Proceedings of the 13th International Conference on Auditory Display, Montréal, Canada, June 26 - 29, 2007

EFFECT OF LARGE SYSTEM LATENCY OF VIRTUAL AUDITORY DISPLAY ON LISTENER'S HEAD MOVEMENT IN SOUND LOCALIZATION TASK

Yôiti Suzuki, Satoshi Yairi and Yukio Iwaya

Research Institute of Electrical Communication and
Graduate School of Information Sciences, Tohoku University,
Katahira 2–1–1, Aoba-ku, Sendai 980–8577, Japan
{yoh@ais, yairi@fir, iwaya@fir}.riec.tohoku.ac.jp

ABSTRACT

Virtual Auditory Display (VAD) technology is expected to enable the development of new communication tools and many other related applications. However, in computer-network-based communications, large latencies can sometimes occur. Therefore, the influence of large system latency (SL), up to 2 s, on VAD-based sound localization tasks was investigated in terms of the precision and time course of sound localization performance by listeners engaged in head movements. A software VAD system developed by the authors on a Linux PC (with SL of 12 ms) was used in the experiments. Listeners were asked to indicate the location of a virtual sound source by moving their heads in order to face the direction of the perceived sound image. Virtual sound sources were presented to the listeners with one of seven amounts of system latency (12, 50, 100, 200, 500, 1000 and 2000 ms). While the latency detection threshold has been estimated as an SL of about 75 ms, no significant influence on accuracy of sound localization was observed for any of the tested SLs. On the other hand, the time to conclude the sound localization increased as the SL increased. Moreover, a remarkable overshoot was observed in the listener's head movement particularly when SL was greater than 500 ms. This strongly suggests that the tolerable SL caused by network communications should be kept smaller than 500 ms for VAD applications.

[Keywords: Binaural Technology, System Latency,
Head Movements, Head-Tracking]

1. INTRODUCTION

Listeners learn to use HRTFs (Head Related Transfer Functions) to aid in the localization of sound sources in the space surrounding them [1, 2]. In virtual auditory display (VAD), a perceived sound source position can be arbitrarily controlled by convolving sound sources with impulse responses corresponding to HRTFs, *i.e.*, HRIRs (Head–Related Impulse Responses) [3]. To prevent cross-talk between the audio signals to be reproduced at the listener's ears, headphones are often employed in applications using VAD systems. In sound localization, dynamic changes in HRIRs caused by the listener's head movements provide one of the most important localization cues [4], particularly when we make voluntary head movements during active listening. In fact, the accuracy of localization is markedly enhanced by allowing listeners to move their heads freely [5, 6, 7, 8, 9, 10, 11, 12]. That is, it is important in a VAD system to reproduce not only static sound information, but also dynamic variation of sound caused by the listeners' movement. To accomplish this interactive processing, a three-dimensional position sensor is usually employed to obtain information about the listener's head position and movement [7, 13, 14]. Appropriate HRIRs are then set according to the position of sound relative to head direction and position. This means, for example, that when a listener's head is rotated, a change in the processing of the source must simulate a virtual sound rotating through an equal angle in the opposite direction of the head rotation, thereby rendering a spatially stable sound source image, fixed in world coordinates. In applications of VAD systems, there inevitably arises a measurable system latency (SL) between detected listener's movement and the required changes in the control of the sound position. Here, SL is the sum of the durations of a number of events, beginning with the time that source position data is updated. These events include the time it takes to interpolate HRIRs, to convolve the HRIRs with a sound source, and to output the data through buffers and D/A. Moreover, network-based transmission delay will become an additional component in the latency of the overall system, and this can be a critical factor in VAD-based telecommunication environments.

To render the virtual world more realistically by VAD, SL should be as short as possible. Indeed, it is important to design SL by taking into consideration the latency detection threshold (DT), which is a minimum delay time for listeners to notice that the output is delayed in reference to the head movement. If the SL of VAD is much longer than DT, then the delay caused by SL will be easily detected by the listener. As a result, the listener feels that the virtual sound image is not fixed in the virtual world, but moves through the world as if in response to head movements, exhibiting a delay in stopping its motion after the listener's head motion has stopped. This never occurs for an actual sound source at a fixed position in the real world. Therefore, it is important to investigate DT and make SL sufficiently smaller than observed DT. Kimura *et al*. [15] used a paired comparison task in their experiments to measure a difference limen (DL) as an estimate of DT. Their results allowed them to estimate DT to be around 80 ms. Sasaki *et al*. [16] examined DT and DL in two experiments and reported both to be about 50 ms. Recently, Brungart *et al*. [17] reported that in their task, using a VAD with a minimum SL of 11.7 ms, the average listener was unable to reliably detect an SL smaller than about 80 ms. They also pointed out that there were large inter-subject differences in measured DL values. In our recent related research [18], we estimated the average DT for SL as being around 75 ms, again with certain inter-subject differences.

Such a small value of SL is easily realizable these days in DSP, or even when using an ordinary PC. In fact, there are several VAD systems for which the minimum system latency is as low as 10 ms [17, 18, 19, 20]. This means that recent VAD system's SL are

sufficiently smaller than DL. Therefore, the biggest problems with SL for VAD systems in the near future may be the transmission delay in networks. Since VAD technology is expected to enable the development of new communication tools and many other related applications, objectionably large latencies will no doubt occur through communication via computer networks. Therefore, in addition to the study of just detectable SL values, it is also important to study the influences of large SL for VAD systems. The present study investigated two important aspects of VAD performance. To begin with, the precision of sound localization was examined under conditions when the SL could be as great as 2 seconds. The study also investigated behavioral details of the listener's head movements during sound localization tasks for a wide range of SL values.

2. EXPERIMENTS

2.1. VAD system

A software VAD system developed by the authors [18] was used in the experiments. The system consisted of a pair of headphones, a magnetic position sensor, and a personal computer (3.06 GHz Pentium 4 CPU, 2 GByte memory) running the Linux (kernel 2.6) operating system. Electrostatic open–back type headphones were used (STAX SRS-2020, earspeaker: SR-202 and driver unit: SRM-212). A Polhemus FASTRAK system was used as the magnetic position sensor having six degrees of freedom (6DOF: relative x, y and z-position and yaw, $pitch$ and $roll$-angle. In this system, the FASTRAK receiver was mounted on the top of the headband of the STAX headphones to acquire the position data at a rate of 120 samples/s. The minimum SL of this VAD system was about 12 ms, including the latency of the position sensor.

2.2. Method

The experiment was performed in a soundproof room. The group of listeners included three young males and two young females, all with normal hearing. The listener stood during experimental sessions, and was asked to localize a virtual sound source by moving his/her head so as to face the sound image. Stimuli were generated by convolving the listener's own HRIRs with a sound source using the VAD system described in Sec. 2.1. The input sound source was pink noise with a sampling frequency of 48 kHz and 16-bit quantization.

A virtual sound source was presented to each listener with one of seven SL values (12, 50, 100, 200, 500, 1000 and 2000 ms). The synthesized sound source was presented at eight initial azimuth angles ranging from 0 to 315 degrees, spaced at an interval of 45 degrees on the horizontal plane (with 0 degrees indicating frontal incidence, and azimuth value increasing in a clockwise manner). The above-mentioned seven SL values were combined in random sequence with the eight initial directions for the sound images, and were presented just once in each session. Each listener participated in three sessions. Listeners were asked to shut their eyes, to first judge the direction of the sound image, and then to turn their head in the direction of the perceived sound image. When the sound image was positioned directly in front via head rotation (i.e., change in yaw), the listener was instructed to nod his/her head (i.e., a change in pitch). This "nodding" action was the signal to be used by listeners to indicate that they had finished a single sound localization trial. Figure 1 shows an example of the time course for yaw and pitch angles of listener's head (Listener 1, at a sound source direction of 90 degrees, and an SL of 2000 ms). In the figure, a dip, which corresponds to the listener's nodding gesture, is clearly observed in the time course of the pitch angle. Point a of this dip is regarded as listener's decision time of localization, and the corresponding angle of yaw (b point) is then recorded as the angle matching the sound source azimuth.

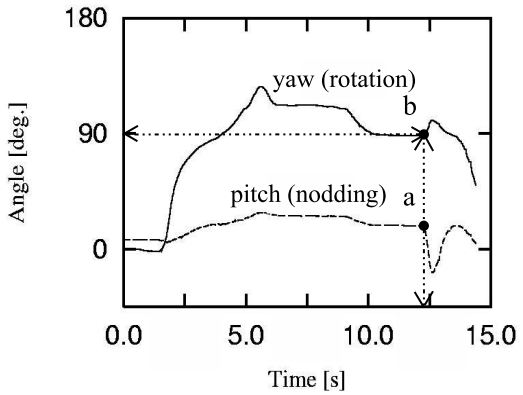

Figure 1: *An example of a listener's head movement (For Listener 1, sound source direction: 90 degrees, SL: 2000 ms).*

3. INFLUENCE OF LARGE SYSTEM LATENCY ON LOCALIZATION

3.1. Influence on localization accuracy

The localization angles averaged over the responses from all listeners are shown in Fig. 2. This figure shows that there is very little influence of SL on the indicated sound source azimuth angles. To examine this, a two-way analysis of variance (ANOVA) was performed. The SL of the VAD and the direction of the sound source were treated as factors, and the listener was treated as a repeated measure. As a result, the main effects, as well as the interaction between the two factors, were not statistically significant. This means that no remarkable influence exists on the localization accuracy provided by the head-tracking VAD system even when there is an SL value much greater than the DL.

3.2. Influence on the time to conclude sound localization

Figure 3 shows the time T required for the sound localization as a function of system latency. The figure plots the averages calculated over all listeners. It is shown that the time T that listeners need to conclude their sound localization increases as the system latency increases. A one-way ANOVA was conducted to examine the effect of changing SL on the time T required for sound localization. Again, the listener was treated as a repeated measure. The main effect of the SL factor was statistically significant ($F(6, 24) = 30.26$, $p < .01$). Tukey's HSD test shows that the time required for the localization under conditions in which SL was 1000 and 2000 ms was significantly longer than that when SL was 12, 50, 100 and 200 ms ($p < .05$). As shown in Fig. 3, the plotted data seem to be well fit by a straight line, with a correlation coefficient of 0.996. Indeed, an equation relating the time T required for sound localization to SL can be formulated as

$$T = 2.02SL + 7.25, \quad (1)$$

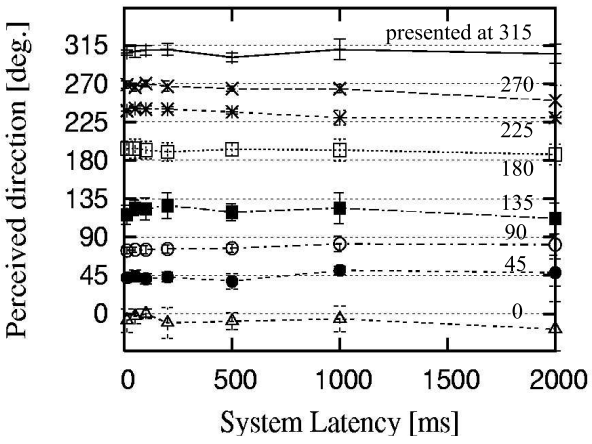

Figure 2: *Average perceived-direction responses as a function of system latency. Error bars show the 95% confidence intervals.*

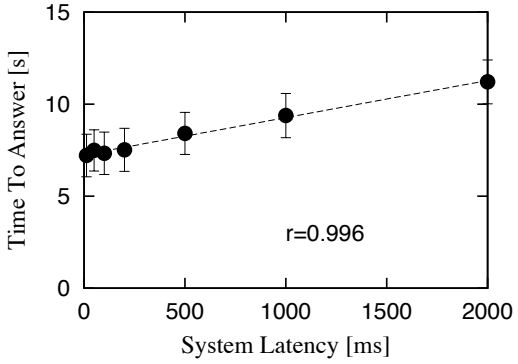

Figure 3: *Averaged time required for listeners to indicate sound source azimuth as a function of system latency. Error bars show the 95% confidence intervals.*

4. HEAD MOVEMENT DURING SOUND LOCALIZATION

4.1. Typical pattern of listener's head movement

The time courses of listener's absolute head rotation and its relative angle to the sound source in aziumth are shown in Fig. 4. This figure shows some typical examples of listener behavior under each system latency condition for the initial sound source direction of 90 degrees. The curves are plotted up until the point in time at which listeners concluded the sound localization. Moreover, as shown in the figure, the relative angle to the sound source generally changes in the opposite direction to listener's head movement with the delay comparable to the system latency. It can also be observed that the features of the head movement change as the system latency increases. When the system latency was long, a remarkable overshoot that exceeds the presented angle was observed in listener's head movement. The amount of the overshoot seems to increase as the system latency increases.

4.2. Analysis of listeners' head movement

To analyze the head movement in detail, the time course of the head movement beginning with the presentation of a virtual sound source, and lasting until the sound localization was concluded, was divided into three temporal sections, s_1, s_2 and s_3, as shown in Figs. 5 and 6. Figures 5 and 6 show the typical time course of the facing angle (azimuth) when no overshoot is observed and when an overshoot is observed, respectively. The time corresponding to the end of each section is called as t_1, t_2 and t_3, hereafter. These three sections are summarized as follows:

1. s_1: From the beginning of the trial until the head begins to move.
2. s_2: From the beginning of the head movement until the head angle approaches the direction of the virtual sound source.
3. s_3: From when the head angle nearly reaches the sound position until the concluding of the sound localization.

In the following subsections, the head movement and the time for localization is analyzed for each of the three temporal sections.

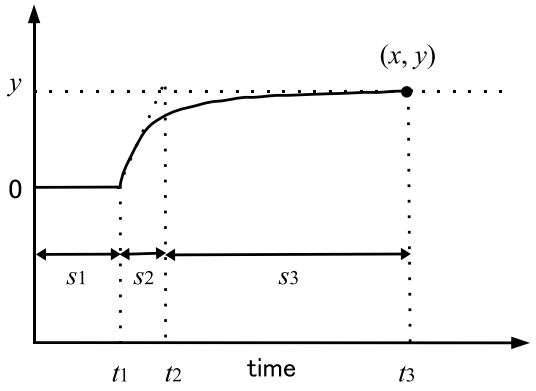

Figure 5: *Diagram showing the division of head movements into three temporal sections (illustrating the case when there is no overshoot).*

4.2.1. Duration of s_1

The boundary point between s_1 and s_2 was determined by visual inspection of the records of the head movement. Since this first temporal section, s_1, covers the time from the beginning of the sound presentation until the time when the head begins to move, it is usual that the head hardly moves during s_1. The duration of s_1 averaged over all five listeners are shown in Table 1 for each system latency and for each sound source direction. A two-way ANOVA was conducted within which system latency and initial sound source direction were treated as factors. In this analysis, the listener was treated as a repeated measure. As a result, no statistically significant main effects or interaction effects were found. This means that in the present task listeners needed about 1.52 s to localize the sound image with a certain confidence, and then afterwards they started to move their heads. This duration appears to be independent of the system latency and the virtual sound source direction.

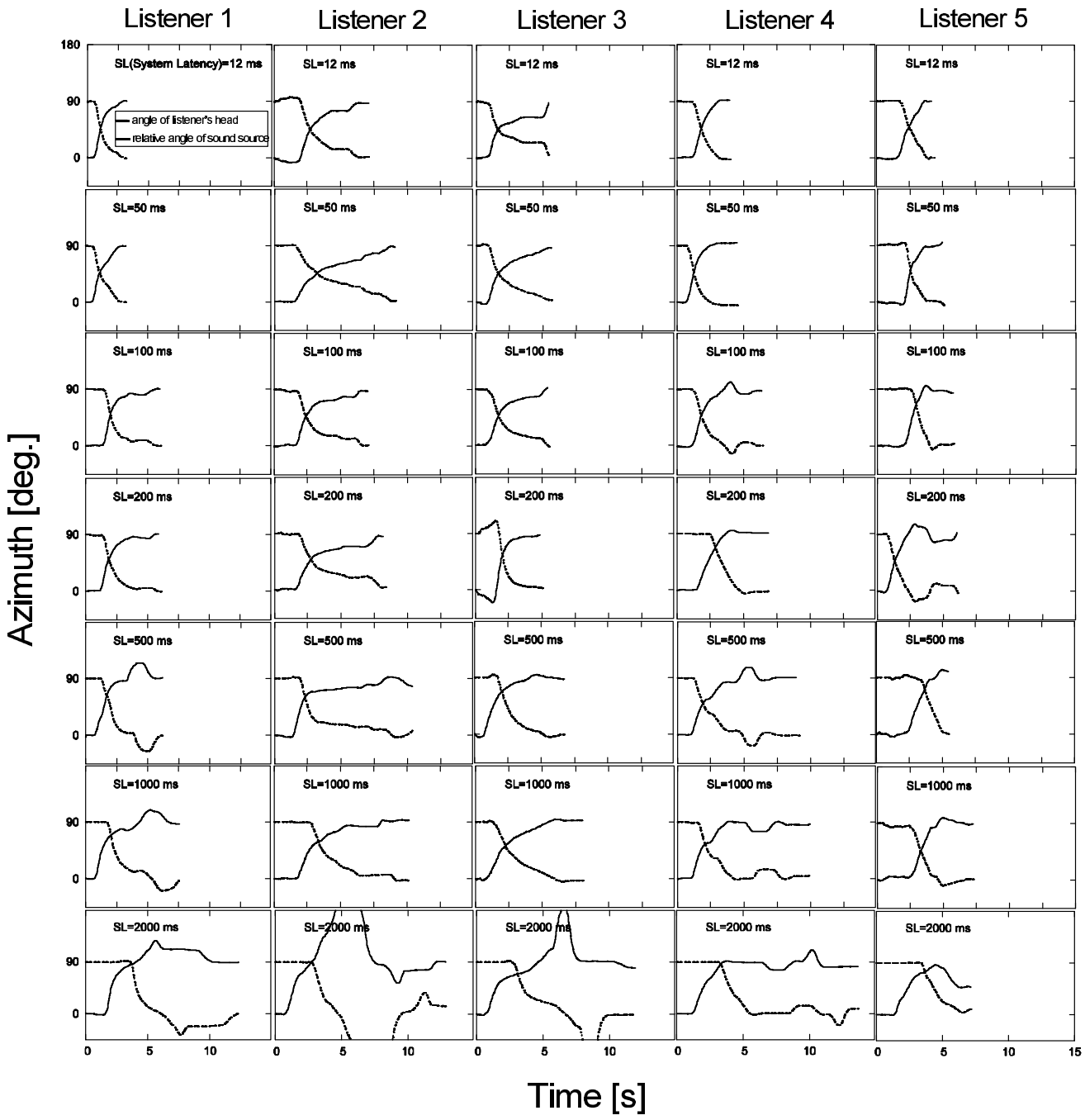

Figure 4: *Example of head movements for all listeners comparing angle of the listener's head and the relative angle of sound source in azimuth (all for an initial sound source direction of 90 degrees).*

Table 1: The averaged duration of s_1 until the head begins to move for five listeners [s].

		System latency [ms]							
		12	50	100	200	500	1000	2000	Average
Sound source direction [deg.]	0	1.49	1.54	1.49	1.66	1.50	1.63	1.64	1.56
	45	1.69	1.31	1.84	1.52	1.57	1.36	1.73	1.57
	90	1.55	1.51	1.36	1.66	1.42	1.60	1.32	1.49
	135	1.36	1.56	1.42	1.91	1.45	1.41	1.42	1.50
	180	1.64	1.48	1.40	1.61	1.70	1.80	1.39	1.57
	225	1.53	1.50	1.35	1.66	1.39	1.42	1.32	1.45
	270	1.43	1.62	1.64	1.46	1.66	1.65	1.26	1.53
	315	1.32	1.55	1.55	1.66	1.50	1.42	1.55	1.51
	Average	1.50	1.51	1.51	1.64	1.52	1.54	1.45	1.52

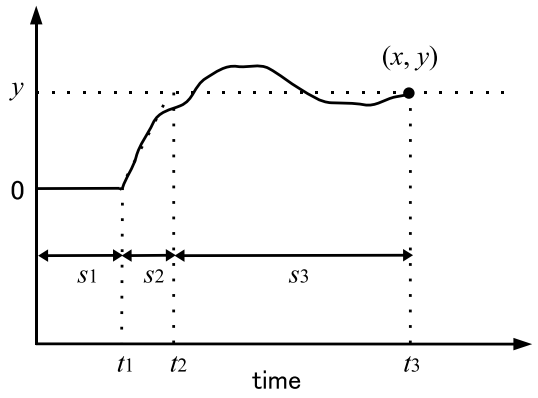

Figure 6: *Section division of head movement (When there is an overshoot).*

4.2.2. Duration of s_2

The boundary from s_2 to s_3 was determined as follows: The typical time course of the facing angle in s_2 seems hardly to depend on whether there was an overshoot or not, and seems to be well fit by a function $\theta(1-e^{-\frac{t}{\tau}})$, where θ is the direction of the virtual sound source, t is time and τ is a time constant. After fitting this function to the observed yaw data, the obtained τ value was used as the duration of s_2, *i.e.*, $(t_2 - t_1)$. The duration of s_2 averaged over five listeners is shown in Table 2 for each system latency and for each sound source direction. A two-way ANOVA, in which system latency and sound source direction were treated as factors, was conducted. In this analysis, the listener was treated as a repeated measure. As a result, no statistically significant main effects or interaction effects were found. Since the duration of s_2 is identical to τ, this result means that the time constant of the head movement in s_2 is independent of the system latency and the sound source direction.

4.2.3. Duration of s_3

The duration of s_3 averaged over listeners is shown in Table 3 for each SL value and for each sound source direction. A two-way ANOVA, in which system latency and sound source direction were treated as factors, was conducted. The listener was treated as a repeated measure. As a result, the main effect of system latency was statistically significant ($F(6, 24) = 30.44$, $p < .01$). The main effect of the direction as well as the interaction was not significant. Tukey's HSD test was then conducted for the multiple comparison. The conditions graphically connected with an underline in Table 3 were not significantly different. The duration of s_3 for the conditions with SL of 1000 ms is significantly longer than that for SL of 12, 50, 100 and 200 ms, and that for the condition of 2000 ms is certainly the longest (significantly so, at $p < .05$). The difference between 500 ms and 1000 ms is not significant. There results show that the duration of s_3 is deeply depend on the system latency.

4.2.4. Amount of overshoot

An overshoot observed in s_3 when system latency is large is one of the most remarkable features of the listener's head movements. In particular, as shown in Fig. 4, the amount of the overshoot seems to increase as the system latency increases. Here, the overshoot is defining as the area within which the listener's head rotation exceeds that needed to bring the presented sound source direction directly in front of the listener. When the sound source direction was 0 degree, the listeners could conclude without rotation, and therefore, the overshoot could not be defined for this sound source direction. The exceeded angles of overshoot for the other directions were averaged over five listeners for each system latency, and these averages are shown in Table 4. If an overshoot was not observed in the trial, the amount of the overshoot was treated as a 0 degree overshoot.

As shown in Table 4, the overshoot grows remarkably when the system latency is greater than 500 ms. A two-way ANOVA in which system latency and sound source direction were treated as factors, was conducted. The listener was treated as a repeated measure. As a result, the main effect of system latency was statistically significant ($F(6, 24) = 39.06$, $p < .01$). The main effect of the direction as well as the interaction was not significant. Tukey's HSD test was conducted for the multiple comparison. The conditions graphically connected with an underline in Table 4 were not significantly different. The overshoot values for the conditions of 1000 ms and 2000 ms were significantly larger than that of 12, 50, 100, 200 and 500 ms ($p < .05$).

Table 2: The averaged duration of s_2 (from initial motion until the head angle nearly reaches the sound source direction) for five listeners [s].

		System latency [ms]							
		12	50	100	200	500	1000	2000	Average
Sound source direction [deg.]	0	–	–	–	–	–	–	–	–
	45	1.19	1.28	1.28	1.11	1.07	1.26	1.48	1.24
	90	1.80	1.70	1.44	1.88	1.26	1.50	1.36	1.56
	135	2.20	1.86	1.80	1.92	2.22	2.00	2.20	2.03
	180	2.28	1.88	2.00	2.08	1.90	2.34	2.32	2.11
	225	1.86	2.26	2.42	1.84	1.84	2.12	2.02	2.05
	270	1.50	1.66	1.42	1.30	1.22	1.40	1.50	1.43
	315	1.38	1.13	1.14	1.26	1.14	1.50	1.04	1.23
	Average	1.74	1.68	1.64	1.63	1.52	1.73	1.70	1.66

Table 3: The average duration of s_3 for five listeners [s].

		System latency [ms]							
		12	50	100	200	500	1000	2000	Average
Sound source direction [deg.]	0	3.76	5.00	4.51	4.91	6.17	7.57	7.96	5.70
	45	3.58	4.78	5.04	4.68	5.85	6.68	7.91	5.50
	90	4.35	4.91	4.70	3.90	5.64	4.82	8.48	5.26
	135	4.02	4.60	4.78	4.35	5.27	6.79	8.40	5.46
	180	4.16	4.18	3.52	4.65	5.22	6.52	8.79	5.29
	225	3.91	3.88	3.73	4.26	4.59	5.84	7.08	4.75
	270	4.53	3.99	3.75	3.62	4.80	5.67	7.98	4.90
	315	4.03	3.92	4.02	4.18	5.48	6.61	8.35	5.23
	Average	4.04	4.40	4.26	4.32	5.38	6.31	8.12	5.26
Result of multiple comparison with Tukey's HSD									–

4.3. Discussion

The above-mentioned analysis shows that the influence of large system latency on head movements appears not in s_1 and s_2, but only in s_3. The average duration of s_1 and s_2 are 1.52 s and 1.66 s, respectively, which are hardly dependent on the SL value or on the sound source direction. In contrast, the averaged duration of s_3 was found to be highly dependent on the SL value, as shown in Fig. 3, and could be well fit by a linear function

$$t_3 = 2.04SL + 4.13. \quad (2)$$

Here, the total time required for the sound localization is

$$T = t_1 + t_2 + t_3 = 1.52 + 1.66 + (2.04SL + 4.13). \quad (3)$$

This, in effect, amounts to $T = 2.04SL + 7.31$. The obtained Eq. 3 well agrees well with Eq. 1. This coincidence shows that the influence of system latency on the duration required for the sound localization task is apparent only within the period termed s_3. The factors of 2.02 in Eq. 1 or 2.04 in Eq. 3 may be regarded as about a factor of 2, which could be explained as follows: Suppose the system latency is a ms, and that change in the relative direction of the sound source always follows the head movement with the delay of a ms (factor 1). Moreover, the sound source will come in front of the listener a ms after the listener has turned to face the perceived position of the sound source (factor 2). As a result of these two factors, the effect of the system latency should make such head rotation last for a duration twice the SL value.

When the system latency is longer than 500 ms, the overshoot becomes remarkable in s_3. In particular, the amount of overshoot when SL is 1000 ms and 2000 ms is about 7 times larger than that for smaller SL values. The origin of the overshoot could be explained as follows: By the end of s_1, listeners were able to perceive the direction of presented virtual sound source. During s_2, listeners pursue the perceived direction with a mechanism modeled by a simple first-order integration circuit. Moreover, in s_2, listeners seem not to re-evaluate the perceived sound source direction. When listeners almost face the perceived direction, if the SL is large, they notice that the virtual sound source is located still further from the facing direction, and therefore decide to turn further in order to reach the sound source direction. This would explain the resulting overshoot. However, at a duration of SL, the listener's motion is reflected and the virtual sound source passes in front of the listener back in the opposite direction. Listeners then naturally turn their heads back in the opposite direction (from the end of the overshoot) in order to face the initially perceived direction, where they may pause for a while to confirm the sound localization (this explaining the flat part after the overshoot).

Table 4: The amount of overshoot [deg.].

		System latency [ms]							
		12	50	100	200	500	1000	2000	Average
Sound source direction [deg.]	0	–	–	–	–	–	–	–	–
	45	3.8	3.8	4.8	7.7	10.9	19.9	43.6	13.5
	90	3.3	3.5	2.9	2.2	4.1	8.2	26.3	7.2
	135	3.5	2.3	6.0	4.9	12.5	37.4	48.0	16.4
	180	0.7	4.3	3.0	6.9	5.6	26.1	26.8	10.5
	225	2.1	2.8	3.3	3.0	3.8	23.5	14.4	7.6
	270	6.8	5.1	4.5	4.9	10.9	26.3	31.8	12.9
	315	6.1	6.5	7.8	6.6	10.4	34.6	27.4	14.2
	Average	3.8	4.0	4.6	5.2	8.3	25.2	31.2	11.7
Result of multiple comparison with Tukey's HSD									–

The overshoot can be regarded as a sign that listeners are being puzzled by the effects of the longest SL values. Long latencies that occasionally appear in network communications would thus cause a big problem in comfortable localization of virtual sound sources. System latencies longer than 500 ms should be avoided in any such VAD applications.

5. CONCLUSION

The research reported in this paper investigated the influence of VAD system latencies of up to 2 s. The experiment employed a localization task in which listeners turned to face the direction of a virtual sound source was performed with a VAD system realized using a Linux PC. As a result, the following issues were clarified with regard to the influence of large amounts of system latency: Large system latency does not have a significant influence on localization accuracy. On the other hand, as the system latency increases, an overshoot in the head movement becomes remarkable, especially when the system latency is longer than 500 ms. Moreover, at such large values of system latency, the time required for the sound localization task increases in a manner proportional to twice the value of the system latency, notably during the final phase in which the sound localization trial is concluded by the listener. This means that the influence of system latency on sound localization is apparent only when latency is longer than 500 ms. In other words, system latency below 500 ms does not have a large influence on listener behavior during the sound localization task, even though the listener may feel a certain sense of incompatibility for moderate latency values that are nonetheless longer than detection threshold (DT), which on average is about 75 ms. Such system latencies are easily realizable using DSPs or PCs that are currently available. Though a large latency may occasionally occur in network communications, the latency value of 500 ms is realizable in many environments. Therefore, the present results suggest that there are great possibilities for future development of effective computer-network-based communication tools and many other VAD applications.

6. REFERENCES

[1] J. Blauert, "Spatial Hearing," The MIT Press, 1983.

[2] P.M. Hofman, J.G.A.V. Riswick and A.J.V. Opstal, "Relearning sound localization with new ears," *Nature Neuroscience*, vol. 1, pp. 417-421, 1998.

[3] M. Morimoto and Y. Ando, "On the simulation of sound localization," *J. Acoust. Soc. Jpn. (E)*, vol. 1, pp. 167-174, 1980.

[4] H. Wallach, "On sound localization," *J. Acoust. Soc. Am.*, vol. 10, pp. 270-274, 1939.

[5] W.R. Thurlow and P.S. Runge, "Effect of induced head movement in localization of direction of sound," *J. Acoust. Soc. Am.*, vol. 42, pp. 480-488, 1967.

[6] W.R. Thurlow, J.W. Mangels and P.S. Runge, "Head movements during sound localization," *J. Acoust. Soc. Am.*, vol. 42, pp. 489-493, 1967.

[7] J. Kawaura, Y. Suzuki, F. Asano and T. Sone, "Sound localization in headphone reproduction by simulating transfer function from the sound source to the external ear," *J. Acoust. Soc. Jpn. (in Japanese)*, vol. 45, pp. 756-766, 1989.
Translated in English as: J. Kawaura, Y. Suzuki, F. Asano and T. Sone, "Sound localization in headphone reproduction by simulating transfer function from the sound source to the external ear," *J. Acoust. Soc. Jpn. (E)*, vol. 12, pp. 203-216, 1991.

[8] S. Perrett and W. Noble, "The effect of head rotations on vertical plane sound localization," *J. Acoust. Soc. Am.*, vol. 102, pp. 2325-2332, 1997.

[9] S. Perrett and W. Noble, "The contribution of head motion cues to localization of low-pass noise," *Percept Psychophys*, vol. 59, pp. 1018-1026, 1997.

[10] F.L. Wightman and D.J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *J. Acoust. Soc. Am.*, vol. 105, pp. 2841-2853, 1999.

[11] M. Kato, H. Uematsu, M. Kashino and T. Hirahara, "The effect of head motion on the accuracy of sound localization," *Acoust Sci & Tech*, vol. 24, pp. 315-317, 2003.

[12] Y. Iwaya, Y. Suzuki and S. Takane, "Effects of listener's head movement on the accuracy of sound localization in virtual environment," in *Proc. of the 18th International Congress on Acoustics*, 2004.

[13] N. Asahi, H. Aoyama and S. Matsuoka, "Headphone hearing system to reproduce natural sound localization," *Technical Report of IEICE (in Japanese)*, EA79-24, 1979.

[14] M. Ohuchi, Y. Iwaya, Y. Suzuki and T. Munekata, "Training effect of a virtual auditory game on sound localization ability of the visually impaired," in *Proc. of the 11th Meeting of the International Conference on Auditory Display*, 2005.

[15] D. Kimura and Y. Suzuki, "An effect of delay time in an auditory display system on the perception of a sound image," *Technical Report of IEICE (in Japanese)*, EA2000-67, 2001.

[16] H. Sasaki, Y. Iwaya, and Y. Suzuki, "Estimation of the detection threshold of latency of localization in a virtual auditory display," in *Proc. of Spring Meeting of Acoustical Society of Japan (in Japanese)*, pp. 531-532, 2003.

[17] D.S. Brungart, B.D. Simpson and A.J. Kordik, "The detectability of headtracker latency in virtual audio displays," in *Proc. of the 11th Meeting of the International Conference on Auditory Display*, 2005.

[18] S. Yairi, Y. Iwaya and Y. Suzuki, "Investigation of System Latency Detection Threshold of Virtual Auditory Display," in *Proc. of the 12th Meeting of the International Conference on Auditory Display*, 2006.

[19] S. Yairi, Y. Iwaya and Y. Suzuki, "Estimation of Detection Threshold of System Latency of Virtual Auditory Display," Applied Acoustics (in press).

[20] J.W. Scarpaci, H.S. Colburn and J.A. White, "A system for real-time virtual auditory space," in *Proc. of the 11th Meeting of the International Conference on Auditory Display*, 2005.