Simultaneity in Interactive Computer Music
- is it always the same, or does it depend e.g. on frequency?

Abstract

The subjective perception of simultaneity (= real time) of two tone signals was tested in 24 subjects. Variables were e.g. delay time, frequency, and rise time. The results show that the Point of Subjective Simultaneity (PSS) differs for different frequencies, and that rise time, or rather envelope shape, can introduce errors into investigations like the present. Thresholds follow a parabolic relationship with frequency, the middle frequencies showing the best temporal acuity.

Real time is important in interactive computer music. Often a human performer plays an acoustic instrument and a computer or effects box manipulates the sound. In the digital domain this setup will produce a delay between input and output. Producers have established their own in-house requirements concerning the maximum allowable delay of their devices. My aim was to expand this a bit by finding a common “Point of Subjective Simultaneity” (PSS) at different frequencies. Thus giving an answer to the simple question: how long time can my algorithm spend working on an audio signal in an interactive computer music system, in order that the listener still hears the computer sound output as simultaneous to the acoustic source. Earlier research has established some related psychoacoustic thresholds, an overview can be found in [Butler, 1992]. Few results, however, study dependence of such thresholds on frequency. This would be an interesting quantity to know, as it could designate a maximum allowable latency through such a computer system, and we would have a figure limiting the concept “real time for music purposes”. See References for some of the classics studies in this field.

1 Experiment

After a pilot experiment was conducted during a visit to CCRMA, Stanford University, in 1995-96 with the assistance of Max Mathews and Craig Sapp, an experiment was set up in Denmark., involving 24 test subjects and a tightly controlled environment. 600 test signals had been prepared on a NeXT Slab computer with the cmusic program (by F. Richard Moore, UCSD) and were presented using a constant stimuli paradigm. Listening levels were approx. 70 Phons, the subjects listening through Beyer DT 990 headphones with appropriate equalization. Tests were run by a custom HyperCard program running on a Macintosh Quadra 840 AV computer, the subjects pacing themselves with no outside interference. Subjects were asked to judge whether a test signal containing two tones, one of which was a reference tone of 620 Hz, actually did start simultaneously, or one of the tones was delayed (labels were simultaneous and delayed). Several parameters varied. Test tones varied between 70 and 12,000 Hz, being a more comprehensive test of frequencies than any other I have read about.

The 24 test subjects could be grouped as: 14 male and 10 female, 7 engineers / EE students and 17 musicology students, 6 of whom specialize in music technology. Genre preferences were noted. All test subjects were instructed to listen to the test signals as many times as they wanted, but were expected to finish 600 tests in about two hours for which they were paid. All subjects except one were under 30 years of age and all had normal hearing (standard audiometry).

Individual statistics for each subject were calculated using generalized linear models (glm) with the S-Plus statistical program at the Math dept. at Aalborg University, DK. First deviance figures were calculated for each parameter and test subject, thus finding the significant parameters involved in the subjects' decision of simultaneous or delayed. Tested parameters were: delay time, linear (log) frequency, squared (log) frequency, cubic (log) frequency, attack time, order of the two tones of the signal, no. of trials of each test signal and the possible fusion the two sine tones involved.

Only delay, linear frequency, squared frequency, attack time and no. of trials were found to be significant parameters, and were included in the final generalized linear model:

$$\log(p/(1-p)) = a + b1*delay + b2*logfrq1 + b3*logfrq2 + b4*attack + b5*trials$$

Next steps were to calculate the constant and the coefficients of the model and thus find an estimate for each subject's (4 variable) psychometric function. Using the figures thus obtained, I calculated one-sided anovas to compare various groups among the test subjects to try to extract significant differences among them. The following is a summary of the results of the investigation. More details can be found (in Danish) in Brandorff [1998].
1.1 Results: Frequency

Aggregate results for all test signals are not brought here, due to space constraints. A simple, logistic function describes the fact that the longer the delay, the greater the percentage of test subjects describe the signal as delayed. As expected.

A main objective of the present investigation was to search for indications that values of PSS are frequency dependent. Therefore a wide selection of frequencies were tested. I found various kinds of frequency dependence. Figure 1 serves as a first approximation to the influence of frequency.

![Correct answers by frequency](image)

*Fig. 1 Percentage of correct answers ("delayed") by frequency. Test frequencies of 220 Hz have been omitted (used only for control purposes).*

In Fig. 1 it is clearly seen, that discrimination decreases below 500 Hz, but also that frequency influence becomes less important as the inter-onset time difference is increased. The shapes are characteristic of the results obtained for many individuals, too. The final results from the four-variable model described go along the same lines, expressed as 75%-threshold values of PSS in Fig. 2:

![Threshold values (75%) by frequency](image)

*Fig. 2 Threshold values of PSS for six frequencies, calculated from the coefficients found by logistic regression with the model mentioned. The figures are averages of the values from 23 test subjects. Lowest thresholds are obtained at 1 or 2 kHz, the very frequencies earlier investigated by Hirsh [1969].*
1.2 Fusion

I anticipated fusion to be a factor in discrimination, but other parameters turned out to be more important. A closer survey revealed that fusion is only important with test tones of 220 Hz and 110 Hz (ref. tone 220 Hz), no other frequency showed signs of increasing thresholds. A few test subjects, however, could be shown to be significantly influenced by fusion (5% level).

This result confirms Bregman’s [1990] statement that fusion between only two participating frequencies is not likely - it takes more partials to dissolve the identity of separate tones, producing the pitch perception of only the fundamental. The feature of phase locking in neural spiking patterns supports his view.

1.3 Attacks - rise time

A somewhat surprising result was found by testing four different rise times: 10, 25, 50, and 100 msec. As seen in fig. 3, long attack times show the best discrimination for short inter-onset time differences. At first this seemed almost self-contradictory to me.

![Graph: Pct. "delayed" vs. inter-onset time difference for 4 diff. rise times]

*Fig. 3. Percentage of correct answers for four values of rise time: 10, 25, 50, and 100 msec. Surprisingly, the longer attacks produce better discrimination, whilst I did expect the opposite result.*

I can offer the following explanation: Using a synthesis program like cmusic, I accept envelopes consisting of line segments, possibly not differentiable at the joints. Sharp corners would produce brief moments of high frequency content, and perhaps, when increasing the rise time the second corner pops out from the shadow of the first, and I suspect that this problem might have polluted more experiments than just mine.

1.4 Threshold values

The main result of the present study must be that average values for the Point of Subjective Simultaneity (PSS) depend on frequency. The exact values calculated are shown in Table 1:

<table>
<thead>
<tr>
<th>Freq</th>
<th>220</th>
<th>440</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>8000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave.</td>
<td>48.4</td>
<td>39.0</td>
<td>33.0</td>
<td>32.3</td>
<td>35.5</td>
<td>42.6</td>
</tr>
</tbody>
</table>

*Table 1 75% threshold values obtained by averaging the individual figures from the logistic regression calculations.*

Minimum values of thresholds are found in the frequency region from 1 to 4 kHz, a region of high information content in human speech communication. I name this a "parabolic" relationship between frequency and threshold value, and that is why I used factor “squared log-frequency”. The region of speech sounds will be the best trained frequency region in most humans.
1.5 Other, minor results

The study also compared the results of different subpopulations among test subjects, showing that Music students could listen more purely to the time parameter than engineering students. Males seemed to be better than women to focus on physical time, not being influenced by other parameter variations. No woman in the test, however, was an engineering student. Pianists, as opposed to other musician groups, seemed to be more willing to label test tones as “simultaneous”.

2 Conclusion

I claim some new results of the present study:

1. Expect influence from frequency on perception of real time Discrimination of inter-onset time differences does depend on frequency. For many individuals - as well as for the average over all test subjects - this relationship is parabolic, indicating the frequency interval between the participating sine tones is not the primary factor - absolute frequency is.

2. Rise times above 50 msecs may produce additional cues in the test signals, as corners (non-differentiable points) on the envelope curve could produce cues that had not been planned as part of the stimuli.

It is hard to say how this could influence real time music performance systems. However, some individuals will claim that two tones with a great difference in frequency do not start simultaneously, even when they do - in the physical sense. So you can not expect to impress everybody with your real time system, no matter what the specifications.

References: