencing in the early 1980's [1]. The purpose of this paper is to give an overview of the techniques for speech acquisition using microphone arrays, focusing on so-called *frequency invariant beamforming* methods.

2. THE SPEECH ACQUISITION PROBLEM

In hands-free speech acquisition applications the desired signal is corrupted by interfering noise, either from independent sources (such as fans, other talkers, etc.) or room reverberation. In general, these interfering signals originate from points in space separate from the location of the desired source, i.e., the talker's mouth. It is this spatial dimension which is exploited by microphone arrays in order to obtain a high-quality speech signal. By forming a directional microphone, the array is able to pass signals originating from some desired location while attenuating signals arriving from other directions. Unlike a fixed directional microphone (such as a boom microphone), the microphone array is able to adapt to changing signal environments and automatically follow the talker as he or she moves about.

In order to make the following discussion more concrete, consider the linear array shown in Fig. 1. For a single source, the time signal received at the nth microphone is given by

$$y_n(t) = s(t - \tau_n), \tag{1}$$

where $s(t)$ is the source signal, and τ_n is the time delay to the nth microphone (relative to the origin $x = 0$). The time delay to each microphone depends on the microphone location as well as the source position. Assuming the source signal is in the farfield of the array (see [2, 3] for nearfield sources), the delay to the nth sensor is given by

$$\tau_n = x_n c^{-1} \cos \theta, \tag{2}$$

where x_n is the location of the nth microphone, c is the speed of wave propagation, and θ is the direction to the source (measured relative to the array axis).

The general idea of the microphone array is to apply a complex weight[1] to each microphone signal, and then sum the resulting

---

[1] More generally, a filter is applied to each microphone signal.

signals to form the desired output, see Fig. 1. Thus, the output of the microphone array is given by

$$z(t) = \sum_{n=1}^{N} w_n \, y_n(t)$$
$$= \sum_{n=1}^{N} w_n \, s(t - \tau_n), \tag{3}$$

where w_n is the complex weight applied to the nth microphone; it is the design of these weights that it is of interest here. The array output is usually expressed in terms of frequency, giving

$$Z(f) = S(f) \sum_{n=1}^{N} w_n \, e^{j 2 \pi f \tau_n}, \tag{4}$$

where $S(f)$ is the Fourier transform of the time signal $s(t)$.

Since we are particularly interested in how the array performs in passing signals from a given direction and attenuating signals from other directions, we will consider its spatial response, or *beampattern*, defined by

$$b(\theta) = \sum_{n=1}^{N} w_n \, e^{j 2 \pi f c^{-1} x_n \cos \theta}. \tag{5}$$

Note that the beampattern is given by the array output, $Z(f)$, after factoring out the common signal term, $S(f)$. For a source at a location θ, the beampattern gives the transfer function to the array output. The typical properties of a beampattern are: (i) signals arriving from a certain spatial region (called the *main beam*) are passed by the beamformer with little or no attenuation; and (ii) signals arriving from other directions (called the *sidelobes*) are attenuated by the beamformer. Thus, the beamformer acts as a spatial filter, and it is this spatial selectivity which makes the beamformer an effective tool in removing unwanted noise from speech. Refer to [4] for more details of beamforming.

If the microphones are equally spaced, then equation (5) reduces to a Fourier series, and digital filter theory can be used to design the array weights. The use of digital filter design tools is a very useful technique for narrowband array design. However, if the array is to be used for broadband signals (such as speech), more sophisticated design techniques are required.

3. BROADBAND BEAMFORMING

The results presented so far have been for a narrowband array operating at a single frequency. In many classical array problems (such as radar or digital communications), this narrowband assumption is valid, and narrowband techniques are effective. However, speech is a broadband signal covering several octaves, and narrowband techniques are ineffective. To understand why this is so, consider again the linear array shown in Fig. 1.

At a given frequency, a large array with many elements has a very narrow main beam, whereas a small array has a broad main beam; the important dimension in measuring an array is its size in terms of wavelength. Thus, for high frequency signals (having a small wavelength) a fixed array will look large, and the main beam will be narrow. However, for low frequency signals (large

wavelength) the same physical array appears small, and the main beam will spread out.

This is illustrated in Fig. 2, which shows the beampattern of an array designed for 1.5 kHz, but operated over a frequency range of 500 Hz to 2 kHz. The frequency variation of this beampattern is unacceptable for speech applications. If an interfering speech signal is present at $60°$ say, then ideally it should be attenuated completely by the array. However, because the beam is broader at low frequencies than at high frequencies, the interfering speech signal will be high-pass filtered rather than uniformly attenuated over its entire band. This "spectral tilt" results in a disturbing artifact in the array output.

One common approach to overcome this problem is to use harmonically nested subarrays [5, 6]. In this case, the array is composed of a set of nested equally-spaced subarrays, each of which is a single-frequency design. The outputs of the subarrays are then combined by appropriate bandpass filtering. The effect of harmonic nesting is to reduce the beamwidth variation to that which occurs within a single octave. A set of subarray filters can be used to interpolate to frequencies between the subarray design frequencies [7, 8].

The first successful design for a broadband array with a constant spatial response in terms of frequency, was presented by Doles and Benedict [9]. They used the asymptotic theory of unequally-spaced arrays to derive relationships between beampattern characteristics and functional requirements on sensor spacings and weightings. This results in an array with filters on each sensor, the outputs of which are summed to form the final beamformer output. These sensor filters create a space-tapered array: at each frequency the non-zero filter responses identify a subarray having total length and spacing appropriate for that frequency. Although their design method provides a frequency invariant beampattern over a specified frequency band, it is based on a specific array geometry and beampattern shape.

Prompted by the work of Doles and Benedict, we derived in [10] a very general design method for broadband frequency invariant beamformers, suitable for linear, planar and even three-dimensional arrays. Our approach was to develop a frequency invariant beampattern property for a theoretical continuous sensor, and then to approximate this sensor by an array of discrete sensors. Thus, the problem of designing a broadband frequency invariant beamformer reduces to one of providing an approximation to a theoretically continuous sensor.

The basic structure of our frequency invariant beamformer is shown in Fig. 3. This block diagram illustrates the important properties of the beamformer. First, the filter response required on each microphone can be separated into two components: (i) the *primary* (beam-shaping) filter, and (ii) the *secondary* (normalization) filter. The primary filters, denoted by $H\left(x_n/x_{\mathrm{ref}} f\right)$ in the figure, perform the role of maintaining a constant beampattern shape as a function of frequency; they have a low-pass characteristic and are all related by a frequency dilation property. The output of each primary filter is multiplied by a constant *spatial weighting term*, g_n (which is dependent only on the microphone locations), and these signals are then summed. Finally, the secondary filter, denoted by αf, (which is simply a differentiator and serves to normalize the peak beampattern response) is common to all microphones and is implemented after the summation.

From an implementation standpoint, the most important of these properties is the dilation property of the primary filters. This property means that all primary filters are derived from a single set of filter coefficients, called the *reference coefficients* [11]. Hence, to change the beampattern it is only necessary to change the reference coefficients, and the beamforming structure ensures that the resulting beampattern will be frequency invariant. This is an important property for adaptive beamforming, since it reduces the number of adaptive parameters which must be updated at each step; an adaptive algorithm exploiting this property
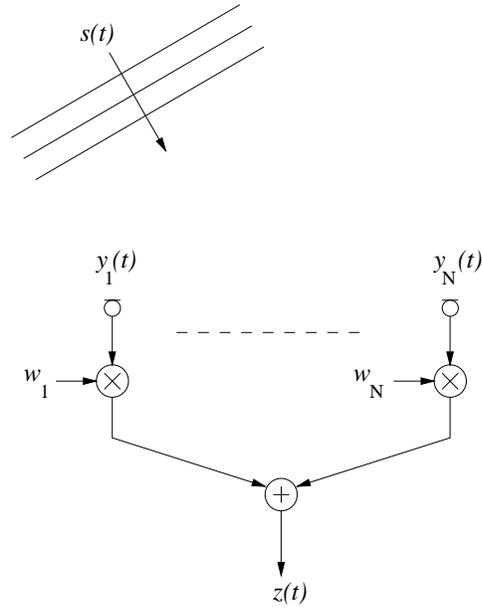
Figure 1: General linear array. The source signal is denoted by $s(t)$, $y_n(t)$ is the signal received at the nth microphone (only the first and last microphone in the array are shown), w_n is the weight applied to the nth microphone signal, and $z(t)$ is the array output.
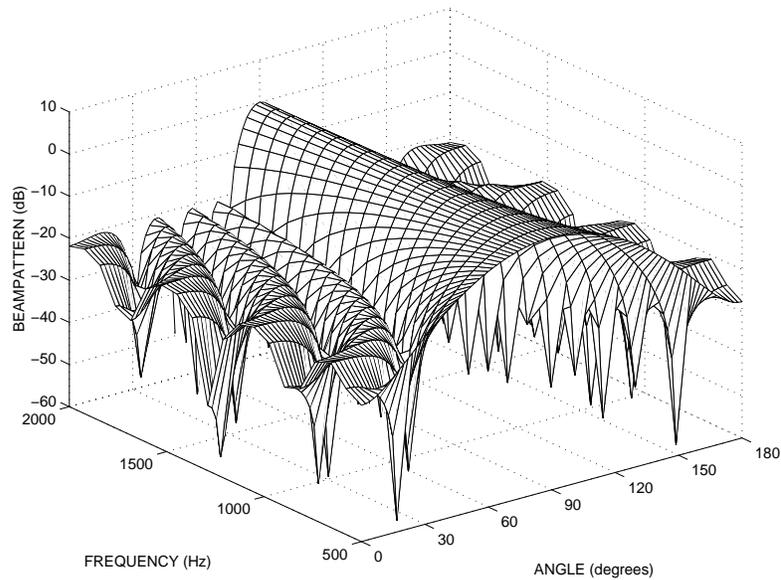

Figure 2: Spatial response of a narrowband array operated over a wide bandwidth.

was presented in [12].

A typical spatial response obtained with a frequency invariant beamformer is shown in Fig. 4. The design is for a frequency range of 300–3000 Hz, requiring 20 microphones. Notice that, as desired, the beampattern is (almost) frequency invariant over the entire band, exhibiting an almost constant main beam, and only slight ripple in the sidelobe region (c.f. Fig. 2 for a naive design).

4. CONCLUSIONS

Microphone arrays provide one of the most effective means of eliminating noise and reverberation from acquired speech. We have given a brief overview of how they may be used to provide a high-quality front-end to systems requiring hands-free speech input.

References

[1] J. Flanagan, J. Johnston, R. Zahn, and G. Elko, "Computer steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Amer.* 78(5), 1508–1518 (Nov. 1985)

[2] R. Kennedy, P. Abhayapala, D. Ward, and R. Williamson, "Nearfield broadband frequency invariant beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-96)*, Atlanta, USA (May 1996), pp. 905–908

[3] R. Kennedy, D. Ward, and P. Abhayapala, "Nearfield beamforming using nearfield/farfield reciprocity," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany (Apr. 1997), vol. 5, pp. 3741–3744

[4] B. Van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.* 5(2), 4–24 (Apr. 1988)

[5] J. Flanagan, D. Berkeley, G. Elko, J. West, and M. Sondhi, "Autodirective microphone systems," *Acustica* 73, 58–71 (1991)

[6] W. Kellermann, "A self-steering digital microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-91)* (1991), vol. 5, pp. 3581–3584

[7] R. Smith, "Constant beamwidth receiving arrays for broad band sonar systems," *Acustica* 23, 21–26 (1970)

[8] J. Lardies, "Acoustic ring array with constant beamwidth over a very wide frequency range," *Acoust. Letters* 13(5), 77–81 (1989)

[9] J. Doles III and F. Benedict, "Broad-band array design using the asymptotic theory of unequally spaced arrays," *IEEE Trans. Antennas Propagat.* 36(1), 27–33 (Jan. 1988)

[10] D. Ward, R. Kennedy, and R. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," *J. Acoust. Soc. Amer.* 97(2), 1023–1034 (Feb. 1995)
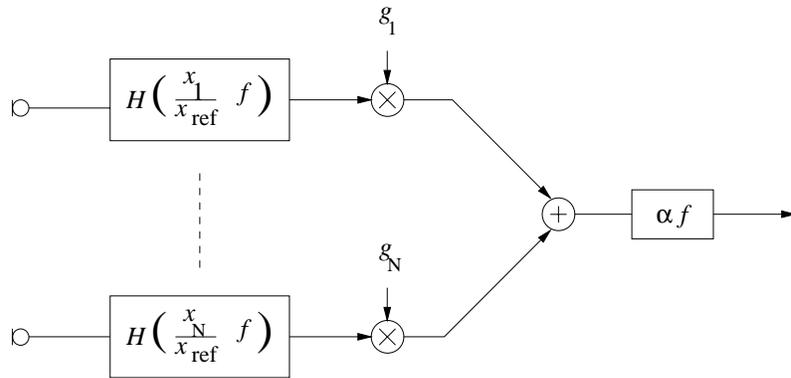
Figure 3: Block diagram of frequency invariant beamformer. The location of the nth microphone is given by x_n, x_{ref} is the location of a reference microphone, $g_n, n = 1, \ldots, N$, are a set of spatial weighting terms dependent only on the array geometry, and α is a normalization constant.

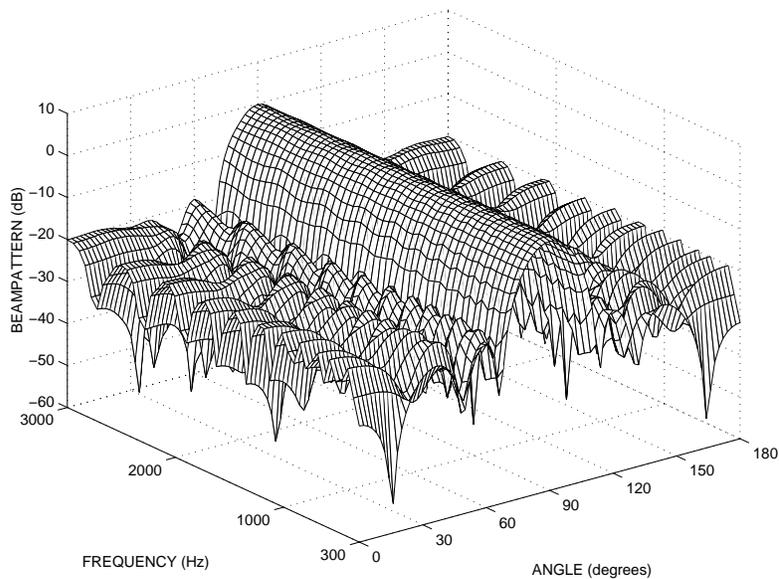
Figure 4: Beampattern of an example frequency invariant beamformer.

[11] D. Ward, R. Kennedy, and R. Williamson, "FIR filter design for frequency-invariant beamformers," *IEEE Signal Processing Lett.* 3(3), 69–71 (Mar. 1996)

[12] D. Ward, R. Kennedy, and R. Williamson, "An adaptive algorithm for broadband frequency invariant beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany (Apr. 1997), vol. V, pp. 3737–3740