Latency Tolerance Enhancement in In-Ear Monitoring Systems

Georgios Marentakis[1], Christian Kranzler[1], Matthias Frank[1], Martin Opitz[2], Alois Sontacchi[1]
[1] *Institut für Elektronische Musik und Akustik, Kunstuniversität, 8010 Graz, Österreich, Email: marentakis@iem.at*
[2] *AKG Acoustics GmbH, 1230 Wien, Österreich*

Introduction

Monitoring systems are used to enhance the perception of either a solo or an ensemble performer when it is compromised due to concert hall acoustics and/or interference from the ensemble instruments. Monitoring systems use either loudspeakers (wedge monitoring) or in-ear monitors. In-ear monitors support better performer mobility and control of the acoustical interference from the other performers. The monitoring signal is however, inadvertently delayed while being processed and transmitted back to the performer. In this work, we try to understand the extent to which the implications due to latency in in-ear monitoring can be moderated by providing room information in the case of vocalists.

The perceptual implications of a delayed monitoring signal depend on its level and, the amount of delay relative to the signal in the acoustic path. At small latencies, comb filtering due to the superposition of the unsynchronized signals yields timbre coloration. At delays above the echo threshold, the monitoring signal will be heard separately as an echo [2]. The exact delay when these phenomena happen depends on the monitoring signal level. In in-ear monitoring, the delayed signal is louder than the direct, a case that has not been explored systematically.

Noson et al. [5] reported that when vocalists sang in an anechoic chamber with or without a single, lateral reflection (presented at variable delays and levels), they objected anechoic singing and preferred a delay in the range of 10-20 ms at -5 dB relative to the direct sound. Marshall et al. [4] also found a strong preference to reflections at -14 dB and 20ms, when presenting musicians with anechoic 'music minus one' recordings and reflections from different directions, delays and levels. Consequently, a temporal preference window for reflections was hypothesized to exist [1].

Lester and Boley [3] found that musicians object increased latency in both in-ear and wedge monitoring systems; less in wedge than in-ear monitors; however in certain cases prefer a small latency to none. Woczcyk et al. [7] used measured impulse responses from superior acoustic spaces to acoustically support musicians on stage through speakers. The perceptual evaluation by a violinist duet, hints to possible improvements due this approach that depend on the placement of the musicians and the position of the loudspeaker array.

The indications in [7], the fact that tolerance to latency was higher for wedge monitoring than in-ear monitoring in [3], together with the fact the vocalists seem to prefer the presence of a reflection relative to anechoic singing indicate that improvement may be expected when room information is provided in in-ear monitoring.

Experiment

An experiment was designed, where vocalists wearing AKG IP2 in-ear monitors and singing on a DPA 4088F Headset, compared five systems providing variable room information levels. A binaural impulse response (BRIR) was measured in the IEM experimental studio using a B&K dummy-head, and a JBL SRX712M stage monitor located on the floor in front of the dummy-head. The space ahead was acoustically damped to simulate open-air stage situations. In the five systems, the monitoring signal was: 1. presented unprocessed diotically, 2. filtered by the first BRIR floor reflection at 0°,-60°, 3. filtered by the first 60 BRIR taps, 4. processed by a late-reverb generator (TC Electronics M5000), 5. filtered by 200 BRIR taps plus late reverb (as in 4). System output loudness was normalized using Zwicker's loudness model and a noise floor was added to all systems to mask noise artifacts. A sparse BRIR version, created by peak-tracking 201 BRIR prominent peaks and setting the first (direct signal) and the unused ones to 0, was used to filter the input signal by a sparse convolution algorithm [6], for Systems 2, 3 and 5.

In all five systems, monitoring signal was presented at 5 delays: 4, 7, 10, 13 and 16 ms. For System 1, latency of 1 ms was also used. To simulate latency, vocalist microphone input was processed by a Behringer Ultradrive Pro DCX2496 Delay Line (minimum latency of 0.8 ms) controlled by Pure Data. Monitoring signal was at +10 dB relative to the acoustic signal level measured on the dummy head.

Ten vocalists, performing regularly, rated the five systems at the tested latencies relative to an analog zero latency system (Reference) in terms of preference, coloration and spaciousness in pairwise comparisons. The position of the reference in the pair was randomized. Participants responded by adjusting a slider in a graphical user interface, implemented in Matlab on a touch-screen. Prior to testing, they listened to the reference and the five systems at minimum and maximum latencies.

Results

Results (Figure 1) are given separately for Preference, Coloration and Spaciousness. Ratings are analyzed by

This study was part of the project AAP, which is funded by Austrian ministries BMVIT, BMWFJ, the Styrian Business Promotion Agency (SFG), and the departments 3 and 14 of the Styrian Government. The Austrian Research Promotion Agency (FFG) conducts the funding under the Competence Centers for Excellent Technologies (COMET, K-Project), a programme of the above-mentioned institutions.

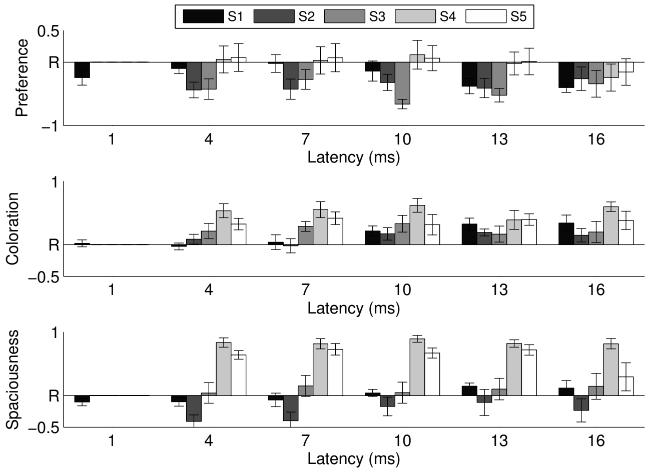

Figure 1: Preference, Coloration and Spaciousness Ratings for the five systems (S1-S5) as function of monitoring system latency

a two-way (System x Latency) within-subjects ANOVA. Following, single-tailed t-tests indicated where preference was rated significantly worse (< 0), coloration significantly higher (> 0) and spaciousness significantly higher (> 0) than the reference ($p < 0.05$).

Preference: At all latencies, Systems 2 and 3 were less preferred than the reference. On the contrary systems 4 and 5 and to a lesser extent 1 were slightly more preferred or at par with the reference. There was a main effect of System, F(4,36) = 3.7, $p = 0.012$, a marginal interaction between System and Latency, F(16,144) = 1.6, $p = 0.06$, and a marginal effect of Latency, F(4,36) = 2.2,$p = 0.079$. T-tests showed that Systems 4 and 5 at all latencies and System 1 at latencies of 4,7 and 10 ms were not different than the reference. System 2 and 3 were overall rated worse than the reference.

Coloration: Systems 1,2 and 3 were found to colorate the timbre less than System 4 and 5. There was significant main effect of System, F(4,36) = 9.2, $p < 0.001$, but no effect of Latency or interaction was observed. Systems 4 and 5 and System 1 above 4 ms colorated sound more than the rest. T-tests showed a tendency for Systems 4 and 5 to colorate the sound more than the reference at all latencies. System 1 did not at 1, 4, 7 ms but did at higher latencies. Systems 2 and 3 did, at 13, and 7 and 10 ms respectively.

Spaciousness: Systems 4 and 5 increased spaciousness perception, while Systems 1,2 and 3 did not. There was a significant effect of System, F(4,36) = 22.5, $p < 0.001$, a significant interaction between System and Latency, F(16,144) = 2.03, $p = 0.014$, and a marginal effect of latency, F(4,36) = 2.2, $p = 0.08$. Pairwise comparisons showed that Systems 4 and 5 yielded significantly higher ratings than all other systems ($p < 0.002$), but no difference between them. There was no difference between the other 3 systems. T-tests showed Systems 4 and 5 provided more spaciousness than the reference at all latencies. Systems 2 and 3 did not. System 1 did yield higher spaciousness than the reference at 13ms. When fitting a general linear model to preference as a function of coloration and spaciousness a coefficient of -0.37 and 0.55 was obtained for coloration and spaciousness respectively, implying a negative correlation of coloration to preference and a higher weight of spaciousness rather than coloration on preference judgements.

Discussion & Conclusions

As preference is at par with the reference for Systems 4, 5, and for System 1 at 4 and 7 ms, we conclude that the negative effects of latency in in-ear monitors for vocalists, can be moderated by providing room information. Furthermore, the preference window hypothesis is supported by the increase in preference between 4 and 7 ms for System 1. This tendency for a local preference maximum is also observed for the other systems, at neighbouring latencies. Apparently, the increased spaciousness at higher latencies for System 1 counteracts the negative impact of the increased coloration. The latencies at which this is observed correspond roughly to what would be caused by a wedge monitor on stage. It appears however, that detailed spatial information including early reflections and late reverberation is necessary for this to be accomplished successfully. This is evidenced by the fact that only System 4 and 5 resulted in preference comparable to the analog system at all latencies but not Systems 2 and 3. Late reverberation works for vocalists almost as well as a full BRIR. Both techniques seem to yield acceptable preference up to about 13 ms. Arguably, increased spaciousness provides a desired effect, which although it does not cancel coloration increases in-ear monitoring preference.

References

[1] Y. Ando. Subjective preference in relation to objective parameters of music sound fields with a single echo. *The Journal of the Acoustical Society of America*, 62(6):1436–1441, 1977.

[2] S. Bech. Perception of reproduced sound: Audibility of individual reflections in a complete sound field. In *Audio Engineering Society Convention 96*, 2 1994.

[3] M. Lester and J. Boley. The effects of latency on live sound monitoring. In *123rd Convention of the Audio Engineering Society*, New York, NY, USA, 2007.

[4] A. H. Marshall, D. Gottlob, and H. Alrutz. Acoustical conditions preferred for ensemble. *The Journal of the Acoustical Society of America*, 64(5):1437–1442, 1978.

[5] D. Noson, S. Sato, H. Sakai, and Y. Ando. Melisma singing and preferred stage acoustics for singers. *Journal of Sound and Vibration*, 258(3):473 – 485, 2002.

[6] Martin Opitz. Method of simulating a room and/or sound impression. US Patent No. 5,544,249, August 1996.

[7] W. Woszczyk and W. Martens. Evaluation of virtual acoustic stage support for musical performance. In *Acoustics '08*, pages 1041–1046, 2008.