Library of Congress Cataloging-in-Publication Data

Noise reduction in speech applications / edited by Gillian M. Davis.
p. cm. — (The electrical engineering and applied signal processing series)
Includes bibliographical references and index.
ISBN 0-8493-0949-2 (alk. paper)
1. Speech processing systems. 2. Telephone systems. 3. Electronic noise—Prevention. 4. Noise control. 5. Signal processing—Digital techniques. I. Davis, Gillian M. II. Series.

TK7882.S65 N65 2002
621.382′8—dc21 2002017483

International Standard Book Number 0-8493-0949-2
Library of Congress Card Number 2002017483
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

form, the headphone has an open construction providing little or no attenuation of the environmental noise. In headsets designed for noisy environments, the headphones are mounted in ear cups with cushions that provide some attenuation.

The microphone is designed primarily to pick up the speech signal, but if the headset is used in a noisy environment, the background noise will also be picked up and transmitted together with the speech. As a consequence, speech intelligibility at the receive end will be reduced, possibly to zero. To increase the speech-to-noise ratio (SNR), it is common to use a directional microphone that has a lower sensitivity to sound incident from other directions than the frontal direction. In addition to this, the microphone electronics are usually equipped with a gate function that completely shuts off the microphone signal if its level drops below a threshold value. The purpose of the gate is to open the channel for transmission only when a speech signal is present.

Headsets are frequently used in noisy environments where they suffer from problems of speech intelligibility. Even if an ear cup–type headset is used, the attenuation is relatively poor for low frequencies. Low-frequency noise has a masking effect on speech, which significantly reduces the speech intelligibility.[1,2] Several cases have been reported in which the sound level of the communication signal was increased to hazardous levels by the user to overcome this low-frequency masking effect.[1,2] Ear exposure to the communication system resulted in hearing damage, such as hearing loss, tinnitus, and hyperacusis.

Passive Headsets

This section discusses headsets based on traditional passive ear defenders. Basic theory for passive ear defenders is introduced and practical issues that influence the performance of passive headsets are presented.

Traditional passive methods to attenuate noise employ barriers to block sound transmission and sound-absorbing materials to absorb the sound energy.[2] A passive ear defender — circumaural or closed-back headset — is based on a rigid ear cup containing sound-absorbing material.[2,3] Principally two ear cups are sealed to the users head via cushions by a spring band over the head. Passive ear defenders may be equipped with loudspeakers and boom- or throat-mounted microphone to provide one- or two-way communication. A closed back passive headset for two-way communication is shown in Figure 12.1.

The transmission ratio, $p_i(f)/p_e(f)$, for a passive headset at the frequency f[Hz] is given by:[3]

$$\frac{p_i(f)}{p_e(f)} = \frac{K_a}{K_a + K_c - (2\pi f)^2 M + j2\pi f R} \tag{12.1}$$

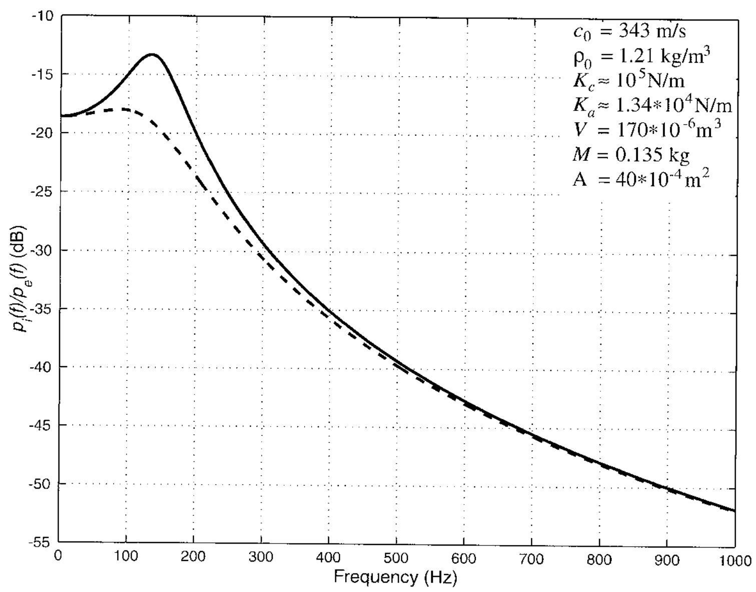


FIGURE 12.2
The transmission ratio between the internal and external sound pressure for two well-designed passive headsets; solid line R = 70 Ns/m and dashed line R = 140 Ns/m. The eigenfrequencies of the two headsets shown in this figure are at frequencies of about 100 and 130 Hz.

headset cup must often be designed to fit most people, the minimum area of the plane surface of the headset cup is limited by the fact that the ear cup must fit most ears. Furthermore, increasing the air volume inside the headset cup must also be limited so that the headset does not become too bulky and uncomfortable to wear.

As mentioned above, the mechanical stiffness of the cushion K_c is also a factor that influences the low-frequency performance of the headset. By selecting a cushion material with larger mechanical stiffness, the headset should theoretically produce greater low-frequency attenuation. However, in practice, the stiffness acting on the ear cup due to the cushion is limited by the layer of flesh underlying the cushion. Furthermore, as the stiffness of the cushion increases, not only does the headset become more uncomfortable to wear but its ability to provide good sealing around the ear also decreases. As a consequence of this, air leakage around the ear increases which prevents further improvement in the low-frequency noise attenuation being realized.

As discussed above, a passive headset designed to provide good low-frequency noise attenuation is likely to cause the wearer considerable discomfort due to its size, weight, and cushion stiffness. Active noise control techniques do not have the same limitations and have been proven to be very successful at improving the low-frequency attenuation achievable with headsets while at the same time allowing them to be comfortable to wear.[4]

where $p_i(f)$ is the sound pressure inside the analog hearing protector without control, K is the amplifier gain, $H(f)$ is the frequency function of the compensation filter, and $C(f)$ is the frequency function of the control path, i.e., the transfer path comprising the loudspeaker, headset cavity, and error microphone. By letting the amplifier gain K assume large values, the magnitude of the denominator in Equation (12.3) becomes large and the sound pressure under control approaches zero.

In practice, however, the performance of an active feedback control system is limited by closed-loop stability requirements, i.e., the Nyquist stability criterion.[7] Physical paths, such as the electro-acoustic response of the loudspeaker and the acoustic path from the loudspeaker to the microphone, introduce time delay due to propagation time, and this will introduce increasing phase shift with frequency and thus limit the performance of the control system. As the net phase shift in the electro-acoustic response of the loudspeaker and the acoustic path from the loudspeaker to the microphone approaches 180°, the feedback becomes positive and the magnitude of the open loop frequency response $KC(f)H(f)$ for the feedback control system must be less than one in order to remain stable. Thus, the frequency range of the control system where the loop gain $|KC(f)H(f)|$ can be large is usually upper limited by the frequency where the net phase shift in the electro-acoustic response of the loudspeaker and the acoustic path from the loudspeaker to the microphone approaches 180°.[7] By using a compensation filter to provide phase lag compensation, the low-frequency loop gain of the feedback control system may be increased as the phase-lag filter attenuates the high-frequency gain. In this way the gain margin (i.e., the maximum factor by which the open loop frequency response for the feedback control system can be amplified without the feedback control system becoming unstable[6,7]) of the open loop frequency response for the feedback control system can be improved, and the phase shift added by the compensation filter can be minimized.[6,7]

Although different compensation filter designs have been reported,[8,9] they have not been described in detail for commercial reasons. It is clear, however, that since the cavity enclosed by the headset is likely to vary between different users, it is important to ensure that the design of the controller used in the active headset is robust to such variations.[6] Robustness of digital controllers regarding variations in the control path is discussed in "Digital Active Noise Control Headsets."

To enable radio communication via the headset loudspeaker of an analog active noise control communications headset, the communication signal may be injected between the error microphone and the compensation filter, as shown in Figure 12.3. This results in a sound pressure under control, $p_c(f)$, given by:[8]

$$p_c(f) = \frac{p_i(f)}{1+KC(f)H(f)} + \frac{KC(f)H(f)p_s(f)}{1+KC(f)H(f)} \tag{12.4}$$

cavity and the pressure to maintain the acoustic seal around the ear necessary to provide passive attenuation can cause substantial discomfort. In contrast, an open-backed headset design, such as the headset type typically used in combination with portable tape and compact disk (CD) players, offers a more comfortable headset solution. This type of headset design does not, however, provide any passive high-frequency attenuation. In addition, it allows large variability in the acoustic path between the loudspeaker and error microphone.[4,10] Thus, the performance of an active open-backed headset typically falls short of that achievable with an active closed-back headset.[4,10]

Analog active noise control headsets typically produce an attenuation of about 20 dB at 100 to 200 Hz, which falls to zero below approximately 30 Hz and above approximately 1 kHz.[5] Higher attenuation of narrowband noise components may be achieved with nonadaptive analog active noise control headsets by using a more sharply tuned compensation filter.[6] However, as narrowband noise components may be time variable and are likely to differ between different environments, an adaptive controller is preferable. Since it is difficult and expensive to implement adaptive controllers that are analog, digital controllers tend to be used for the control of nonstationary and stationary narrowband noise.[6]

Digital Active Noise Control Headsets

This section covers digital active noise control headsets based on both adaptive feedforward and adaptive feedback control algorithms for active noise control headsets. The discussions of both types of algorithms are based on the well-known filtered-x least mean squares (FXLMS) algorithm. Since this algorithm was originally defined for feedforward control applications, this section begins by introducing active noise control headsets of the feedforward type.

Feedforward Control Systems

Feedforward control systems are theoretically more robust than feedback control systems. Feedback controllers are generally designed based on a model of the system to be controlled.[7] Variation in the control path may cause the feedback to become positive and lead to instability of the control system, i.e., the Nyquist stability criterion is violated.[7] In contrast to feedback control systems, a feedforward system is not based on a feedback control signal that may introduce this positive feedback and thereby instability.

In contrast to the feedback systems discussed above and in "Feedback Control Systems," feedforward control systems rely on the availability of a reference signal that contains information about the frequency content of the noise to be controlled.[11] The attenuation achievable is related to the amount of information about the noise to be controlled in the reference signal. The reference signal is processed by an adaptive digital control system prior to feeding the loudspeaker. For the control of broadband noise, a broadband be continuously adjusted to point toward the noise source in order to continue to fulfill the causality condition. Unfortunately, both this requirement for continual manual adjustment of the boom, as well as the need for the boom microphone in the first place, may cause an active headset based on a digital controller for broadband applications to be impractical to use. If the causality constraint is *not* fulfilled, the system can efficiently only reduce more deterministic noise, e.g., tonal noise, for which it is always possible to find a correlation.

We will continue our discussion on feedforward active noise control headsets by introducing some important feedforward adaptive control algorithms suitable for active noise control headsets.

The human ear responds mainly to the mean square value of the pressure it perceives. Consequently, the "quantity" or "cost" function that most adaptive active control systems are designed to minimize is the mean square value of the error microphone signal, which is proportional to the acoustic energy.[6,11]

A digital feedforward control system is illustrated in Figure 12.4. In both broadband and narrowband applications, the control filter is commonly based on a transversal filter, i.e., finite impulse response (FIR) filter, steered by the well-known FXLMS algorithm. This algorithm is developed from the least mean squares (LMS) algorithm and is based on a gradient search method that relies on the optimization technique known as the *method of steepest descent*.[6,11] The FXLMS algorithm is given by:[6,11]

$$\begin{aligned} y(n) &= \boldsymbol{w}^T(n)\boldsymbol{x}(n) \\ e(n) &= d(n) + y_C(n) \\ \boldsymbol{w}(n+1) &= \boldsymbol{w}(n) - \mu \boldsymbol{x}_{\hat{C}}(n)e(n) \end{aligned} \tag{12.6}$$

where μ is the adaptation step size and

$$\boldsymbol{x}_{\hat{C}}(n) = [\sum_{i=0}^{I-1} \hat{c}(i)x(n-i), \ \ldots \ , \sum_{i=0}^{I-1} \hat{c}(i)x(n-i-M+1)]^T \tag{12.7}$$

is the filtered reference signal vector, which usually is produced by filtering the reference signal $x(n)$ with an FIR-filter estimate $\hat{c}(i),\ i \in \{0,1,\ldots,I-1\}$ of the physical path between the loudspeaker and error microphone (i.e., the control path), and M is the length of the adaptive FIR filter, $y(n)$ is the output signal from the control filter, $\boldsymbol{w}(n) = [w_0(n),\ldots,w_{M-1}(n)]^T$ is the control filter weight vector, $\boldsymbol{x}(n) = [x(n),\ldots,x(n-M+1)]^T$ is the reference signal vector, $d(n)$ is the noise to be controlled, $e(n)$ is the estimation error, i.e., error microphone signal, and $y_C(n)$ is the output of the control path.

In practice, the elements in the filtered reference signal vector $\boldsymbol{x}_{\hat{C}}(n)$ are produced by filtering the reference signal, $x(n)$, with an FIR-filter estimate

The algorithm is robust, however, to errors in the estimate of the control path.[6,11] For example, in the case of narrowband reference signals, the algorithm will converge even for phase errors in the estimate of the control path of up to 90° provided that the step size μ is sufficiently small.[6,11] Furthermore, phase errors smaller than 45° will have only a minor influence on the algorithm convergence rate.[6,11]

In order to ensure stable action of the FXLMS algorithm, it has been found that the step size μ should be selected according to:

$$0 < \mu < \frac{2}{(\Delta + M)E[x_{\hat{C}}^2(n)]} \tag{12.8}$$

where $E[x_{\hat{C}}^2(n)]$ is the power of the filtered reference signal and Δ is the number of samples corresponding to the overall delay in the control path.[6,11] In practice, however, the power of a reference signal obtained by, for example, a reference microphone, might be time varying. If the reference signal has a time-varying power, it follows from Equation (12.8) that the upper limit for the step size μ is time varying. Variations in reference signal power influence the performance, e.g., the stability and convergence speed, of the FXLMS algorithm.[6,11] A common way to improve the performance of the FXLMS algorithm, regarding variations in the power of the reference signal, is to replace the fixed step size μ with a time varying step size $\mu(n)$ in the FXLMS algorithm (Equation [12.6]) according to:

$$\mu(n) = \frac{\mu_0}{\varepsilon + M\hat{E}[x_{\hat{C}}^2(n)]} \tag{12.9}$$

Here μ_0 is a step-size parameter typically less than two, $\hat{E}[x_{\hat{C}}^2(n)]$ is an estimate of the power of the filtered reference signal, and ε is a small positive number added in order to avoid division by zero if $\hat{E}[x_{\hat{C}}^2(n)] = 0$. By using the time-varying step size given by Equation (12.9) in the FXLMS algorithm, the normalized FXLMS algorithm is obtained.[11]

The mean power of the filtered reference signal vector can be updated according to different update laws.[11] One recursive update law for estimating the signal power is given by[11]

$$\hat{E}[x_{\hat{C}}^2(n+1)] = \hat{E}[x_{\hat{C}}^2(n)] + \frac{x_{\hat{C}}^2(n+1) - x_{\hat{C}}^2(n-L+1)}{L} \tag{12.10}$$

where L is the block length.

The stability and convergence properties of the FXLMS algorithm are related to errors in the estimate of the control path. One efficient way to improve the robustness to errors in the estimate of the forward path is to use the leaky FXLMS algorithm that is defined by[6,11] introduce a normalized version of the complex FXLMS algorithm by using a time-varying adaptation step size

$$\mu_f(n) = \frac{\mu_0}{\varepsilon + \hat{E}[\,|x_f(n)|^2]\,|\hat{C}_f|^2} \tag{12.13}$$

in Equation (12.12). As in the case of the normalized FXLMS algorithm, the estimate of mean power of the reference signal $\hat{E}[\,|x_f(n)|^2]$ may be updated according to different update laws.[11]

Feedback Control Systems

As discussed above, feedforward control systems rely on the existence of some prior knowledge of the noise to be controlled. This knowledge is provided by a reference signal that drives the control loudspeaker through the controller. Generally speaking, the ideal active controller is of this feedforward type provided, of course, that a reference signal highly correlated with the undesired acoustic noise is available.[6,11]

Active feedforward control is typically well suited to applications where it is simple and practical to obtain a reference signal of the noise requiring cancellation. Such a situation is typically found in helicopter and aircraft cockpits, and consequently the active headsets used in these environments are often of the feedforward type. In some headset applications, however, the generation of a suitable reference signal may be impractical or too costly. For example, there might be a large number of different noise sources with reference signals produced by tachometers or optical or inductive sensors, all of which have to be fed to the headset controller.

In such a situation, the use of feedback rather than feedforward control, for which no reference signal is required, has an obvious advantage. Furthermore, although feedforward control systems are theoretically more robust than feedback control systems (see "Feedforward Control Systems"), the performance of the feedforward controller is highly dependent on the *quality* of the reference signal, and in many cases a feedback system may perform equally well or better than a system with feedforward control.

The performance of a feedback controller in broadband applications is largely determined by the delay in the feedback loop. To obtain high performance, a small delay in the feedback loop is required.[6] This delay affects the length of the prediction interval of the controller, i.e., how far into the future the controller has to produce an estimate of the error signal.[6] Due to the inherent delay in digital controllers associated with their processing time, A/D- and D/A-conversion processes, analog anti-aliasing, reconstruction filtering, and analog feedback controllers, as discussed in "Active Noise Control Headsets," are usually preferred for use in broadband applications. For example, the hardware cost of a fast digital controller that introduces a

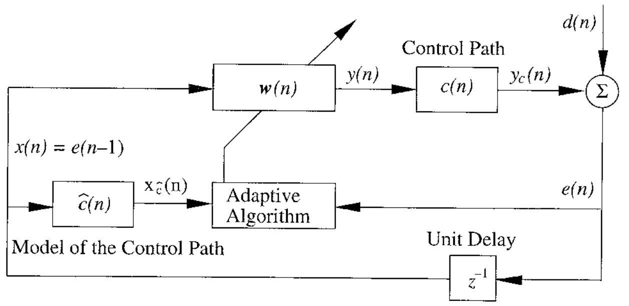


FIGURE 12.6
Block diagram of the feedback FXLMS controller.

$y(n)$ is the output signal from the control filter, and $e(n)$ is the error microphone signal. The relationship between the delayed estimate of the noise to be controlled $\hat{d}(n-1)$ and the reference signal $x(n)$ describes the fact that we are dealing with an adaptive digital filter in a feedback control system.

This type of controller is based on feedback cancellation; it uses an estimate of the control path to cancel the feedback.[6] If the controller is provided with a good estimate of the control path, it will act as an adaptive feedforward predictor.[6] However, differences between the estimate of the control path and the actual control path reduce the stable region of operation of the adaptive control system and this algorithm may suddenly become unstable.[6] Using the leaky FXLMS algorithm (Equation [12.11]) instead of the FXLMS algorithm (Equation [12.6]) in the IMC controller improves robustness to differences between the estimate of the control path and the actual control path.[6] This is an important issue, especially when it comes to open-backed headsets, as the control path may be subject to large variations in this type of headset, as discussed in "Analog Active Noise Control Headsets."[10]

Another adaptive feedback controller that might be suitable for narrowband applications is the feedback FXLMS algorithm. A block diagram of such a controller is shown in Figure 12.6. This algorithm is defined by adding Equation (12.15)[13]

$$x(n) = e(n-1) \tag{12.15}$$

to the equations defining the FXLMS algorithm (see Equation [12.6]). Here the relation between the delayed error signal $e(n-1)$ and the reference signal $x(n)$ describes the fact that we are dealing with an adaptive digital filter in a feedback control system. As in the case of the IMC controller (see Equation [12.14]), a leakage factor in the weight adjustment equation — the leaky FXLMS algorithm (Equation [12.11]) is used instead of the FXLMS algorithm (Equation [12.6]) in the feedback FXLMS algorithm — improves the robustness of the algorithm to differences between the estimate of the control path and the actual control path.[13] In contrast to the IMC controller (Equation [12.14]), this algorithm does *not* rely on cancellation of the feedback path.

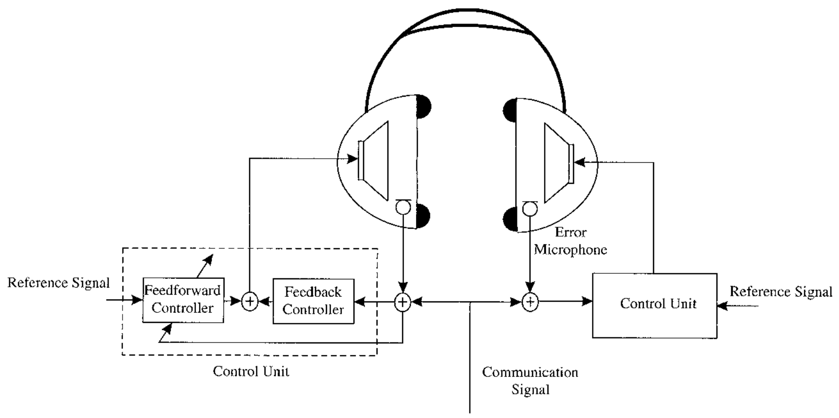


FIGURE 12.7
The principle of a closed-back hybrid headset based on digital feedforward and analog feedback active noise control techniques with communication signal injection.

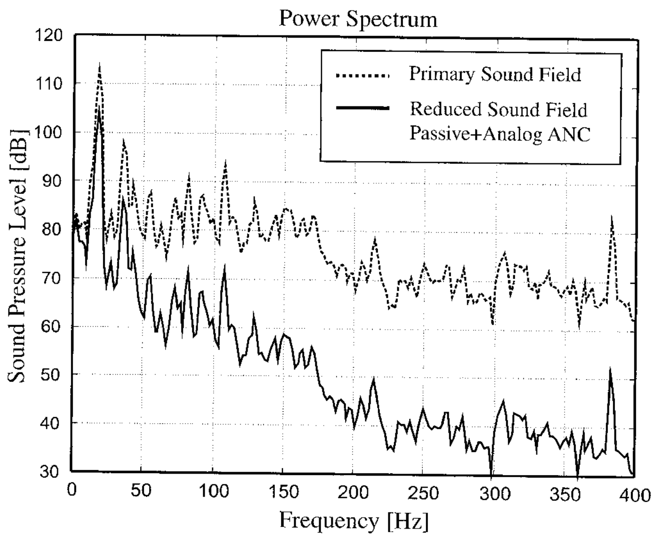
illustrated in Figure 12.8(b). By combining the passive headset with analog feedback active noise control, an improved overall low-frequency performance can be obtained, as shown in Figure 12.8(c). A combination of feedforward and feedback control results in significant attenuation of both the tonal components and low-frequency broadband noise, as illustrated in Figure 12.8(a).

Speech Enhancement for Headset Applications

Thus far, this chapter has focused on passive and active control of the noise inside the ear cups of headsets. Although these techniques reduce the environmental noise for the headset wearer, in the case of a communications headset, the noise picked up by the intercom microphone remains and reduces communication quality. This section focuses on two techniques for reducing the amount of background noise picked up by the intercom microphone and transmitted with the speech: spectral subtraction and a new in-ear technique.

Spectral Subtraction

Spectral subtraction is a broadband noise reduction method well suited for use in speech communication systems in severe noise situations such as intercom systems in boats, motorcycles, helicopters, and aircraft.[15] It is an efficient and robust background noise reduction technique that can be used in combination with the conventional active noise control techniques



(c)

FIGURE 12.8 *Continued.*
(c) Solid line; reduced sound pressure level inside the ear cups after the *analog feedback* controller has been switched on.

discussed earlier in this chapter. Spectral subtraction is based on a fast Fourier transform subtraction technique.[15] Two steps are required to remove noise: Step 1 is a noise level estimation step based on data gathered during speech inactive periods, i.e., collection of information about the type of noise to be removed. Step 2 is a spectral subtraction step involving subtraction of the noise from the speech using the estimated noise level from Step 1. Figure 12.9 illustrates schematically this spectral subtraction technique.

The SNR improvement achievable with spectral subtraction techniques is normally substantial. From experience, a rule of thumb is that the SNR improvement in headset applications achievable is of the same order of magnitude as the SNR before reduction. One disadvantage of spectral subtraction is, however, that sometimes background distortion of the processed signal may occur in the form of musical tones.[15] Such distortion, if heard at all, depends on the type of spectral subtraction scheme used, the level of noise reduction achieved, and the degree to which the background noise is nonstationary.[16] Spectral subtraction is commonly used on the transmit side of the communication channel, but the technique can be used on the receive side as well.

In-Ear Microphone

A common approach to achieving good SNR in an ordinary communication headset is to mount the microphone on a boom close to the mouth. In severe

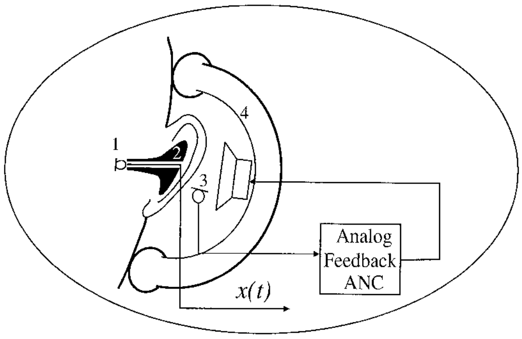


FIGURE 12.10
The ear microphone (1) is mounted in a plug (2). In this case, the ear microphone is combined with an analog active noise control (ANC) feedback headset (4) using a separate error microphone (3).

Conclusions

Headsets for speech communication are used in a wide range of applications. A communication headset usually consists of a pair of headphones and a microphone attached to the headset with an adjustable boom. In its simplest form, the headset has an open construction with little or no attenuation of the environmental noise. In headsets designed for noisy environments, the headphones are mounted in ear cups with cushions that provide some attenuation. The microphone is primarily designed to pick up the speech signal, but if the headset is used in a noisy environment, the background noise will also be picked up and transmitted together with the speech.

Passive headsets produce good attenuation of noise above their eigenfrequency, typically in the order of 40 dB above 500 Hz. Analog feedback active noise control techniques for use in headset applications have received considerable attention. These techniques have proved to be very successful at improving attenuation at frequencies below 1000 Hz by up to 20 dB and at the same time enabling comfortable noise canceling headsets to be designed. For the control of narrowband noise both digital feedback and digital feedforward controllers enable further attenuation to be achieved compared with analog feedback controllers. Digital feedback and feedforward controllers may be used in combination with analog feedback controllers or on their own.

In headset applications such as intercom systems, the background noise picked up by the boom microphone will be transmitted together with the speech. In order to enhance the speech transmitted by such systems, spectral subtraction is typically used. It is an efficient and robust broadband background noise reduction technique, which will not interfere with conventional active noise control systems used to improve the low-frequency attenuation