Audio Engineering Society

Convention Paper

Presented at the 129th Convention
2010 November 4–7 San Francisco, CA, USA

Parametric Modeling of Human Response to a Sudden Tempo Change

Nima Darabi[1], U. Peter Svensson[1], Jon Forbord[1]

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

ABSTRACT

A human-computer interactive subjective test was arranged in which 12 users took part in a sensorimotor synchronization experiment. Their task was to follow a suddenly changing metronome by hand-clapping and finger-tapping, trying to adopt to the suddenly changing tempo. Up-sampled recorded trials with different interpolation methods were used to measure their internal timekeeper's tempo in response to each tempo change. An iterative prediction error minimization method was applied to the step response signals, to identify the underlying human users' phase correction process related to these sensorimotor synchronization tasks. Experimental data indicated that the identified system is stimuli-dependent as well as individual, while not varying considerably by the task of hand clapping or finger tapping. We concluded that the sensory and cognitive process was dominant in this experiment comparing to the motion process, i.e. the mean of generating impulses. Fit-ratio comparisons showed that a delayed second order underdamped system (P2DU) could fit all the observations very well, while adding a third real pole or a zero, into P3DU or P3DUZ, would only slightly improve the model, at the cost of complexity. Concluding, the P2DU model constitutes the best trade-off between complexity and accuracy of the model. The related parameters for different stimuli (step size) and both tasks were extracted and reported.

[*] *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

This study is done as a part of a wider project aiming at modeling human behavior in musical interaction over networks. In our overall approach the network is reduced to its imperfections like bandwidth limitation, delay, and packet-loss pattern. Such simplified networks are emulated and examined through subjective experiments such as real musical interaction trials under the influence of these network limitations. In a wider scope both the auditory and visual aspects of these experiments are taken into account. A study of the collected data will lead us to find a quantified strategy-based human behavior model for this specific application. In order to model musicians' interaction over a network, we have previously studied couples of ensemble hand-clappers under the influence of delay and have quantified the strategy they have taken during the course of a synchronization trial. Exploring the differences from pair to pair we realized that there must be a series of experiences to discover the individual abilities of synchronization for each performer.

This calls for a new series of subjective studies to reveal the possibly relevant individual features of a human subject, among those the role of built-in sensory or motor memory capabilities are thought to play an important role.

In this paper we have developed a test setup, as well as a method to analyze individual responses to a tempo change in a synchronization task. The general proposed methodology is then examined on an uncomplicated task of synchronization with a suddenly tempo changing rhythm. The stimuli are also chosen to be the simple case of a temporally regular (isochronous) sequence of clicks. We are then aiming at modeling the behavior underlying the related rhythmic behavior.

2. PREVIOUS RESEARCH

Previous research on *Sensorimotor Synchronization* (SMS) is possibly the most relevant for modeling human rhythmic behavior in musical performances.

2.1. Studies on SMS

Sensorimotor synchronization (SMS) is defined as the rhythmic synchronization between a timed sensory stimulus and a motor response or, in other words, the rhythmic coordination of perception and action [1]. This is also classified as a form of referential behavior [5], describing a behavior of any kind, which is subject to some stimulus (input) that is used to produce a response (output).

Studies on rhythm and perception started in the early years of experimental psychology with studies by Wilhelm Wundt (known as the founder of the first psychophysical laboratory back in the 1890's) [10]. These studies focused on the different sensations occurring while a subject was listening to a metronome, with tension building up before a beat and the release of tension after a beat [38]. There are also examples of studies in the early 20th century, but the most important pioneers were Paul Fraisse starting from the 1950's [1] and John Michon in his 1967 PhD dissertation [3]. Michon did pioneering work on rhythmic perturbations including step changes, ramps and sinusoids and even sums of sinusoids. Later, Jeff Pressing presented reviews on SMS [1, 4, 5]. Bruno. H. Repp also has an extensive review of literature on SMS [1] including the established scientific terminology that is used in this paper.

In studies of SMS there are two main theoretical approaches: information processing and dynamic systems theory [1]. Information-processing theory approach describes the rhythmic responses and stimuli, as discrete time series and aims at describing hypothetical internal processes underlying the behavior [1]. Dynamic systems theory approaches deal with continuous movement tasks, such as circle drawing, in phase-space and are concerned with the mathematical description of observable synergies [1, 7]. These different approaches have led to the distinction between event-based (discrete) and emergent timing (continuous) [7]. In Studying a paradigm that deals with an in nature discrete task of generating discrete events, e.g. hand-clapping or finger-tapping, the continuous timing are usually of little interest in the literature. Therefore the information processing approach with discrete timings has been dominant in scenarios similar to what's covered in this research.

In contrast to the more common information-processing analysis, our approach in this study will be dynamics system theory by applying system identification: we have considered a notion of an *internal timekeeper*, an *inner clock*, that is a central aspect of dynamic modeling in the SMS-literature. The inner clock represents the continuous internal tempo that can be triggered/driven by external stimuli such as sudden tempo changes in an SMS task. How the internal timekeeper's tempo, as the output, is influenced by the input stimuli will then be described by a transfer function, identified by analyzing the experimental measurements of subjects tapping to a suddenly changing metronome.

2.2. Step Change and Sudden Stimuli

Sudden changes in stimuli have been particularly of interest to the researchers of SMS. A step change, in the

terms of this paper, is a sudden change in Interonset interval (IOI) being isochronous before and after the step. Several studies on step changes in SMS have shown an initial overshoot or overestimation of the new tempo before the subject synchronizes with the new tempo within 4-5 taps [3][6][8][9].

Michon [3] suggested a model of the period correction process (no phase correction) as an ideal predictor where the current IOI is a function of the two last IOIs. The model could take the initial overshoot into account, but to account also for the gradual adaption to the new tempo after the overshoot, he had to add another parameter. Mates took the next step including a linear phase correction and period correction process [8][9]. This model can explain the initial overshoot, but it is not clear which of the error correction processes are most effective. Hary and Moore [2] studying subliminal step changes (10ms) reported no overshoot. This was attributed to that the period correction process occurred only very gradually in combination with a mixed resetting.

For this paper, we are not interested in the inner workings of the tempo perception processes in the brain. We are merely considering the actually presented input and output data's relationship to each other in a black-box approach. This paper also aims at shifting the temporal axis from the "clap instance domain", as is done with most similar studies of SMS (Michon [3], Mates [8][9], Pressing [5]), to the time domain. This approach gives us continuous signals to process and thus offers its own set of problems to be solved.

As an example, despite the literature, in our approach inter-onset interval (IOI), the time between two consequential onsets, will not be a parameter that we describe the model on. The reason is that in a continuous model IOI only shows the sampling frequency in which subjects' internal (central) timekeeper is measured. By double tapping with the same tempo while IOI is halved, we will get the same system behind the time analysis.

3. SUBJECTIVE TEST SETUP

In this study, we have had 12 subjects perform a simple task of hand-clapping as well as finger-tapping in one-to-one (1:1) in synchronization with the auditory sequence of clicks generated by a computer application. The tempo of isochronous generated sequence was frequently changing among 100, 140 and 180 claps per minute. All trials were recorded and upsampled with different interpolation methods to estimate the internal timekeeper's tempo (subjects' step response as system's output) in response to each tempo step change (as input).

3.1. Participants

Totally 12 participants took part in the measurement. They were students and staff with no auditory or mental disorder and in both groups they were chosen to be within the range of 23 and 33 years old in which human's rhythmic synchronization error between perception and action is reported be the smallest [11]. No qualification regarding musical background was considered. Of the 12 participants, 5 were men and 7 were women

3.2. Application and the test setup

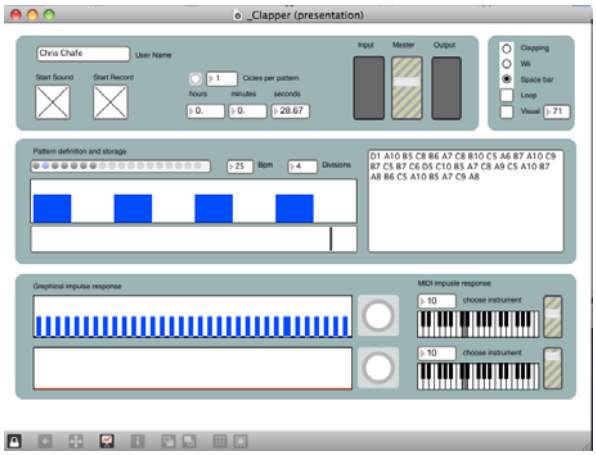
The computer-clapper application shown in Fig 1.a was implemented by MAX/MSP 5.11 and was run on MAC OS 10.6.1. The total system delay was estimated to be a few milliseconds. Subjects' task in general was to clap/tap with the sequence of rhythmic impulses given by a computer-clapper application, in an anechoic chamber. (Fig 1.b)

For the tapping sessions the subjects were encouraged to use the wrist to lead the discrete action, to take an abrupt, pulsed action and to suddenly release the downward force on the laptop computer keyboard's space button to decrease the asynchrony compared to smooth continuous movements [12].

For the clapping trials, participants were instructed to clap by taking the fingers-to-palm position of the right hand relative to the left hand, forming a right angle with a natural curvature. This naturally common configuration is one of the eight different configurations Repp has proposed [13][14]. It is called the A3 clapping mode and has the highest center frequency [15].

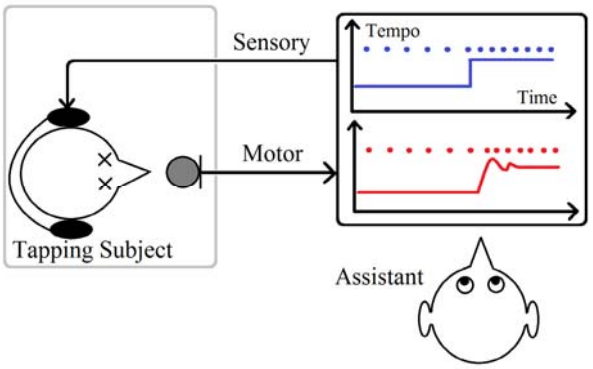
During each trial, the examined subject was blindfolded not to be distracted by visual stimuli and was supposed to make a click sound by hand-clapping (and in a similar round with finger-tapping) to the stimuli. All impulses either generated by the application or detected from user clapping (by a real-time peak detection algorithm), were saved in two separate sections of an XML file for post-processing. The isochronous given sequence of clicks was changing suddenly from time to

time during the course of a trial so that all tempos as well as all possible tempo steps occur equally many times, for each trial of a subject.



(a)

(b)



(c)

Figure 1. The test setup: (a) The test application interface (b) A subject while taking the interactive listening/acting test while monitored by the examiner (c) Illustration of the sensorimotor task

3.3. Stimuli

Assume that $G = (V, V \times V)$ is a complete non-looped directed graph in which V, set of vertices, consists of all of n tempi covered in one trial. A single *Eulerian circuit* (or *Euler tour*) is a sequence of $n.(n-1)$ tempo steps which covers all $n.(n-1)$ tempo steps (edges of a tempo transition graph) once and covers all of n tempos (vertices) equally $n-1$ times, ignoring the very first tempo as the starting point. We know that for each n, there exists a number of $n!$ of such a path.

For each single running trial, the computer application constructs a random so-called Euler tour that started with one tempo, covered all of the possible tempo changes (directed edges) and ended up with the first tempo (the starting vertex). The application kept the tapping tempo for a while (randomly between 7 to 15 musical barometers, so that subjects can't predict when the change is coming) and then suddenly jumped to the next isochronous sequence with a new tempo. The client randomly picked one such Eulerian circuit for each trial. To include repetitions, so that all subjects hear the same case k times, the application combined k Euler tours to give a pattern for the tempo change plan during the extent of the trial.

We used a setting of k=10 (ten repetitions for each tempo step change), and n=3 (for three tempos: A=100, B=140, C=180 claps per minute) for the dataset analyzed in this paper. An example of a randomly generated pattern would then look like BCBACBCABACBCBCBACACABABCABACABCBACBC ABACBCBCBACACABABCABACAB that was broken into 4 sessions (each 6 minutes) with refreshing breaks in between. Each session started from the ending tempo of the last session.

3.4. Trial consistency monitoring

It turned out that the analysis is highly sensitive to missed claps and wrongly detected ones. In order to have a smooth proper dataset, the test was carefully monitored and also some post processing routines were performed:

Once any user-generated impulse (by clapping or hitting the spacebar) was detected, a secondary different click sound was immediately generated by the application in addition to the correspond stimulus click made by the computer metronome. All of the twelve subjects reported that hearing two synchronous click sounds

(clapping/tapping physical sound and the "secondary" immediately reflected computer-generated click) was not confusing. Therefore, the actual task was to keep "the simultaneous pair of the physical response and its immediate reflection by the computer" synchronized with the "frequently and suddenly changing isochronous sequence of clicks given by the computer application".

During clapping sessions, an assistant was monitoring the application's interface to make sure that neither some weak claps were missing nor possible additional impulses were detected. The latter would be caused due to the high sensitivity of the automatic real-time peak detector in the echo-free room that could easily detect noisy body moves of the blindfolded users. The assistant examiner would then decide to repeat the test, in the case of observing many missed or wrongly detected impulses within a trial.

4. POST PROCESSING

4.1. Enrolling the missed/unexpected impulses

In addition to monitoring the session, the unexpected additional impulses were removed in a preliminary post processing stage, and also "fictive" claps were interpolated where they were expected but not observed those potentially missed weak claps.

After all this, the noisy and inaccurate nature of the experiment that irregularly samples the subjects' inner clock's tempo by a low sampling frequency of few claps per second, turned out to be very sensitive to the human uncertainty. This was handled by using averaging.

4.2. Averaging

Each trial covered every possible tempo step in our set of three different tempos, ten times. Therefore each trial was broken into its $3\times(3\text{-}1)\times10=60$ tempo steps that were grouped into three negative and three positive step changes, each repeated 10 times per trial. For each of those steps, 10 repetitive vectors of timestamps were grouped and correspondingly averaged. Six averaged vectors of timestamps thus represented each user's averaged clap onsets, made in response to each tempo step change. These vectors were the discrete data studied to reveal the underlying system behind a subject's rhythmic behavior in this specific synchronization scenario.

4.3. Upsampling and inter-sample behavior

As has been discussed earlier, the main assumption of our approach is that a human subject has a central timekeeper representing his/her continuous internal tempo. In an SMS task, the central timekeeper tracks the external source (such as a sudden tempo change) to cope with. In our subjective test we are measuring this timekeeper from time to time with a non-regular (varying) sampling frequency of the user's handclaps. In other words, we are hypothesizing that inside humans there is some internal tempo circuitry, which is continuous in nature, and that clapping/tapping is effectively sampling its continuous process. The consequent challenge is thus the fact that we are provided with few data-points for each step response. In other words, not only the time resolution of the observed to-be-identified system's behavior is low, but also is sampled non-regularly. In order to get a high-resolution signal regularly sampled signal to account for a sustainable identification process by use of various signal processing techniques, we need to reconstruct the step response signal with a considerably higher artificial sampling frequency than that of hand-clapping. The process is then called upsampling.

In the upsampling process, to reconstruct a continuous-time signal we need to know how to set the values in between two observed clap/taps. Different interpolation methods could be used each forcing their own assumptions regarding the intersample behavior of the central timekeeper. We examined staircase, linear, shape-preserving cubic, and cubic spline interpolations (each assuming a different intersample behavior) which all fitted our observed IOI data-points in the time domain, while they behaved differently in the frequency domain. This difference has an impact on the identified system depending on the system identification type used.

From an information theoretical point of view, this challenge can be formulated by a misleading degree of freedom that upsampling provides us with: Taking the derivative of the upsampled step response to get the impulse response, applying an FFT transform to take the data to the frequency domain, and plotting the resulted frequency response in logarithmic scale we get the Bode plot. The pool of data-points in the frequency domain can contain no more information about the system to extract than our limited vectors of averaged timestamps. Performing this task we observed that different interpolation methods have different artifacts in

frequency ranges outside of a *retrievable band*, i.e. the frequency range of the Bode plot in which the retrievable information underlying the system does exist. We define this band by its lower and upper boundaries.

The lower boundary is related to the signal's length. After how much time, or how many clap instances, does the underlying "period" error correction process accomplish the tempo tracking, and a user's response converging to the new tempo? After how long time does the "phase" error correction process minimize the asynchronous I/O difference, and so the asynchrony between input and output converges to an uncertainty threshold? As the working memory in auditory perception is reported between 7 and 10 seconds [16], each step won't take more time than 10 seconds after the changing stimuli in our experimental set up. Inversing the signal length, the lower boundary of the retrievable band is then around a tenth of Hz, which is also the same as the upsampling frequency divided by the number of samples. This lower boundary shows that no important information/pattern underlying the process is there to retrieve lower than a tenth of one Hz.

As for the upper bound, related to the high-frequency information content of the signal, we assume that the clapping frequency is the actual frequency of sampling the inner clock. Therefore, the size of the IOIs determined by the clapping tempos limits the upper-bound of the retrievable band. According to the Nyquist theorem this boundary can not exceed half of the clapping frequency and thus varies from 0.8 to 1.5 Hz correspond to the tempo range of 100 to 180 claps per minute.

Fig 2 shows that in the frequency domain upsampling the original recorded signal with different interpolation methods, has an impact on the Bode plot (and consequently on the identified system's parameters), outside of this defined retrievable band. This is why it's needed to distinguish which intersample behavior assumption is presumed in interpolating the new samples between tap instances.

Eventually, at any tapping timestamp the tempo was calculated by inversing inter-onset interval (IOI), the time between two sequential detected clap onsets, and was thus measured in a unit of claps per minute. Upsampling was applied to these tempo signals (example in Fig 3 for a BC step) that were those used to proceed with the system identification process.

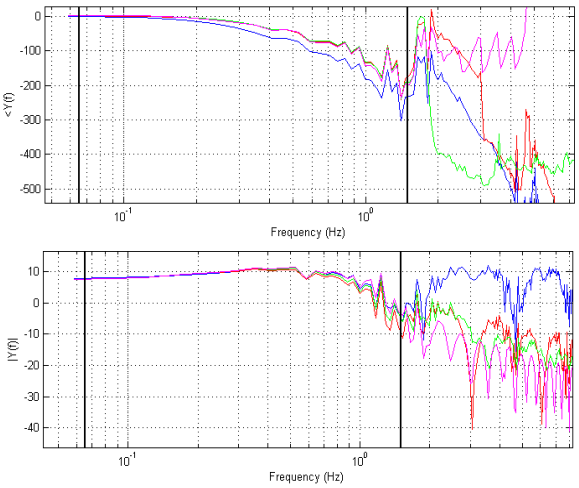
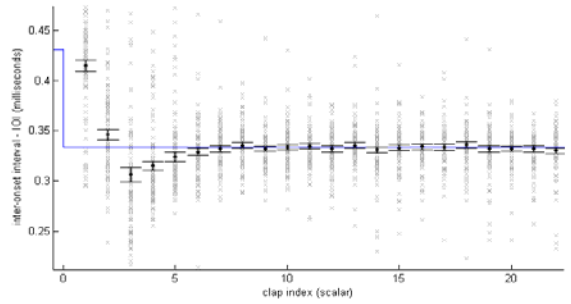
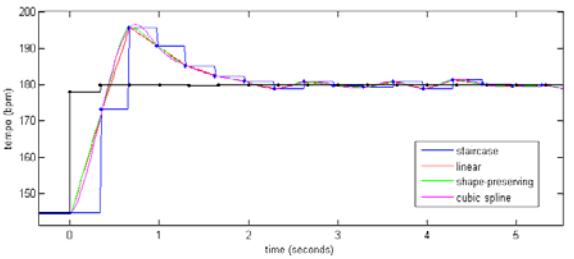


Figure 2 Upsampling the recorded signal with different interpolations, has an impact on the amplitude and the phase of the Bode plot outside of a retrievable band, which is marked by vertical solid lines.



(a)



(b)

Figure 3 (a) All subjects' measured average IOI response to a tempo step change of type BC (from 140 to 180 claps per minute)in the finger-tapping task). A total number of 120=12 persons×10 repetitions are averaged to obtain each data point. (b) The related tempo response, upsampled by staircase, linear, shape-preserving cubic, and cubic spline interpolations.

4.4. System Identification

Time-domain averaged observations for many user responses to the tempo steps covered in this experiment show a considerable overshoot, as reported firstly by Michon [3]. The converging step response to the new tempo after oscillations encouraged us to consider that a damping oscillator can identify the system.

Model parameters are estimated by the System Identification Toolbox in MATLAB© using iterative prediction-error minimization method (the function `pem`). This routine enables us to fit different models varying in the degree of complexity, i.e. the number of extractable parameters. The more complex a model, the more accurately it fits the observation. The goal is then to find a trade-off between simplicity and accuracy. We choose "the simplest good" model (the one with the least number of parameters while having a good fit), or in other words, "the best simple" model (that fits the observations best with a limited number of parameters).

Table 1 shows different transfer functions of 8 models described by their Laplace transforms. A pole-zero plot could also visualize the parameters extracted by fitting each model to the observed data. The number of parameters represents the complexity of the model and the best fit ratio (*FIT*) is a measure of accuracy. The *FIT* is calculated by equation 1 in percent, where Y is the observed output, $\bar{Y}$ is its mean and $\hat{Y}$ is the model output:

$$FIT = (1 - \frac{\sum (Y - \hat{Y})^2}{\sum (Y - \bar{Y})^2}) \times 100\% \qquad (1)$$

In table 1 K_p is a scaling constant. As the task is 1:1 and the stationary output settles and approaches the input, its identified value for this experiment gets values very close to 1. We then forced it exactly to 1 in order to exclude this parameter. T_d is the system's time delay (denoted by D in the model name) and adding it to the model would delay the same response by shifting the output along the time axis. T_{p1} and T_{p2} are two parameters describing the first and the second "real" poles of the system. They appear in P1, P2 and P2D models, the cases that do not let the system go underdamped. Another alternative to the models P2 and P2D with the same number of parameters are P2U and P2DU. They use two complex poles symmetrically placed around the real axis of a pole-zero plot. These two poles are determined by T_ω, the undamped angular frequency and ζ, the damping ratio. Basically T_ω will influence the frequency of the oscillation, and ζ will define how fast the system oscillations will be damped.

Model	Laplace Transfer Function
P1	$G(s) = K_p \frac{1}{1+T_{p1}s}$
P2	$G(s) = K_p \frac{1}{(1+T_{p1}s)(1+T_{p2}s)}$
P2D	$G(s) = K_p \frac{1}{(1+T_{p1}s)(1+T_{p2}s)} e^{-T_d s}$
P2U	$G(s) = K_p \frac{1}{1+2\zeta T_\omega s+(T_\omega s)^2} e$
P2DU	$G(s) = K_p \frac{1}{1+2\zeta T_\omega s+(T_\omega s)^2} e^{-T_d s}$
P3DU	$G(s) = K_p \frac{1}{(1+2\zeta T_\omega s+(T_\omega s)^2)(1+T_{p3}s)} e^{-T_d s}$
P2DUZ	$G(s) = K_p \frac{1+T_z s}{1+2\zeta T_\omega s+(T_\omega s)^2} e^{-T_d s}$
P3DUZ	$G(s) = K_p \frac{1+T_z s}{(1+2\zeta T_\omega s+(T_\omega s)^2)(1+T_{p3}s)} e^{-T_d s}$

Table 1 The transfer function written as the Laplace transform for a variety of 8 models used to identify the system underlying human rhythmic response.

These parameters are identified by the System Identification Toolbox across different dimensions of the dataset.

In practice there could be as many as five varying dimensions in our analysis. Three of them (i.e. subjects, task, and stimuli) are the independent variables of the experimental setup as they vary during the experiment. Two others are chosen in the analysis (interpolation method and model definition). Our methodology is first to find out which model fits the observation best and which upsampling interpolation would presume to describe the intersample behavior. Then we will choose a specific model/interpolation to study the results based on the experimental independent variables.

5. RESULTS

Fig 4, as an example, compares how different model outputs fit to an averaged observation over all hand-clapping step responses of type BC (from 140 to 180 claps per minute). A total number of 120 (twelve subjects, ten repetitions each) step responses have contributed to this example. Figure 4(b) plots the related parameters in terms of poles and zeros in a complex plane (z-plane). Table 2 accordingly contains those extracted parameters and their *FIT* value.

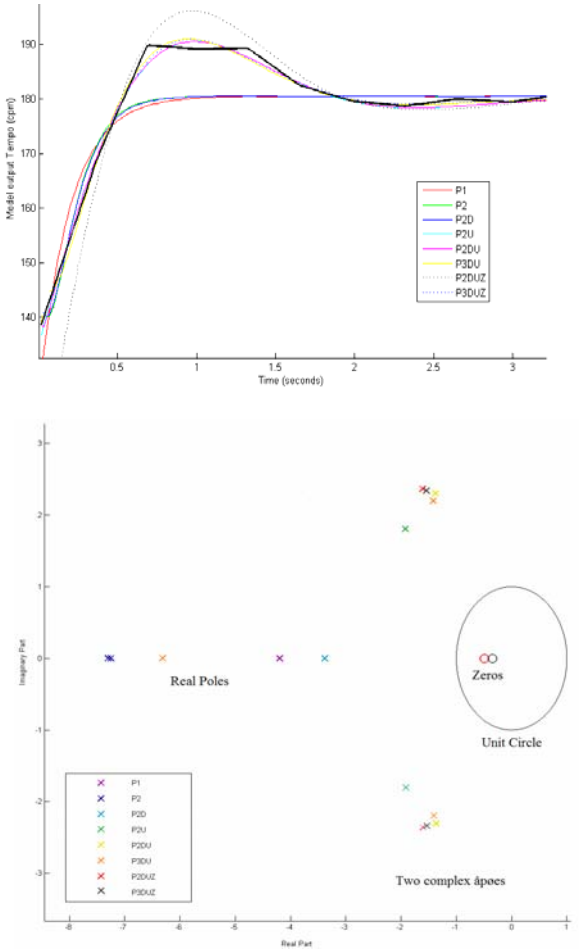


Figure 4 (a) 8 different model outputs fitting upsampled linearly interpolated average of observations over all clapping step responses of type BC. The observed signal shows a considerable overshoot as well as a slight undershoot (b) Each model's extracted parameters in the Z-plane. The numeric values for this case are reported in Table 2.

	P1	P2	P2D	P2U	P2DU	P3DU	P2DUZ	P3DUZ
T_ω	-	-		0.380	0.381	0.357	0.380	0.338
ζ	-	-		0.519	0.527	0.586	0.562	0.581
T_{p1}	0.195	0.106	0.111	-	-	-	-	-
T_{p2}	-	0.107	0.111	-	-	-	-	-
T_d	-	-	0	-	0.001	0.324	0.096	0
T_z	-	-	-	-	-	-	0.906	9.256
T_{p3}	-	-	-	-	-	0.685	-	0.516
ζ	-	-		-	-	-	-	-
FIT	49%	53%	53%	87%	87%	88%	87%	88%

Table 2 model parameters of averaged clapping step responses of type AB over all persons and their repetitions, for the task of hand-clapping extracted by MATLAB's system identification toolbox.

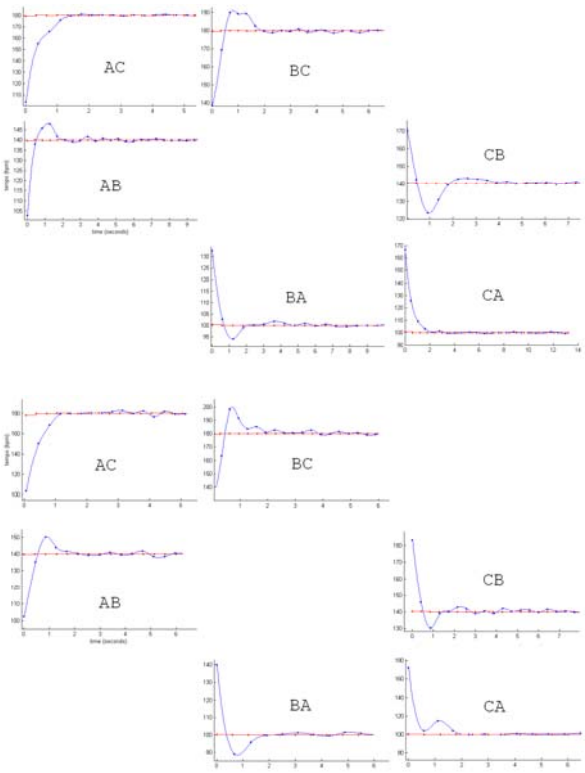


Figure 5 Tempo response upsampled by cubic spline method plotted for six steps for both tasks. Measurements regarding 12 subjects with 10 repetitions are averaged to shape each tempo response. (a) hand-clapping data, (b) finger-tapping sequences.

5.1. Time Responses

Average users' response to tempo steps for all 6 steps and 2 tasks are given in Fig 5. We can observe responses with no overshoot, with overshoot but not considerable undershoot and also systems that include both. Choice of a good model then should depend on the step type. AC and CA both in Clapping and tapping experiments show a response without overshoot. That could be due to the fact that for these tests, one tempo is close to twice the other one. So by almost doubling or halving the tapping pace, after few claps, subject can adapt to the new tempo without an overshoot. The rest of the trials show a considerable overshoot and then settle without undershoot or with a minor one. From the time data it's observed that the more the overshoot the higher the chance of an undershoot appearing. We could conclude that an overshoot which is not high enough, will not be followed by an "undershoot". This is in accordance with Repp's finding regarding the lack of an "overshoot" for subliminal sudden changes [6]. Repp also argues that for a period correction process to take place, there has to be an awareness of that the step change has taken place [6]. This means that continuing oscillations around the destination tempo observed in the time domain, do not have to be systematic as they are not perceived.

5.2. Choice of the model

To choose a good model we need to apply as many parameters as needed. Therefore, we should know which features of the observed signal are captured by any of the parameters.

Starting with one real pole (P1), an overshoot feature is not captured and it is then too simplistic for a general model as it's also shown in Fig 6 that this model gives a low *FIT* for most cases. Adding another real pole increases the fit ratio for model P2 due to allowing the simulated signal to overshoot, however, the system identification toolbox does not necessarily decide to include an overshoot as it might not give the best fit, i.e. including overshoot would have the cost of not converging to the input early enough. A P2 system can also have undershoot, if its two poles can be complex. Adding the parameter U will allow the system not to limit its poles to the real axes of the complex plane, without increasing the number of needed parameters to describe the system. P2U is then a two-pole system that can go underdamped.

Including the parameter D is also expected to improve the model fit: Users can't predict the time of the sudden change and thus they must respond to the tempo change by a delay of at least their reaction time. It means that their attack in the quickest response would at least be delayed by their reaction time. Repp has argued that while phase correction is a lower cognitive process, period correction is a higher cognitive process [6]. Considering a high level cognitive process involved in sudden temporal change tracking would increase the expected delay time even more than the reaction time. D is then decided to account for the model and this decision can be confirmed both by its increasing impact on *FIT* as seen in Fig 6, as well as by the positive values detected as the system's delay in table 3.

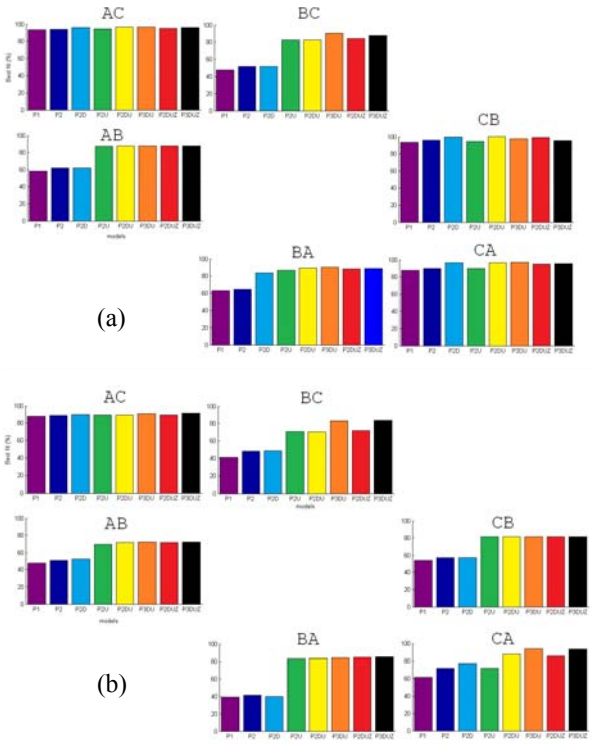


Figure 6 Values of the *FIT* for the 8 models applied to 6 tempo steps upsampled by linear interpolation (a) Clapping trials, (b) Tapping trials

To include another parameter, at the cost of increased complexity, an additional real pole or a zero would be the options. Adding each of them has shown to slightly improve the model, independent of the interpolation used. Mostly, adding a third parameter (P3DU) improves the system better than an additional third pole

(P2DUZ). Including them both still gives a mildly better fit, but the system identified by P3DUZ model could be more complex than needed in many related applications to this modeling. Figures 6 and 7 show the model comparison measured by the *FIT* value (equation 1) separately for all six steps of two tasks of hand clapping and finger-tapping.

5.3. Choice of the interpolation

Another comparison should also give us the best choice of the interpolation. Figure 7 shows that linear interpolation does not only force a weaker presumption to the inter-sample behavior but also shows a slightly better fit even than the higher order interpolations)shape-preserving and cubic spline). Therefore, the linear interpolation is also decided to be the final choice.

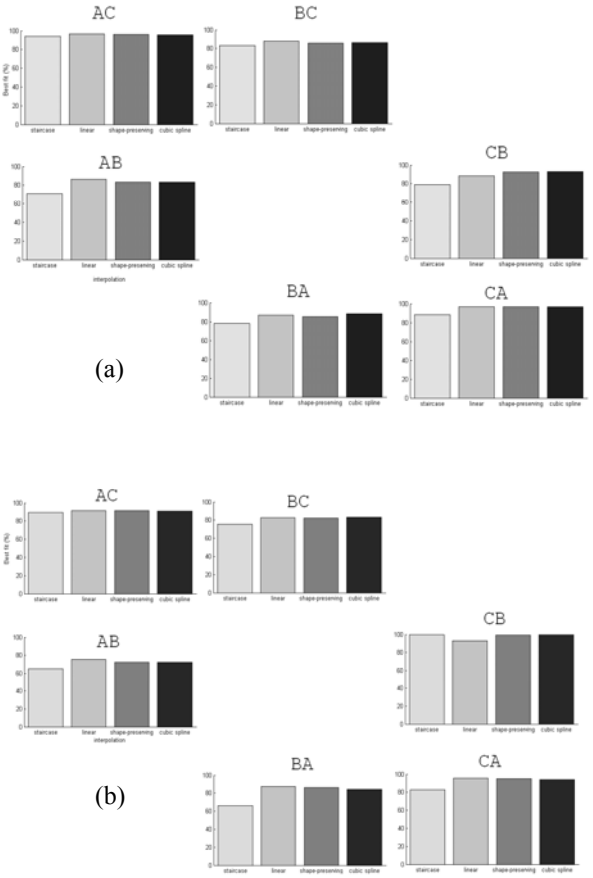


Figure 7 Best fit for the P2DU model applied to 6 tempo steps upsampled by 4 different interpolation schemes (a) Clapping trials, (b) Tapping trials

5.4. Across different tempo step changes

In table 3 the P2DU model parameters of linearly upsampled tempo step responses are reported for three negative and three positive tempo changes as well as two tasks of hand-clapping and finger-tapping. Each parameter set is estimated based on the averaged data over 120 trials (12 subjects, 10 repetitions). The related pole-zero plots are given in Fig 8.

(a)	AB	BC	AC	BA	CB	CA
T_ω	0.288	0.415	0.106	0.396	0.346	0.278
ζ	0.371	0.412	0.360	0.468	0.526	0.204
T_{p3}	0.336	0.159	0.275	0.002	0.480	0.238
T_d	0.5	0.228	0.43	0.401	0.5	0.5
(b)	**AB**	**BC**	**AC**	**BA**	**CB**	**CA**
T_ω	0.305	0.137	0.475	0.492	0.364	0.336
ζ	0.514	0.639	0.001	0.678	0.437	0.774
T_{p3}	0.967	0.734	0.327	0.093	0.184	0.140
T_d	0.483	0.5	0.5	0.5	0.50	0.442

Table 3 P3DU model parameters of linearly interpolated tempo step response reported for all 6 different tempos, (a) clapping trials and (b) tapping trials

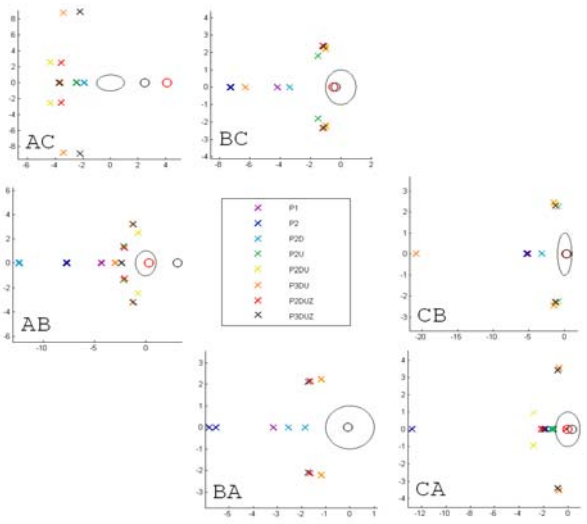


Fig 8 Model parameters in the complex plane. 8 models are fit to 6 steps; each averaged over 12 subjects and their 10 test repetitions per person. The third distant pole of P2DUZ or P3DUZ is not plotted in some cases.

5.5. Model validation

Verification of an estimated model for a specific observation by comparing the model output to the observed output, even given the best fit does not assure us that the model would work for other similar sets of observation. The statistical approaches to verify the model by exposing it to as much data as possible are skipped in this paper. The similarity of the estimated SMS model parameters for two different tasks of hand-clapping and finger-tapping, however, is an indicator that the results are reliable.

6. CONCLUSION

We showed that the phase correction process underlying adaptation to an auditory sudden tempo change of 100, 140, and 180 claps per minutes could be modeled very well by a delayed under-damped second order system (P2DU). Adding an additional pole (P3DU), giving a third order system, would give us a slightly better fit in some cases. In total four one-dimensional parameters were needed to determine our identified a system: delay, 2 poles (could be complex or real), and an optional real pole. These parameters show to vary from person to person and they were also very dependent on the step size: Where the tempo gets nearly halved or twice (100 to 180 claps per minutes and reverse) an overshoot does not appear while there is a big overshoot (an over-reaction to the tempo change) for the cases where the tempo gets 1.3 or 1.4 times less or more (100 to 140, 140 to 180 claps per minutes and vice versa). A high overshoot could be followed also by an overshoot.

System parameters for each tempo step showed high similarities between hand-clapping and finger-tapping tasks while the same task on two different tempo steps could highly influence some of the estimated parameters. That means that the system's tracking behavior behaves similar for these two types of generating clicks while behaves different across the step change dimension (stimuli). We then concluded that the sensory and cognitive process was dominant in this experiment comparing to the acting part.

7. FUTURE WORKS

Studying the system across another of its dimensions (subjects) would show whether the model parameters are individually dependent or not. By averaging all normalized step responses performed by each person (each stimuli and its repetitions) we get different individual sets of parameters from person to person. But how considerable these differences are compared to what they could be set randomly? To address this question a suggestion is to extract the model parameters across two dimensions (subjects and repetitions). For a specific task/stimuli, we identify any model parameter for each of the unique step responses rather than their averaged signal. This will give us, for each average identified parameter, a matrix with the size of number of subjects times number of repetitions. Mean of the variances of its rows (a measure of "intra-subject" variation) divided by mean of the columns' variances (a measure of "inter-subject" variation) will give us a degree of individuality for each parameter of the model, per task, per stimuli.

Other scenarios are also recommended to study by this approach. So far we have only applied it to a simple sensorimotor task of in-phase tapping in one-to-one to an auditory sequence of clicks. For the sensory part, visual and tactile stimuli can be used rather than auditory and as of the motor part in addition to finger-tapping and hand-clapping, one may study variety of musical instruments, game devices, hitting different weights against a pad or even movement without contact. The in-phase task could also be anti-phase, in 2:1, 1:2, 1:3 etc. or any other pattern of interest.

Residuals of the model output comparing to the observed step response are not discussed either. How substantial are the residuals? Are they noisy random errors caused by measurement uncertainty or systematic with some observable patterns? Besides, possible non-linearity was not taken into account either, as we didn't detect considerable systematic time-varying parameters. Could including non-linearity help to get a more accurate model? Addressing these questions would account for better descriptions underlying human rhythmic behavior.

It should be mentioned that as a result of our focus on tempo rather than asynchrony, this paper is only aiming at revealing the system behind period error correction (correction processes behind tempo tracking) and has not yet considered the phase error correction (that deal with reducing temporal asynchrony). In adaptation to a changing metronome both are taking part to the synchronization. As an example, a tempo tracker that acts based on the model in this paper can mimic a human's tempo changing behavior very well, while it might still play off-beat/off-phase. A period error

correction routine has to be added to the definition of the system probably in the form of a feedback, in order to improve the model.

The underlying process of the tempo perception in its perceptive/cognitive level is not explored in this study. On the contrary, the whole individual-medium chain from sensory to action in a sensorimotor synchronization task is identified in a black-box approach. This could be broken down into more details like the sensory part (auditory system, internal brain and nerve system) and the motor part (muscles, and external equipment) to reveal each of these two block's transfer function. Studying how the same subject has different model parameters by using different physical mediums is a way to proceed.

Finally, the discovered transfer function of P2DUZ seems to be related to the built-in auditory and active memory abilities involved in SMS. A lower damping factor in tracking a sudden tempo change indicates more "inertia" from a musical performance point of view. Explaining the identified systems behind human rhythmic behavior by an analogous RLC circuit or a mass-spring-friction system, instead of the transfer function one can build an equivalent model to the human behavior. Therefore, as we have mass, spring constant, and friction in mass-spring systems or inductance, capacitance, and resistance in series RLC circuits, there might be a chance to define and quantify an equivalent to some built-in constant or time-varying perceptive/cognitive parameters of an individual based on its observed behavior in a SMS task.

8. ACKNOWLEDGEMENT

We would like to thank Jordi Puig, the former researcher and the current PhD candidate for the implementation of the computer-clapper, the test interface used in this study and upcoming studies.

9. REFERENCES

[1] B.H. Repp, "Sensorimotor synchronization: A review of the tapping literature". *Psychonomic Bulletin and Review*, 12(6), pp. 969–992 (2005)

[2] D. Hary, G.P. Moore, "Synchronizing human movement with an external clock source, *Biological Cybernetics*, 56, pp. 305–311 (1987)

[3] J. A. Michon. "Timing in temporal tracking", A brief summary of the PhD dissertation in 1967 from Leiden University, *Van Gorcum, Assen, The Netherlands*, (2008)

[4] K. Overy, Robert Turner, "The rhythmic brain", *Cortex*, 45, pp. 1–3 (2009)

[5] J. Pressing, "The referential dynamics of cognition and action", *Psychological Review*, 106(4), pp. 714–747, (1999)

[6] B. H. Repp, "Processes underlying adaption to tempo changes in sensorimotor synchronization", *Human Movement Science*, 20, pp. 277–312 (2001)

[7] B. H. Repp, S. R. Steinman, "The rhythmic brain", *Journal of Motor Behavior*, 42(2), pp. 111–126, (2010)

[8] J. Mates, "A model of synchronization of motor acts to a stimulus sequence, i. timing and error corrections", *Biological Cybernetics*, 70, pp. 463– 473, (1994)

[9] J. Mates, "A model of synchronization of motor acts to a stimulus sequence. ii. stability analysis, error estimation and simulations", *Biological Cybernetics*, 70, pp. 475–484, (1994)

[10] R. W. Rieber, D. K. Robinson, "Wilhelm Wundt in History: The Making of a Scientific Psychology (Path in Psychology)", *Springer,* ISBN 978-03-0646-599-4 (2001)

[11] K. Drewing et al., "Sensorimotor Synchronization across the life span", *International Journal of Behavioral Development*, 30 (3), pp. 280-287 (2006)

[12] M.T. Elliot, A.E. Welchman, A.M. Wing, "Being discrete helps keep to the beat", *Exp Brain Res* 192, pp. 731-737 (2009)

[13] B. H. Repp, "The Sound of two hands clapping: An exploratory study", *Journal of Acoustical Society of America*, Am. 81(4), pp. 1100-1109 (1987)

[14] L. Peltola, C. Erkut, P. R. Cook, V. Välimäki, "Synthesis of Hand Clapping Sounds", *IEEE Transaction on audio, speech, and language processing,* 15 (3), pp. 1021-1029 (2007)

[15] A. Jylhä, C. Erkut, "Inferring the hand configuration from hand clapping sounds", in *Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, (2008)