MUSICAL EFFECTS OF LATENCY

Teemu Mäki-Patola
Laboratory of telecommunications software and multimedia
Helsinki University of Technology

ABSTRACT

Guidelines for making musical instruments have so far been highly intolerant towards latency. Two to ten milliseconds is usually suggested as the maximum latency in instrument design. However, perception of latency (subconscious or conscious) is a complex issue and depends on several factors, such as instrument sound, musical piece, familiarity of the instrument and properties of the human perception. This article reviews the earlier latency related research and argues that also larger latencies are acceptable in many situations as they do not increase the errors in playing.

1. INTRODUCTION

Until lately music instrument design has been highly intolerant towards latency. Professional pianists may perceive latencies of under 10ms (Finney 1997). Thus, this amount is often suggested as the maximum latency for a music controller (Finney 1997, Freed et al. 1997). However, in the light of latest research this may not be a good generalization beyond keyboard instruments.

Naturally, it would be convenient if the sound of an instrument reacted instantly to its user's actions but sometimes this is not possible. Physical sound models, new interface technology and gesture recognition, for instance, introduce some delay between the control action and the sound reaction. Yet, these new technologies offer vast possibilities for new kinds of musical expression and are thus highly interesting for musical instrument design and research. Instead of taking latencies under 10ms as a force major condition for any instrument a better understanding of the issue needs to be established. We need to know the effects of latency and how much latency actually impairs the control.

This article reviews research on the subject of latency and the closely related perception of event simultaneity focusing on the point of view of music and instrument design. To understand the effects of latency it is important to understand how different variables affect us to perceive events as happening simultaneously or asynchronously.

Understanding the human perception in isolated tests conducted in laboratory environment and playing music especially together with other people are quite different situations. The perceptual basics are the same but the situations differ by additional parameters that affect the total perception. We start our review by familiarizing ourselves with the individual physiological properties of the human perception and continue towards a more musical understanding.

2. PERCEPTION OF SIMULTANEITY

Although much research, the issue of simultaneity perception is not yet fully understood, especially with stimulus from two different sensory modalities, e.g. auditory and tactile senses. However, there is well enough information to form a good understanding of the main contributing mechanisms.

A classical experiment conducted by Michotte and reported by Card, Moran and Newell (1983) shows that humans perceive two events as connected by immediate causality if the delay between the events is less than 50ms. This is ok as a basic memory rule but to be more precise different stimuli behave differently.

Levitin et al. (1999) concludes the range -25 to 42ms as the threshold inside which aural and tactile feedback events are perceived as simultaneous. The negative time means that the audio event precedes the tactile event. In the case of aural and visual feedback the range is -41 to 45ms. The study used a more than 25% level of forced choice answers labeling the events as asynchronous as a definition of detection, unfortunately without statistical reasoning for this.

By temporal precision hearing is the most accurate of our senses in simultaneity perception. Touch is less accurate but still more accurate than sight (Levitin et. al. 1999; Repp 2003). Beyond the individual senses the time precision of simultaneity perception involving multiple types of stimuli is lower than that of the same kind of stimuli.

Subjects trying to tap along with a metronome tend to tap 30ms ahead of time by average without noticing the asynchrony (Aschersleben 2002). Many people tap even as much as 80ms prior to the metronome beat. This effect has been known already for over hundred years. The amount of anticipation depends on the involved feedbacks. Auditory-only feedback produces perfect synchronization, haptic-only feedback large anticipations and together they produce relatively small anticipations. Combined stimuli with delay added to the auditory feedback causes the anticipations to increase with the amount of introduced delay (Stenneken et al. 2003, Aschersleben and Prinz 1997).

If the subjects are informed about the size and direction of the asynchrony in the tapping experiment they can be trained to tap in synchrony. However, they then report to subjectively delay their taps, i.e. tap too late to produce the required objective synchrony (Aschersleben 2002, 2000). With a strong musical background the negative asynchrony is smaller but still there.

If the pacing metronome is changed to be produced by the subject's taps in the middle of the test causing the subject to tap in perfect synchrony the subject tends to speed up his taps (Fraisse and Voillaume 1971). This suggests that he then perceives to be constantly late. Similarly, in an experiment with pairs of musicians clapping together with varying auditory delays too small latencies caused the subjects to accelerate their tempo (Gurevich et al. 2004). 11.5ms was found as the best performing latency causing the subjects to keep the given tempo well. The subjects clapped as near to a microphone as possible and heard each other through head phones.

These seeming anomalies in behaviour lead us after some important and general human psychophysical characteristics. Before going into them we quickly review a few underlying physiological properties.

The A-beta nerve fibers, which carry information related to the touch sensation, transmit the signal at the speed of 35 to 75m/s (Kandel, Schwartz and Jessell 1991). On a distance of one meter this causes a transit delay of about 14 to 28ms. After this time the sensation signal has arrived to the sensory cortex in the brain. Auditory stimulus takes 8 to 10ms to reach the auditory cortex (Kemp et al. 1973) and visual stimulus 20 to 40ms to reach the visual cortex (Marshall et al. 1943). Reaction time studies conclude 140-160ms reaction times for sound, 155ms for touch and 180-200ms for visual stimulus (Kosinski). These are times it takes for instance to push a button when something is heard or seen. Thus, the reaction times contain the perception of an event, decision and sending and receiving the motor command.

Two models have been proposed to explain the tapping behaviour: The nerve-conduction hypothesis (Paillard-Fraisse hypothesis) and the sensory accumulator model (Aschersleben 2002, Gehrke 1995). The idea behind the nerve-conduction hypothesis is that the synchrony between events needs to be established at the subject's central representational level for the events to be perceived as being in synchrony. Based on the reviewed physiological facts the tactile feedback from the tap takes longer to travel to the brain and to be perceived than the auditory signal. Thus, the tap needs to precede the metronome click to allow synchrony between the perceived events at the central level of the brain. Similarly, in the clapping experiment when the claps are in synchrony the performer feels that he is late because the tactile feedback of the claps processes slower.

The nerve-conduction hypothesis is supported by physiological facts and by several tapping studies designed to test it. For instance, when the subjects attempt to tap with their foot instead of a finger the anticipation is about 45ms larger (Aschersleben 2002).

The idea behind the second model, the sensory accumulator model, is also that synchrony is established at the central level. However, instead of the nerve conduction times it stresses the importance of the time required to generate the central representations. A stronger signal from the physical world event causes the sensation to cumulate faster over a perception threshold. Most of human nerve cells function in this manner. The strength of excitation transfers to density of signals, which are integrated over a period of time. For instance, tapping with larger amplitudes produces significantly smaller asynchrony than tapping with small amplitudes (-25ms vs. -60ms) (Aschersleben 2002). Larger amplitude increases the tactile force and thus the amount of tactile stimulation causing the sensation threshold to fill up faster.

Both of the models used to explain the tapping behaviour are based on physiological facts and contribute to its emergence. The two models should be used together to better understand the behaviour of the human simultaneity perception. What is not yet known is the time it takes to perceive and compare the events after they reach the brain.

3. LATENCY PERCEPTION AND PERFORMANCE IN MUSICAL CONTEXT

Experiments suggest that we can tap a steady beat with as low as 4ms variations in inter tap intervals (Rubine and McAvinney 1990). We can also begin to compensate variations of that size (Repp 2000) and consciously detect timing variations of the size of 6ms in monotonic, isochronous sequences (Friberg and Sundberg 1995). If the variations are cyclic and slightly higher, about 10ms, we even begin to spontaneously perform together with them. We correct our tapping more efficiently than by just simply adapting after detecting each variation (Thaut, Tian and Azimi-Sadjadi 1998). However, this spontaneous adjusting seems to happen subconsciously.

Rhythmic perception shows strong evidence to be based on comparison between the actual and expected time of each sound attack (Schulze 1978). Tapping out of phase with the metronome does not seem to affect the tracking precision (Repp 2001). This suggests that the perception of rhythm is not based on auditory cues related to the small, 10 to 20ms differences in attack times of close sounds (Lago and Kon 2004). Instead such variations seem to be perceived subconsciously as kind of musical characteristic, the so-called feel of the music (Lago and Kon 2004).

Although variations of 20ms in audio feedback delay with tactile feedback are not consciously noticed they are compensated for similarly than we can adjust tapping to the slightly disturbed beat sequence (Wing 1977). We create an estimate for the time of feedback, detect the difference and attempt to correct it (Aschersleben 2002, Lago and Kon 2004). As our motor system does not react instantaneously we must issue motor commands ahead of time in order to perform on time. It seems natural that as a consequence we are good in calibrating how much ahead of time the commands need to be issued, even under changing circumstances.

Although that the subconscious noticing precision is high, asynchronies of up to 50ms in supposedly simultaneous notes are common in a normal musical performance. Similar asynchronies are common even in chamber music (Rasch 1979).

In a study of professional percussionists the average flutter of the hits ranged between 10 and 40ms, between 2-8% of the associated tempo in a normal drum playing (Dahl 2000). The relative size of the flutter increased with smaller tempos. This suggests that in normal percussion playing the inter-onset-intervals of the consecutive onsets vary quite much.

In piano performance the delay between pressing the key and the onset of the note is about 100ms for quiet notes and 30ms for forte notes (Askenfelt and Jansson 1990). The hammer hits the string sometime on the keys way down. This rather high latency does not seem to bother anyone. Actually the 30ms should work well together with the delayed perception of touch.

In piano pieces the notes of the melody are typically played about 30ms before the supposedly simultaneous notes. The effect is called the melody lead. Instead of being subconsciously introduced by the performer to highlight the melody line it seems to results from the dynamic differences between voices (Goebl 2001). The perceptual effect of the melody lead appears to be small (Goebl and Parncutt 2003). It is likely that the pianists who notice latencies of less than 10ms in the sound feedback do it through noticing a difference in the feeling of the instrument. The instrument behaves and feels different than what the performers are used to from their practice with the instrument. As a contrast the latencies of church organs may be several hundreds of milliseconds. Yet even they can be played well when practiced under the same circumstances.

Also the physics of the sound force some asynchronies to music. 10ms of latency is introduced already by the limited speed of sound when band partners play three meters apart from each other. Similarly parts of an orchestra audience hear the sound from different orchestra sections with additional asynchronies of even 40ms because of their distance difference to the sections. Yet this does not seem to cause any annoyance.

The acoustical properties of the environment also affect the sound perception. If the pacing signal's duration in a tapping experiment is lengthened the subject will not synchronise his hits to the onset of the signal anymore (Vos et al. 1995). Instead the signal will have a perceptual center, the location of which differs from the onset. The room acoustical properties cause the sounds to reach the perceiver's ears along several reflection paths. This may affect the perceptual center of each sound.

When there is a delay in the sound feedback of an instrument we tend to match its sound with external sounds regardless of the tactile feedback (Finney 1997, Dahl and Bresin 2001). It is the discrepancy between different feedbacks that makes the task harder and introduces errors. Pianists can play well also without any aural feedback. Introducing latency to the instrument changes the feeling of the instrument and makes the learned inner prediction model unreliable. The instrument does not behave as expected anymore and basically needs to be relearned.

With a continuous sound instrument with audio-only feedback 30ms was found as the just noticeable instrument latency when comparing to a reference with zero latency (Mäki-Patola and Hämäläinen 2004a). The noticing was subjectively highly uncertain. The subjects felt they were guessing, yet statistically they performed better. The subjects felt that they started to notice the latencies around 60ms. Younger subjects detected latencies more accurately than older ones.

4. LATENCY TOLERANCE IN MUSIC

We have seen that although our subconscious time perception can be accurate even relatively large asynchronies do not seem to cause problems in music. Latencies causing asynchronies of up to at least 30ms in external events may be normal and acceptable under most circumstances as they do not seem to impair performances with traditional musical instruments (Lago and Kon 2004). Asynchronies of this magnitude are actually used by the ear to identify simultaneous tones (Rasch 1978). When the tones are too close together they mask each other and may be perceived as a single tone.

If the asynchronies have a musical role contributing to the feel of the music also this seems to be small. Experiments show that variations in artificially added asynchronies have a minor impact (Goebl and Parncutt 2003). Also because of the influence of tactile and kinaesthetic sensations that accompany the action, the performers do not seem to have a high precision in controlling note asynchronies (Lago and Kon 2004).

Our study with continuous sound instruments with audio-only feedback suggests that with such instruments even latencies as high as 60ms do not increase errors in playing (Mäki-Patola and Hämäläinen 2004b).

Latency tolerance in music is also highly dependent on the piece of music and the instrumentation (Sawchuk et al. 2003). Somewhat surprisingly, the performers of the collaborative playing study tolerated latencies of 100ms with a piano sound but only 20ms with an accordion sound in the same piece.

The presented results strongly suggest that in many situations it is reasonable to relax the strict conditions of latency tolerance from the values below 10ms. However, more research is needed about the tolerance in different situations. It may be concluded that interesting new technologies should be studied for sound control although that they introduce somewhat high latencies. Also, as the hardware improves and the computation power grows the latencies will reduce.

5. REFERENCES

Aschersleben, G. 2002. Temporal control of movements in sensorimotor synchronization. *Brain and Cognition* 48: 66–79.

Aschersleben, G. 2000. Knowledge of results and the timing of actions *Paper No. 1/2000*. Munich: Max Plank Institue for Psychological Research.

Aschersleben, G., Prinz, W. 1997. Delayed auditory feedback in synchronization. *Journal of Motor Behavior* 29(1): 35–46.

Askenfelt, A., Jansson E. V. 1990. From touch to string vibrations. I: Timing in the grand piano action. *Journal of the Acoustical Society of America* 88(1): 52–63.

Card, S. Moran, T. Newell, A. 1983. *The Psychology of Human- Computer Interaction.* Lawrence Erlbaum Associates.

Dahl, S. Bresin, R. 2001. Is the Player More Influenced by the Auditory Than the Tactile Feedback from the Instrument. *Proceedings of the Conference on Digital Audio Effects.*

Dahl, S. 2000. The Playing of an Accent – Preliminary Observations from Temporal and Kinematic Analyses of Percussionists. *Journal of New Music Research* No. 3: 225-233.

Freed, A. Chaudhary, A. Davila, B. 1997 Operating Systems Latency Measurement and Analysis for Sound Synthesis and Processing Applications. *Proceedings of the International Computer Music Conference.*

Finney, S.A. 1997. Auditory Feedback and Musical Keyboard Performance. *Music Perception*, vol. 15, No. 2: 153-174.

Fraisse, P. Voillaume, C. 1971. Les repèred du sujet dans la synchronisation et dans la pseudo-synchronisation [Subjects's reference point during synchronization and pseudo-synchronization]. *L'Annèe Psychologique*, 71: 359-369.

Friberg, A., Sundberg, J. 1995. Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustical Society of America* 98(5): 2524–2531.

Gehrke, J. 1995 Sensorimotor synchronization: The intensity of afferent feedback affects the timing of movements *Paper No. 15/1995*. Munich: Max Planck Institute for Psychological Research.

Goebl, W. 2001. Melody lead in piano performance: Expressive device or artefact? *Journal of the Acoustical Society of America* 110(1): 563–572.

Goebl, W., Parncutt R. 2003. Asynchrony versus intensity as cues for melody perception in chords and real music. In R. Kopiez, A. C. Lehmann, I. Wolther, and C. Wolf (Eds.), *Proc. of the 5th Triennial ESCOM Conference*, Hanover, Germany, pp. 376–380. Available at: http://www.oefai.at/cgi-bin/get-tr?paper=oefai-tr-2003-11.pdf (V:27.2.2005)

Gurevich, M., Chafe, C., Leslie, G. Tyan, S. 2004. Simulation of Networked Ensemble Performance with Varying Time Delays: Characterization of Ensemble Accuracy. *Proceedings of the International Computer Music Conference* (ICMC 2004): 29-32.

Kosinski, R. J. A Literature Review on Reaction Time. http://biae.clemson.edu/bpc/bp/Lab/110/reaction.htm (Visited 27.2.2005)

Lago, N.P., Kon, F. 2004 The Quest for Low Latency. *Proceedings of the International Computer Music Conference* (ICMC 2004): 33-36.

Levitin, D. J., Mathews, M. V., and MacLean, K. 1999. The perception of cross-modal simultaneity. *Proceedings of the International Journal of Computing Anticipatory Systems*, Belgium, 1999.

Mäki-Patola, T. Hämäläinen, P. 2004 Effect of Latency on Playing Accuracy of two Gesture Controlled Continuous Sound Instruments Without Tactile Feedback. *Proceedings of the 7th International Conference on Digital Audio Effects* (DAFx'04), 5-8 Oct 2004, Naples, Italy: 11-16.

Mäki-Patola, T. Hämäläinen, P. 2004 Latency Tolerance for Gesture Controlled Continuous Sound Instrument without Tactile Feedback. *Proceedings of the International Computer Music Conference* (ICMC 2004), 1-6 Nov 2004, Miami, USA: 409-416.

Praamstra, P. Turgeon, M., Hesse, C.W., Wing, A.M., Perryer, L. 2003. Neurophysiological correlates of error correction in sensorimotor-synchronization. *NeuroImage* 20: 1283-1297.

Rasch, R. A. 1979. Synchronization in performed ensemble music. *Acustica* 43: 121–131.

Rasch, R. A. 1978. The perception of simultaneous notes such as in polyphonic music. *Acustica* 40: 21–33.

Repp, B. H. 2003. Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits and costs of interval subdivision. *Journal of Motor Behavior* 35(4): 355–370.

Repp, B. H. 2001. Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance* 27(3): 600–621.

Repp, B. H. 2000. Compensation for subliminal timing perturbations in perceptual-motor synchronization. *Psychological Research* 63: 106–128.

Rubine, D. and P. McAvinney 1990. Programmable finger tracking instrument controllers. *Computer Music Journal* 14(1): 26–40.

Sawchuk, A. A. Chew, E. Zimmermann, R. Papadopoulos, C. Kyriakakis, C. 2003. From Remote Media Immersion to Distributed Immersive Performance. *Proceedings of the 2003 ACM SIGMM workshop on Experiential Telepresence.*

Stenneken, P., Cole, J., Paillard, J., Prinz, W., Aschersleben, G. 2003. Anticipatory timing of movements and the role of sensory feedback: Evidence from deafferented patients. Available at: http://jacquespaillard.apinc.org/deafferented/pdf/stenneken-et-al-ms-03.pdf (V: 27.2.2005)

Thaut, M. H., B. Tian, and M. R. Azimi-Sadjadi 1998. Rhythmic finger tapping to cosine-wave modulated metronome sequences: Evidence of subliminal entrainment. *Human Movement Science* 17: 839–863.

Vos, P. G., Mates, J., Kruysbergen, N. W. 1995. The perceptual centre of a stimulus as the cue for synchronization to a metronome. *The Quarterly Journal of Experimental Psychology*, 48: 1024-1040.

Wing, A. M. 1977. Perturbations of auditory feedback delay and the timing of movement. *Journal of Experimental Psychology: Human Perc. and Performance* 3(2): 175–186.