Audio Engineering Society

Convention Paper

Presented at the 125th Convention
2008 October 2–5 San Francisco, CA, USA

Quantifying the strategy taken by a pair of ensemble hand-clappers under the influence of delay

Nima Darabi[1], Peter Svensson[1], and Snorre Farner[2]

[1] Centre for Quantifiable Quality of Service in Communication Systems*, NTNU,
Trondheim, O.S. Bragstads plass 2E, NO-7491, Norway
darabi@q2s.ntnu.no
svensson@q2s.ntnu.no

[2] IRCAM, Analysis/Synthesis Group, Paris, 1 place Igor Stravinsky, FR-75004 , France
farner@ircam.fr

ABSTRACT

Pairs of subjects were placed in two acoustically isolated rooms clapping together under an influence of delay up to 68 ms. Their trials were recorded and analyzed based on a definition of compensation factor or CF. This parameter was calculated from the recorded observations for both performers as a discrete function of time and thought of as a measure of the strategy taken by the subjects while clapping. Increasing the delay CF was shown to be increased linearly as it is desired to avoid tempo decrease for such high latencies. Theoretically a critical value for CF was defined as tempo over measure (or beat) duration and was used to explain why very short latencies may lead to a tempo acceleration in accordance with Chafe effect.

* *Centre for Quantifiable Quality of Service in Communication Systems, Centre of Excellence* appointed by The Research Council of Norway, funded by the Research Council, NTNU, and UNINETT. http://www.ntnu.no/Q2S/

1. INTRODUCTION

One of the prerequisites of understanding the effects of musical interaction over the Internet is learning about psychophysical effects of network latency on music as a form of human interaction. To achieve this, studying strategies taken by participants in a musical collaboration seems to be important to find out how the persons react to a delay and finally to develop a model for the behavior.

A number of research groups have presented studies of the psychophysical influence of latency on musical synchronization over the Internet. They have focused on various acoustic environments, or used a variety of musical mediums with different tempos or delays. Their results mainly show that beyond a critical delay (12 – 18 ms) the mean tempo decelerated close to linearly as function of the delay and also timing imprecision and subjective dissatisfaction of ensemble performance were observed to increase with the delay. Below this critical limit, no added imprecision was measured and no synchronization difficulties were reported; however, the counterintuitive result is that such shorter delays produced a modest acceleration, which means that moderate amounts of delay might be beneficial to tempo stability [1][2][3].

The strategies taken by subjects during ensemble playing have been offered little attention. Various researchers have reported remarkable differences between individual pairs, as considerable second moment properties were evident, but mean values of the observed factors were studied in the end, and some information was lost at that point. Individual differences might be the information that is needed to evaluate strategies.

The goal here is to quantify such strategy parameters in duo hand-clapping as well as to explore how it will influence the effect of delay on tempo.

2. HAND-CLAPPING EXPERIMENTS

The present analysis is based on further analysis of recordings that were made in a previous experiment [1]. Here we describe a summary of the experiment but further details can be found in Ref. [1].

Pairs of subjects were set to perform simple ensemble playing by clapping rhythmical patterns with their hands, in three different acoustic conditions. We have only considered the last two ones (VA and VR), where the same subjects (11 pairs) participated under both conditions. To conclude the result 5.2, we also included recordings performed in RR condition, in which another 11 pairs of subjects took part:

• real reverberant conditions (RR): the subjects were placed symmetrically at different distances (2 m – 23 m) in a large lecture hall with a reverberation time around 1.2 s at mid-frequencies,

• virtual anechoic conditions (VA): one subject was placed in an anechoic room and the other one in an acoustically well-damped room, monitoring each other via microphones and headphones through filters delaying the sound with the same delay time as in the RR case,

• virtual reverberant conditions (VR): the subjects were placed as in the VA conditions, but in addition to the delay, a computer also added artificial reverberation simulating that of the lecture hall with the clappers in the same positions as in the RR experiments.

In each condition, the subjects were placed blindfolded standing in front of the microphone to enhance their concentration on the aural input and to avoid knowing their positions. The RR, and the VA/VR experiments were conducted at separate occasions with different pairs of subjects.

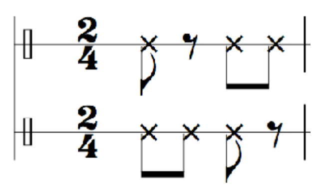

Figure 1. Complementary rhythms used in the hand-clap experiment, from [1] and [2].

To initiate each trial, both clappers were simultaneously given the tempo by an initiating recorded clapping sequence (recording of six 2/4-measures with a clap at each 1/4-note including a count-down before start). This ensured a symmetric start, i.e., avoided biasing one of the subjects to lead and the other to lag. They were asked to listen to the other and perform one of two complementary hand-clapping rhythms (see Figure 1) together in ensemble, while keeping the tempo as steady as possible.

For the two VA/VR conditions, this led to 36 trials for each of the 11 pairs of subjects that participated (in total 396 trials). The clappers were manually stopped after about 15 s, and the number of 1/8-notes finally recorded varied from 30 to 79 with a mean of 50. Tempo was changed by playing different recordings, which were constructed by recording a clapping/count-down sequence in an anechoic room and editing the recording by repositioning the claps corresponding to the wanted tempos. The order of the trials was randomized.

Assuming that musical experience might influence the results, when recruiting subjects, it was avoided to pair persons without music experience with persons having some experience with an instrument.

3. METHODOLOGY

In the present study we have focused on strategies taken by participants in a musical interaction. Our approach is to suggest a model with some numerical parameters, and to fit those parameters to the observed hand-clapping data. We aim at a model which can change over time, i.e., that a person's behavior can change during the extent of a session. It is argued below that our set of data permits a single-parameter model only per person. This approach leads to the solving of a system of linear equations. Each clapping instant for each person will provide the system with a new equation and one new equation is informative enough to calculate exactly one more unknown variable, which is a single strategy parameter of the related person assigned to that clapping instant.

4. COMPENSATION FACTOR

When a person detects that the other one has a falling tempo (that is, the incoming clap is later than the anticipated time instant), then the person will have to make a decision: how early should he/she clap? Our suggested parameter describes exactly that: to what degree does a person preempt.

Our definition of the *compensation factor,* or CF, is such that a value of 0 means that a person is clapping exactly when he/she expects the other person's clap to happen. A positive value means that a person claps a bit earlier than he/she expects the other person's clap to be heard, and a negative value means that a person claps even later than his/her expectation of the other side. The scaling of the CF is such that a value of 1 corresponds to one measure of the musical score.

We can formulate this definition so that we can use it to analyze the recordings: In the event detection for a pair of clapping partners, A and B, each clapping measure (consisting of two beats as in Figure 1) can be marked with k time stamps. First we construct a fourth virtual clap to fill the silent slot in each measure. This virtual clap was simply constructed as the average of the surrounding claps. When $k = 1$ we use the average of the four clapping times (three real ones and one virtual) as a time stamp for that measure. When $k = 2$ we use the first clap of each beat (two time stamps per measure). And when $k = 4$ we take all 3 claps of a measure plus the fourth virtual one. The i^{th} time stamp of A and B (supposed to be close to each other) are occurring at T_i^A and T_i^B. It should be noted that those are the times that they occur at the source. We made our definition based on the case $k = 1$: we define Δ_i as the timing difference of two sequential time stamps i and i-1, which here is length of each measure i:

$$\Delta_i^A = T_i^A - T_{i-1}^A \qquad (1)$$

$$\Delta_i^B = T_i^B - T_{i-1}^B \qquad (2)$$

We assume that for instance person A makes the decision about when he/she should clap to start the new measure (what T_i^A is supposed to be) based on such a process: he/she receives the last barometers from B at $T_{i-1}^B + \tau$. Having a memory of the person B's tempo at the last moment (Δ_{i-1}^B) and expecting the same tempo, he/she anticipates that T_i^B will arrive at $T_{i-1}^B + \tau + \Delta_{i-1}^B$. In the case that he/she wants to synchronize his/her muscles to perform a clap (at T_i^A) and what is heard by his/her ears (anticipated at T_i^B) he/she will set these to as equal and will clap at $T_{i-1}^B + \tau + \Delta_{i-1}^B$. Such a strategy will, however, lead to a rapid tempo decrease; rather, he/she tries to compensate for the delay and clap a bit earlier, to preempt. This amount of preempting in comparison with the last remembered measure duration (interval between last two measures of the other side) can be taken as a measure of the compensation factor (

C_i^A) for the person A at measure *i*. Using this definition for both persons A and B, we get this formulation:

$$T_i^A = T_{i-1}^B + \tau + \Delta_{i-1}^B (1 - C_i^A) \quad (3)$$

$$T_i^B = T_{i-1}^A + \tau + \Delta_{i-1}^A (1 - C_i^B) \quad (4)$$

Based on an observed clapping session, τ is known and constant, all T_i^A/T_i^B are known, and the C_i^A and C_i^B (for *i* = 3, 4, 5,…) are unknowns. These two equations related to measure *i* provide us with enough information to calculate those two unknown strategy parameters and capture the CF at each clapping measure. This definition was based on *k* = 1. For other values of *k* we have to normalize the *y*-axis (divide it by *k*) so that the scales could be the same.

4.1. Critical compensation factor

Consider a specific clapping trial with delay τ and an assumed constant pair of CF for both sides A and B over all the measures (denoted C^A and C^B). What is the condition of tempo stability? To find an answer we rewrite (3) and (4):

$$T_i^A = 2\tau + (2 - C^A) T_{i-1}^B + (C^A - 1) T_{i-2}^B \quad (5)$$

$$T_i^B = 2\tau + (2 - C^B) T_{i-1}^A + (C^B - 1) T_{i-2}^A \quad (6)$$

Instead of T_{i-1}^B and T_{i-2}^B in (5) we put values from (6):

$$T_i^A = (2 - C^A)[\tau + (2 - C^B) T_{i-2}^A + (C^B - 1) T_{i-3}^A] + 2\tau + (C^A - 1)[\tau + (2 - C^B) T_{i-3}^A + (C^B - 1) T_{i-4}^A]$$

Or in a Matrix form:

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ -1 & 2 & 2 & -4 \\ 2 & -3 & -3 & 4 \\ -1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} C^A C^B \\ C^A \\ C^B \\ 1 \end{bmatrix} \begin{bmatrix} T_i^A & T_{i-2}^A & T_{i-3}^A & T_{i-4}^A \end{bmatrix} - 2\tau = 0 \quad (7)$$

Tempo stability means that the measure duration is constant for both persons over all the measures, i.e. $\Delta_i^A = \Delta_i^B = \Delta$. We can replace $\begin{bmatrix} T_i^A & T_{i-2}^A & T_{i-3}^A & T_{i-4}^A \end{bmatrix}$ with $\begin{bmatrix} T_{i-4}^A + 4\Delta & T_{i-2}^A + 2\Delta & T_{i-3}^A + \Delta & T_{i-4}^A \end{bmatrix}$ in (7) and expand the result. Such long expansion surprisingly has a short result:

$$C_A + C_B = 2\frac{\tau}{\Delta} \quad (8)$$

This is the condition for which the trial will have a stable tempo. We call the achieved "delay time over measure duration" (both with the same unit) as C_τ or *critical CF*:

$$C_\tau = \frac{\tau}{\Delta} \quad (9)$$

If both persons (on average) have a higher CF than C_τ, then the tempo is supposed to be sped up. If both have lower CFs than this, then the tempo decreases during the sequence.

4.2. Some observed facts about calculated CFs

What do these compensation factors look like? It really depends on the analyzed trial. A sample is shown in Fig. 2, the calculated compensation factors for a pair of musicians in a reverberant hall with the starting given tempo of 86 bpm and under the influence of a 44 ms delay.

Taking a look at different trials some facts are observed about compensation factor diagrams:

- Choosing a higher *k* gives a better temporal resolution, but the compensation factor will change rapidly in a short time which does not seem to be in accordance with reality. Since the middle beats may be shifted more or less depending on the delay, it can be argued that choosing *k* = 2 is a better than *k* = 4. We observed that choosing *k* as 2 or 4, the plotted compensation factor follows a kind of periodical pattern which does not seem desired, we prefer to take just one time stamp per measure; that is, *k* = 1. Doing this, we get very few data points to work with but we remove much of the (quasi-) stochastic variations.

- For each clapping trial the compensation factor curves get shifted up and down depending on the considered amount of delay.

- One can define an order for calculating the compensation factor, referring to how many measures back in time that a person averages across when "mentally computing" the expected next clapping time. Our observation of the different orders showed that calculating orders more than 1 do not offer more information.

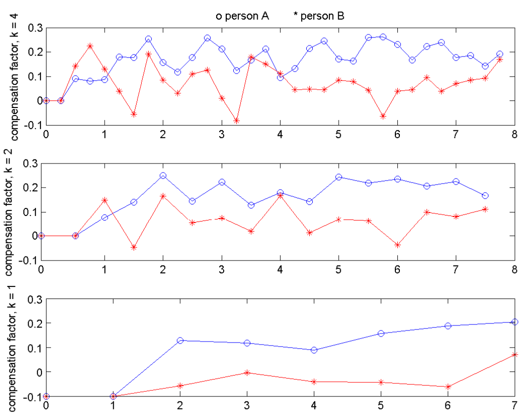

Figure 2 Compensation factors based on a recorded clapping trial for a pair of persons A (○) and B (*) with the starting given tempo of 86 bpm and the delay of 44 ms with k time stamps per clapping measure: top: $k = 4$, middle: $k = 2$, bottom: $k = 1$

5. RESULTS

5.1. CF increases linearly with delay

Fig. 3 shows CF as function of delay. For each of the 6 delays there are 132 data points (11 pairs, 2 persons, 3 tempos, and 2 virtual conditions, VA and VR). Each data point stands for one person's average of CF during one specific trial. The mean and standard error of CF for each delay are shown too in Fig. 3. A straight line was fit to all 792 data points, giving the best fit for $CF = 1\text{e-}3*(0.5639.\tau+0.8391)$ which is plotted as a continuous curve in Fig. 3. The two dashed lines show critical CFs ($1\text{e-}3*0.7167.\tau < CFcritical < 1\text{e-}3*0.7833.\tau$) for the highest and lowest given tempos (86 and 94bpm). In a clapping trial, if the average of two person's overall CFs exceeds the dashed lines then the tempo will increase in that sequence.

One should consider that for a tempo increase to occur, one data point (referring to one person) that exceeds the critical CF is not enough - the other's average CF should be taken into account too. The fact that the fitted line is below the dashed lines, for all but very short delays, illustrates that the tempo will tend to sink for most trials and subjects. It is interesting, however, that the participants apparently choose to preempt/pre-clap to a large extent – but a bit short of what would be needed to keep the tempo constant. There is a built-in aversion against choosing too high CF: the apparent lack of synchronisation gets worse the earlier oneself claps.

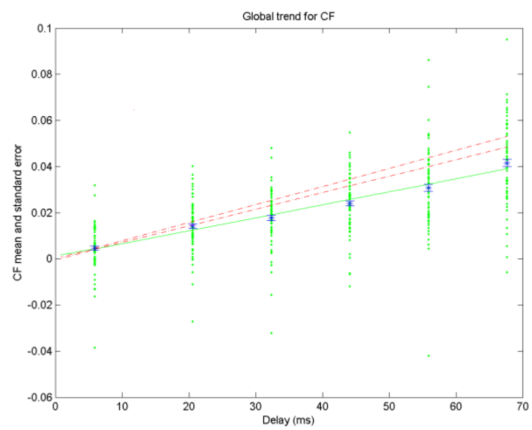

Figure 3 Global trend of CF as function of delay shows that this factor increases linearly as a function of delay, fitting with $CF=1\text{e-}3*(0.5639.\tau+0.8391)$ which is quite below the dashed lines (critical CFs related to two our tempo limits 86 and 94 bpm).

5.2. CF can explain Chafe effect

The concept of critical CF can be used to theoretically describe the recently observed phenomenon, the "Chafe effect": Chafe et.al. have, in their hand-clapping experiment, shown that short delays (< 11.5 ms) produced a modest, but surprising acceleration [2]. This can be formulated in another way: that moderate amounts of delay are beneficial to improve the collaboration [3] which means tempo stability. Based on the data set used in present study this effect was confirmed again showing that with the delays about 15 to 23 ms the tempo increased during the sequence [1]. It seems that, generally, when there is no, or a very low, delay and a tempo increase is observed during the collaboration (while the subjects are supposed to keep the tempo constant), we are facing this effect.

Why can we observe acceleration at the minimal latency? Chafe hypothesized it as a reason of some intrinsic tendency to speed up, because we have got used to latencies caused by the air that naturally separates performers [2]. It means that when there is a very low latency (delays less than Chafe's critical bound), performers still have the tendency to compensate their suspected delay which is more than the existing delay. Such an internal compensation tendency is captured by showing that the average of CF tends to exceed critical CF for very low tempos.

Fig. 4 shows the mean of all compensation factors for the latency 5.9 ms (which was the shortest delay used in our dataset, and also less than observed Chafe's critical bound), for each individual. All the recordings in VA, VR, as well as RR conditions have been taken into account. Each person's average is then based on 6 values. To make more sense we grouped persons into musicians and non-musicians. 24 persons of all participants (12 pairs) were considered as musicians (persons 1-24), and 20 of them (10 pairs) were non-musicians (persons 25-44).

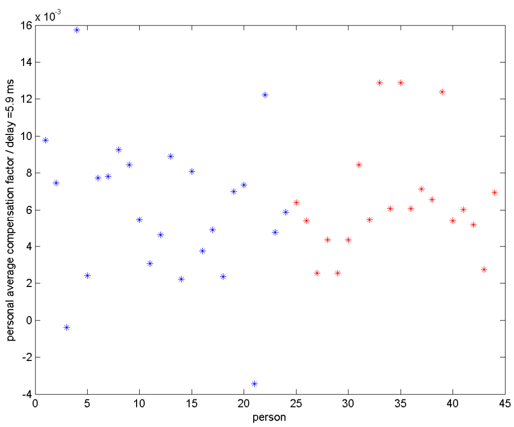

Figure 4 The CF average for each person, for the delay of 5.9 ms. The positive tends to exceed critical CF for the latency of 5.9 ms showing that for such an almost non-delayed clapping session, performers are likely to speed up for compensating a suspected delay which is more than what exists.

The plot shows a personal average CF that mainly exceeds the critical CF related to this latency. To calculate the critical CF from (9), we find the measure duration by 60**b/t,* in which *b* stands for the number of beats per measure (based on the time signature in Figure 1 it is 2), and *t* is the tempo (the starting given tempo is 86, 90 or 94 bmp here). Thus, for the delay 5.9 ms, $C_{5.9}$ will be between $4.2*10^{-3}$ and $4.6*10^{-3}$. Most data points in Fig. 4 exceed these critical values, which mean that the tempo should increase – and this was indeed observed in the previous test. One surprising observation is that both musicians and non-musicians seem likely to have intrinsic compensation.

6. CONCLUSION

A parameter called *compensation factor* is suggested to capture the amount of early clapping which performers do to compensate the tempo decrease caused by latency. We tested this parameter on a data-set from previous hand-clapping experiments, by applying a system of equations based on the theoretical definition of the compensation factor, to the observed clap times detected by performing peak detection algorithms on the recording trials [4]. Solving this system, the compensation factor was extracted and plotted as a discrete function of time and were thought of as a measure of the strategy taken by the users while clapping.

We also showed that CF is linearly increasing as a function of delay and concluded that increasing delay gives more tendencies to clap earlier, leading to a higher compensation factor.

We theoretically calculated a critical CF and proved that when two persons have higher CF than this boundary then the tempo will be increased and if they have less, the rhythm slows down. We used this definition to explain the "Chafe effect" which implies that very short latencies produce a modest acceleration. To do this we showed that for such short delays the average of CF for different persons are more than is needed and this could be the reason that we observe acceleration at minimal latency.

7. FUTURE WORK

Further work is needed using the present approach to study the possible role played by subjective strategies (taken by participants in a delayed musical collaboration) on how latency will influence observed factors (tempo, imprecision, judgment, etc.).

In future experiments it might studied if the measured strategy parameters correspond to the self-judgment of the participants or not. To do it as a kind of studying the relationship between physical stimuli and the subjective percepts, one can design some questionnaires asking about the strategies participants take and give them after each trial. It would be an interesting observation if measured parameters and self-evaluated strategies are in accordance.

Other possible ways of defining strategies that people take should be studied. Further studies based on the present approach, but with more possible parameters added, could instruct the musicians precisely to practice to adopt more to network delays. Furthermore, deeper studies on how these quantified strategies are influenced by the others may lead into finding mathematical models behind rhythmic aspects of musical collaboration and synchronization.

Although it seems that some subject pairs will learn how to adopt better to network delays; it should be investigated, however, how this would relate to our parameters. Also, related instructions which could be generalized for musicians willing to interact well at a distance should be investigated.

Studying the strategy parameters can also be used to design a musical computer collaborator, as a simplified case a computer clapper that tries to imitate the strategies that people use when clapping. Analysis of the recordings based on such a computer-human clapping session, instead of a human-human interaction has a clear benefit: The strategy of one side (the computer) is known and we would have more information to extract the strategy of the other side (human). It means that our proposed model for equation system is informative enough to be solved for two strategy parameters of one human side, instead of one parameter of two human sides.

Such models, moreover, can be generalized to more interconnected nodes than two. At that point, it could be a great interdisciplinary achievement to show these model's susceptible relation to "models of synchronization in crowds" which have previously emerged in another scientific branch, theory of complex networks. (All these planned efforts can be done with respect to this future horizon).

8. REFERENCES

[1] S. Farner, A. Solvang, A Sæbø, P. Svensson, “Ensemble hand-clapping experiments under the influence of delay and various acoustic environments” *in Preprint no. 6905, 121th AES convention*, San Francisco, CA, USA, Oct. 2006.

[2] C. Chafe and M. Gurevich, “Network time delay and ensemble accuracy: effects of latency, asymmetry” *in Preprint no. 6208, 117th AES convention*, San Francisco, CA, USA, Oct. 2004.

[3] N. Schuett, “The Effects of Latency on Ensemble Performance”, Research Sponsored by National Science Foundation, CCRMA, Department of Music, Stanford University, Stanford, California 94305, Advised by Chris Chafe, May 2002.

[4] W. A. Sethares, “Rhythm and Transforms”, Springer, ISBN: 978-1846286391, July 2007.