Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

A comparison of two active-speaker-detection methods suitable for usage in noise dosimeter measurements

Fredric Lindstrom[a], Keni Ren[b], Kerstin Persson Waye[a] and Haibo Li[b]

[a]Dept. of Environ. Medicine, The Sahlgrenska Acad. of Gothenburg Univ., Box 414, 405 30 Gothenburg, Sweden
[b]Umea University, Dept. of Applied Physics and Electronics, Teknikhuset, Box 414, 901 87 Umea, Sweden
kerstin.persson-waye@amm.gu.se

Measuring noise exposure in a working environment is often done by using standard noise dosimeters. This method is suitable for the evaluation of many working environments. However, in some situations the worker uses his/her voice a large amount during the day, e.g. teachers in a pre-school environment. Thus, in these situations regular dosimeter measurements will not correspond to the actual noise exposure, since they will include a large amount of sound originating from the workers own voice. In order to provide correct measurements, methods that can detect when the worker's own voice is active are required. This paper presents a study of two such methods; the binaural and the throat microphone method. The results of the study suggest that the throat microphone method is more suitable for challenging environments, e.g. pre-schools.

1 Introduction

Several studies have shown that children and personnel at schools and pre-schools are exposed to a high level of noise and that the exposure time may be many hours per day [1-4]. High noise levels have been found to interfere with children's reading comprehension, recognition memory [5], study performance [6], well-being, and coping strategies [7, 8]. Children are also considered to be at risk regarding damage to the hearing caused by noise exposure [9, 10]. Frequently reported symptoms among personnel are "tired ears", tiredness and distress [3]. When evaluating noise exposure using a regular wearable dosimeter, the dosimeter will register all sounds reaching the microphone including the wearer's own speech. In environments with a moderate noise level and where people use their voice extensively during the working day, e.g. in a pre-school environment, the wearer's own voice will give a significant contribution to the measured dose. In order to achieve reliable and more specific measurements of the actual noise exposure, it is necessary to develop methods, which can determine when the person wearing the dosimeter is speaking. Methods to detect own speech activity in noisy environments suitable for wearable measurement equipment have been developed in the field of voice research. The two most well used methods are based on the use of binaural recording [11, 12] and throat microphone recording [13, 14]. We denote these methods the binaural method and the throat method, respectively. In this paper, we compare the active speech detection performance of these two methods.

2 Methods

2.1 The binaural method

In the binaural method two regular omnidirectional microphones are placed on both sides of the subject's head, close to the ears, at equal distance from the mouth. The two signals are recorded using two separate channels, i.e. a left and a right channel. In a signal processing procedure two new signals are created by adding the left and the right channel and by subtracting the right channel signal from the left, then the ratio between the sum and the difference is calculated. This ratio is denoted the self-to-other ratio (S/O). Thanks to the equal distance of the two microphones to the mouth, sound originating from the subjects mouth will result in a high S/O, while broad spectrum sound originating from locations where the distance to the two microphones are not equal will result in a low S/O. Thus, when high levels of S/O are detected it is assumed that the user is speaking and during these periods noise evaluation should be omitted. Note, however, that noise originating from a location right in front of the user, will fulfil the equidistant condition and will thus also result in a high S/O level. This is an intrinsic drawback with the binaural method.

The above only gives a summarized presentation of the processing in the binaural method. For a more elaborate description see e.g. [12].

2.2 The throat method

The throat method also uses a two channel structure. One channel is used to record sound using a regular omnidirectional microphone. The other channel is used to register skin vibrations at the neck, i.e. using an accelerometer as a throat microphone. Since the accelerometer only registers vibrations at the skin, it will only detect sounds originating from the user, i.e. it is assumed that external noise will not cause the skin at the throat to vibrate. Thus, whenever a significant energy level is detected at the throat microphone channel it is assumed that the user is speaking. A more detailed description of the throat microphone evaluation can be found in e.g. [13, 14].

3 Test procedure and evaluation method

For both methods the signals were recorded using a wearable M-Audio Microtrack 2-channel recorder. The same type of microphone, Panasonic WM-61A, was used for both methods, and in the throat method a Knowles Acoustic BU-27135 accelerometer was used. The testing procedures were performed in a soundproof studio and in a pre-school environment. In the pre-school study three female pre-school teachers wore both the binaural microphone system and the throat microphone system during a work day. The analyzed sessions comprised durations of 180 seconds. The sessions contained noise sources, in the form of children's talk and sound generated from the children play as well as sections of the teacher's voice. In order to examine the performance of the methods in a controlled environment some laboratory studies were performed. In the laboratory study a single test person first pronounced a standard sentence in a quite environment. Then, a second recording was done where the test subject stayed in the same position, but kept quiet. During this second recording noise were emitted into the room using a loudspeaker. The signals from the first quite session were then added together with the noise signals in post processing. This allowed a calculation of a ground truth using only the first recording. The data processing parts of the evaluated methods were implemented in MATLAB^{TM}. Both methods used the noisy signals as input and the output of the methods could then be compared to the ground truth. In the pre-school environment recordings, this procedure could of course not be used. In these situations the ground truth was instead achieved by listening through the recorded data and sections where the participant's voice was active was marked manually. The methods were evaluated using Probability-of-Detection (Pd) and Probability-of-False-Alarm (Pf) measures, i.e. the probability that the method correctly detects a situation of active speech and the probability that the method detects active speech when in fact the subject is silent. In the evaluation, the parameters of the binaural and the throat methods were tuned so that the probability of detection was 99% or more, (i.e. Pd>0.99), in all situations. Then, the performance of the method could be evaluated by studying the corresponding value for the probability of a false alarm, i.e. the methods are tuned so that they both fulfil Pd>0.99 for a certain noise source, the better method is then given as the one who has the lowest value of Pf.

In a first laboratory test, pink noise was used as a noise source. The intensity of the pink noise signal was amplified so that the signal to noise ratio (SNR) was 20dB. The position of the loudspeaker emitting the noise was varied using two different positions: right in front of the subject at a distance of about 1.5m from the test subjects head, a position denoted "front" and from a position at a 90 degree angle, a position denoted "side", see figure 3.

In a second test, speech was used as noise signal. In this test the loudspeaker was kept in the side position. The intensity of the speech signal was varied to three different levels so the signal to noise ratio was >40dB, 20, and 10 dB respectively.

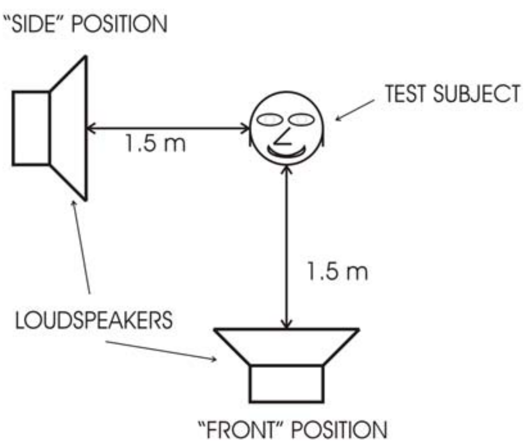

Figure 1. Loudspeaker positions in laboratory tests

4 Result

4.1 Comparing the binaural and throat methods using pink noise

With the binaural method, the Pf was 3%, when the noise came from the front direction and Pf was 1% when the noise came from the side direction, see table 1. This is in accordance with earlier presentations of the binaural method and the assumption that a noise coming from a front direction yields a higher S/O due to the equal distances from the noise source to the two microphones. For the throat method the Pf was less than 0.1% independent of the noise source location, i.e. as expected the external noise did not affect the accelerometer signal.

BINAURAL METHOD		
	Front	Side
Pf	0.031	0.011
Pd	0.994	0.994
THROAT METHOD		
	Front	Side
Pf	<0.001	<0.001
Pd	1	1

Table1. Pf and Pd when using pink noise from the front and side direction. (SNR 20 dB)

4.2 Comparing the methods using speech noise

The result of the comparison of the methods using speech noise is summarized in Table 2. For the >40dB SNR condition the binaural method performed quite well with a Pf as low as 0.2%. However, when the noise levels were increased the performance got worse, Pf increased to about 6% for SNR 20dB and to about 48% for SNR 10dB. For the throat method the Pf kept at a low level <0.1% for conditions all SNR conditions. This test shows that the binaural method might not be suitable for low SNR conditions.

Pf		
SNR(dB)	**BINAURAL**	**THROAT**
>40	0.002	<0.001
20	0.064	<0.001
10	0.487	<0.001
Pd		
SNR(dB)	**BINAURAL**	**THROAT**

>40	1	1
20	0.994	1
10	0.994	1

Table2. Pf and Pd using speech noise from the side direction under different SNR levels

4.3 Pre-school environment recorded data

Table 3 shows the results from the evaluation of the two methods in a pre-school environment. The probability of false alarm was 0.3% for the throat method and 8% for the binaural method.

BINAURAL	
Pf	0.082
Pd	0.991
THROAT	
Pf	0.003
Pd	1

Table 3. Result of the pre-school evaluation

5 Discussion

In the laboratory tests were speech was used as a noise source and the SNR was 10-20dB there was a large difference between the methods, where the throat method had the superior performance. Also for pink noise the performance was better for the throat method and as expected the difference was larger if the disturbing noise was coming from the "front" direction.

The evaluation in the pre-school environment showed a superior performance for the throat method. In a realistic environment, such as a pre-school, one can expect that low SNR values might occur quite frequently. For example if a child is screaming in a close proximity of the teacher or in situations where the background noise level is high, the SNR can be very low, i.e. less than 0dB. A situation where the disturbing signal is coming from a front position is also likely to occur frequently in other types of workplaces, for example, when two persons are standing face to face and talking to each other.

Precautions are necessary in drawing firm conclusions from the findings, due to the low number of participants. However, it can be concluded that there are situations where the throat method outperforms the binaural method. The laboratory studies verify the intrinsic problem in the binaural method with sounds coming from a front direction. However, the laboratory study also suggests that under beneficial conditions the binaural method might perform quite well.

The throat method requires an accelerometer to be firmly attached to the subject's neck. This might imply protocol problems if the subjects are reluctant to wear this type of sensor. For the binaural method the microphones can be attached to glasses or other devises attached to the ears not requiring band-aid or glue for fixation.

6 Conclusion

In this paper, two different methods for detecting speech activity in various noise environments were evaluated. The comparison showed that both methods can be suitable when working in environments where the noise levels are low. For more challenging environments, such as the pre-school environment evaluated in this study, the results from the comparison suggests that the throat method is more suitable.

Acknowledgment

The authors thank the Swedish Research Council (VR), the Swedish Research Council (FORMAS), and Swedish Council for Working Life and Social Research (FAS) for the funding.

References

[1] E. Sala, E. Airo, R. Johansson, P. Olkinuora, J. Pentti, "Noise exposure and NIHL in day care center teachers", *Proc. of the 8th International Congress on Noise and Public Health Problem*, Rotterdam, Netherlands, (2003)

[2] A. Bertilsson, A.-C. Hagaeus, Y. Sandqvist, K. Skagelin, M. Björkman, L. Barregård, *Report on sound level measurements in Schools and pre-schools in Lidköping och Skara kommun 2000-2003* (in Swedish). Lidköpings kommun, Skaras kommun och Västra Götalandsregionens Miljömedicinska centrum, (2003)

[3] U. Landström, B. Nordström, A. Stenudd, L. Åström "The importance of the number of children for the noise level and pre-school personnel experience". (in Swedish). *Arbetslivsrapport* 6, 1-43 (2003) ISSN 1401-2928.

[4] "Noise and indoor environment". *Noise and indoor environmental factors in Danish child care centers (in Danish)* BUPL, forbundet for paedagoger og klubfolk, (1998)

[5] S. Stansfeld, B. Berglund, C. Clark, I. Lopez Barrio, E. Fisher, E. Öhrström, M. M. Haines, J. Head, S. Hygge, I. Van Kamp, B. F. Berry, "Aircraft and road traffic noise and children's cognition and health: a cross-national study", *Lancet* 365, 1942-1949 (2005)

[6] M. M. Haines, S. A. Brentnall, S. A. Stansfeld, E. Klineberg, "Qualitative responses of children to environmental noise" *Noise Health* 5, 19-30 (2003)

[7] M. M. Haines, S. A. Stansfeld, R. F. Job, B. Berglund, J. Head, "A follow-up study of effects of chronic aircraft noise exposure on child stress responses and cognition" *Int J Epidemiol* 30, 839-845 (2001)

[8] P. Lercher, G. W. Evans, M. Meis, W.W. Kofler, "Ambient neighbourhood noise and children's mental health" *Occupational and Environmental Medicine* 59, 380-386 (2002)

[9] W. Passchier-Vermeer, *Noise and health of children. Leiden: TNO report PG/VGZ/2000.042*, (2000)

[10] G. W. Evans, "Child development and the physical environment", *Ann Rev Psychol.* 57, 423-451 (2006)

[8] M. Södersten, S. Granquist, B. Hammarberg, A. Szabo, "Vocal behavior and vocal loading factors for preschool teachers at work studied with binaural recordings", *Journal of Voice* 16, 356-371 (2002)

[9] S. Granqvist, "The self-to-other ratio applied as a phonation detector for voice accumulation", *Log Phon Vocol* 28, 71-80 (2003)

[10] P. S. Popolo, K. Rogge-Miller, J. G. Svec, I. R. Titze "Technical considerations in the designing of a wearable voice dosimeter" *Proceedings of 144th Meeting of the Acoustical Society of America*, Cancun, Mexico (2002)

[11] P. S. Popolo, J. G. Svec, and I. R. Titze, "Adaptation of a Pocket PC for use as a wearable voice dosimeter". *Journal of Speech Language and Hearing Research* 48, 780-791 (2005)