13 Acoustic Echo Cancellation for Beamforming Microphone Arrays

Walter L. Kellermann

University Erlangen – Nürnberg, Germany

Abstract. Acoustic feedback from loudspeakers to microphones constitutes a major challenge for digital signal processing in interfaces for natural, full-duplex human–machine speech interaction. Two techniques, each one successful on its own, are combined here to jointly achieve maximum echo cancellation in real environments: For one, acoustic echo cancellation (AEC), which has matured for single-microphone signal acquisition, and, secondly, beamforming microphone arrays, which aim at dereverberation of desired local signals and suppression of local interferers, including acoustic echoes. Structural analysis shows that straightforward combinations of the two techniques either multiply the considerable computational cost of AEC by the number of array microphones or sacrifice algorithmic performance if the beamforming is time-varying. Striving for increased computational efficiency without performance loss, the integration of AEC into time-varying beamforming is examined for two broad classes of beamforming structures. Finally, the combination of AEC and beamforming is discussed for multi-channel recording and multi-channel reproduction schemes.

13.1 Introduction

For natural human–machine interaction, acoustic interfaces are desirable that support seamless full-duplex communication without requiring the user to wear or hold special devices. For that, the general scenario of Figure 13.1 foresees several loudspeakers for multi-channel sound reproduction and a microphone array for acquisition of desired signals in the local acoustic environment. Acoustic signal processing is employed to support services such as speech transmission, speech recognition, or sound field synthesis offered by communication networks or autonomous interactive systems. Such hands-free acoustic interfaces may be tailored for incorporation into a wide variety of communication terminals, including teleconferencing equipment, mobile phones and computers, car information systems, and home entertainment equipment.

For signal acquisition, microphone arrays allow spatial filtering of arriving signals and, thus, desired signals can be enhanced and interferers can be suppressed. With full-duplex communication, echoes of the loudspeaker signals will join local interferers to corrupt the desired source signals. Beamforming, however, does not exploit the available loudspeaker signals as reference information for suppressing the acoustic echoes. This is accomplished by acoustic
Fig. 13.1. Acoustic interface for natural human-machine communication.

echo cancellation (AEC) algorithms [1–3]. For discussing the combination of AEC with microphone arrays, the concept of AEC is first reviewed in Section 13.2 and beamforming methods are categorized in Section 13.3 with respect to the properties determining the interaction with AEC. Then, generic concepts for the combination of AEC and beamforming are discussed in Section 13.4. Structures for integrating AEC into beamforming are investigated in Section 13.5. Finally, the extension from single-channel reproduction to the case of multiple reproduction channels is outlined.

13.2 Acoustic Echo Cancellation

The concept of AEC is first considered for the case of a single loudspeaker and a single microphone according to Figure 13.2. To remove the echo from the microphone signal \( x(n) \) (with \( n \) denoting discrete time), AEC aims at generating a replica \( \hat{v}(n) \) for the signal \( v(n) \), which is an echoed version of the loudspeaker signal \( u(n) \). Aside from the echo \( v(n) \), \( x(n) \) contains components originating from local desired sources and local interferers, \( s(n) \) and \( r(n) \), respectively. Introducing the residual echo

\[
e(n) = v(n) - \hat{v}(n),
\]

the estimate for the desired signal \( \hat{s}(n) \) can be written as:

\[
\hat{s}(n) = x(n) - \hat{v}(n) = s(n) + e(n) + r(n).
\]