

Importance and Representation of Phase in the Sinusoidal Model

Tue Haste Andersen, Kristoffer Jensen
Department of Computer Science, University of Copenhagen
email: {haste,krist}@diku.dk

November 2004*

Abstract

In this paper work is presented on the representation and perceptual importance of phase. Based on a standard sinusoidal analysis/synthesis system, the phase alignment of the sound components is analyzed. A novel phase representation, partial period phase, is introduced, which characterizes phase evolution over time with an almost stationary parameter for many musical sounds. The proposed partial period phase representation is used to control the phase when synthesizing sounds. Sounds synthesized with varying amount of phase information are compared in a listening experiment with 11 subjects. It is shown that phase is of great importance to the perception of sound quality of common harmonic musical sounds, but indications are found that phase is not of importance to the slightly inharmonic piano sounds. In particular, the sound degradation is large for low-pitched sounds, approaching *Slightly Annoying* when no phase information is used. In addition, a model based on the partial period phase representation has a significantly better perceived sound quality than sounds with random phase shifts.

0 Introduction

Synthetic sounds are more and more common in contemporary music. Sampling and algorithmic synthesis techniques are used not only for artistic reasons, but also to improve sound quality and reduce production costs. Sampling provides good sound quality, but is limited to the timbre of the recorded sound, with few parameters to control the timbre. One solution to this problem is to use parametric analysis/synthesis methods. These methods combine models with good sound quality and large parameter space, allowing near real-time sound transformations, for changing the perceptually important characteristics [1]. The additive (sinusoidal) sound model [2] is the most widely used parametric analysis/synthesis method in sound modeling research. The additive parameters form a convenient representation for the deterministic part of the sound, allowing for easy transformations of perceptually important characteristics such as duration, pitch and timbre.

Literature studies using harmonic tones show that the perceived sound quality is influenced by synthesizing tones with the same harmonic content, but with different phase shifts. However, it is not clear to what extent these effects are important to the perception of natural occurring sounds.

We propose a novel phase representation, partial period phase, describing phase evolution over time. The partial period phase representation is a convenient way to

*Final accepted manuscript, paper to appear in J. Aud. Eng. Soc., 52(11):1157-1169.

represent phase in the additive model. The representation solves several problems with comparison of consecutive phase values in time, resulting in near constant partial phase trajectories.

Using the additive analysis/synthesis framework, we conducted a listening experiment where resynthesized sounds were compared with original recorded sounds. These sounds were resynthesized with phase information, without phase information, and with two different models of phase. The perceived degradation was judged by 11 subjects, and compared using analysis of variance.

In the following section, previous work related to perception of phase is overviewed. Section 2 gives an overview of the sinusoidal analysis technique and section 3 describes the improved phase representation. Section 4 presents the listening experiments and conclusions are given in section 5.

1 Previous work

In this section, literature showing the relevance of phase in musical sound modeling is briefly reviewed. From the experiments reported here it is clear that phase is of importance to the perception of complex tones, and especially to the perception of timbre. The question remains as to what extent phase is of importance to the quality of naturally occurring sounds, and if so, how phase can be modeled in sound transformations and pure synthesis.

1.1 Does phase affect timbre?

One of the first experiments concerning the perception of timbre of complex tones in relation to phase was conducted by von Helmholtz [3]. Using a special technique he was able to generate complex tones consisting of eight sinusoids (partials) with variable phase and fundamental frequencies of 120 and 240 Hz. Helmholtz concluded that “the changes in timbre are not distinct enough to be observed after a few seconds required to alter the phases; anyhow these changes are too small to transform one vowel in another”, and “harmonics beyond the sixth to eighth give dissonances and beats, so it is not excluded that, for these higher harmonics, a phase effect does exist.” These conclusions have often been interpreted as indicating that phase did not have any influence on timbre [4], even though later experiments showed otherwise [5, 6].

Plomp and Steeneken [4] conducted a number of experiments involving complex tones with ten harmonic partials and equal spectral envelope, but different phase shifts. The most important finding was that the maximum effect of phase on timbre perception occurs when a tone containing harmonic partials that all start at sine phase (0°) is compared to one where the partials alternate between sine phase and cosine phase (90°). The effect of reducing the level of each successive partial by 2 dB was greater than the maximum phase effect described above. Also the effect of phase on timbre appeared to be independent of sound level.

Patterson [7] presented psycho-acoustic experiments involving alternating phase (APH) waves, that is, harmonic partials in which even partials start in cosine phase, while odd ones start in cosine phase + D° . It was found that the value of D leading to a just noticeable difference (JND) between a sound with partials in cosine phase, and an APH sound was lower for sounds with high bandwidth, low repetition rate and high signal level. The signal duration was found to have no, or very little, effect on the JND.

Progressively improved models, using summary autocorrelation [8, 9], auditory imaging [10], and models including the behavior of early cortical stages, using the summary measure of spectrogram [11], have provided explanations for the observed effect of phase.

Alcántara et al. [12] studied the influence of phase on identification of vowel-like sounds. The “vowel” were created by increasing the level of three pairs of successive harmonic partials. They found better identification when the components had cosine starting phase than when they had random phase, and poorer performance for weaker stimuli. Pressnitzer and McAdams [13] studied the influence of phase on roughness perception, and found that roughness is linked to shapes of the waveforms at the output of the simulated auditory filter. Roberts et al. [14] showed that phase shifts could influence stream segregation for rapid sound sequences. Gockel et al. [15] studied the influence of phase on loudness and forward masking produced by harmonic complex tones. They found that a tone with components added in cosine phase was louder but was a less effective forward masker than a tone with components added in random phase.

Whereas phase changes are detectable in controlled situations, as even polarity change was found to be audible in two-component signals [16], the discrimination of phase changes in individual components often requires specific phase alignment, such as cosine phase [17].

1.2 Importance of phase in transients

Patterson and Green [18] used Huffman sequences, in which the phases can be varied independently of the energy spectrum, to assess the discrimination of phase changes in transients. They found that phase changes could be discriminated reliably, for some stimuli waveforms, for durations above 5 ms.

Wakefield et al. [19] conducted a study of the perception of transients using filtered noise, where a two-interval forced-choice adaptive psycho-physical procedure was used to find the just noticeable difference (JND) between a given sound and a copy of the sound, where the magnitude spectrum was smoothed and the phase spectrum held constant. The surprising result was that the JND depended strongly on the phase pattern used. It was concluded that “the effect for short duration signals is greater than what the (sparse) literature on the auditory perception of transients would suggest.” The perception of clicks and chirps was further investigated by Uppenkamp et al. [20]. The up-chirps used by Uppenkamp et al. are signals constructed to contain the same frequencies as clicks, but where the phase is manipulated to compensate for the spatial dispersion along the cochlea. Up-chirps should therefore reach maximum amplitude at the same moment in time at all places of the basilar membrane. They compared the perceived ‘compactness’ of clicks to that of chirps, and found that clicks were perceptually more compact than up-chirps, but that down-chirps, that is up-chirps reversed in time, sounded more compact than up-chirps. Even though up-chirps are aligned in time at the basilar membrane output, they have a longer within-channel impulse response than down-chirps and clicks. This suggests that “the perceived ‘compactness’ of a sound is apparently more determined by the fine structure of excitation within each peripheral channel than by between-channel phase differences.”

1.3 Phase models

Schroeder [21] reported a number of effects related to sounds with up to 31 harmonic partials. Most interesting is the reported strong dependence of timbre on

peak factor. The peak factor can be minimized via an analytical approximation equation [22]. The synchronization index model (SIM) of Leman [23] employs a functional model of the auditory periphery, and a method of predicting the roughness of a sound. This model was used by Tind and Jensen [24] to devise a propagation formula of the phase shifts that control the roughness output of the SIM. By basing the propagation formula on the roughness prediction for three partials, they obtained a correspondence between the roughness control parameter and the predicted roughness for complex harmonic sounds. They concluded that there exists a (non-unique) phase shift for a given perceptual roughness of complex harmonic sounds.

2 Sinusoidal analysis/synthesis

This study is based on the analysis by synthesis methodology [25], by use of additive (sinusoidal) analysis/synthesis techniques. In the additive framework, sounds are modeled as a sum of sinusoids with time-varying amplitudes, frequencies and sometimes also phase shifts.

The Short-Time Fourier Transform (STFT) [26] is a related technique that can be used for analysis/synthesis and transformations of sounds [27]. In the STFT, overlapping blocks of the windowed sound are Fourier transformed, modified and inverse Fourier transformed. However, for harmonic sounds or sounds with strong partials, the frequency components between the strong partials are masked. Because a large number of the frequency components in the STFT are masked, the number of parameters used to model the sound can be greatly reduced. This is the assumption in the additive model that is used in this work. The additive model is chosen for two reasons: First, it is well suited for further high-level modeling of musical sounds. Second, the additive model is being used in many research and development prototypes today [28, 29, 30, 31], and thus it provides a stable framework for exploration in perception of natural sounds. The additive model has, however, several shortcomings. The transients are often smeared in block-based analysis/synthesis and noise is not well represented.

Several methods exist for determining the time-varying amplitudes and frequencies of the harmonic partials. Already in the last century, musical instrument tones were divided into their Fourier series [3]. Early techniques for the time-varying analysis of the additive parameters are presented by Matthews et al. [32] and Freedman [33]. Today, the most common technique for the additive analysis of musical signals is based on STFT analysis [2]. In order to retain the noise components, several noise models of musical sounds have been presented, including the residual noise model in the Fast Fourier Transform (FFT) [28, 2], the bandwidth-enhanced additive synthesis [34, 1], and the narrow band basis functions (NBBF) in speech models [35]. In order to improve the frequency, and in particular the time resolution, i.e., to better retain the transient behavior of percussive musical instruments, time-frequency based [1, 36] methods could be used, and the time and frequency reassignment method [37] has recently gained popularity [38, 34]. Ding and Qian [39] have presented an interesting method for improving the time resolution, fitting a waveform by minimizing the energy of the residual. This was improved and dubbed adaptive analysis by Robel [40].

We used a software package developed by the authors [1, 41], previously used in explorations of timbre of musical instruments. It is based on the classic peak-picking method, where overlapping blocks are windowed, and the amplitudes, frequencies and phases are found from interpolated peaks of the magnitude of the

FFT. This method has been shown to work well in forming stable sinusoidal tracks over time from analyzed recordings of instrument sounds. The synthesis quality of a comparable analysis method, when phase information is not used, has previously been measured to be equal to, or better than *perceptible, but not annoying*, when compared to the original recorded monophonic sounds [1].

2.1 Analysis

For each sound under analysis, the fundamental frequency ω_0 is estimated using autocorrelation [42]. This method, which is applicable only to monophonic quasi-harmonic sounds, is used to determine the fixed block size used in the analysis of the given sound. For each block of sound, k , under analysis, a new local measure of the fundamental frequency, ω_0^k , is calculated, again by use of autocorrelation. From this measure, a FFT is performed and a search for peaks is done near the regions of the quasi-harmonic frequencies. The amplitude A_i^k and frequency ω_i^k are stored for each partial i and time frame k .

In order to retain more of the additive noise components, a method inspired by the NBBF [35] has been employed here, in which sinusoids are estimated in between the harmonic partials if the fundamental frequency is above 400 Hz. This method essentially retains noises such as hammer noise, or the additive noise in wind instruments.

Peaks from adjacent blocks are connected to form sinusoidal tracks. The system has been extended to output not only the amplitude and frequency of the tracks, but also the phase for each block, θ_i^k . To model the phase over time, high precision of the estimated phase values is required. To achieve this, it was found necessary to extend the length of each analysis block from 2.8 periods of the fundamental period length to 4 periods. By doing this, the time resolution is affected, and thus the sound quality of transient sounds is degraded.

2.2 Synthesis

The sound is synthesized using the analysis parameters in the following way:

$$s(n) = \sum_{i=0}^N A_i(n) \cos(\theta_i(n)), \quad (1)$$

for N partials, where $\theta_i(n)$ denotes the time varying phase for partial i and sample index n . In practice, the values of $A_i(n)$ used in the synthesis are obtained by linear interpolation of the measured amplitude values between the block boundaries. Two methods for finding the phase, θ_i are used:

S_a : Synthesis without measured phase information

S_b : Synthesis with measured phase information

When synthesizing sound without the measured phase information θ_i^k , the phases of the sinusoidal tracks are found by the cumulative sum of the interpolated frequency values over time:

$$\theta_i(n) = \sum_0^n \omega_i(n). \quad (2)$$

When synthesizing the sound using phase information, the phase trajectory is interpolated in such a way that boundary conditions are satisfied. This can be done by cubic interpolation [2] of the phase using the measured frequency and phase

values. In this way, the measured phase and frequency values are preserved at the block boundaries, but oscillating frequency tracks can occur between the block boundaries [39]. A solution to this problem is to use quadratic interpolation where the phase and frequency values cannot be preserved at the boundaries. Instead a weighting factor is used to determine the importance of the estimated phase relative to the frequency [39]. However, no degradation caused by the oscillations in frequency has been found in this work; therefore the cubic interpolation method is used.

3 Phase representation

The goal of additive phase modeling is to improve sound quality in pure synthesis models such as the Timbre Engine, based on the timbre model [31], to improve sound quality in time/frequency scaling of signals, and finally to gain a better understanding of the perception of musical signals. We chose to investigate the phase as a function of time, and thus a convenient representation of the phase trajectories over time is needed.

3.1 Visualization

The phase is rarely considered or even visualized in sound modeling literature. A way to visualize the phase is to use a spectrogram, but plotting phase instead of energy of the discrete Fourier transform. Figure 1 demonstrates this for two types of musical sounds. On the top is plotted a stationary part of a soprano voice, where the phase progresses in a coherent way through time and frequency. The attack of a guitar note is plotted in the lower part of the figure. Here the phase evolution over time is less coherent. The goal of the phase representation presented in this section is to describe stable sounds, such as the sustained part of most sounds from musical instruments, using a few parameters.

3.2 Phase delay

The phase values obtained from the Discrete Fourier Transform, $\theta(\omega)$, and thus also the values used in additive analysis, are specified as the phase shift in radians for each sinusoidal component. Another way to represent phase is as *phase delay* [43]:

$$P(\omega) = -\frac{\theta(\omega)}{\omega}, \quad (3)$$

where $\theta(\omega)$ is the phase at frequency ω , and $P(\omega)$ expresses the time delay in seconds relative to the center of the frame. The magnitude, phase and phase delay as function of frequency of a stable part of a saxophone sound are shown in figure 2. Phase delay is not a common way to represent phase in the sinusoidal model. However, it is shown here as it is used as the basis of the relative phase delay described in the following section.

3.3 Relative Phase Delay

Waveform preservation when performing time or pitch scaling of harmonic sounds can be achieved without the use of a specific phase representation, by using analysis step sizes exactly equal to the fundamental period length. However, in many cases it is not convenient to have forced non-constant step sizes during analysis, and thus another phase representation is needed. To overcome this problem, while

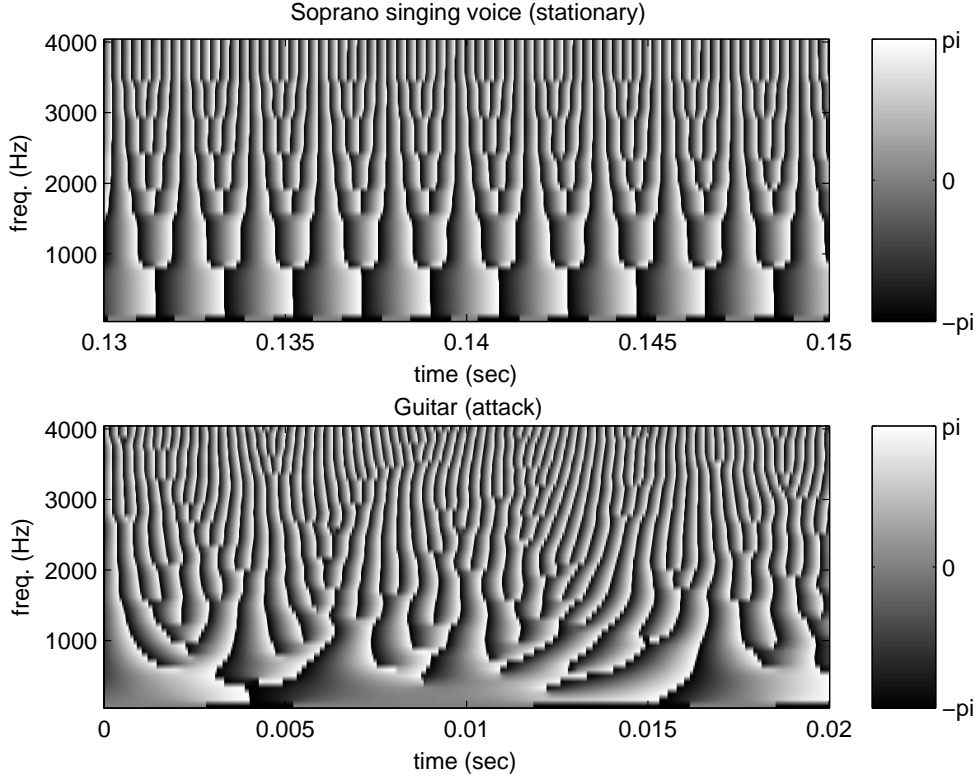


Figure 1: Phase as a function of time and frequency. The brightness represent the phase in radians between $-\pi$ and π . Sustained part of a soprano voice (top), and attack of a guitar (bottom). The fundamental frequency of both sounds is approximately 500 Hz.

still being able to preserve the shape of the waveform when doing time or pitch scaling, Di Federico proposed a representation, relative phase delay (RPD) [44], based directly on additive parameters, by representing the phase trajectories as phase delays relative to the phase delay of the first partial. When performing time scaling, the amplitudes and frequencies of the partials are left untouched, but the phase values of the fundamental are updated using a propagation formula. After the new phase values of the fundamental are found, the phase values of the other partials are changed, based on their position relative to a fixed point in the fundamental period.

RPD is based on the definition of phase delay from equation (3), and it is defined as

$$\tau_{i,k} = \frac{\theta_{i,k}}{\omega_{i,k}}, \quad (4)$$

where i is the index of the partial and k is the analysis frame index. τ expresses the distance in time between the analysis frame center and a specific point in the partial period.

The relative phase delay is defined for the partials as the difference between the phase delay of the fundamental and the partial i :

$$\Delta\tau_{i,k} = \tau_{i,k} - \tau_{1,k}. \quad (5)$$

Since the relative phase delay, $\Delta\tau_{i,k}$ for $i = 2 \dots N$ is defined relative to a fixed point in the fundamental period, the overall waveform characteristics are preserved

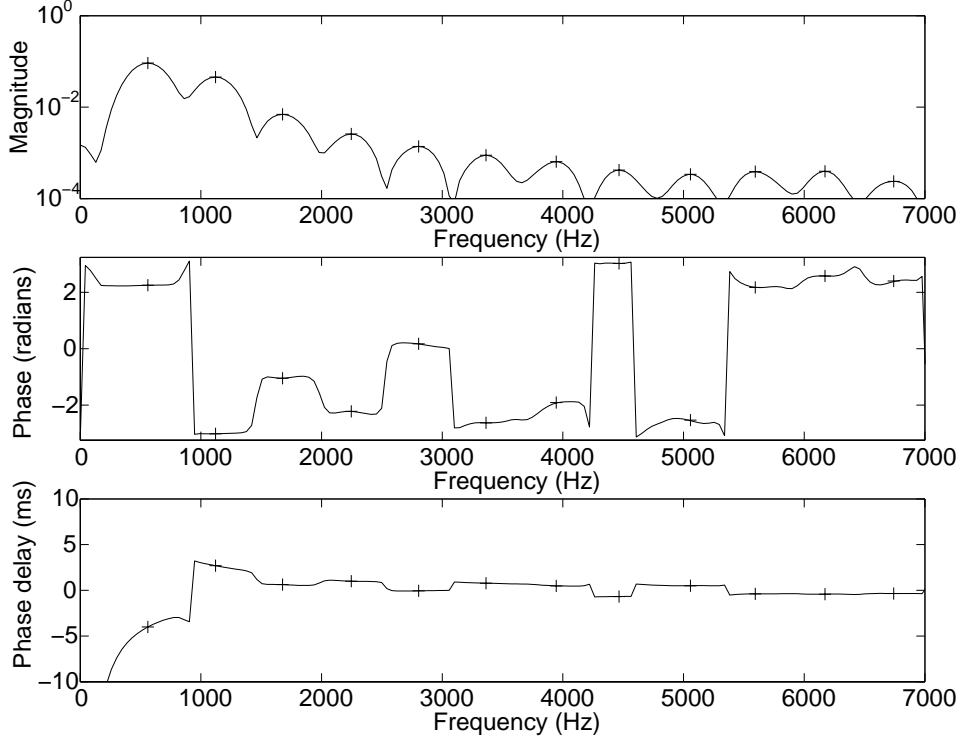


Figure 2: Magnitude (top), phase (middle), and phase delay (bottom) for one analysis frame of the sustained part of a saxophone sound. Spectral peaks are identified in the magnitude plot, and marked with ‘+’.

and thus the phase delay of the fundamental can be chosen at random. Having modified the phase delay of the fundamental, phase delays of the other partials are converted back into phase values:

$$\theta_{i,k} = \text{mod} \left\{ \left(\frac{\theta_{1,k}}{\omega_{1,k}} + \Delta\tau_{i,k} \right) \omega_{i,k}, 2\pi \right\} \quad k = 2, \dots, N, \quad (6)$$

where $\theta_{1,k}$ is the phase of the fundamental.

Relative phase delay is a representation that works well for harmonic sounds, in that the waveform characteristics are preserved. However, to actually use this representation it is necessary to take phase wrapping into account. Another problem with the relative phase delay is that the phase delay calculated from the wrapped phase values approaches zero as the frequency increases. This makes it difficult to compare relative phase delays and plot them visually. Finally, if the sound is slightly inharmonic, a drift in the relative phase delays will occur for the partials. To show this, imagine a nearly harmonic signal with two sinusoids of start phase 0, one with frequency of 110 Hz, and one with frequency of 225 Hz. The sound is analyzed using a step size of 100 ms. In the first analysis frame, the relative phase delay between the first and second partials is 0. At the next frame, $t = 0.1\text{s}$, the phase of each partial is:

$$\begin{aligned} \theta_{0,1} &= \text{mod}(0.1\text{s} \cdot 2\pi \cdot 110\text{Hz}, 2\pi) = 0 \quad \text{and} \\ \theta_{1,1} &= \text{mod}(0.1\text{s} \cdot 2\pi \cdot 225\text{Hz}, 2\pi) = \pi \end{aligned} \quad (7)$$

The phase delay of the first partial is $\tau_{0,1} = 0/110\text{Hz} = 0\text{s}$. The phase delay of the second is evaluated at an integer multiple of the “fundamental frequency”,

$2 \cdot 110\text{Hz} = 220\text{Hz}$, $\tau_{1,1} = \pi/220\text{Hz} \approx 0.014\text{s}$, and thus a drift in the relative phase delay of the second partial has occurred. If the second partial had been at frequency 220 Hz, no drift would have occurred, and the relative phase delay representation would have given a usable result. This drift can be demonstrated on synthetic and recorded signals [41].

3.4 Fundamental Period Phase Representation

To overcome some of the problems of the relative phase delay, an improved phase representation is proposed. One of the goals that was achieved with the RPD was that phase values between frames could be compared. This is usually possible only in a frame-based analysis when the frame size is exactly an integer multiple of the fundamental period. In this case, the phase value is measured at the same point in the waveform period for successive frames, and is thus comparable between frames. In the RPD this problem was overcome by using phase delays. Another way to make the phase values comparable between frames is used in the fundamental period phase representation. In this representation, the measured phase values of the fundamental, $\theta_{k,0}$, for a given block k , are corrected by a linear change in phase, corresponding to a time difference $\Delta t_{k,0}$ at the measured frequency, $\omega_{k,0}$. $\Delta t_{k,0}$ is defined as the time difference between a fixed point in the fundamental period, that is, a point in the period which is the same between frames, and the point where $\theta_{k,0}$ is measured. The point where $\theta_{k,0}$ is measured is dependent on the step size R_a . More formally $\Delta t_{k,0}$ can be found by the following formula, where $L_{k,0}$ represents the length of the fundamental period in the k th frame. The modulus function is used to ensure that the phase correction stays in the interval between $-\pi$ and π :

$$\Delta t_{k,0} = \left(1 - \text{mod} \left\{ \frac{R_a - \Delta t_{k-1,0}}{L_{k-1,0}}, 1 \right\} \right) L_{k,0}, \quad (8)$$

noting that $L_{k,0}$ can be found by knowing the frequency of the fundamental $\omega_{k,0}$:

$$L_{k,0} = \frac{2\pi}{\omega_{k,0}}. \quad (9)$$

The corrected phase for the fundamental $\phi_{k,0}$ can now be found:

$$\phi_{k,0} = \theta_{k,0} + \Delta t_{k,0}\omega_{k,0}. \quad (10)$$

The phase values of the other partials are corrected using the same time difference $\Delta t_{k,0}$ that was used in correcting the fundamental:

$$\phi_{k,i} = \theta_{k,i} + \Delta t_{k,0}\omega_{k,i}. \quad (11)$$

This representation is called fundamental period phase representation, and is equivalent to the RPD, apart from the fact that the RPD uses phase delays measured in time to represent the phase differences, whereas radians are used in the fundamental period phase representation. This means that now we have a representation preserving the waveform characteristic and allowing for comparison of phase values between frames, as in the RPD representation. Furthermore, the new way of representing the phase solves the phase unwrapping problem of the RPD.

3.5 Partial Period Phase Representation

To correct for drifting phase values in inharmonic sounds, an improvement to the fundamental period phase representation is proposed, the partial period phase

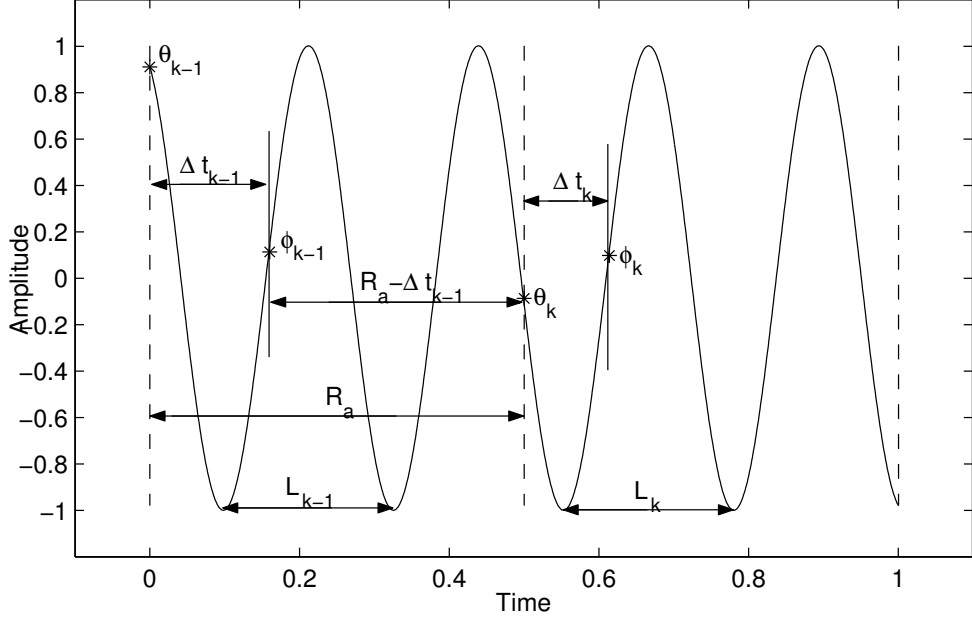


Figure 3: Schematic drawing of partial period phase value for one partial. The vertical dotted lines represents the block boundaries, with θ as the measured phase value for each block. ϕ is the PPP value, found by knowing the frequency of the partial, and the step size R_a . As seen in the figure ϕ_k and ϕ_{k-1} are at the same position in the partial period, even though the phase values, θ_k and θ_{k-1} are measured at different positions in the partial period.

(PPP), in which the phase is expressed relative to the same point between frames of the partial period instead of relative to a point in the fundamental period. The method presented here bears some similarity to the phase propagation employed in STFT based phase vocoders [45] when time or pitch scaling a signal.

Equations (10) and (8) then give:

$$\phi_{k,i} = \theta_{k,i} + \Delta t_{k,i} \omega_{k,i}, \quad (12)$$

where k is the frame number, i is the index of the partial, and $\Delta t_{k,i}$ is the time difference between the point at which the phase value is measured, and the corrected value (see figure 3):

$$\Delta t_{k,i} = \left(1 - \text{mod} \left\{ \frac{R_a - \Delta t_{k-1,i}}{L_{k-1,i}}, 1 \right\} \right) L_{k,i}. \quad (13)$$

Figure 4 shows an example of the difference between the fundamental period phase representation and the PPP representation. A segment of a piano sound is analyzed, and the corrected phase values for the first five partials are shown in the figure. The piano sound is known to have stretched harmonic frequencies [46], and thus effectively demonstrates the problem with RPD and fundamental period phase representation. The partial period phase representation (bottom) is clearly superior to the fundamental period phase representation (top), removing the phase drift caused by non-harmonic partials.

All phase representations presented here preserve the phase information and thus no degradation in sound quality results from use of these representations. By

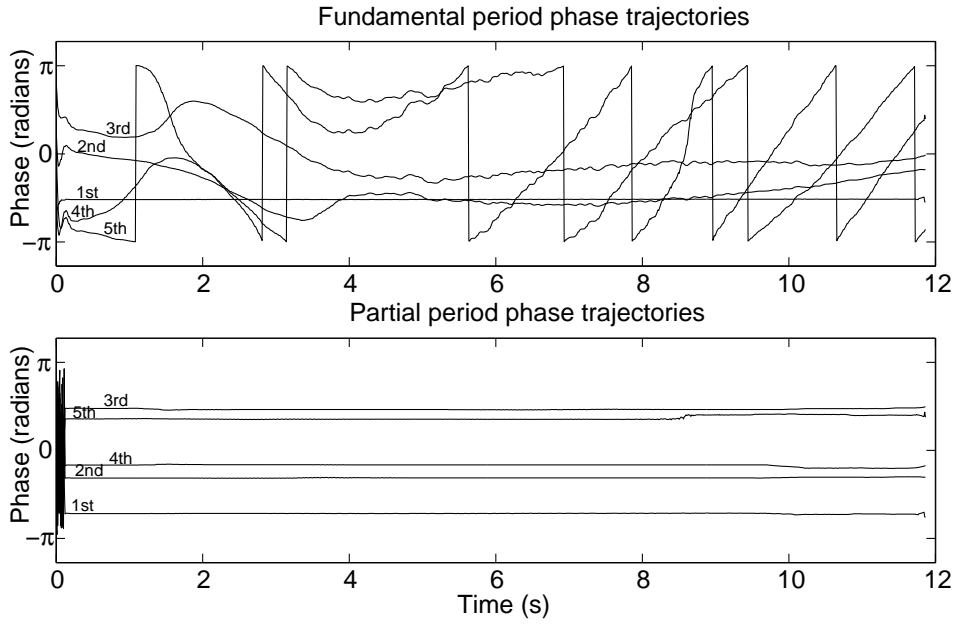


Figure 4: Fundamental period phase (top) and partial period phase (bottom) for the first five partials of a piano sound. The noise at the start of the sound is due to the transient sound of the piano attack, for which no stable frequency information can be estimated. The partial period phase is clearly superior to the fundamental period phase, in that the phase trajectories are nearly constant over time for the sustained part of the sound.

substituting R_a in equation (13) with the synthesis frame length R_s we obtain the phase values used in the cubic interpolation, when synthesizing:

$$\theta_{k,i} = \phi_{k,i} - \Delta t_{k,i} \omega_{k,i}. \quad (14)$$

Using equation (10), all partials in a given analysis frame are analyzed relative to the same time in the fundamental period. In the partial period representation of equation (12), each partial phase value is evaluated relative to the same point in the last analysis frame of the particular partial. In a transient or low energy part of the sound, the estimation of the partial frequencies is likely to fail, resulting in new absolute phase values, and thus occurrence of transients has to be taken into consideration when using the partial period phase representation. In practice the phase representation and modeling should be coupled with a transient detector, to signify in which portions of the sound the phase values are comparable.

4 Experiment: The importance of phase

The purpose of this experiment is to determine how important phase is with regard to sound quality, when synthesizing monophonic singing voice or other musical instruments. Recorded instrument sounds are analyzed using the method described in section 2, followed by modification of the phase trajectories using the partial period phase representation. Finally the sounds are resynthesized from the modified analysis parameters, and compared in a listening experiment.

4.1 Sound reproduction methods

Five conditions were used in a repeated measures full factorial experiment. In each condition the original sound was compared to one of the following sounds:

1. Original sound (ORG)
2. Synthesized sound, with full phase information, maintaining absolute and relative phase (ARP)
3. Synthesized sound, with phase information, maintaining relative phase (RP)
4. Synthesized sound, constant partial period phase, approximating absolute phase values in the stationary part (AP)
5. Synthesized sound, no phase information (NP)

For the synthesized sounds, method S_b described in section 2.2 was used, except for the sound with no phase information, where method S_a was used.

In ARP all phase information is preserved, and thus no modification to the phase information is done between analysis and synthesis.

To change the partial phase trajectories in RP and AP, we use the partial period phase representation. In synthesizing RP the relative phase shift between each partial is preserved, but the absolute phase value of each partial is discarded. This is accomplished by randomizing the start phase of each phase trajectory in the partial period phase representation.

In AP, the partial period phase trajectory is approximated with a constant, κ_i . The start phase is chosen so that the phase is close to the phase of the original sound in the stationary part of the sound. This method poses a problem as to how the attack should be synthesized. The attack can be synthesized with random start phases as in NP, but then the phases have to be interpolated at the start frame of the stationary sound, k_s , to the phase value κ_i . Another alternative is to synthesize the attack with full phase information as in ARP, to minimize the phase difference between the attack and stationary part at frame k_s . This last method is chosen to avoid artifacts introduced by interpolation between frame $k_s - 1$ and k_s . The selection of k_s was done by manual segmentation. κ_i was approximated by the following function:

$$\kappa_i = \sum_{k=k_s}^{k_e} \frac{\phi_{i,k} A_{i,k}}{\sum_{k=k_s}^{k_e} A_{i,k}}, \quad (15)$$

where $\phi_{i,k}$ is the partial period phase for partial i at frame k , k_s is the first frame of the stationary part of the sound, and k_e is the ending frame. The function weights the phase values in the high-energy portion of the sound more highly than phase values in low energy portions of the sound. As the AP synthesis method retains the waveform of the high-energy harmonic partials, the resulting AP sounds have a waveform that resembles the original waveform.

Sound is synthesized with no phase information (NP) using synthesis method S_a described in section 2.2. The start phase of each trajectory is randomized.

4.2 Sounds

The sounds used in the experiment were bass clarinet, bass trombone, cello, piano, and singing voice. The instruments were chosen to represent a broad spectrum of acoustic instruments, and because of their extended pitch range.

Instrument	f_1 (Hz)	f_2 (Hz)	f_3 (Hz)	f_4 (Hz)	f_5 (Hz)
Bass clarinet	59.2	109.7	176.0	295.0	469.2
Bass trombone	58.8	117.0	237.6	263.2	469.7
Cello	65.6	98.3	131.2	196.1	525.4
Piano	49.4	97.9	132.3	263.2	526.6
Singing voice	82.5	109.5	216.7	273.7	391.0

Table 1: Reference sounds used in listening experiment.

Score	Impairment
5.0	Imperceptible
4.0	Perceptible, but not annoying
3.0	Slightly annoying
2.0	Annoying
1.0	Very annoying

Table 2: Scale used in listening experiment

The singing voice is probably the best known musical instrument, the piano is chosen for its percussive, slightly inharmonic sound, the cello has a particular sound because of the jitter introduced by the bow-string interaction. The bass clarinet and bass trombone are examples of reed and lip driven wind instruments. These instruments are believed to be representative of today’s music instruments, and in particular, the different aspects of timbre in common instruments that could influence the perception of phase. For all instruments, recordings of five fundamental frequencies were used, as shown in table 1. The length is approximately two seconds for most sounds, and they were played *forte* for the most part.

4.3 Procedure

We used the double blind triple stimulus with hidden reference method [47], for assessing perceptual differences between original and synthesized sounds. This method was used in previous studies to measure the sound quality of timbre models based on analyzed sounds without phase information [1]. In these experiments it was found that the sound quality was dependent on the fundamental frequency of the synthesized sounds. In general, resynthesized sounds with high fundamental frequency were rated to have less degradation than resynthesized sounds with low fundamental frequency.

For each sound, the subject first heard a reference sound, in this case the original recorded sound, followed by two sounds in random order, where one was the reference, and the other was the experimental sound, a sound from one of the five conditions described above. The subject was then asked to rate the degradation of the two sounds relative to the reference. The degradation of the sounds was rated on a scale from 1 to 5, where 1 corresponds to *very annoying* and 5 corresponds to *imperceptible*. One decimal could be used in the rating, and one of the sounds had to be given the score 5. The full scale is shown in table 2. The subjects were allowed to listen to the sounds as many times as found necessary.

Statistical analysis of the results was done using repeated measures analysis of variance with synthesis method, instrument type, and fundamental frequency as independent variables. The individual levels were compared using a post-hoc

least significant difference (LSD) test at a 0.05 significance level. The results were evaluated according to a measure of degradation, which is the difference between the rating of the reference, S_r and the rating of the processed sound S_p :

$$d = S_p - S_r \quad (16)$$

The degradation measure takes into account the rating of the reference sound, to compensate for any erroneous rating by the subjects. No statistically significant variation between subjects in the rating of the reference sounds was found ($F_{10,2739} = 1.58$, $p = 0.106$). In the following, sound quality is defined by the degradation d ; a low degradation is assumed to be synonymous with high sound quality.

All sounds were played back over Beyer Dynamic DT990 headphones, connected to a M-Audio firewire 410 sound interface. 11 subjects, most of them male in their mid twenties, listened to each condition two times, and thus 250 sets of sounds were presented to each subject. The experiment took about an hour, and thus time was the limiting factor for the number of repetitions, reproduction conditions and instrument types used in the experiment.

4.4 Results and discussion

After the experiment the subjects were asked to comment on the experiment. Many stated that the perceived difference between reference and processed sound was due to changes in the sustained part of the sound. The sounds used in the experiment all had soft attacks and no transients, except for the piano that has a fast attack when the hammer hits the string. For the piano one subject commented on a perceived difference in the attack.

The degradation varied significantly across condition ($F_{4,84} = 60.7$, $p = 0.001$). Figure 5 shows the mean degradation and standard error of mean for each of the reproduction types. Pairwise comparison showed that the individual levels of reproduction were significantly different from each other ($p \leq 0.001$). The order of the mean degradations ranged, as expected, from *Imperceptible* towards larger degradation of the sound quality, as phase information was removed. One exception is that the mean degradation of relative phase (RP) was lower than that of absolute phase (AP). In synthesizing RP far more phase information is used than in AP. Even though AP is rated lower than ARP, the results show that AP, the model using the partial period phase representation, can indeed retain some of the perceptually important phase information. Another explanation for the finding may be the fact that the attack in AP is identical to the attack in ARP and thus close to that of the original sound.

Variation in degradation across fundamental frequency group was also significant ($F_{4,84} = 82.5$, $p < 0.001$). Figure 6 shows the mean degradations for the fundamental frequencies f_1 to f_5 as defined in table 1 for the different levels of reproduction. The sound quality improved as a function of frequency, except for f_5 which was significantly lower than for f_3 and f_4 , at $p \leq 0.024$. The interaction of reproduction and fundamental frequency was significant ($F_{16,336} = 30.1$, $p < 0.001$), with an unexplained large mean degradation in the AP and NP reproductions for f_5 compared to f_3 and f_4 . For ARP and RP, we see a clear relationship between fundamental frequency and perceived sound quality, where low fundamental frequency results in larger degradation. It seems that ARP and RP retain the phase relations that are important when modeling the noise between the harmonic partials in the high pitched sounds. The noise is modeled using additional non-harmonic sinusoids, which make “the noise take on a tonal quality that is unnatural

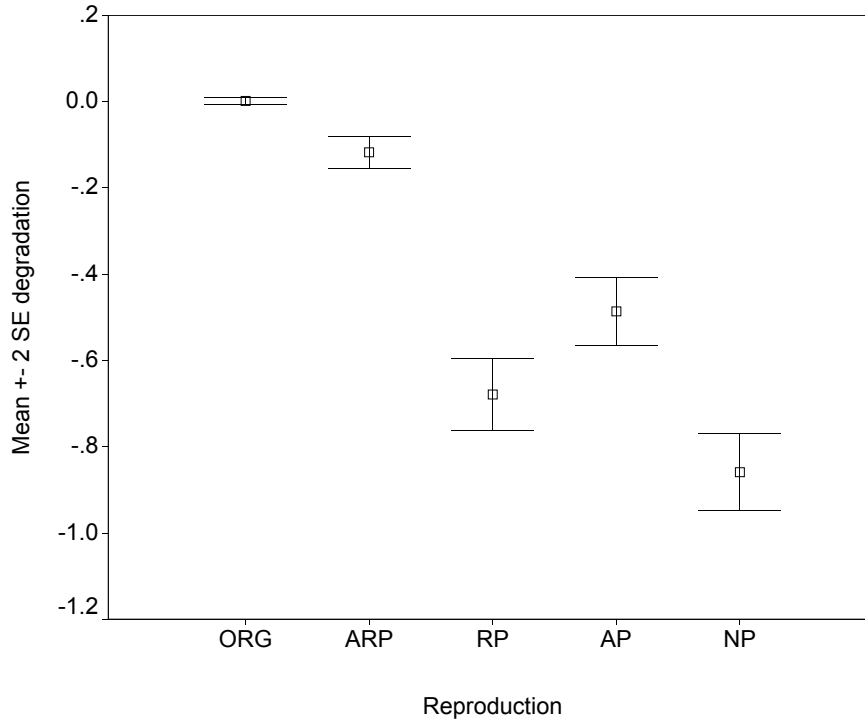


Figure 5: Mean degradation for different reproduction types.

and annoying”, if the phase information is not used [2]. This only applies to the high-pitched tones, however.

Mean degradation as a function of instrument type and reproduction is shown in figure 7. The degradation varies significantly across the type of instrument ($F_{4,84} = 54.4$, $p < 0.001$). In general degradation was lower for cello and piano than for the rest of the instruments. The interaction of instrument type and reproduction was significant ($F_{16,336} = 28.4$, $p < 0.001$). For bass trombone and bass clarinet, AP gave a lower degradation than the other reproduction methods, with the exception of ORG and ARP.

Piano gave the largest degradation for ARP reproduction, which is most likely due to errors in the reproduction of the attack. The transient caused by the hammer string interaction in the piano attack is the only fast transient that occurs in the instrument selection included in this experiment. Because of the window used in the block-based analysis, smearing does occur, which is harmful to the modeling of fast transients. This may explain the higher degradation in ARP for piano. A within subject analysis of the degradation for the piano sounds reveals a significant difference between reproduction type ($F_{4,84} = 4.5$, $p = 0.002$). However, pairwise comparison of the four different synthesized reproduction types reveals no significant difference between any of them. The piano tones are slightly inharmonic, and thus the relative phase shift is not constant over time. This change in relative phase shift may explain why no significant difference in degradation is observed between the synthesis methods, since the non-harmonic relationship prevents stable auditory waveforms that would influence the perceived sound quality [22, 13]. In addition, the relatively short transient of approximately 5 ms [48] of the piano attack makes phase changes less perceptible [18].

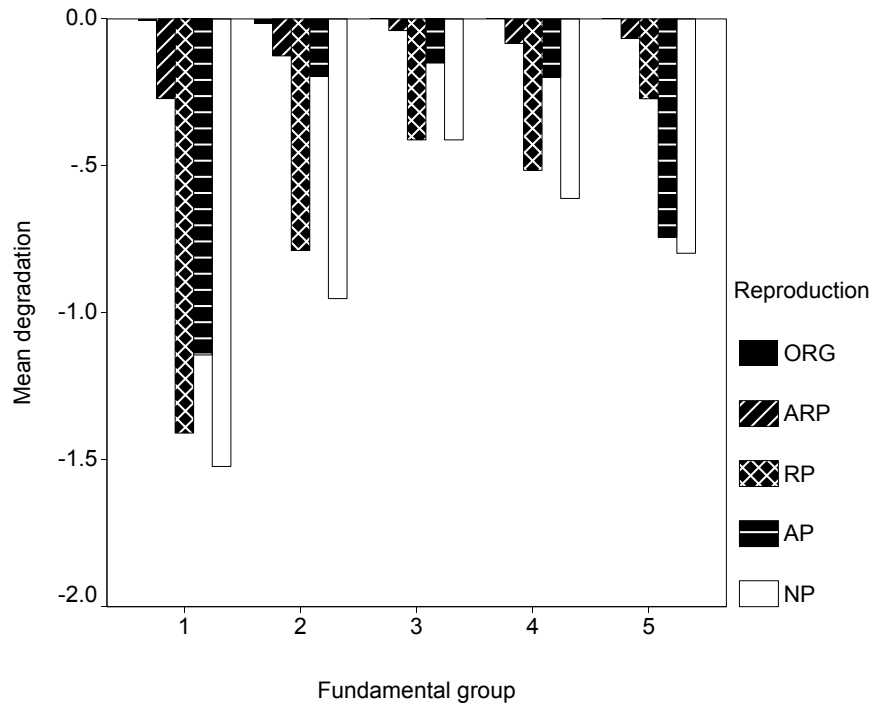


Figure 6: Mean degradation for different reproduction types, as a function of fundamental frequency. A high fundamental frequency group corresponds to a high fundamental frequency.

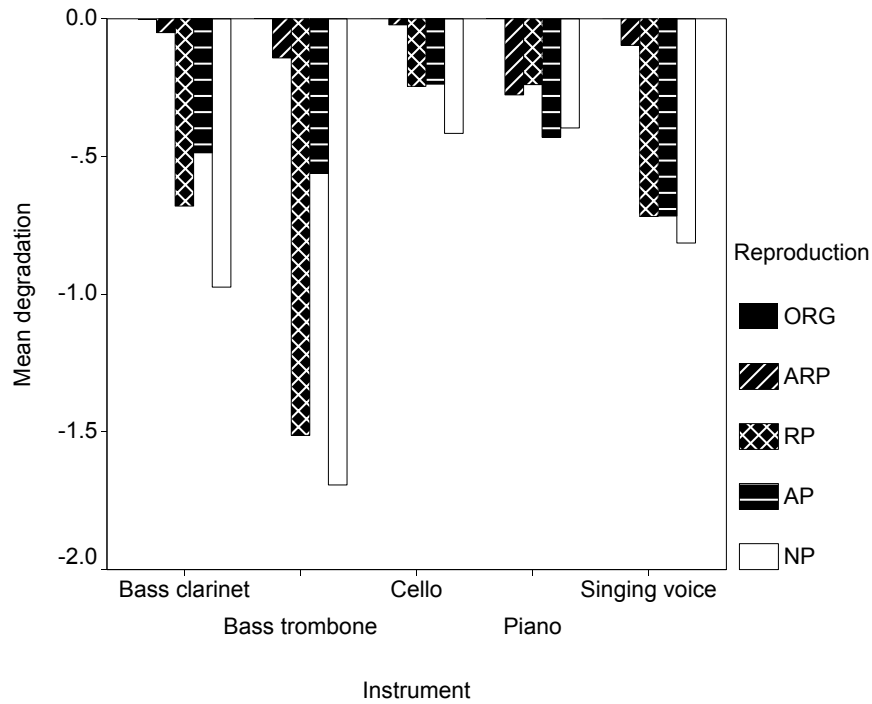


Figure 7: Mean degradation for different reproduction types, as a function of instrument type.

5 Conclusions

We have presented an improved representation, partial period phase, for use in the sinusoidal analysis/synthesis framework. The proposed representation makes phase values from consecutive frames comparable, while solving problems with phase unwrapping. The representation results in stable phase trajectories for the stationary part of harmonic and inharmonic sounds. Partial period phase is of direct use in additive analysis/synthesis, where phase is modeled together with frequency and amplitude, when transforming or synthesizing sounds.

An experiment was conducted where synthesized sounds were compared to original recorded sounds. The results of the experiment show that the inclusion of phase alignment enhances the sound quality of the analysis/synthesis system. A significant change in mean degradation was found between synthesis without and with phase, going from *perceptible, but not annoying* to *imperceptible*. This result is in agreement with literature on auditory perception of complex tones. A significant effect of fundamental frequency was found, resulting in degradation approaching *Slightly Annoying* for sounds with fundamental frequencies below approximately 100 Hz synthesized without phase (NP).

By use of the partial period phase representation, a phase model (AP) is proposed, where the sustained part of the sound is modeled by a constant partial period phase trajectory. The experiment shows that this model is significantly better than when discarding phase alignment information (NP), or when maintaining the relative phase shift (RP) but discarding the absolute alignment of the partials.

For the piano, synthesis with full phase information (ARP) was worse than for the other instruments, which is most likely caused by the smearing of the fast transient in the attack. No significant difference was found between the different synthesis methods of the piano sound.

6 Acknowledgment

The authors would like to thank the reviewers, Brian C.J. Moore and one anonymous reviewer, for helpful comments and suggestions. We would also like to thank the subjects participating in the experiment.

References

- [1] K. Jensen. *Timbre Models of Musical Sounds*. Ph.D. dissertation, Department of Computer Science, University of Copenhagen, 1999. DIKU Tryk, Technical Report no. 99/7.
- [2] R. J. McAulay and T. F. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal processing*, ASSP-34(4):744–754, August 1986.
- [3] H. Helmholtz. *On the Sensation of Tone*. Dover Publications, Inc., New York, 1954. Second English Edition, based on the Fourth German Edition of 1877.
- [4] R. Plomp and H. J. M. Steeneken. Effect of phase on the timbre of complex tones. *J. Acoust. Soc. Am.*, 46:409–421, 1969.
- [5] R.C. Mathes and R.L. Miller. Phase effects in monaural perception. *Journal of the Acoustical Society of America*, 19:780–797, 1947.

- [6] J.L. Goldstein. Auditory spectral filtering and monaural phase perception. *Journal of the Acoustical Society of America*, 41:458–479, 1967.
- [7] R. D. Patterson. A pulse ribbon model of monaural phase perception. *J. Acoust. Soc. Am.*, 82:1560–1586, 1987.
- [8] R. Meddis and M. J. Hewitt. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: pitch identification. *J. Acoust. Soc. Am.*, 89:2866–2882, 1991.
- [9] R. Meddis and M. J. Hewitt. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: phase sensitivity. *J. Acoust. Soc. Am.*, 89:2883–2894, 1991.
- [10] R.D. Patterson, M.H. Allerhand, and C. Giguere. Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform. *J. Acoust. Soc. Am.*, 98:1890–1894, 1995.
- [11] R.P. Carlyon and S. Shamma. An account of monaural phase sensitivity. *J. Acoust. Soc. Am.*, 114:333–348, 2003.
- [12] J.I Alcántara, I. Holube, and B.C.J. Moore. Effects of phase and level on vowel identification: Data and predictions based on a nonlinear basilar-membrane model. *J. Acoust. Soc. Am.*, 100:2382–2392, 1996.
- [13] D. Pressnitzer and S. McAdams. Two phase effects on roughness perception. *J. Acoust. Soc. Am.*, 105(2):2773–2782, 1999.
- [14] B. Roberts, B.R. Glasberg, and B.C.J. Moore. Primitive stream segregation of tone sequences without differences in F0 or passband. *J. Acoust. Soc. Am.*, 112:2074–2085, 2002.
- [15] H. Gockel, B.C.J Moore, R.D. Patterson, and R. Meddis. Louder sounds can produce less forward masking: Effects of component phase in complex tones. *J. Acoust. Soc. Am.*, 114(2):978–990, 2003.
- [16] R.A. Greiner and D.E. Melton. Observations on the audibility of acoustic polarity. *J. Audio Eng. Soc.*, 42(4):245–253, April 1994.
- [17] B.C.J. Moore and B.R. Glasberg. Difference limens for phase in normal and hearing-impaired subjects. *J. Acoust. Soc. Am.*, 86:1351–1365, 1989.
- [18] J. H. Patterson and D. M. Green. Discrimination of transient signals having identical energy spectra. *J. Acoust. Soc. Am.*, 48:894–905, 1970.
- [19] G.H. Wakefield, L.M. Heller, L.H. Carney, and M. Mellody. On the perception of transients: Applying psychophysical constraints to improve audio analysis and synthesis. In *Proceedings of the International Computer Music Conference*, pages 225–228, 2000.
- [20] S. Uppenkamp, S. Fobel, and R.D. Patterson. The effects of temporal asymmetry on the detection and the perception of short chirps. *Hearing Research*, 158:71–83, 2001.
- [21] M.R. Schroeder. New results concerning monaural phase sensitivity. *J. Acoust. Soc. Am.*, 31:1579, 1959.

- [22] M. R. Schroeder. Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. *IEEE Transactions on information theory*, 16:85–89, 1970.
- [23] M. Leman. Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (sim). In *Proc. of the Conf. on Digital Audio Effects (DAFX-00)*, pages 125–130, 2000.
- [24] E. Tind and K. Jensen. Phase models to control roughness in additive synthesis. In *Proceedings of the International Computer Music Conference, Miami, USA*, 2004. Accepted for publication.
- [25] J-C. Risset and D. L. Wessel. Exploration of timbre by analysis and synthesis. In D. Deutsch, editor, *Psychology of Music*. Academic Press, 1982.
- [26] M.R. Portnoff. Implementation of the digital phase vocoder using the fast fourier transform. *IEEE Transactions on ASSP*, 24:243–248, 1976.
- [27] J. B. Allen. Short term spectral analysis, synthesis and modification by discrete fourier transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-25:235–238, 1977.
- [28] X. Serra and J. Smith. Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4):12–24, winter 1990.
- [29] K. Fitz and L. Haken. Sinusoidal modeling and manipulation using lemur. *Computer Music Journal*, 20(4):44–59, 1996.
- [30] X. Rodet. The additive analysis-synthesis package. Technical report, IRCAM, Paris, France, July 2004. <http://www.ircam.fr/equipes/analyse-synthese/DOCUMENTATIONS/additive/index-e.html>.
- [31] K. Jensen. The Timbre model. In *Workshop on current research directions in computer music, Barcelona, Spain*, pages 174–186, 2001.
- [32] M. V. Matthews, J. E. Miller, and E. E. David. Pitch synchronous analysis of voiced speech. *J. Acoust. Soc. Am.*, 33(2):179–186, February 1961.
- [33] M. D Freedman. Analysis of musical instrument tones. *J. Acoust. Soc. Am.*, 41(4):793–806, 1967.
- [34] K. Fitz and L. Haken. On the use of time-frequency reassignment in additive sound modeling. *Journal of the Audio Engineering Society*, 50(11):879–893, 2002.
- [35] J. S. Marques and L. B. Almeida. New basis functions for sinusoidal decompositions. In *Proceedings of the 8th European Conference in Electrotechnics (EUROCON'88)*, pages 48–51, June 1988.
- [36] P. Guillemin. *Analyse et modélisation de signaux sonores par des représentations temps-frequence linéaires*. PhD thesis, Université d'Aix-Marseille II, 1994.
- [37] F. Auger and P. Flandrin. Improving the readability of time frequency and time scale representations by the reassignment method. *IEEE Transactions on Signal Processing*, 43:1068–1089, 1995.

- [38] S. Borum and K. Jensen. Additive analysis/synthesis using analytically derived windows. In *Proceedings of the Digital Audio Effects Workshop, Trondheim, Norway*, pages 125–128, 1999.
- [39] Y. Ding and X. Qian. Processing of musical tones using a combined quadratic polynomial-phase sinusoid and residual (QUASAR) signal model. *J. Audio Eng. Soc.*, 45(7/8):571–584, July/August 1997.
- [40] A. Röbel. Adaptive additive synthesis of sound. In *Proceedings of the International Computer Music Conference, Berlin, Germany*, pages 256–259, 1999.
- [41] T. H. Andersen. Phase models in real-time analysis/synthesis of voiced sounds. Master’s thesis, Department of Computer Science (DIKU), University of Copenhagen, Denmark, January 2002.
- [42] L.R. Rabiner. On the use of autocorrelation analysis for pitch detection. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-25:24–33, 1977.
- [43] A. Papoulis. *Signal Analysis*. McGraw-Hill, New York, 1977.
- [44] R. Di Federico. Waveform preserving time stretching and pitch shifting for sinusoidal models of sound. In *Proceedings of the COST-G6 Digital Audio Effects Workshop*, pages 44–48, 1998.
- [45] J. Laroche and M. Dolson. New phase-vocoder techniques for real-time pitch shifting, chorusing, harmonizing, and other exotic audio modifications. *Journal of the Audio Engineering Society*, 47(11):928–936, 1999.
- [46] H. Fletcher, E. D. Blackham, and R. Stratton. Quality of piano tones. *J. Acoust. Soc. Am.*, 34(6):749–761, 1962.
- [47] Methods for the subjective assessment of small impairments in audio systems, including multichannel sound systems. Technical report, International Telecommunication Union, Geneva, Switzerland, March 1994. ITU-R 8510, Recommendation.
- [48] J. Bensa, K. Jensen, and R. Kronland-Martinet. A hybrid resynthesis model for hammer-string interaction of piano tones. *EURASIP J. Applied Signal Processing*, 7:1021–1035, 2004.