IN-EAR MICROPHONE TECHNIQUES FOR SEVERE NOISE SITUATIONS

Nils Westerlund, Mattias Dahl, Ingvar Claesson

Blekinge Institute of Technology
Research report No. 2005:12

In-Ear Microphone Techniques for Severe Noise Situations

Research Report, November 2005

In-Ear Microphone Techniques for Severe Noise Situations

Nils Westerlund   Mattias Dahl   Ingvar Claesson

November 2005

Abstract

Today, noise reduction methods are nearly ubiquitous although often not noticed by laymen. Cellular phones, some hands-free headsets and ear-phones offered by some airlines during long haul flights all utilize noise reduction algorithms. However, there are some situations in everyday life where ordinary noise reduction algorithms do not suffice; situations where the surrounding noise sound pressure level is too high to be efficiently attenuated by ordinary algorithms. Personal communication is then partly or totally prohibited by this noise. Examples of such situations may be motorcycle riding or attending a concert. In addition, many occupations, foremost industrial work, expose people to very high sound pressure levels. Still, these people need to be able to communicate safely. This report describes a technique where an ear-mic, i.e. a small microphone for communication purposes, is placed inside the auditory canal where it picks up bone conducted speech from the user's speech organ. The report describes three different approaches and usage areas: First, a basic approach where combination effects of an ear-mic and a pair of Active Noise Control (ANC) equipped ear-muffs are investigated. Second, this approach is used to improve a speech recognition system. The third approach is to connect a well-known noise reduction algorithm—the spectral subtraction—in cascade with the previously described ear-mic/ANC-solution in order to achieve extreme noise suppression.

Contents

Chapter 1

Introduction

Human noise exposure is a twofold problem. On the one hand, the noise itself may be harmful to humans and our hearing system. Noise of high sound pressure level is annoying and during periods of long exposure, it could cause fatigue, vertigo, nausea and loss of concentration. On the other hand, noise prohibits reliable personal communication. During personal communication the masking effect due to surrounding noise can significantly decrease speech intelligibility and quality.

Considering personal hearing protection, ANC is an attractive alternative and/or complement to passive reduction of unwanted noise since the latter implies heavy and bulky sound absorbers. ANC is an effective way of cancelling noise at frequencies below approximately 1 000 Hz [1]. Today, ear-muffs equipped with ANC are commercially available off-the-shelf.

Still, problems occur when trying to *communicate* in a noisy environment. If the communication takes place via some communication channel using acoustical microphones as transducers, these will pick up the surrounding noise, and add it to the desired speech signal. The noisy speech signal will be transmitted to the far-end side and, due to masking effects, the far-end talker will possibly have difficulties understanding the speech. Indeed, this may in fact force the far-end user to increase the volume in the communication system to levels that are damaging to the human ear [1].

The idea that a microphone for communication purposes can be placed at other locations of the body than in front of the mouth, is by no means a new one. Different locations of the body, such as the throat or forehead, have been used to attach transducers (in these cases tactile microphones) [2, 3]. The External Auditory Canal (EAC) has also been investigated as an alternative microphone location [4, 5, 6, 7, 8]. This research aim at preventing the noise from entering the communication system in the first place.

This report describes a technique where a small microphone for communication purposes is placed inside the auditory canal picking up bone conducted speech from the user's speech organ. Throughout this report this custom microphone is denoted "ear-mic". The report describes three different approaches

- First, a basic approach where combination effects of an ear-mic and a pair of ANC equipped ear-muffs are investigated.
- Second, this approach is used to improve a speech recognition system.
- The third approach is to connect a spectral subtraction noise reduction algorithm in cascade with the previously described ear-mic/ANC-solution in order to achieve extreme noise suppression. This hybrid solution offers the possibility for the user to communicate via some communication channel in the most extreme noise situations.

1.1 Combining Ear-Mic and ANC

The effects of combining an ANC equipped pair of ear-muffs with an ear-mic have previously not been thoroughly investigated. An audiometric earphone system which employs ANC implemented in a foam plug is introduced in [9] but this solution is not intended for communication purposes. This motivates chapter 3 of this report where the advantages of combining a pair of ANC equipped ear-muffs with an ear-mic are evaluated. It turns out that not only is the user's hearing protected by both passive and active noise control. Also, the indirect low-pass filtering that the ANC system implies functions as an equalizer of the low-pass in-ear speech signal.

1.2 Improving Speech Recognition Reliability

Since the 1950's, a system for fool-proof speech recognition has been a distant goal for researchers in the field. This problem was earlier considered rather easily solved, but has now proven to be a difficult task since the human communication is a more subtle process than initially was anticipated. Nevertheless, advances in computer technology and computational power have to some degree compensated our lack of knowledge about the human speech and hearing system [10, 11].

One of the major challenges of the speech recognition problem is to make the system robust to background noise. The fact that speech recognition accuracy is highly sensitive to noise, is indeed unfortunate since many everyday tasks that would benefit from a speech recognition system, are performed in noisy environments. For example, flight and car driving safety would probably improve if some of the functions in the airplane or car could be managed by speech regardless of the surrounding noise situation.

In chapter 4, a Hidden Markov Model (HMM) speech recognition system is used in combination with an ear-mic/ANC-solution. Since the combination of ANC and ear-muffs attenuates broadband noise inside the EAC, the speech recognition accuracy should increase. This increase in speech recognition reliability can also be viewed upon as an objective measure of the speech quality and intelligibility increase.

1.3 Hybrid Noise Reduction

In a personal communication situation, such as during inter-com, it is desirable to obtain both protection for the near-end user as well as enhanced speech quality for the far-end user in noisy environments with high sound pressure levels. To achieve that, chapter 5 presents an approach where three noise reduction methods are combined and applied to the speech signal inside the EAC:

- **Passive absorbers** are usually used to attenuate high frequency noise. A problem is that passive attenuation of low frequencies implies heavy and bulky absorbtion materials. A typical application, where passive absorbers are used to attenuate noise, is a pair of ordinary ear-muffs [12].

- **ANC** in headsets dates back to the 1950's, and a good active headset will effectively combine low frequency active attenuation with high frequency passive attenuation to provide high attenuation of the exterior noise at a wider frequency range [13]. ANC can be both feedforward [14], feedback [15] or a combination of both [1].

- **Spectral subtraction** is a non-linear, yet straightforward way of reducing unwanted broadband noise acoustically added to a speech signal [16]. The method estimates the magnitude frequency spectrum of the underlying clean speech by subtracting an estimate of the noise magnitude spectrum from the noisy speech magnitude spectrum. The noise estimate is obtained during speech pauses. A number of enhancements of the original algorithm have previously been proposed [17, 18, 19].

In a noisy environment, the combination of an ear-mic with both passive ear-muffs and ANC protects the near-end user from high sound pressure levels. However, the speech signal picked up by the ear-mic inside the ear will be contaminated by some remaining noise. A post processing spectral subtraction can enhance the speech quality and intelligibility before communicating it to the far-end user.

In other words, the two first methods, i.e. passive and active noise reduction, both protect the user from potentially hazardous noise levels as well as improve transmitted speech quality and intelligibility. The third method—spectral subtraction—functions as a pure noise reducer.

Chapter 2

Background

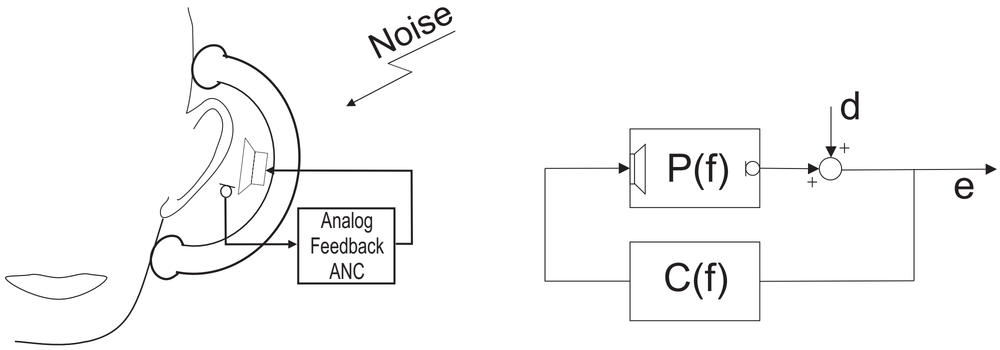
This chapter describes the basic components that are used throughout this report. A brief overview of passive and active ear-muffs are given as well as a description of the measurement setup that forms the basis of all measurements in the report. Also, an overview of bone conduction is given.

2.1 Passive/Active Noise Control Ear-Muffs

In a pair of passive ear-muffs, tight fit and good seal around the ear are crucial for the performance. To achieve this, the ear-muff cushions need to be soft and flexible. However, these soft cushions allow the ear-muff shell to vibrate when exposed to external sound, and this vibration is perceived by the user. In addition, some direct air leakage between ear-muff cushions and skull will always be present in practice. Basically, to achieve maximum noise attenuation, the ear-muff mass should be large, the cushion should provide a high resistance to air leakage, and the ear-muff inner volume should be large [12]. Obviously, the fulfillment of these demands is constrained by practical issues. Another factor that limits the maximum ear-muff noise attenuation performance is noise perception through bone conduction (see section 2.2.1).

Regarding frequency characteristics, low frequency attenuation is mainly affected by imperfect seal, and attenuation of higher frequencies is determined by the high frequency dynamics of the shell [13].

A pair of active ear-muffs equipped with *feedforward ANC* employs an external reference microphone to detect surrounding noise. This noise is then filtered by a controller and transmitted to the loudspeaker inside the shell of the ear-muff. An internal error microphone is used to adapt the controller. The well known Least Mean Square (LMS) algorithm is by far the most widely used method for adapting the digital filter.



(a) (b)

Figure 2.1: (a) Feedback ANC in an ear-muff. (b) The corresponding block scheme where $P(f)$ represents the channel transfer function, $C(f)$ represents the control transfer function, d is the primary noise and e is the resulting noise.

Feedback ANC implemented in a pair of ear-muffs, employs an internal error microphone placed inside the shell. The input signal from this microphone is phase shifted, weighted, and fed back to a loudspeaker also placed inside the shell, in order to create a noise cancelling sound field, see Fig. 2.1a. In Fig. 2.1b, the corresponding feedback control system block scheme is shown, with $P(f)$ representing the channel, i.e. the response from the loudspeaker input to the microphone output, and $C(f)$ representing the analogue controller. The noise inside the ear-muff shell is denoted by d whereas e represents the error signal. The active response, $G_A(f)$, of the system to the right in Fig. 2.1 can then be written as [13]

$$G_A(f) = \frac{e}{d} = \frac{1}{1 - C(f)P(f)} \tag{2.1}$$

Also note that signal delay is a critical issue in both feedforward and feedback active noise control.

2.2 The Ear-Mic

A custom ear-mic had to be manufactured in order to pick up bone conducted speech from the EAC of the user. Narrow canals were drilled in a pair of custom molded, acrylic ear-plugs and flexible plastic microphone probes tube were inserted into these canals. The probe lengths were adapted to reach as deep as possible inside the EAC without running the risk of damaging the tympanic membrane of the test person. This resulted in a probe length of 25 mm. The probe outer diameter was 3 mm, and inner diameter was 2 mm. The effects of using microphone probes are thoroughly investigated in [20]. In this study an optimal placement of the probe tube was found to be 1–3 mm from the tympanic membrane.

Two high-quality small-size microphones (Sennheiser) were attached to the exterior end of the probes, see Fig. 2.2.

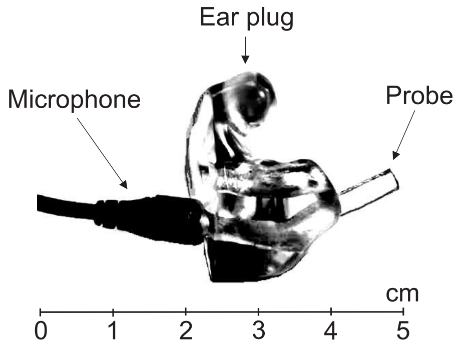


Figure 2.2: Custom-made acrylic ear-plug with microphone and microphone probe attached.

2.2.1 On Bone Conduction

Bone conducted sound in the human skull is a well known but not fully understood phenomena that anyone may evaluate simply by occluding one's ears and then speak. Although the acoustic transmission path from the mouth to the ear is practically eliminated, it is still possible to hear one's voice. This is mainly due to bone conduction of the speech signal from the speech organs to the tympanic membrane and cochlea.

Bone conducted sound for hearing aid purposes is a large research area. A tactile transducer, i.e. an exciter, may be attached to a patient's temporal bone via titanium screws and sound may be converted to mechanical vibrations and picked up by the human hearing organ. This is denoted Bone Anchored Hearing Aid (BAHA) [21].

In this report, the bone conducted sound plays an important role since the major part of the sound that is picked up by the ear-mic inside the EAC, has been transferred from the speech organ via the skull. The muscles and tissues covering the skull are assumed to transmit a small or even negligible amount of sound energy. Also, they are assumed to have a muffling effect, i.e. to attenuate higher frequencies.

A large difference between BAHAs and the approach adopted in this report is that with BAHAs the skull is excited from the exterior while in this report, the skull is assumed to be excited from the interior by the speech organ. Nevertheless, it can probably be assumed that the same characteristics of bone conducted sound are present in both cases.

Transmission of sound through the skull to the cochlea is clearly a complex phenomenon. However, research performed on patients with skin penetrating titanium implants in the temporal bone, i.e. skulls *in vivo*, has shown that bone conducted sound possesses the desirable property of linearity [22]. This implies that no distortion such as harmonics etc. is present in the ear speech signal. Furthermore, the resonances of the skull are relatively highly damped (2.6–8.9%) [23]. Hence, the ear speech signal should not be significantly altered. Nevertheless, as previously stated, the muscles, tissues and cavities of the human head are assumed to have a low-pass filtering effect on the bone conducted speech.

2.3 Basic Measurement Setup

An illustration of the basic measurement setup used throughout this report is shown in Fig. 2.3. The speech signal, $s(t)$, originating from the human speech organ propagates mainly through the skull. The external acoustic path, i.e. the air borne speech sound transmission, is assumed to be negligible due to the attenuation by the ear-muffs. The high frequency part of the surrounding noise $w(t)$ is attenuated by the passive absorbers in the ear-muffs, but the speech signal inside the EAC is still contaminated by the low frequency components of the noise that passive absorbers are unable to attenuate efficiently. However, the ANC attenuates those remaining noise components. The resulting speech signal $x_E(t)$ picked up by the ear-mic, is properly bandlimited and sampled forming a digital output signal $x_E(n)$. This signal can now be transferred to, for instance, Matlab for further processing.

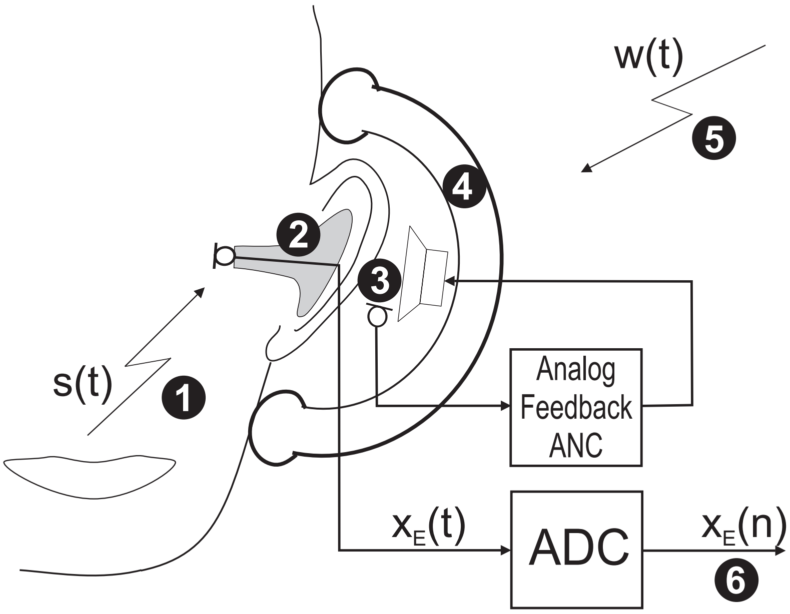


Figure 2.3: Measurement setup and signal paths. (1) Bone conducted speech signal. (2) Ear-mic inserted in custom-made acrylic earplug. (3) ANC reference microphone and loudspeaker. (4) Passive ear-muff. (5) Surrounding noise. (6) Digital output speech signal.

Chapter 3

Ear-Mic and ANC Combination

As described in section 2.1 a pair of ear-muffs that employs ANC for noise reduction, substantially reduces the influence of the low frequencies inside the ear-muff cap. This implies an indirect high-pass filtering of the sound in the EAC. This chapter shows that not only is it possible to use the in-ear speech signal for communication purposes. In addition, the above mentioned high-pass filtering property is also convenient when combining an ANC headset with an ear-mic for communication purposes, since the speech signal inside the EAC is a low-pass filtered version of the speech signal at the mouth. Hence, the in-ear speech signal is to some extent restored by the implicit ANC high-pass filtering, the quality of the speech signal in the auditory canal is improved and the speech intelligibility is increased. By that, combining an active ear-muff with an ear-mic serves two purposes: Protecting the user from harmful noise and enables the user to communicate over some channel using the bone conducted speech signal in the EAC.

Note also that this indirect high-pass filtering does not introduce any additional delay on the speech signal. This is advantageous since one of the application areas this equipment is intended to be used in is a mobile communication system. Since the channel equalizer is designed to operate in such a large system, it is desirable to reduce the delay caused by the filtering and in this way minimize the total delay introduced by the whole system.

3.1 Data Acquisition and Method

Two different ear-muffs were evaluated: A pair of modified Hellberg ear-muffs and Bose QC-1 ear-muffs. Both the Hellberg and the Bose ear-muffs were equipped with a switch to enable/disable the ANC. Three different noise environments were used: No noise (speech only), noise from the outside of

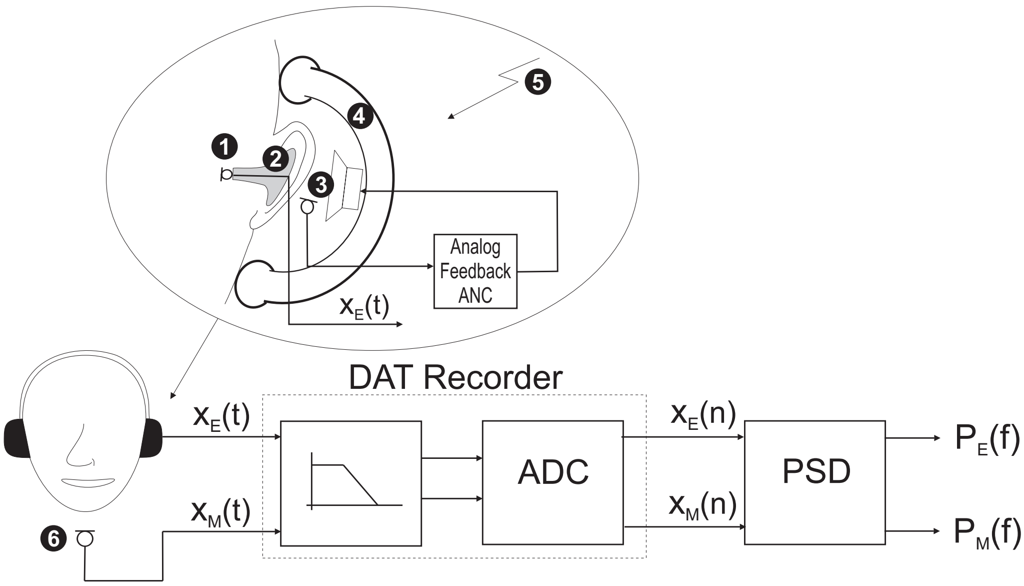


Figure 3.1: Measurement setup and signal paths. (1) Ear-mic. (2) Custom moulded acrylic ear-plug. (3) ANC reference microphone and loudspeaker. (4) Ear-muff. (5) Surrounding noise. (6) Reference microphone at mouth.

a helicopter and noise from a helicopter cockpit. These two types of noise are normally dominated by tonal components in the lower frequency range (< 500 Hz) [24]. A high-end loudspeaker was used to reproduce the noisy environment. All measurements were performed in an ordinary office space room with dimensions 4.15×3.45 m (13.6×11.3 ft).

The basic measurement setup from section 2.3 was used with the addition of a reference signal, $x_M(t)$, taken from in front of the mouth. Fig. 3.1 illustrates the measurement setup and signal paths. The analog speech signals from the mouth and the EAC, $x_M(t)$ and $x_E(t)$ respectively, were recorded to a DAT recorder at a sampling frequency of $F_s = 48$ Khz. The resulting digital signals $x_M(n)$ and $x_E(n)$ were transferred to a MATLAB-file and the bandwidth was reduced to 4 kHz ($F_s = 8$ kHz). Welch's method was used to calculate the power spectral densities (PSD) of the signals, $P_M(f)$ and $P_E(f)$.

The basic idea is that a spectral weighting function (i.e. a gain function) $G(f)$ could be calculated using the expression

$$G(f) = \frac{P_M(f)}{P_E(f)} \tag{3.1}$$

The function $G(f)$ could then be used to weigh the PSD of the speech signal

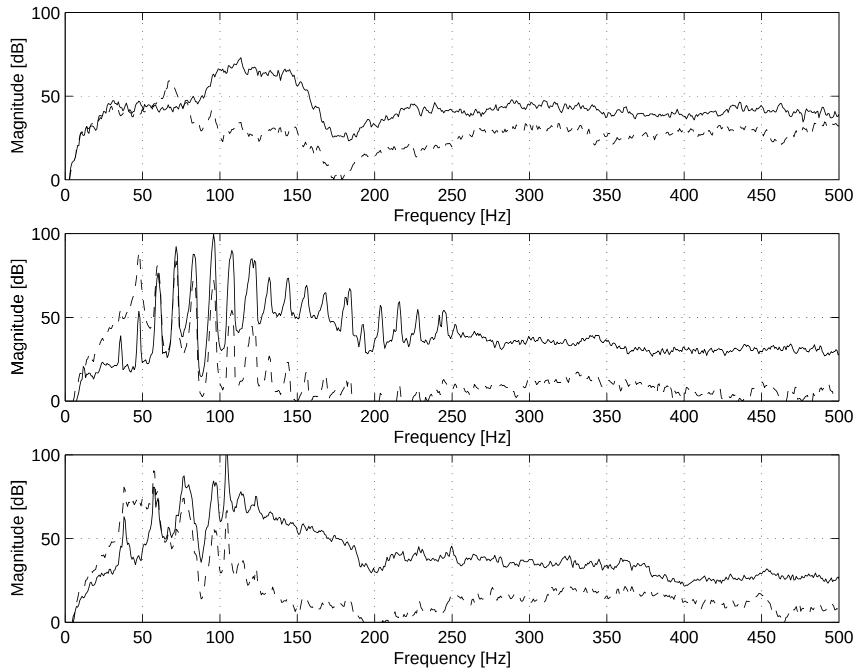


Figure 3.2: Power spectral densities of speech signal inside the EAC using Hellberg ear-muffs. Solid line: ANC disabled. Dashed line: ANC enabled. (Upper plot) Speech with no surrounding noise. (Mid plot) Speech and noise from the outside of a helicopter. (Lower plot) Speech and noise from helicopter cockpit.

inside the EAC to a PSD more like the one at the mouth and in that way improve the quality of the speech.

3.2 Results

Fig. 3.2 and Fig. 3.3 show the PSDs for the sound inside the EAC. It is clear that the low frequency components of the signal inside the EAC are partially attenuated when the ANC is enabled. This low frequency damping was perceived as a quality and intelligibility increase during informal listening tests.

To further improve the speech signal quality and intelligibility, gain functions, $G(f)$, were calculated according to (3.1). The gain functions with ANC enabled and disabled for both Hellberg and Bose ear-muffs, are plotted in Fig. 3.4. Application of $G(f)$ on the speech signal resulted in an additional quality and intelligibility increase.

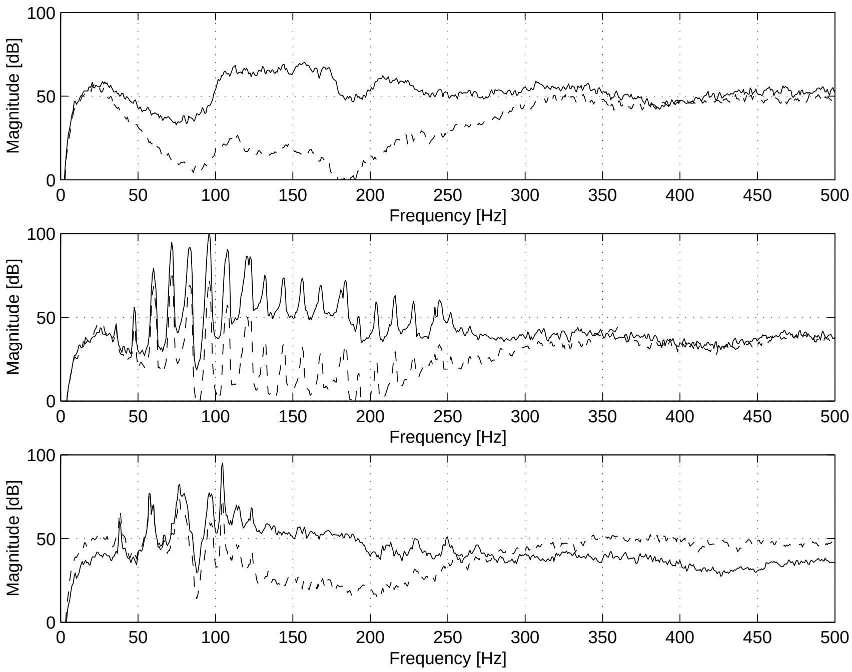


Figure 3.3: Power spectral densities of speech signal inside the EAC using Bose ear-muffs. Solid line: ANC disabled. Dashed line: ANC enabled. (Upper plot) Speech with no surrounding noise. (Mid plot) Speech and noise from the outside of a helicopter. (Lower plot) Speech and noise from helicopter cockpit.

3.3 Conclusions

Since the bone conducted mouth-to-ear channel represents a fairly simple low-pass system and the active noise control system attenuates low frequencies, the speech picked up inside the EAC becomes more intelligible when the active noise control is enabled. This is because the speech is less dominated by low frequency components. The use of a gain function calculated from the PSDs of the input and output signal can also improve the speech quality.

This initial study serves as a basis for the subsequent chapters in this research report.

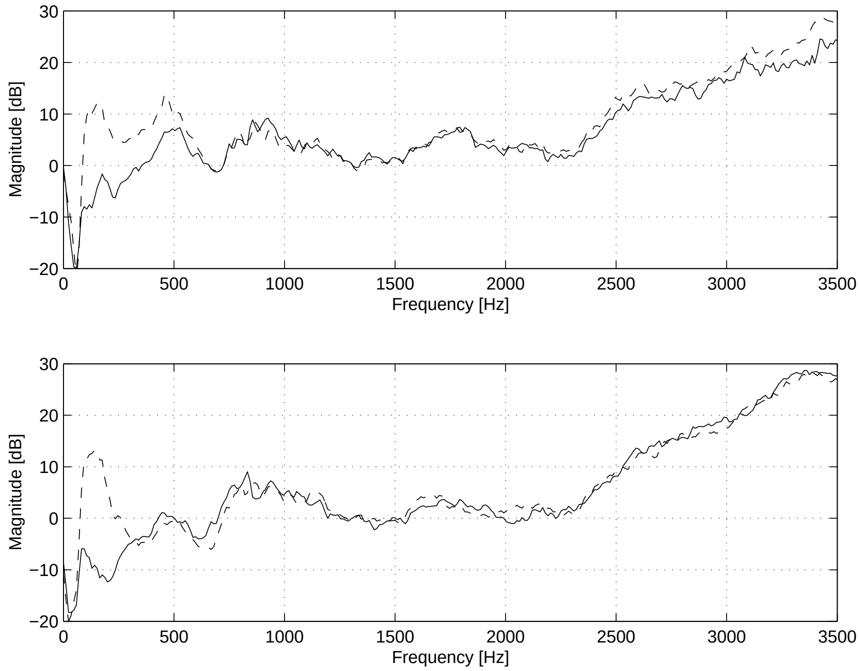


Figure 3.4: (Upper plot) Gain function for Hellberg ear-muffs. Solid line: ANC disabled. Dashed line: ANC enabled. (Lower plot) Gain function for Bose ear-muffs. Solid line: ANC disabled. Dashed line: ANC enabled.

Chapter 4

Improving Speech Recognition Reliability

This chapter examines whether it is possible to increase speech recognition robustness in noisy environments by means of an ear-mic/ANC-combination.

4.1 On Speech Recognition

Speech recognition can be of interest in many situations. For example, many functions in a vehicle can be voice controlled and voice controlled functionality obviously improves safety since the user can stay focused on operating the vehicle at hand. Hence, speech recognition is a growing field in vehicular technology. If the speech intelligibility and by that the speech recognition robustness could be increased in a speech recognition controlled system implemented in a vehicle, the safety would be further improved.

Speech recognition is traditionally performed in two separate steps: Initially, a number of words are chosen (i.e. a training set) and these words are pronounced and presented to the speech recognition system repeatedly. This is referred to as the *training phase*. When the speech recognition system has been trained, new occurrences of the chosen words are presented to the system in order to perform speech recognition. This is referred to as the *test phase*. The terms "training phase" and "test phase" will be used to describe these different steps of speech recognition for the remainder of this chapter.

In this report, the Hidden Markov Model Toolkit (HTK) is used. HTK is a toolkit for building Hidden Markov Models (HMMs). These can be used to model any time series and the core of HTK is similarly general-purpose. However, HTK is primarily designed for building HMM-based speech processing tools, in particular recognizers. For an exhaustive survey on the speech

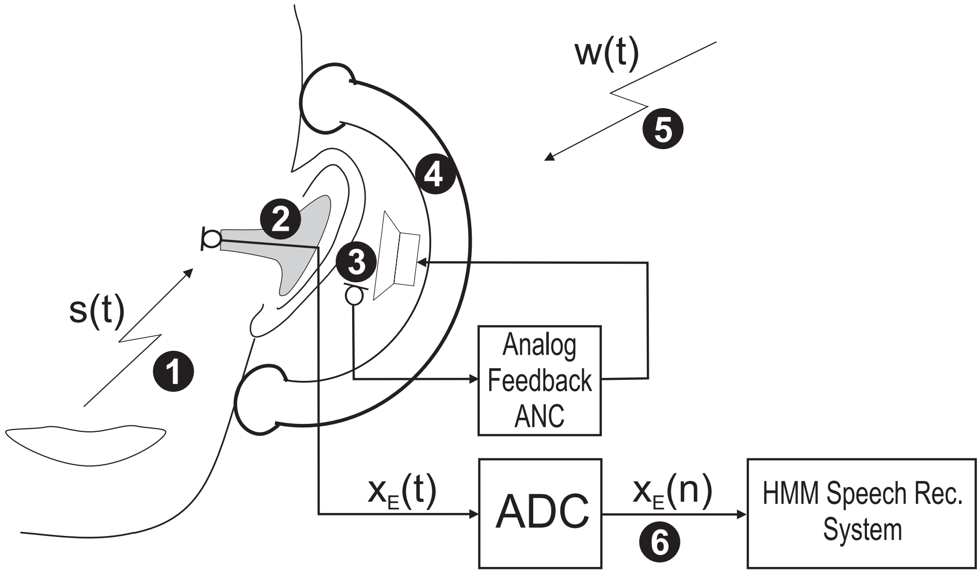


Figure 4.1: Measurement setup and signal paths. (1) Bone conducted speech signal. (2) Ear-mic inserted in custom-made acrylic earplug. (3) ANC reference microphone and loudspeaker. (4) Passive ear-muff. (5) Surrounding noise. (6) Digital output speech signal from ear-mic.

recognition subject, see [25]. For more information on the HTK, see [26].

4.2 Data Acquisition and Method

The basic measurement setup described in section 2.3 was used: The ear-mic described in section 2.2 was inserted into the EAC of the test person and the ANC equipped ear-muffs were then fitted onto the head. The final digital output signal $x_E(n)$ with a bandwidth equal to 4 kHz was used as input to the HMM speech recognition system. See Fig.4.1.

The training phase of the speech recognition system was performed with no surrounding noise, i.e. $w(t) = 0$. For evaluation purposes, two training sets were recorded: One with the ANC disabled and one with the ANC enabled. A total of five different words repeated 20 times each by a male speaker, were used in the training phase.

In order to test the speech recognition capabilities in a noisy environment, recordings were made in a car travelling at a speed of 90 km/h. The five words used in the training phase were pronounced repeatedly and randomly a total of 30 times with the ANC disabled as well as enabled. The same male speaker was used both in the training phase and the test phase. As a demonstration of the impact of the ANC on the noisy signal recorded inside the EAC, a time plot and a spectrogram of such a signal are shown in Fig. 4.2.

It should also be mentioned, that the speech recognition system used in this paper is highly sensitive to surrounding noise. Recordings made with the microphone placed in a more orthodox way, e.g. placed as a hands-free microphone, result in very poor Signal-to-Noise Ratio (SNR) and hence a speech recognition rate that is virtually zero and of no practical use or interest.

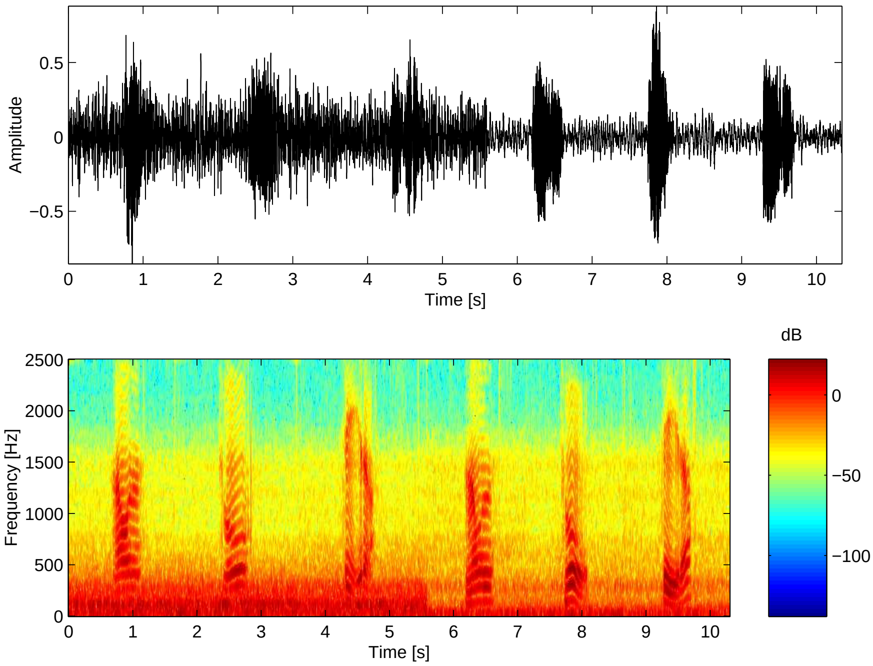


Figure 4.2: Microphone signal recorded inside the EAC. (Upper plot) Time plot of sequence where the ANC first was disabled and then enabled after approximately 5.5 s. (Lower plot) Spectrogram of sequence from upper plot.

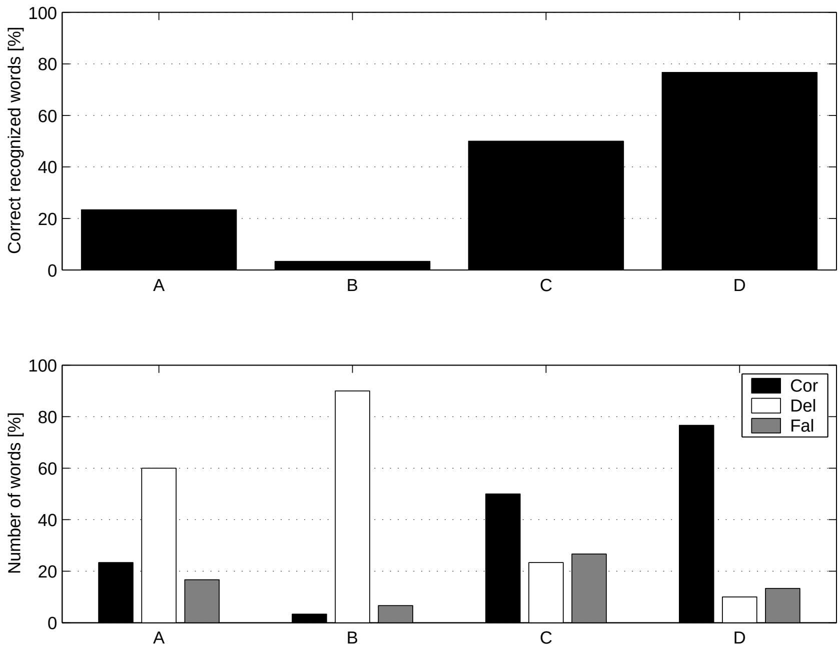


Figure 4.3: Speech recognition results. (A) Both training and testing performed with ANC disabled. (B) Training performed with ANC enabled, testing performed with ANC disabled. (C) Training performed with ANC disabled, testing performed with ANC enabled. (D) Both training and testing performed with ANC enabled. Lower plot legend: Cor — Number of correct recognized words. Del — Number of words that existed in input but never was detected. Fal — Number of words falsely recognized.

4.3 Results

The results are plotted in Fig. 4.3 and tabulated in Tab. 4.1. It is logical to train and test using the same prerequisites, i.e. if speech recognition will be performed with ANC enabled, the training should also be performed with ANC enabled and vice versa. With this in mind, bar A and bar D in Fig. 4.3 are the most interesting ones. When employing ANC, the recognition rate increases by more than 50 percentage units. Apparently, the ANC has a considerable effect on the speech recognition capabilities in a noisy environment. Hence, from the view of the HMM speech recognizer, the speech quality is substantially improved.

	Trained with ANC off						**Trained with ANC on**					
	Cor		Del		Fal		Cor		Del		Fal	
Tested, ANC off	*23.3%*	7	*60.0%*	18	*16.7%*	5	*3.3%*	1	*90.0%*	27	*6.7%*	2
Tested, ANC on	*50.0%*	15	*23.3%*	7	*26.7%*	8	*76.7%*	23	*10.0%*	3	*13.3%*	4

Cor = Number of correct recognized words.
Del = Number of words that existed in input but never was detected.
Fal = Number of words that was falsely detected.
All figures are given both in percent (*italic font*) and absolute number.

Table 4.1: Speech recognition results using a total of 30 words as input in each test scenario. The training phase words were recorded in a silent environment, and the testing phase words were recorded in a car travelling at a speed of 90 km/h.

4.4 Conclusions

A speech recognition system based on a Hidden Markov Model is highly sensitive to noisy input signals. The presence of noise will severely degrade the performance of the system. However, when using an ear-mic combined with a pair of ANC equipped ear-muffs in a noisy environment, the number of correct recognized words increases noticeably. Not only does the method increase speech recognition reliability but this result can also be interpreted as an increase in speech quality and intelligibility.

Chapter 5

Ear-Mic Hybrid Noise Reduction

This section presents an approach for performing personal communication noise reduction in severely disturbed environments. The noise reduction is achieved by using three different noise reduction methods: High frequencies are attenuated by passive absorbers, low frequency components are attenuated by employing active noise control and finally, a broadband noise reduction is achieved by using spectral subtraction. In addition to this, the user's hearing is protected from harmful noise levels by the ANC ear-muffs.

5.1 The Noise Reduction Methods

Passive and active noise reduction in ear-muffs were described in chapter 2. These two noise reduction methods will now be complemented by a broadband spectral subtraction [16, 11]. This frequency domain method is based on the Fast Fourier Transform (FFT) and is a non-linear, yet straightforward way of reducing unwanted broadband noise acoustically added to a signal. The noise bias is estimated in frequency domain during speech pauses, and then subtracted from the noisy speech spectra. The quality of the noise bias estimate is crucial for the final result and a Voice Activity Detector (VAD) is needed to detect non-speech activity. Properly tweaked spectral subtraction is a powerful tool for noise reduction in acoustic signals, and a number of enhancements of the original algorithm have been proposed previously to reduce delay and so-called musical tones [17, 18, 19]. Efforts have also been made to reduce or eliminate the dependence on VADs [27, 28].

Spectral subtraction is based on some assumption regarding signal characteristics: The noise is assumed to be additive and uncorrelated to the speech signal. A slow varying noise environment is acceptable as long as there is enough time to calculate a new estimate of the noise magnitude spectra and to apply this estimate before the noise characteristics have changed. Furthermore, the speech is assumed to be short-time stationary.

Often the conventional spectral subtraction equation is written as

$$\hat{S}_N(f) = H_N(f)X_N(f) \tag{5.1}$$

where $\hat{S}_N(f)$ is the N-point estimate of the clean speech magnitude spectra, $X_N(f)$ is the N-point noisy speech magnitude spectra and

$$H_N(f) = \left[1 - k \cdot \frac{|W_N(f)|^a}{|X_N(f)|^a}\right]^{\frac{1}{a}} \tag{5.2}$$

where $W_N(f)$ is the N-point estimate of the background noise magnitude spectra calculated during non-speech activity. The function $H_N(f)$ is denoted the weighting function. The parameter a decides whether to use a power spectral subtraction ($a = 2$) or a magnitude spectral subtraction ($a = 1$ or some other preferred value). The parameter k adjusts the noise reduction. A larger k reduces the noise level more than a smaller k does, but the resulting speech will be more distorted.

5.2 Data Acquisition and Method

The basic measurement setup described in section 2.3 was used, see Fig. 5.1. Further noise reduction is obtained by applying the spectral subtraction weighting function, $H_N(f)$, to the output signal $x_E(n)$, forming the final enhanced signal $x_{E,ss}(n)$.

The hybrid noise reduction method is aimed at extremely noisy environments e.g for usage during helicopter flights. Hence, in this evaluation two different noise environment were used: Helicopter rotor blade noise (exterior noise) and noise from a helicopter cockpit (interior noise). The interior and exterior helicopter noise are dominated by tonal components in the lower frequency range ($<$ 500 Hz).

The effect of the ANC in the ear-muffs has been evaluated in chapter 3. Still, some low frequency background noise is present in the speech signal inside the EAC. The spectral subtraction algorithm reduces this noise significantly and in addition, the two systems, i.e. the ANC and the spectral subtraction, overlap in the frequency domain. The spectral subtraction algorithm described in section 5.1 was used with $a = 1$ and $k = 2$. The length of the FFTs were 128 and a hanning window was used to reduce the effects due

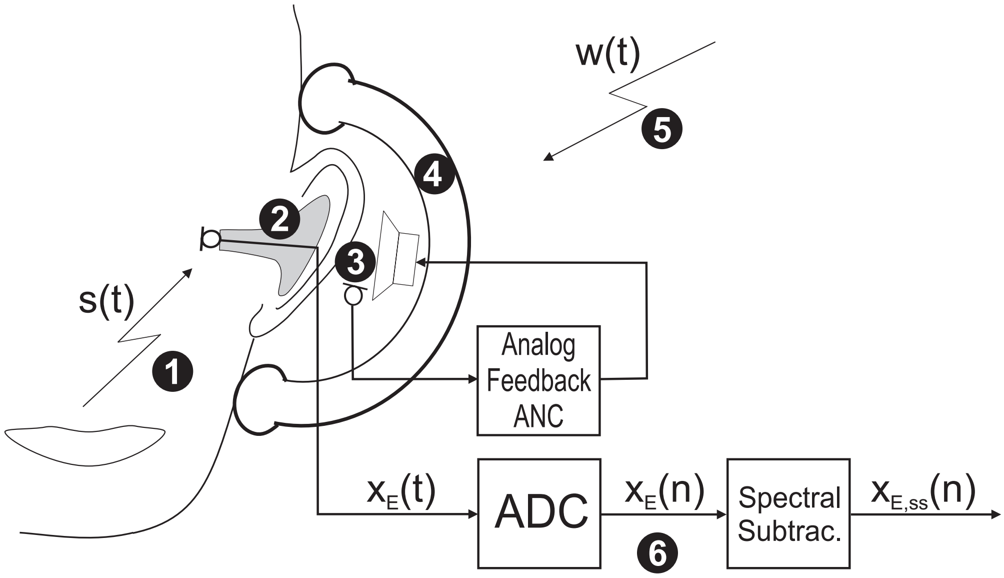


Figure 5.1: Measurement setup and signal paths. (1) Bone conducted speech signal. (2) Ear-mic inserted in custom-made acrylic earplug. (3) ANC reference microphone and loudspeaker. (4) Passive ear-muff. (5) Surrounding noise. (6) Digital output speech signal processed by spectral subtraction algorithm.

to framing of the signal. Furthermore, the weighting function $H_N(f)$ was exponentially averaged to mitigate the effects of spectral subtraction musical tones.

5.3 Results

A time plot of the speech signal inside the EAC with interior helicopter noise is shown in Fig. 5.2. The noise suppressing effect due to the spectral subtraction is obvious and a informal subjective listening test reveals that the speech distortion caused by the noise reduction is acceptable. The SNR increase due to the usage of ANC ear-muffs is illustrated in Fig. 5.3.

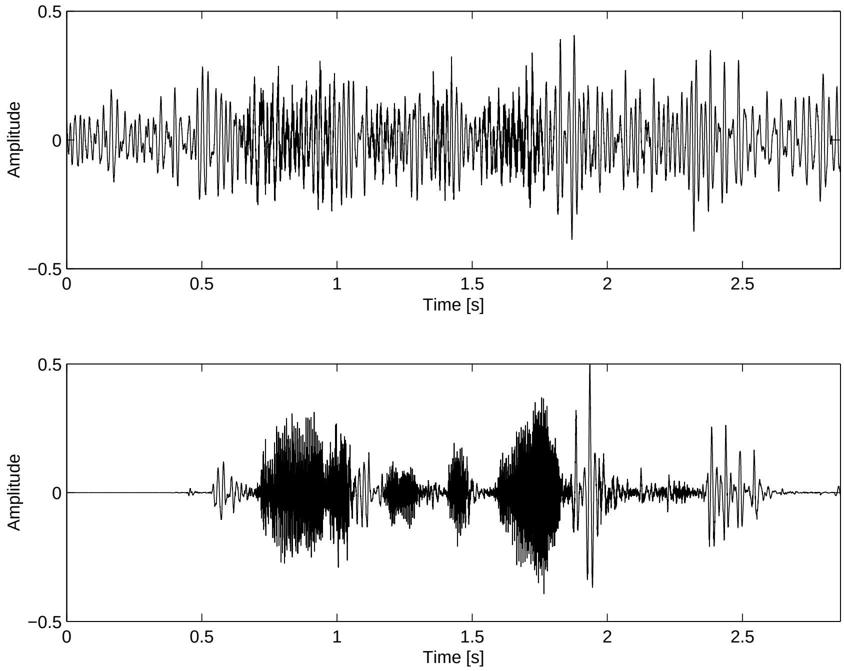


Figure 5.2: (Upper plot) In-ear speech signal before spectral subtraction. (Lower plot) In-ear speech signal after spectral subtraction.

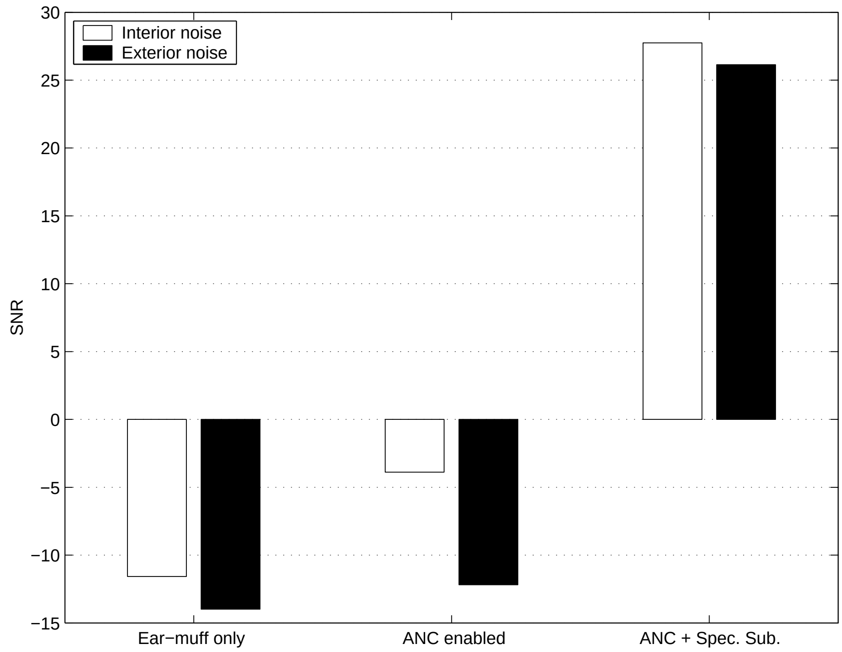


Figure 5.3: The SNR improvement for both interior and exterior helicopter noise.

5.4 Conclusions

In a severely disturbed environment, with high surrounding noise sound pressure levels, an ear-mic combined with a pair of ear-muffs equipped with ANC is superior to a conventionally placed microphone in front of the mouth. Not only is the microphone protected from the surrounding noise both by passive absorbers and by the ANC, but also protected against mechanical damage.

Furthermore, a spectral subtraction algorithm, even in its simplest form, significantly reduces the remaining background noise level.

Altogether, the hybrid noise reduction system that combines an ear-mic, passive and active noise reduction and a spectral subtraction algorithm, offers several advantages in an environment severely disturbed by noise and the method is widely applicable in numerous situations.

Bibliography

[1] M. Winberg, S. Johansson, T. Lagö, and I. Claesson, "A new passive/active hybrid headset for a helicopter application," *International Journal of Acoustics and Vibration*, vol. 4, no. 2, pp. 51–58, 1999.

[2] D. W. Martin, "Magnetic throat microphone of high sensitivity," *Journal of the Acoustical Society of America*, vol. 19, no. 1, pp. 43–50, 1947.

[3] H. M. Moser and H. J. Oyer, "Relative intesities of sounds at various anatomical locations of the head and neck during phonation of the vowels," *Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 275–277, 1958.

[4] R. D. Black, "Ear-insert microphone," *Journal of the Acoustical Society of America*, vol. 29, no. 2, pp. 260–264, 1957.

[5] H. J. Oyer, "Relative intelligibility of speech recorded simultaneously at the ear and mouth," *Journal of the Acoustical Society of America*, vol. 27, no. 6, pp. 1207–1212, 1955.

[6] H. Ono, "Improvement and evaluation of the vibration pick-up-type ear microphone and two-way communication device," *Journal of the Acoustical Society of America*, vol. 62, no. 3, pp. 760–768, September 1975.

[7] S. Aoki, K. Mitsuhashi, and Y. Nishino, "Super-compact microphone/receiver unit for noisy environments," *Journal of the Acoustical Society of Japan*, vol. 20, no. 5, pp. 381–383, 1999.

[8] S. Aoki, K. Mitsuhashi, Y. Nishino, and T. Sakurai, "Noise-suppressing compact microphone/receiver unit," *NTT-Review*, vol. 10, no. 6, pp. 102–108, 1998.

[9] B. Rafaely and M. Furst, "Noise-suppressing compact microphone/ receiver unit," *IEEE Trans. Speech and Audio Proc.*, vol. 4, no. 3, pp. 224–230, 1996.

[10] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, New Jersey, 1993.

[11] J. R. Deller Jr., J. G. Proakis, and J. H. L. Hansen, *Discrete time processing of speech signals*, Macmillan Publishing Company, 1993.

[12] E. A. G. Shaw and G. J. Thiessen, "Acoustics of circumaural earphones," *Journal of the Acoustical Society of America*, vol. 34, no. 9, pp. 1233–1246, 1962.

[13] B. Rafaely, "Active noise reducing headset — an overview," in *Proc. of Internoise*, August 2001.

[14] S. M. Kuo and D. R. Morgan, *Active noise control systems*, John Wiley & Sons, Inc., New York, 1st edition, 1996.

[15] B. Rafaely, J. Garcia-Bonito, and S. J. Elliot, "Feedback control of sound in headrests," in *Proc. of Active '97*, 1997, pp. 445–456.

[16] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust. Speech and Sig. Proc.*, vol. ASSP-27, pp. 113–120, April 1979.

[17] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust. Speech and Sig. Proc.*, vol. ASSP-32, pp. 1109–1121, December 1984.

[18] O. Cappé, "Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor," *IEEE Trans. Speech and Audio Proc.*, vol. 2, no. 2, pp. 345–349, 1994.

[19] H. Gustafsson, S. Nordholm, and I. Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," *IEEE Trans. Speech and Audio Proc.*, vol. 9, no. 8, 2001.

[20] P. A. Hellström and A. Axelsson, "Miniature microphone probe tube measurements in the external auditory canal," *Journal of the Acoustical Society of America*, vol. 93, no. 2, pp. 907–919, 1993.

[21] B. Håkansson, A. Tjellström, and U. Rosenhall, "Hearing thresholds with direct bone conduction versus conventional bone conduction," *Scand. Audiol.*, , no. 13, pp. 3–13, April 1984.

[22] B. Håkansson, P. Carlsson, A. Brandt, and S. Stenfelt, "Linearity of sound transmission through the human skull in vivo," *J. Acoust. Soc. Am.*, vol. 99, no. 4, pp. 2239–2243, April 1996.

[23] B. Håkansson, A. Brandt, P. Carlsson, and A. Tjellström, "Resonance frequencies of the human skull in vivo," *J. Acoust. Soc. Am.*, vol. 95, no. 3, pp. 1474ñ1481, March 1994.

[24] T. L. Lagö, "Frequency analysis of helicopter sound in the AS323 super puma," Research Report ISRN HKR-RES-96/8-SE, Blekinge Institute of Technology, 1996.

[25] L. Rabiner and B.H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, 1993, ISBN 0-13-015157-2.

[26] Webpage, "HTK speech recognition toolkit," `http://htk.eng.cam.ac.uk/`, Last visited on August 18, 2005.

[27] R. Martin, "Spectral subtraction based on minimum statistics," in *SIGNAL PROCESSING VII: Theories and Applications*, M. Holt, C. Cowan, P. Grant, and W. Sandham, Eds., pp. 1182–1185. European Association for Signal Processing, 1994.

[28] H. G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," in *Proc. IEEE. Int. Conf. Acoustics, Speech, Signal Processing*, 1995, vol. 1, pp. 153–156.

IN-EAR MICROPHONE TECHNIQUES FOR SEVERE NOISE SITUATIONS

Nils Westerlund, Mattias Dahl, Ingvar Claesson

Printed by Kaserntryckeriet AB, Karlskrona 2005.

ISSN 1101-1581
ISRN BTH-RES–12/05–SE