EFFECT OF LATENCY TIME IN HIGH FREQUENCIES ON SOUND LOCALIZATION

Victoria Evelkin and Israel Cohen
{sevelkin@t2, icohen@ee}.technion.ac.il

Electrical Engineering Department
Technion - Israel Institute of Technology

ABSTRACT

Tactical military headsets allow hearing protection and communication between soldiers using surround technology. In this paper, we investigate the effect of latency time between low and high frequencies on the ability of a person to correctly identify an external sound source direction with tactical military headsets. The latency time between low and high frequencies results from a digital processing unit. Low frequencies from an external sound source are not processed; therefore, they are received in the ear canal before the digitally processed higher frequencies. Two experiments were conducted using non-individualized Head Related Transfer Function and headphones for trained and untrained volunteers. The experiments were done for two types of sources: human speech and white Gaussian noise, and were performed with latency times of zero, 20msec, and 40msec applied to frequencies below 20Hz. The experimental results show that the errors in sound localization accuracy in both experiments are fewer for signals without latency time, compared to processed signals with latency times between low and high frequencies.

1. INTRODUCTION

Due to recent developments in tactical military headsets, the effects of latency times between low and high frequencies on a person's ability to correctly identify the source direction need to be investigated. Tactical military headsets allow communication between soldiers in extremely noisy surroundings. The system supports spatial hearing with headphones. In the headset, the current processing is analog. In order to replace it with a digital processing unit, we have to investigate the effect of latency times between low and high frequencies, since the processed sound has to be delayed before it is transmitted to the ear canal due to digital processing. Low external signal frequencies are not processed; therefore, they are received in the ear canal before the digitally processed higher frequencies.

It is generally accepted that people can only hear frequencies within the range of 20Hz-20KHz. In spite of this, previous studies have shown that most people are, indeed, able to hear frequencies below 20Hz, but hearing acumen becomes less sensitive with decreasing frequencies [1]. The Broadband noise can seriously affect sound localization [2]. It is unclear if infrasound noise can also affect

Figure 1. Tactical in-ear headsets

sound localization. Because people are less sensitive to infrasound than to sound at higher frequencies, it is assumed that the influence of infrasound on an observer's localization of sounds is negligible. We decided to test this assumption in the context of latency times.

A human's ability to detect the direction of a sound source relies on two components: interaural time differences (ITD) and interaural level differences (ILD), where the time differences are due to the differences in the distances that sound has to travel to each of the ears, and level differences are caused by the structure of the pinnae and shadows of shoulders, torso, and head. ITD is dominant in low frequencies below 800Hz, while

ILD is dominant in high frequencies above 1600Hz [3-5]. Also, for variant frequencies for the same angle, ITD and ILD also vary.

Head Related Transfer Function (HRTF) is a well-known way of producing spatial sound in headphones. By placing microphones in the ears of a manikin or within the ear canals of a person, we can record a binaural sound coming from the stimuli. Then, HRTF can be extracted from the measurements for each ear. The convolution of a source signal $x(t)$ with HRTF will give a desired spatial hearing effect of the source in headphones, i.e., $y_{right}(t)$ and $y_{left}(t)$ are transmitted to the right and left headphones, respectively, where

$$y_{right}(t) = \{x * h_{right}\}(t) \qquad (1)$$
$$y_{left}(t) = \{x * h_{left}\}(t)$$

978-1-4673-4681-8/12/$31.00 ©2012 IEEE

In our experiment, we used a non-individualized HRTF [6]. There are studies that show the accuracy of non-individualized HRTF. The rates of errors reported by using a non-individualized HRTF on average error angle in azimuth in low elevation is reported ~23º for non-experienced listeners [6]. In addition, a relatively high percent of front/back confusion is recorded for non-individualized HRTF compared to individualized HRTF. A minimum audible angle for the white noise was observed on average 7º-9º for a non-individualized HRTF [7].
The aim of our investigation is to determine the effects of latency times between high and low frequencies occurring in the tactical military headsets on the ability of a person to correctly identify the source direction. We describe two different experiments that show whether or not there is a relation between the delay applied on low frequencies, and the accuracy of the source localization of a virtual sound. We show that in both of experiments the average errors in sound localization is greater for signals with applied delay, compared to the original signals without any delay.
This paper is organized as follows. In Section 2, we present tactical military headsets with hearing protection and ambient sound hear-through. In Section 3, we describe the experimental methods we used for the experimental section. In section 4, we present the results of two experiments. Finally, in Section 5, we discuss the results of our experiments.

2. TACTICAL MILITARY HEADSETS

A tactical headset is a communications headset worn by members of law enforcement, military, and similar organizations for tactical operations. Besides other advantages, the system supports sound localization. A typical in-ear tactical headset is shown in Figure 1. The in-ear headset consists of a hear-through path by which the processed sound is received in the ear canal, an earplug and a microphone; a control box provides noise reduction of external sound sources. The sound is acquired at the microphones and passes to the control box, where signal processing takes place. From the control box, the signal continues through the hear-through path to the ear canal, while the earplugs significantly reduce the noise that comes from outside in the 500Hz range and above. The system also provides active noise reduction in the 20Hz-500Hz range. The system does not provide protection for lower frequencies.
In analog systems, the time that it takes to process the sound is insignificant. However, switching to a digital processing unit, the time that takes for the signal to propagate from the microphones to the hear-through path, through the control box, will increase, so that the part of the signal below 20Hz will notably precede the part of the signal above 20Hz. Hence, a latency time is inevitable between low frequencies from external sound sources and high frequencies processed in the digital processing unit.

3. EXPEREMENTAL METHODS

A latency time was applied for frequencies above 20 Hz by using the system illustrated in Figure 2, where H(ω) is a transfer function

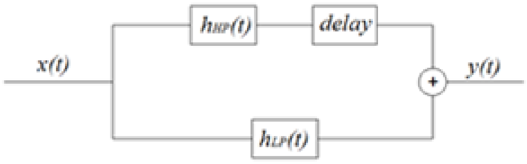

Figure 2. The scheme of the filtering and delaying the source.

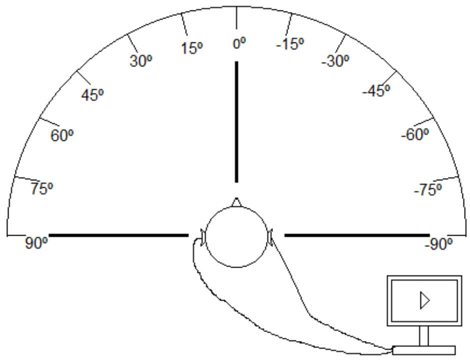

Figure 3. The measurements system.

of an IIR (Infinite Impulse Response) 1st order Chebyshev filter, with a pole in 20Hz.

$$y(t) = \{x * h_{HP}\}(t - \tau) + \{x * h_{LP}\}(t) \qquad (2)$$

$h_{LP}(t)$ is the impulse response of the low-pass filter, and $h_{HP(t)} = \delta(t) - h_{LP(t)}$ is the impulse response of the corresponding high-pass filter. The signal $x(t)$ represents a sound source (speech or white noise). The signal $y(t)$ is processed with an HRFT according to equation (1), and played to the listener as a test sound.

Two different experiments were conducted. In both of them, we used a non-individualized HRTF measured with a KEMAR dummy head microphone by Bill Gardner and Keith Martin at the MIT Media Lab. The errors of this HRTF were previously described [8]. In both experiments we used Technics RP F880 headphones with 5Hz-30KHz support.
The experiments took place in a silent room. A listener was seated on a chair where in front we drew a semicircle with 15º difference angles and lines for 0º, ±90º for better perception of the directions. The listener had to look straightforward. All sounds were played through the headphones. The experimental system is shown in Figure 3.
Two different sources were used – speech and white Gaussian noise. The duration of each signal was 1.875sec. For each type of source, three different latency times were applied: zero delay, 20msec delay, and 40msec delay. Both experiments included the following steps: 1) Reference signals from seven directions, called "reference" directions, and without any latency time between low and high frequencies, were played to the listener. 2) A "test" signal with one of the latency times in low frequencies was played with

three repetitions, where the direction of the test sound was randomly chosen to be one of the seven reference directions. 3) The seven reference signals were played again. 4) The test signal was repeated three times for better perception of the virtual localization. 5) The listener had to determine to which one of the seven reference signals the test signal was most similar.

The identification of the test sound direction relied only on comparing between the direction of the test sound and the direction of the reference sounds, without explicitly identifying the test sound direction on the semicircle. This task was done for each one of four quarters – for azimuth angles from 0º to 30º; from 180º to 150º; from 180º to 210º, and from 360º to 330º – where 0º is in front of the listener. For all of the directions, the elevation was 0º. The front/back angles were chosen due to reported relatively small errors in front [7, 8] and front/back [6] sound localization compared to side angles without consideration of front/back confusions. At the request of the listener, the three repetitions of the test signal and seven reference angles could be repeated once more. There was no feedback given about the correct source direction during the experiment.

At the end of the experiment, the listener had to determine the range of the reference angles' directions and show it on the scale of the half circle (the first and the seventh reference angles) for I and IV quarters.

3.1. Non trained listeners

In the first experiment, 15 non trained volunteers (9 male and 6 female), with a range of ages from 22 to 30 years old, participated. All of them had no history of hearing problems.

3.2. Trained listeners

In the second experiment, 4 volunteers (2 male, 2 female) were trained to determine sound localization. Three of them took part in the first experiment. The training lasted for 3 hours with non-individualized HRTF and sources of both types (speech and white Gaussian noise), with no latency time between low and high frequencies. During the training, the procedure was the same as in the first experiment; only this time, after each listener's answer, feedback about the correct answer was given. After 3 hours of training, the first experiment was conducted.

4. EXPERIMENTAL RESULTS

The collected data from the non-trained listeners about the average degree of error is shown in Table 1. The statistical information of the errors is shown in Table 2.

During the first experiment, 7 of the 15 listeners reported front-back confusions. One of these listeners could not hear the rear angles at all. This result corresponds, in general, to the results shown in [6] for non-individualized HRTF, despite the fact that in our experiment we did not consider each angle by itself, but the whole quarter set (reference sounds, 3 times of the test sounds, and one more reference sounds). Also, all the listeners reported about the error between reference sound directions, as they should have been perceived from 0º to 30º to the range they actually heard. Thus, 7 reported that the range they heard was 0º-90º; 4 reported about a 0º-75º range; two reported a 0º-45º range; one a 0º-110º range; and one a -30º-60º range. In this way, an average error in angle (azimuth) between played sound and listener's perception is calculated to be about 35º.

The collected data from the trained listeners and statistical information is presented in Tables 3 and 4, respectively.

Table 1. The error of the speech and white noise sources with zero, 20msec and 40msec with non-trained volunteers.

	speech			white noise		
	error (º)			error (º)		
	40msec	*20msec*	*no-delay*	*40msec*	*20msec*	*no-delay*
AZ	3.75	3.75	2.5	6.25	5	2.5
BL	2.5	2.5	3.75	3.75	5	6.25
NK	1.25	3.75	1.25	2.5	8.75	5
EZ	2.5	6.25	2.5	8.75	6.25	2.5
KR	5	6.25	1.25	7.5	5	3.75
MR	3.75	5	3.75	0	2.5	2.5
TT	2.5	1.25	1.25	3.75	3.75	2.5
AS	1.25	5	3.75	5	2.5	5
AC	2.5	1.25	3.75	1.25	3.75	0
OK	3.75	3.75	0	3.75	3.75	7.5
ZP	3.75	3.75	3.75	5	2.5	3.75
SK	5	2.5	3.75	3.75	5	3.75
SB	5	1.25	5	2.5	5	1.25
SG	3.75	2.5	2.5	2.5	1.25	3.75
EA	2.5	5	1.25	2.5	1.25	2.5

Table 2. An average error and variance for speech and white noise sources for non-trained volunteers

	speech			white noise		
	40msec	*20msec*	*no-delay*	*40msec*	*20msec*	*no-delay*
average	3.25	3.58	2.7	3.92	4.08	3.5
variance	1.52	2.87	2	5.33	3.9	3.62

5. DISCUSSION AND CONCLUSION

As can be seen from the Tables, with regard to the non-trained listeners for the speech source, 9 of the 15 volunteers had a better performance when the latency time was zero than when it was 20msec or 40msec. Also, 9 out of the 15 volunteers had fewer mistakes in the localization with zero latency time. The average error is smaller for the zero latency time by 17% - 25%. For the white noise signal, 7 out of the 15 volunteers had a better performance with zero latency time than with 20msec or 40msec latency time; 7 out of the 15 volunteers had fewer mistakes in the localization when the latency time was zero, as well. An average error is also smaller for non-delayed white noise by 10% - 15%. There is also an interesting result that for the 20msec latency time the error is greater than that for 40msec delay, for both types of sources.

Table 3. The error and the correct localization of the speech and white noise sources with 20msec and 40msec delay and without a delay for trained volunteers

	speech						white noise					
	correct localization (%)			*error (º)*			*correct localization (%)*			*error (º)*		
	40msec	*20msec*	*no-delay*	*40msec*	*20msec*	*no-delay*	*40msec*	*20msec*	*no-delay*	*40msec*	*20msec*	*no-delay*
LH	50	50	100	3.75	2.5	0	25	25	75	3.75	3.75	1.25
OK	50	75	100	2.5	1.25	0	75	100	75	1.25	0	1.25
TT	50	25	50	2.5	3.75	2.5	50	75	75	2.5	2.5	1.25
SB	25	25	100	3.75	3.75	0	100	50	75	0	2.5	1.25

Table 4. Average of the error in (º) and variance for voice and white noise source for trained volunteers

	speech			white noise		
	40msec	*20msec*	*no-delay*	*40msec*	*20msec*	*no-delay*
average	3.12	2.81	0.62	1.87	2.19	1.25
variance	0.52	1.43	1.56	2.6	2.47	0

Most of the listeners reported that they found it more challenging to identify directions of the white noise source rather than the speech source. We think that it could be a reason for greater rates of error associated with white noise compared with the speech. The results also indicate that training improved the listener's ability to identify sound localization with non-individualized HRTF. Yet, the error in delayed sound is larger than in sound without any delay for trained listeners from the second experiment. Also, the reference angle's range comprehended by trained listeners is closer to the theoretical range we processed with HRTF (0º to 30º). In this way, 2 of 4 listeners reported about 0º to 45º after the second experiment, although before they reported on 0º to 90º range.

In our work, the measurements were based on a comparison of two sounds to reduce the influence of other factors, such as errors of non-individualized HRTF and front/back confusions. Because of the fact that this technique is not similar to regular sound localization measuring systems, the errors are different and relatively small. Here, we investigated the degradation of localization due to latency time, while comparing a test signal to reference signals. In order to understand the meaning of the errors, we asked the listeners to report the range of reference angles they heard.

This study has an important role, since it shows the importance of analyzing the side effects of infrasound frequencies effects during the development of tactical military headset systems. We showed that latency time between high frequencies (the processed signal that is received from the hear-through path) and the low frequencies (infrasound that comes from outside) can unexpectedly affect the accuracy of sound localization when using tactical military headsets.

6. REFERENCES

[1] H. Moller, C. S. Pedersen, "Hearing at low and infrasonic frequencies", Noise Health 2004; 6:37-57 June 2004

[2] M. D. Good, R. H. Gilkey, "Sound localization in noise: the effect signal-to-noise ratio", The Journal of the Acoustic Society of America, 99(2):1108-17, Feb. 1996

[3] J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners", Annual Reviews, Psychology, 42:135-59, 1991

[4] S. S. Stevens, E. B. Newman, "The localization of pure tones", Proceedings of the National Academy of Science USA, 20(11): 593-596, Nov. 1934

[5] E.A. Macpherson, J.C. Midllebrooks, " Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited" , Journal of the Acoustical Society of America, 111(5 Pt 1):2219-36, May 2002

[6] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman, "Localization using nonindividualized head-related transfer functions " Journal of the Acoustical Society of America, 94(1):111-23, July 1993.

[7] G. Wersényi, "Localization in a HRTF-based Minimum-Audible-Angle Listening test for GUIB applications" Electronic Journal "Technical Acoustics",www.ejta.org, Jan. 2007

[8] O. Grafals, N. Gupta, G. Cremades, A. B. Barreto, M. Adjouadi "Evaluation of digital sound specialization accuracy over commodity audio channels in a personal computer" the 1999 Computing Research Conference, University of Puerto Rico - Mayaguez, Dec. 1999