Synthesis of Hand Clapping Sounds

Leevi Peltola, Cumhur Erkut, Perry R. Cook, *Member, IEEE*, and Vesa Välimäki, *Senior Member, IEEE*

***Abstract*—We present two physics-based analysis, synthesis, and control systems for synthesizing hand clapping sounds. They both rely on the separation of the sound synthesis and event generation, and both are capable of producing individual hand-claps, or mimicking the asynchronous/synchronized applause of a group of clappers. The synthesis models consist of resonator filters, whose coefficients are derived from experimental measurements. The difference between these systems is mainly in the statistical event generation. While the first system allows an efficient parametric synthesis of large audiences, as well as flocking and synchronization by simple rules, the second one provides parametric extensions for synthesis of various clapping styles and enhanced control strategies. The synthesis and the control models of both systems are implemented as software running in real time at the audio sample rate, and they are available for download at at http://ccrma-www.stanford.edu/software/stk and http://www.acoustics.hut.fi/go/clapd.**

***Index Terms*—Acoustic resonator filters, acoustic signal processing, control systems, emotions, signal synthesis.**

I. Introduction

CLAPPING of hands is a very popular audible activity in every culture [3]. Its most common function is to show approval and favor, although it may also be used as a rhythmic instrument. Despite this popularity, there are only a few studies about the analysis of handclaps. In his pilot study, Repp [3] has analyzed the sound of hand clapping as an individual sound-generating activity. His study has showed that the configuration of hands and the clapping rate provide perceptually important cues. These results are in accordance with other perceptual studies, which show that the properties of objects, such as size, material, hardness and shape, can be surprisingly well estimated from their sound (see [4] for a review). Néda and his colleagues have investigated the tendency of a large number of people to synchronize their clapping rates into a rhythmic applause and found that the synchronization is achieved by the period doubling of the clapping rhythm [5], [6]. They have later introduced a revised model for generation of synchronized applause *events* [7]. However, a parametric *sound synthesis* model that incorporates these findings remained to be developed.

Manuscript received June 8, 2005; revised June 5, 2006. This work was supported by the Academy of Finland under Projects 104934 and 105651. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael Davies.

L. Peltola was with the Laboratory of Acoustic and Audio Signal Processing, Helsinki University of Technology, FI-02015 TKK, Espoo, Finland. He is now with EKS, 00120 Helsinki, Finland (e-mail: lpeltola@cc.hut.fi).

C. Erkut and V. Välimäki are with the Laboratory of Acoustic and Audio Signal Processing, Helsinki University of Technology, FI-02015 TKK, Espoo, Finland (e-mail: Cumhur.Erkut@tkk.fi; Vesa.Valimaki@tkk.fi).

P. R. Cook is with the Department of Computer Science and Department of Music, Princeton University, Princeton, NJ 08544-2087 USA (e-mail: prc@cs.princeton.edu).

Digital Object Identifier 10.1109/TASL.2006.885924

In general, sound synthesis is used by musicians and the movie industry, as well as in virtual reality applications and computer games. Synthetic hand clapping can be used, for example, in sport games when modeling the applause given by the audience or to make the applause in live recordings more intense. It can also be used as a feedback in virtual reality and computer games; a user can be rewarded with an enthusiastic applause or negative feedback can be given with a bored applause.

Applause is also one of the basic general musical instrument digital interface (general MIDI)[1] sounds, and a handclap is also one of the percussion sounds of MIDI channel number 10. However, nearly all MIDI implementations of applause are based on one or a few pulse-code modulation (PCM) recordings, rather than a flexible and convincing parametric synthesis model.[2] Some historical analog and sample-based drum machines have included handclaps as one of the many percussion "voices." Some of these devices (such as the Roland TR Series[3]) attempted to provide "chorused" versions of single claps by multitriggered envelopes controlling filtered noise. Our algorithms also use filtered noise to excite simple resonant filters to model the sound of single handclaps. However, we are more interested in modeling the behavior (statistics) of ensembles of individual clappers in audience settings, under different emotional and social settings.

There are many studies on physics-based synthesis of everyday sounds. Cook introduced the Physically Informed Stochastic Event Modeling (PhISEM) algorithm for the synthesis of complex multidimensional multiple particle models, such as maracas, ice cubes in a glass, or rain drops [8], [9]. He also developed an analysis/synthesis system for walking sounds [10]. Fontana proposed a synthesis and control model for crushing, walking, and running sounds [11], [12]. Lukkari and Välimäki developed a synthesis and control model for a wind chime [13]. Interactive multimedia implementations based on physics-based synthesis and control models of everyday sounds have been reported [14]–[16].

Common to all these works is that the synthesis (sound generation) and control (event generation, usually stochastic) models are separated. The synthesis part generally is a simplified model of a single sound emitting process [4], for example modal synthesis can be used for sound generation [17]. These grains of sound are then combined using a higher-level control model that is also based on the physics of the sound event in question.

A similar strategy has been recently used in [18], where a granular synthesizer was controlled by a pulse-coupled network of spiking neurons, although the event and sound generation

[1]MIDI Manufacturers Association, Inc., 2004.

[2]Except in early analog synthesizers and drum machines, such as the Roland TR Series.

[3]http://www.roland.com/about/en/development-history.html

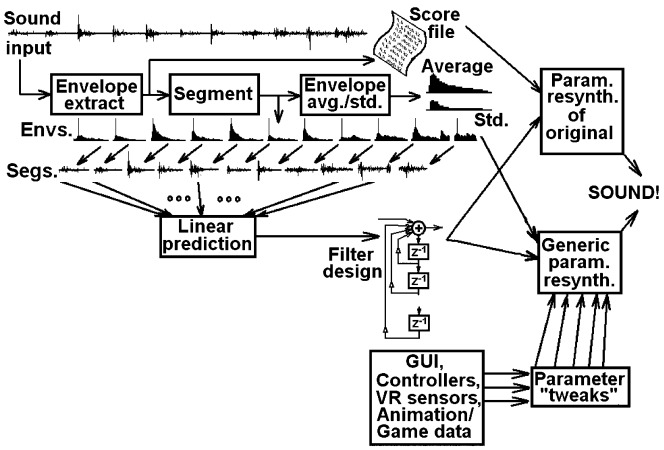


Fig. 1. General system architecture for clapping analysis/synthesis.

models were physically not matched. The advantage of physics-based synthesis is that the models are easy to control, as the physical parameters can be tuned. Furthermore, the control parameters are based on the physics of the model; thus, they can be interactively controlled by sensors. The physics-based sound model can be realized in several degrees of fidelity. For example, simplified synthesis models can be used in portable devices such as mobile phones or electronic games. Models with higher details are used if more fidelity is needed and more computational power is available. Another potential application area of physics-based sound and control models is in structured audio coding [19], which was introduced as part of the MPEG-4 standard.

This paper presents two physics-based synthesis and control systems for synthesizing hand clapping sounds. The first one, the ClapLab, is a general system for clapping analysis/synthesis implemented in STK [1]. It has been demonstrated in various conferences and workshops [20], [21], but a technical description of the system was not published previously. The second system, the ClaPD, provides parametric extensions for synthesis of various clapping styles and enhanced control strategies for one clapper, as well as an ensemble of clappers [2]. ClaPD is implemented in the Pure Data (Pd) environment [22]. In order to synthesize accurate hand-clapping sounds as part of a more complex display representing a listening environment, artificial reverberation has been optionally included within both systems.[4] These two systems are presented consecutively in this paper. Finally, the conclusions are drawn and further research directions are indicated.

II. ClapLab

The block diagram of Fig. 1 shows the general system architecture for clapping analysis/synthesis. This is a simplified form of the walking analysis/synthesis system described in [10].

[4]ClaPD uses *Freeverb* ~ [23], which implements the standard Schroeder–Moorer reverb model [24], [25] and consists of eight comb filters and four all-pass filters on both stereo channels. The filter coefficients on the left and right channels are slightly different to create a stereo effect. ClapLab uses the built-in Chowning–Schroeder reverberator *JCRev*, constructed of four comb filters and three all-pass filters per channel.

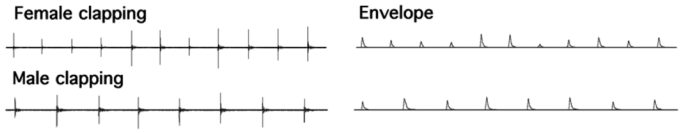


Fig. 2. Envelopes of claps.

To perform analysis, a monophonic (one person) clapping sound file is input to the system, which first does envelope extraction as

$$e(n) = (1 - b(n))|x(n)| + b(n)e(n-1) \tag{1}$$

where

$$b(n) = \begin{cases} b_{\text{up}}, & \text{if } |x(n)| > e(n-1) \\ b_{\text{down}}, & \text{otherwise.} \end{cases} \tag{2}$$

In this "envelope follower," the input signal is first rectified (absolute value) then the rectified signal is passed through a nonlinear one-pole filter. If the rectified input is greater than the current output of the filter, a rapid "attack" tracking coefficient b_{up} is used. If the rectified input signal is less than the current filter output, a slower "release" coefficient b_{down} is used. The filter gain coefficient g is always set to $(1.0 - b)$, to keep the total dc gain of the filter equal to 1.0. Typical values for a 22 050-Hz sample rate clapping/walking file are $b_{\text{up}} = 0.8$ and $b_{\text{down}} = 0.995$. The envelope signal $e(n)$ is sampled at 100 Hz. Fig. 2 shows the results of processing signals through this envelope follower. The top sound file is that of female clapping and the lower sound file is of male clapping.

After envelope extraction, the system estimates the average frequency of events (claps) using three algorithms: autocorrelation, average magnitude difference function (AMDF), and center-clipped zero-crossing detection [26]. These three algorithms "vote" based on their confidence (peak-to-noise ratio in autocorrelation, null-to-energy ratio in the AMDF, etc.) and a decision is made as to the average frequency of events. For well-recorded single person clapping sounds with minimal background noise, usually all methods agree within a very small rounding error. Using this estimate, the system then marks all event beginnings in the envelope file, and also builds a table of the individual event boundaries in the original audio file. A threshold on positive-going spikes in the derivative of the envelope is used to define event beginnings. The mean and standard deviation of event period is stored.

The individual segmented audio files are then processed by low-order (two or four poles) linear prediction [27], to determine the resonance of the claps. Means and standard deviations of resonance frequencies and Q values are computed and stored (analysis verified that clapping periods followed a normal distribution). Fig. 3 shows the spectrum of a single clap, and the second-order LPC filter to that clap. The residual (error) signal after LPC is a short, exponentially decaying burst of noise.

For nonparametric resynthesis, the system can regenerate the original clapping by concatenating the individual PCM clapping event segments. It can also speed up or slow down the clapping by simply changing the spacing between clapping events (by overlap/add if faster clapping is desired). The system can also

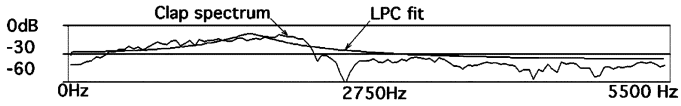


Fig. 3. Spectrum of a single clap and the second-order LPC filter to that clap.

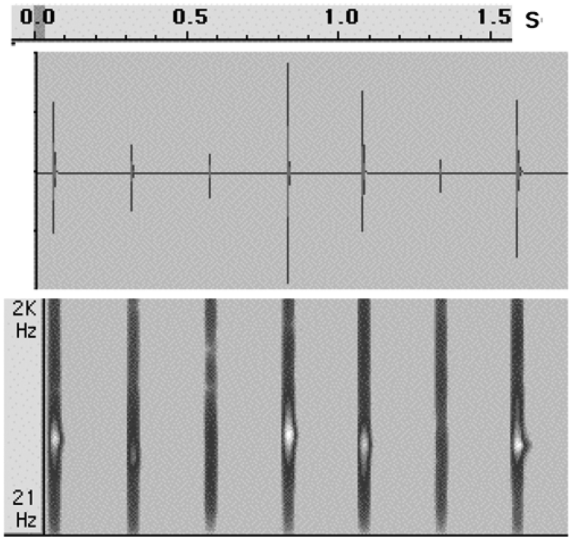


Fig. 4. Waveform and spectrogram of parametrically generated clapping.

concatenate the claps in random order, resulting in a soundfile that can be played longer without sounding repetitive.

For parametric synthesis, a simple exponentially decaying noise source[5] is used to excite a resonant filter whose parameters are controlled by the average LPC resonance and Q parameters (and standard deviations) previously determined. The following resonant filter is used in ClapLab:

$$y(n) = A_0 x(n) + 2R\cos(\theta)y(n-1) - R^2 y(n-2) \quad (3)$$

where A_0 is a gain factor that makes the magnitude response unity at resonant frequency, R is the pole radius, and θ is the pole angle.

The clapping rate is controlled by the average clapping frequency and standard deviation, for a given person to be modeled. Of course, all parameters can be scaled or replaced by values obtained from sliders, sensors, or data from a simulation (like a game or virtual reality system). Fig. 4 shows the waveform and spectrogram of parametrically generated clapping.

Eight subjects (four males and four females) were enlisted to record clapping. Table I shows the average clapping period and standard deviation, and the average and standard deviation of resonant filter center frequency for all eight subjects.

A. *Efficient Parametric Synthesis of Large Audiences*

Based on the parameters from Table I, a variety of simulated clapping "characters" can be easily generated. For large audiences, the system does not actually require one filter/noise-source per "person." A fairly small number of virtual "clappers"

[5]The envelope is the impulse response of a one pole filter with radius at 0.95 at 22 050-Hz sampling rate. This corresponds to a 60-dB decay time of 6 ms.

TABLE I
CLAP STATISTICS

Subject	Mean Clapping Period (s)	STD (s)	Mean Center Freq. (Hz)	STD (Hz)
M1	.256	.0093	1203	278
M2	.327	.0083	435	40
M3	.276	.0128	3863	1009
M4	.265	.0060	1193	243
F1	.238	.0061	1519	219
F2	.284	.0077	2243	775
F3	.298	.0100	1515	239
F4	.285	.0116	1928	764

can be constructed, and their filter settings reset in "round robin" fashion (least recently used filter is reallocated for the next clap). This is the same technique employed in the Physically-Inspired Stochastic Event Model (PhISEM) [8]. The overall noisy nature of large audience applause masks most artifacts that arise from reusing the least recently used filter. For the case of a very large audience with no synchronization (see Section III), a counter is not required for each clapper. Rather, a Poisson event probability can be used at each time step to determine if a clap should occur. This is also the technique used in PhISEM, where it is outlined that the Poisson event waiting time can be computed at a fairly arbitrary sample rate (once the probability is established for a given time interval). The result is more effective, especially in stereo or multichannel, where claps are being located at specific locations, if a separate data structure (counter, spatial location, means, and standard deviations of frequency, amplitude, and filter parameters) is used for each clapper. Artificial stereo reverberation further enhances the crowd effect.

B. *Flocking and Synchronization*

One interesting aspect of applause, and other "flocking-like" behaviors, is that sometimes the clappers fall in and out of synchrony. A simple and novel mechanism to model this effect was added to the ClapLab, called the "Affinity Knob." This slider ranges from 0.0 to 1.0, to control the relative synchronization between clappers in an audience. A very simple algorithm was employed to implement affinity. At each clap event, each individual clapper looks at the "master clapper" and drives his clapping timing toward that by setting his next clap event to HIS_NEXT_ EVENT + affinity$*$(MASTER_NEXT_EVENT $-$HIS_NEXT_EVENT). The master clapper still exhibits the slight random variations in timing. The "slave" clappers exhibit as much timing randomness as their servitude level allows, ranging from their full randomness with affinity $= 0$, to no randomness with affinity $=$ 1.0. All clappers exhibit their same randomizations of amplitude, frequency, and resonance. Thus, even with affinity set to 1.0, all clappers will clap at exactly the same times, but those times will exhibit the random timings of the master clapper, and each individual slave clapper will still have random timbral qualities. Fig. 5 shows clapping spectrograms with affinity set to 0.0, 0.25, 0.5, 0.75, and 1.0.

The same flocking algorithm can be applied to other collections of sound producers, such as birds, crickets, or even musical instruments. ClapLab is implemented in STK [1] and runs in real time at audio sample rates.

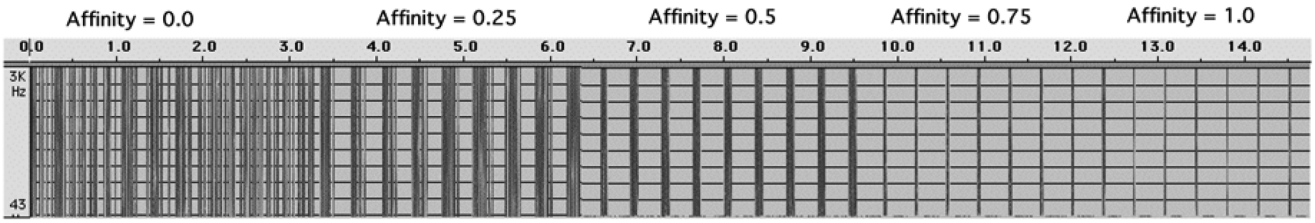


Fig. 5. Spectrograms of 10 clappers at different affinity levels.

III. Synthesis of Various Monophonic Clapping Styles

Acoustically, a hand clap corresponds to a rapid formation and excitation of a cavity between two hands. Motivated by Repp's study [3], which relates the spectral features of various clapping styles to the configuration of hands and indicates that the configuration of hands provide perceptually important cues, we have decided to simulate the cavity resonance separately by including a hand configuration parameter in the synthesis model.

For this purpose, we have conducted measurements, where the hand configuration has been a controlled variable. In these measurements we have followed Repp's suggestions for hand configurations ("modes") that correlate well with the measured spectral features in his study [3]. By assuming that the highest peak of the spectra corresponds to the cavity resonance, we have conducted a high-order LPC analysis to extract it, and allocated the second-order resonator of the ClapLab in (3) for parametric resynthesis of this cavity resonance. This strategy has caused two problems: 1) the overall spectral characteristics of a single hand-clap within the whole frequency band could not be accounted for, and 2) the excitation of the cavity resonator by an exponentially decaying noise burst resulted in a synthetic signal with an unrealistic attack. These problems are because of the relatively high Q-factor of the cavity resonator; an unrealistic attack does not occur in the technique described in Section II, which typically facilitates a lower-Q resonator.

For a simple solution of both problems, we have focused on the excitation signal: the first problem was solved by band-pass filtering the excitation noise, and the second by applying an envelope on it. The first problem could alternatively be solved by tuning another resonator as explained in Section II and running it in parallel with the cavity resonator. However, this strategy would probably over-parameterize the model and increase the computational load in the case of multiple clappers. This section provides a detailed summary of our measurements, analysis, and synthesis model.

A. Measurements

The measurements were made in an anechoic chamber. Two AKG 480B microphones were positioned at the distance of 1 m from the subjects' hands. The first microphone was positioned directly at the front of the subject, whereas the second was at the angle of 60° on the left side of the clapper.[6] The microphone signals were sampled at 44 100 Hz by a Digigram VX Pocket V2 soundcard.[7] The movement of the body when clapping hands caused some vibration on the floor, which was removed from the recordings using a fifth-order Chebyshev Type I high-pass filter with the cutoff frequency of 100 Hz. The mean center frequencies in Table I, as well as our primary investigations on the collected data showed that the frequencies of interest for clapping are higher than this cutoff frequency and, thus, are not affected by the high-pass filtering.

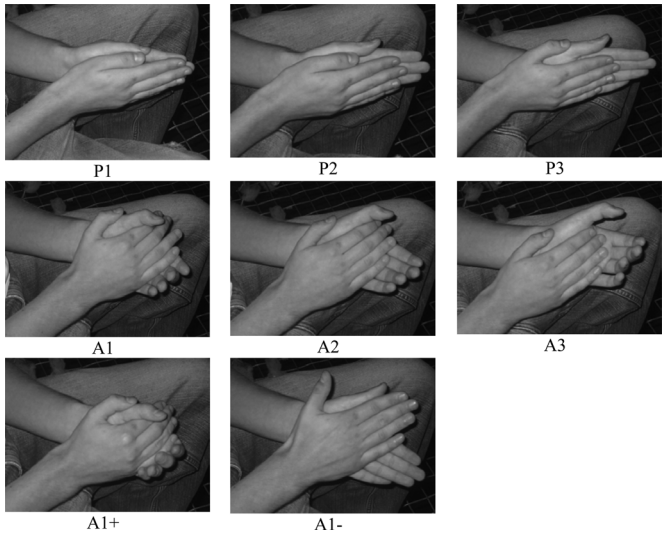


Fig. 6. Different hand clapping modes after Repp [3].

The test was made with three subjects. A sequence of five claps in each clapping mode was recorded. The positioning of hands was photographed for each clapping mode. These photos can be seen in Fig. 6.

A short description of each clapping mode is as follows. In P (parallel) modes, the hands are kept parallel and flat, so that in P1 each finger of the left hand is aligned with the corresponding finger of the right hand. The position of the right hand is varied from palm-to-palm (P1) to fingers-to-palm (P3), P2 corresponding to their midpoint. In A (angle) modes, the position of the right hand is varied in a similar fashion; from palm-to-palm (A1) to fingers-to-palm (A3), but the hands formed a right angle. Finally, the curvature of hands are varied in A1 mode to result a flat $(\mathrm{A1}-)$ or a very cupped $(\mathrm{A1}+)$ configuration, compared to A1.

B. Analysis

In this application, a moderate prediction error can be tolerated in favor of a smooth spectral characteristic of the LPC encoding filter around the cavity resonance. LPC encoder poles have been inspected for several filter orders to estimate a reasonable tradeoff between accuracy and spectral smoothness. Analysis filters of orders around 70 gave good results for our pur-

[6]The second microphone has been considered for probing the directional characteristics of the hand claps. These characteristics were not used afterwards and were left out for further analysis.

[7]http://www.digigram.com/products/VXpocket.html

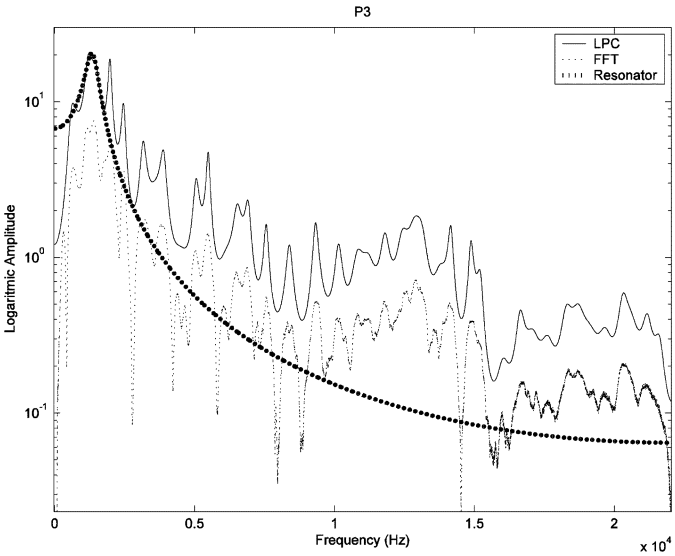


Fig. 7. LP spectrum of a single clap, a resonator fitted to model it and for comparison the FFT spectrum of the clap.

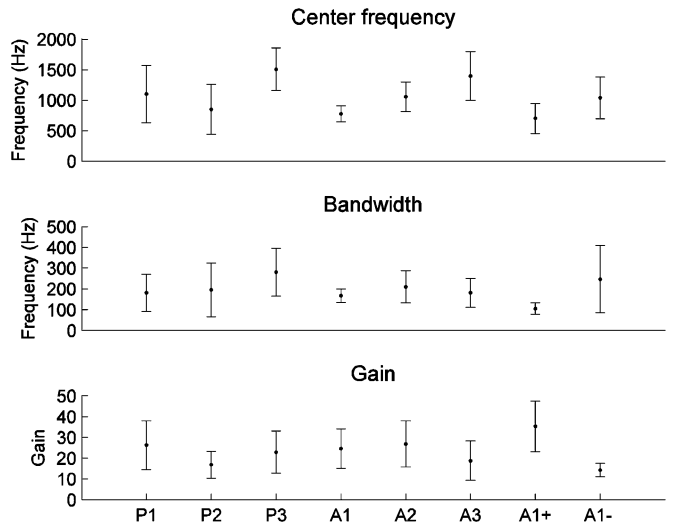


Fig. 8. Mean value and standard deviation of center frequency, bandwidth, and gain.

poses. Fig. 7 presents an example of the LP spectrum of a single hand clap in mode A2, the frequency response of the resonator fitted to the most significant peak (cavity resonance), and the fast Fourier transform (FFT) of the signal with a small offset for comparison.

Comparing our results to Repp's [3], we can observe the similarities in spectra. Modes where the palms strike to each other $(\mathrm{P2}, \mathrm{A1}, \mathrm{A1}+)$ have a spectral peak below 1 kHz, and if the hands are clapped so that fingers of the other hand struck to the palm of the other hand $(\mathrm{P3}, \mathrm{A3}, \mathrm{A1}-)$, the spectral peak is closer to 2 kHz. Fig. 8 illustrates the mean value and the standard deviation of the center frequency f_c, the bandwidth B, and the gain g of the strongest resonance for each clapping mode. Here, the gain is a dimensionless scaling factor that tunes the A_0 coefficient of the resonator in (3), so that the LP and FFT spectra match at the peak level. The statistics are estimated from each clap of each subject (i.e., 15 claps in total). As the number of

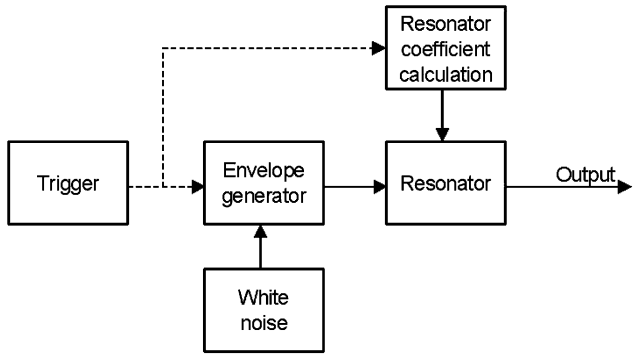


Fig. 9. Block diagram of synthesis of single hand clap.

events is so small, it is difficult to analyze the exact shape of the distribution. A Gaussian distribution was assumed to obtain indicative results.

Once the cavity resonances of hand clapping recordings are extracted, they can be filtered out from the original signal. This spectral flattening technique, called inverse filtering [29], is widely used in speech synthesis and coding. The resonator filter is inverted, i.e., the numerator and denominator are transposed, and then the inverted filter is applied to the original signal.

The resulting excitation signals were collectively modeled as a band-pass-filtered noise burst with an exponential attack and (optional) decay envelope. The parameters of these blocks are extracted by examining the average spatio-temporal characteristics of the inverse-filtered excitation signals and tuned by hand. The following parameters gave good results for the band-pass filter $H_{\mathrm{BP}}(z)$ and the envelope $e(n)$

$$H_{\mathrm{BP}}(z) = \frac{1 - z^{-2}}{1 + 0.2z^{-1} + 0.22z^{-2}} \tag{4}$$

$$e(n) = \begin{cases} 0.99^{(140-n)}, & n \leq 140 \\ 0.99^{(n-140)}, & 140 < n \leq 600 \text{ (optional).} \end{cases} \tag{5}$$

The short (3.2-ms) exponentially-rising attack segment roughly corresponds to the dynamic formation of the cavity. The decay is generally handled by the cavity resonator, although in some cases the cavity resonator provides a shorter decay time compared to our measurements. Usually, this difference is not perceived; however, we included an optional extension of the excitation decay time to compensate it.

C. Parametric Resynthesis in ClaPD

A simplified[8] block diagram of the synthesis process can be seen in Fig. 9. When the system is triggered, new cavity resonator coefficients are calculated based on the mean value and deviation obtained from the analysis (see Fig. 8). New coefficients are updated for each clap so that there is some variation between each clap. A trigger also launches the envelope generator which passes the enveloped noise signal to the resonator.

IV. Control Model for One Clapper

In Repp's research [3], the average onset-to-onset interval (OOI) among his 20 subjects was 250 ms ranging between 196

[8]The band-pass filter $H_{\mathrm{BP}}(z)$ and mode selection are not shown in the block diagram.

TABLE II
ONSET-TO-ONSET INTERVALS OBTAINED FROM MEASUREMENTS

	Natural	Enthusiastic	Bored
Average OOI	403 ms	323 ms	612 ms
Minimum OOI	316 ms	232 ms	362 ms
Maximum OOI	610 ms	547 ms	829 ms
Standard deviation of OOI within a sequence	25 ms	21 ms	29 ms

and 366 ms. Standard deviation varied between 2.8 and 13.6 ms (1% and 5%) with average of 6.8 ms (2.7%). There was a small difference on clapping rates between genders. Males clapped slightly slower (average OOI $=$ 265 ms) than females (average OOI $=$ 236 ms). Articles by Néda *et al.* [5], [6] give also some information on clapping rates. They measured the clapping rate of 73 subjects. First, the subjects were asked to clap naturally as they would after a good performance. Then they were asked to clap in the manner they would do during the synchronized applause. Average OOI of natural clapping was roughly 250 ms and in synchronized mode about 500 ms.

Although these OOI values are in accordance with the ClapLab parameters presented in Table I, we wanted to verify the average OOI of 250 ms for tuning our control model, and enhance it by the statistics of basic expressions, such as the level of enthusiasm. This section presents our experiments and the resulting statistical control model of one clapper.

A. Statistics of a Clapping Sequence

During the acoustically-based "mode" measurements outlined in Section III, some emotionally-based experiments for statistical control modeling were also made. First, the subjects were asked to clap their hands naturally, as they would after an average performance. Then they were asked to make a very enthusiastic and very bored sequence of hand clapping. This procedure was repeated twice. After the different modes of clapping were recorded, the subjects were asked to repeat a sequence of normal, enthusiastic and bored clapping, but this time in clapping mode A2.

The results obtained from our measurements are presented in Table II. These results can only give some direction for the clapping rates with different levels of enthusiasm, as the number of subjects was only three, and only six sequences of clapping was measured for each clapper. Thus, we should pay attention to clapping rate measurements by Repp [3] and Néda *et al.* [5], [6]. Basic statistics for the implementation can be derived from these combined results.

The clapping rate was chosen to vary between 4.17 Hz for enthusiastic clapping $(\mathrm{OOI} = 240\ \mathrm{ms})$ and 2.5 Hz for bored clapping $(\mathrm{OOI} = 400\ \mathrm{ms})$. Even though the measurement data shows that the OOI for bored clapping averages to the 600 ms, it sounds more realistic if the clapping rate is a little bit faster. Especially in the case of bored clapping, our subjects tend to exaggerate the slow clapping rate.

During the analysis, we have observed systematic OOI fluctuations in clapping sequences. For example, at the start of a sequence it takes some time for subjects to find their convenient clapping frequency. This explains why the variation of the OOI is usually larger at the start of a sequence. An example of a clapping sequence, where the variation of OOI is larger at the start and at the end of a sequence can be seen in Fig. 10. The height of a bar in the figure indicates the time interval between claps (OOI). In this sequence, the subject was asked to clap naturally. The clapping rhythm is also disturbed just before the end of a sequence.

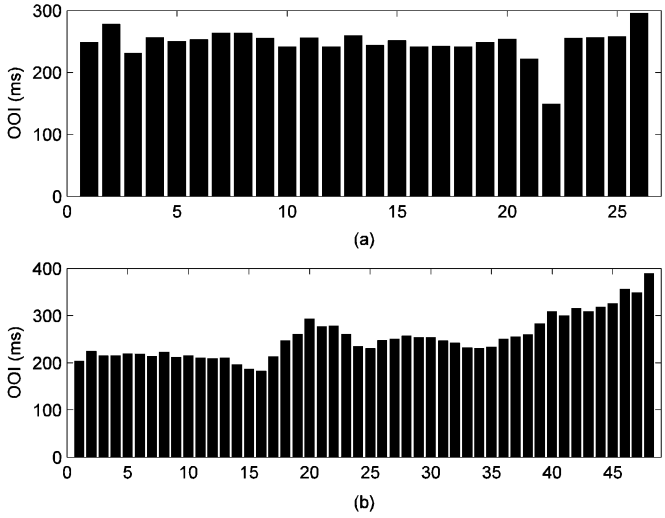


Fig. 10. (a) Example of the OOIs of a clapping sequence where the variation of OOI is larger at the start. (b) Example of the OOIs of a clapping sequence with decreasing tempo at the end (Final Ritard).

The variation of the clapping rate is not entirely random; it resembles the musical performance rules *accelerandi* (Engl. accelerating) and *rallentandi* (Engl. slowing down) [30], [31]. Such rules can also be used in the control model of a clapping sequence to model the fluctuation of clapping rate. Especially at the end of a clapping sequence, the tempo seems to slow down a little. This indicates the *Final Ritard* performance rule that was used in the control model of walking and running sounds to model the decreasing tempo that precedes the stopping [12], [30]. This phenomenon can be seen very well in long enthusiastic clapping sequences where the subject has problems in maintaining the fast clapping rate. This decay of the clapping rate caused by becoming exhausted can also be considered as one reason for the transformation to synchronized clapping. Néda *et al.* [5], [6] proposed that the audience needs to double their natural clapping rate so that the synchronization can be found.

In the single-clapper control model within the ClaPD, the user can control the length of a clapping sequence as well as the level of enthusiasm. The clapping rate varies between OOI $=$ 240 ms for enthusiastic clapping and OOI $=$ 400 ms for bored clapping. The variation of OOI is slightly exaggerated to 10% of the clapping rate and assumed to have a triangular distribution. It is doubled for the first 2 s of a clapping sequence to model the time for people to find their convenient clapping frequency. Also, the *Final Ritard* is implemented so that during the last third of the sequence, OOI is increased after each clap by 2% of the original clapping rate.

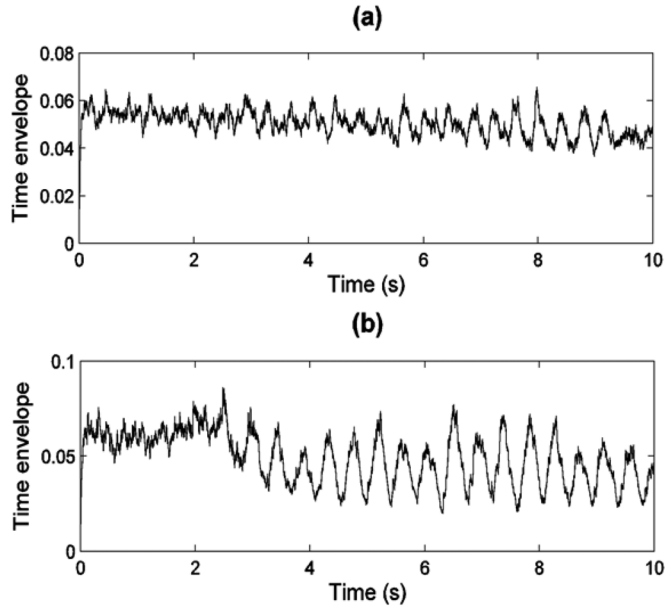


Fig. 11. (a) Envelope of a synchronized applause extracted from a live recording. (b) Envelope of a synthetic synchronized applause ($K = 0$).

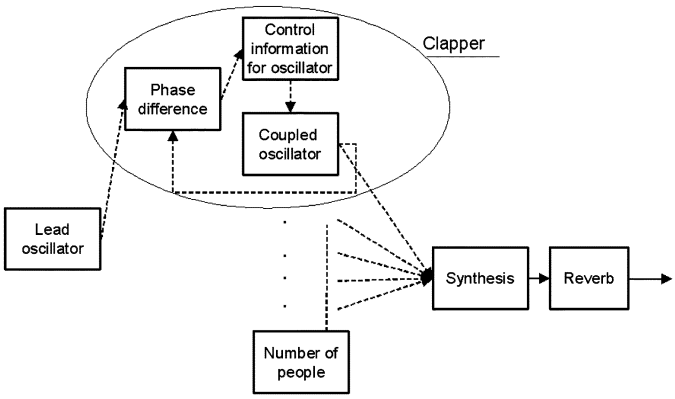


Fig. 12. Simplified block diagram for a synchronization model.

V. Ensemble Control Modeling

Néda and his colleagues have explained that the synchronization is achieved by the period doubling[9] of the clapping rhythm [5], [6]. One example of a synchronized applause (extracted from a live recording) is illustrated in Fig. 11(a). Some fluctuation can be seen in the envelope from the start of the clip. The synchronization becomes audible at about 2 s and it takes about 2 s to find a very clear synchronization. The OOI for synchronized clapping is roughly 400 ms in this example.

A simplified block diagram for a control model that incorporates the period doubling, as implemented in ClaPD, can be seen in Fig. 12. Each clapper is modeled individually and is aware of its current OOI (measured in milliseconds). The user can control the number of clappers.

In the asynchronous mode, each clapper runs independently with its own natural clapping rate. The individual OOIs are drawn from a symmetric triangular distribution Tri(220,70) so that OOI is between 150 and 290 ms. This rate is slightly faster than those proposed by Néda *et al.* [5], [6] because our measurements have indicated faster clapping rates and these faster rates produce more convincing synthetic applauses.

[9]This feature does not exist in many other systems that are known to synchronize [32].

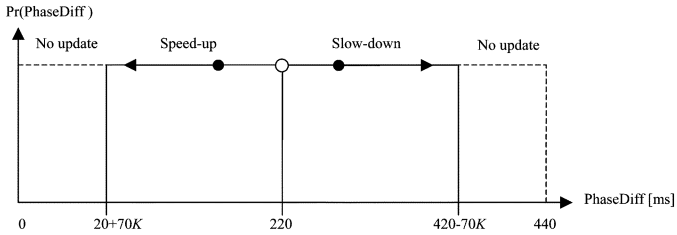


Fig. 13. Acceleration/deceleration determination ranges on the phase cycle.

In the synchronized mode, each clapper aims to clap around the same rate (frequency-locking) and absolute time (phase-locking) with the lead oscillator. Each clapper calculates its phase difference (measured in milliseconds) with the lead oscillator. Since the lead oscillator has a constant OOI of 440 ms, the phase difference can be considered a uniform distribution U(0,440) with the mean of 220 ms. If a clapper is trailing behind of the lead oscillator (phase difference < 220 ms) its clapping rate is accelerated. Similarly, if the clapper is ahead of the lead oscillator (phase difference < 220 ms) its clapping rate is slowed down. These operations are called *entrainment* [32]. A parameter[10] $K \in [0, 1]$ determines the entrainment range in the phase cycle and weights the acceleration/deceleration curve (a function of phase difference and current OOI) during the entrainment. The acceleration/deceleration ranges are depicted in Fig. 13. Within its range, the acceleration is calculated by

$$\mathrm{OOI}_{\mathrm{next}} = \mathrm{OOI}_{\mathrm{lead}} + \frac{K}{2}(\mathrm{OOI}_{\mathrm{current}} - \mathrm{OOI}_{\mathrm{lead}}) - \frac{1}{c_1 + c_2 K}\mathrm{PhaseDiff} \quad (6)$$

and the deceleration by

$$\mathrm{OOI}_{\mathrm{next}} = \mathrm{OOI}_{\mathrm{lead}} + \frac{K}{2}(\mathrm{OOI}_{\mathrm{current}} - \mathrm{OOI}_{\mathrm{lead}}) + \frac{1}{c_1 + c_2 K}(\mathrm{OOI}_{\mathrm{lead}} - \mathrm{PhaseDiff}) \quad (7)$$

where c_1 and c_2 are constants. In our experiments, $c_1 = 3$ and $c_2 = 4$ provided a good match to the observed dynamics of synchronized applause. Finally, when the mode is switched back to asynchronous, a clapper is decoupled from the lead oscillator, and its clapping rate is sped up by

$$\mathrm{OOI}_{\mathrm{next}} = \frac{1}{c_1 + c_2 K}\mathrm{OOI}_{\mathrm{current}} \quad (8)$$

until the natural clapping rate $(150 < \mathrm{OOI} < 290$ ms$)$ is achieved.[11] Again, the constants are tuned by hand; $c_1 = 1.3$ and $c_2 = -0.25$ provided good results. The synthetic results of our control model are convincing and the process of finding synchronization and losing it sound realistic. An example of a

[10]K roughly corresponds to (1-*affinity*) parameter of the ClapLab; $K = 0$ indicates a good synchronization.

[11]These algorithms can be further inspected from the source code *cosc.c* in ClaPD.

synthetic clapping sequence with synchronization is illustrated in Fig. 11(b) for $K = 0$. The sequence starts in asynchronous mode; after approximately 2 s, the user triggers the synthetic mode, and the control model enforces the entrainment.

VI. Conclusion and Future Work

We have presented two physics-based synthesis and control systems for synthesizing hand clapping sounds. A technical description of the ClapLab has been provided. ClapLab is implemented in the Synthesis Toolkit (STK), and is available for download [1].

Parametric extensions for synthesis of various clapping styles and enhanced control strategies for one clapper, as well as an ensemble of clappers are introduced. The extended synthesis and control models are implemented as a Pd library (ClaPD). ClaPD is also available for download [2].

The measurements reported in this paper were conducted with a relatively limited number of subjects. More reliable values for model coefficients may be obtained from a larger set. The perceptual testing of the algorithms and implementations presented in this paper is also an important future direction.

The noise envelope generator in ClaPD was originally designed to create isolated claps. When modeling multiple clappers, the control messages can arrive frequently and irregularly, and the clap that is currently being processed may be discarded when a new event arrives. Although this defect did not cause any audible effects while testing the ensemble control model for multiple clappers, a better scheduling algorithm may prevent potential problems.

Even though the results of stochastic ensemble control modeling are satisfying, there are many possibilities for future studies. For example, the globally coupled two-mode stochastic oscillators [7] can be implemented. The dynamics of the ensemble control model could be further improved by relaxing the probabilistic nature of the control algorithm and implementing simple nonlinear dynamic models based on phase-coupled oscillators [33], [34]. Alternatively, the pulse-coupled oscillator network reported in [18] may be considered. Note that these models generally assume acoustically identical oscillators and concentrate on their rate variations. However, as shown in [3] and verified by the measurements presented in this paper, each clapper can alter the acoustics of their clap. Incorporating such acoustical variations in social clapping and looking for systematic strategies seems to be yet another interesting and important future direction.

Acknowledgment

The authors would like to thank D. Puri, who began the Princeton clapping project as undergraduate independent work in 2001.

References

[1] P. R. Cook and G. P. Scavone, "The synthesis toolkit STK," in *Proc. Int. Comput. Music Conf.*, Beijing, China, Oct. 1999, pp. 164–166 [Online]. Available: http://ccrma-www.stanford.edu/software/stk/

[2] L. Peltola, "Analysis, parametric synthesis, and control of hand clapping sounds," M.S. thesis, Helsinki Univ. Technol., Espoo, Finland, 2004, [Online]. Available: http://www.acoustics.hut.fi/publications.

[3] B. H. Repp, "The sound of two hands clapping: An exploratory study," *J. Acoust. Soc. Amer.*, vol. 81, no. 4, pp. 1100–1109, Apr. 1987.

[4] D. Rocchesso and F. Fontana, Eds., *The Sounding Object*. Firenze, Italy: Edizioni di Mondo Estremo, 2003.

[5] Z. Néda, E. Ravasz, Y. Brechet, T. Vicsek, and A. L. Barabási, "Physics of the rhythmic applause," *Phys. Rev. E*, vol. 61, pp. 6987–6992, 2000.

[6] ——, "The sound of many hands clapping," *Nature*, vol. 403, pp. 849–850, 2000.

[7] Z. Néda, A. Nikitin, and T. Vicsek, "Synchronization of two-mode stochastic oscillators: A new model for rhythmic applause and much more," *Phys. A: Statist. Mech. Applicat.*, vol. 321, pp. 238–247, 2003.

[8] P. R. Cook, "Physically informed sonic modeling (PhISM): Synthesis of percussive sounds," *Comput. Music J.*, vol. 21, no. 3, pp. 38–49, 1997.

[9] P. R. Cook, "FOFs, wavelets, and particles," in *Real Sound Synthesis for Interactive Applications*. Natick, MA: A.K. Peters, 2002, pp. 149–168.

[10] ——, "Modeling Bill's gait: Analysis and parametric synthesis of walking sounds," in *Proc. Audio Eng. Soc. 22nd Conf. Virtual, Synthetic, and Entertainment Audio*, Helsinki, Finland, 2002, pp. 73–78.

[11] F. Fontana, "Physics-based models for the acoustic representation of space in virtual environments," Ph.D. dissertation, Univ. Verona, Verona, Italy, 2003.

[12] F. Fontana and R. Bresin, "Physics-based sound synthesis and control: Crushing, walking and running by crumpling sounds," in *Proc. XIV Colloquium Musical Informatics*, Firenze, Italy, 2003 [Online]. Available: http://profs.sci.univr.it/~fontana/paper/21.pdf

[13] T. Lukkari and V. Välimäki, "Modal synthesis of wind chime sounds with stochastic event triggering," in *Proc. 6th Nordic Signal Process. Symp. (NORSIG)*, Espoo, Finland, Jun. 2004, pp. 212–215 [Online]. Available: http://wooster.hut.fi/publications/norsig2004/61_LUKKA.PDF

[14] D. Rocchesso, R. Bresin, and M. Fernström, "Sounding objects," *IEEE Multimedia*, vol. 10, no. 2, pp. 42–52, Apr./Jun. 2003.

[15] M. Rath and D. Rocchesso, "Continuous sonic feedback from a rolling ball," *IEEE Multimedia*, vol. 12, no. 2, pp. 60–69, Apr./Jun. 2005.

[16] F. Avanzini, S. Serafin, and D. Rocchesso, "Interactive simulation of rigid body interaction with friction-induced sound generation," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 1073–1081, Oct. 2005.

[17] J.-M. Adrien, "The missing link: Modal synthesis," in *Representations of Musical Signals*, G. De Poli, A. Piccialli, and C. Roads, Eds. Cambridge, MA: MIT Press, 1991, pp. 269–297.

[18] E. R. Miranda and J. Matthias, "Granular sampling using a pulse-coupled network of spiking neurons," in *EvoWorkshops 2005*, F. Rothlauf, Ed., Berlin, Germany, Lecture Notes in Computer Science 3449, pp. 539–544.

[19] B. L. Vercoe, W. G. Gardner, and E. D. Scheirer, "Structured audio: Creation, transmission, and rendering of parametric sound representations," *Proc. IEEE*, vol. 86, no. 5, pp. 922–940, May 1998.

[20] P. R. Cook, "Physics-based sound synthesis for graphics and interactive applications," presented at the Proc. ACM SIGGRAPH, San Diego, CA, 2003, Course Notes #36.

[21] ——, "Physics-based synthesis of sound effects," in *Proc. Game Developer's Conf.*, San Jose, CA, 2003 [Online]. Available: http://www.gamasutra.com/features/gdcarchive/2003/Cook_Perry.ppt

[22] M. Puckette, "Pure data: Another integrated computer music environment," in *Proc. 2nd Intercollege Comput. Music Concerts*, Tachikawa, Japan, 1996, pp. 37–41 [Online]. Available: http://pure-data.iem.at

[23] O. Matthes, Freeverb ~ 2004 [Online]. Available: http://www.akustische-kunst.de/maxmsp

[24] M. R. Schroeder, "Natural sounding artificial reverberation," *J. Audio Eng. Soc.*, vol. 10, no. 3, pp. 219–224, 1962.

[25] J. A. Moorer, "About this reverberation business," *Comput. Music J.*, vol. 3, no. 2, pp. 13–28, 1979.

[26] L. Rabiner, M. Cheng, A. Rosenberg, and C. McGonegal, "A comparative performance study of several pitch detection algorithms," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. AASP-24, no. 5, pp. 399–418, Oct. 1976.

[27] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

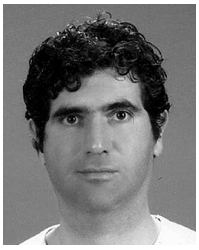
[28] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, "Frequency-zooming ARMA modeling of resonant and reverberant systems," *J. Audio Eng. Soc.*, vol. 50, no. 12, pp. 1012–1029, 2002.

[29] J. O. Smith, III, "Physical audio signal processing: Digital waveguide modeling of musical instruments and audio effects," Stanford, CA, 2004 [Online]. Available: http://ccrma.stanford.edu/~jos/pasp04/

[30] A. Friberg and J. Sundberg, "Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners," *J. Acoust. Soc. Amer.*, vol. 105, no. 3, pp. 1469–1484, 1999.
[31] R. Bresin, A. Friberg, and S. Dahl, "Toward a new model for sound control," in *Proc. Conf. Digital Audio Effects*, Limerick, Ireland, 2001, pp. 45–49.
[32] S. Strogatz, *Sync: Rhythms of Nature, Rhythms of Ourselves*. Baltimore, MD: Penguin, 2003.
[33] Y. Kuramoto and I. Nishikawa, "Statistical macrodynamics of large dynamical system. Case of plane transition in oscillator communities," *J. Statist. Phys.*, vol. 49, pp. 569–605, 1987.
[34] J. A. Acebrón, L. L. Bonilla, C. J. P. Vicente, F. Ritort, and R. Spigler, "The Kuramoto model: A simple paradigm for synchronization phenomena," *Rev. Mod. Phys.*, vol. 55, no. 1, pp. 137–185, 2005.

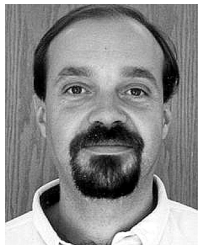
Leevi Peltola was born in Jyväskylä, Finland, in 1980. He received the M.Sc. degree in electrical and communications engineering from the Helsinki University of Technology (TKK), Espoo, Finland, in 2004.

He is currently a software designer at EKS, Hesinki, Finland.



Cumhur Erkut was born in Istanbul, Turkey, in 1969. He received B.Sc. and the M.Sc. degrees in electronics and communication engineering from the Yildiz Technical University, Istanbul, Turkey, in 1994 and 1997, respectively, and the Dr.Sc.(Tech.) degree in electrical engineering from the Helsinki University of Technology (TKK), Espoo, Finland, in 2002.

From 1998 to 2002, he worked as a Researcher at the Laboratory of Acoustics and Audio Signal Processing, TKK. He is currently a Postdoctoral Researcher in the same institution, where he contributes to the research projects "Sound to Sense, Sense to Sound" (S2S2) [FP6, FET, IST-2004-03773] and "Modeling and Perception of Sound Sources" (MAPS) [Academy of Finland, 105651]. His primary research interests are model-based sound synthesis and musical acoustics.



Perry R. Cook (S'94–M'90) attended the Conservatory of Music, University of Missouri at Kansas City (UMKC), from 1973 to 1977, studying voice and electronic music and received the B.A. degree in music and the B.S. degree in electrical engineering from UMKC in 1985 and 1986, respectively, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1990.

He worked as a Sound Engineer and Designer from 1976 to 1981. He continued as Technical Director of the Stanford Center for Computer Research in Music and Acoustics, until joining the faculty of Princeton University, Princeton, NJ, in 1996, where he is now Professor and Associate Chair of Computer Science, with a joint appointment in Music. His primary research interests are the acoustics and simulation of the voice and musical instruments, human perception of sound, and interactive devices for expressive musical performance.

Dr. Cook was the recipient of a 2003 Guggenheim Fellowship to write a new book on the subject of technology and the voice.

Vesa Välimäki (S'90–M'92–SM'99) was born in Kuorevesi, Finland, in 1968. He received the M.Sc., the Licentiate of Science, and the Doctor of Science degrees in technology, all in electrical engineering, from the Helsinki University of Technology (TKK), Espoo, Finland, in 1992, 1994, and 1995, respectively. His doctoral dissertation dealt with fractional delay filters and physical modeling of musical wind instruments.

He was a Postdoctoral Research Fellow with the University of Westminster, London, U.K., in 1996. From 1996 until 2001, he held the position of Senior Assistant at the TKK Laboratory of Acoustics and Audio Signal Processing. He was on leave as a Postdoctoral Researcher under a grant from the Academy of Finland for part of this period. During the academic year 2001–2002, he was Professor of signal processing at the Pori unit, Tampere University of Technology, Pori, Finland. Since 2002, he has been Professor of audio signal processing in the Department of Electrical and Communications Engineering, TKK. He was appointed Docent in signal processing at the Pori unit of the Tampere University of Technology in 2003. His research interests include sound synthesis, musical acoustics, and digital filter design.

Prof. Välimäki is a member of the Audio Engineering Society, the Acoustical Society of Finland, and the Finnish Musicological Society. He was the Secretary of the IEEE Finland Section in 2000 and 2001. He is an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, and he is a Guest Editor of the March 2007 special issue of the IEEE *Signal Processing Magazine* on signal processing for sound synthesis.