Sensitivity of Human Hearing to Changes in Phase Spectrum

Article *in* Journal of the Audio Engineering Society · November 2013

CITATIONS
36

READS
5,796

3 authors, including:

Sascha Disch
Fraunhofer Institute for Integrated Circuits
75 PUBLICATIONS **1,119** CITATIONS
SEE PROFILE

Ville Pulkki
Aalto University
307 PUBLICATIONS **6,444** CITATIONS
SEE PROFILE

Sensitivity of Human Hearing to Changes in Phase Spectrum

MIKKO-VILLE LAITINEN,[1] *AES Student Member,* **SASCHA DISCH**[2]**, AND VILLE PULKKI,**[1] *AES Fellow*
(mikko-ville.laitinen@aalto.fi) (sascha.disch@iis.fraunhofer.de) (ville.pulkki@aalto.fi)

[1]*Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland*
[2]*Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany*

Human ability to perceive differences in sounds due to the modification of the phase spectrum is studied in this article. Formal listening tests were arranged using synthetic harmonic complex signals. The tests confirm that humans can perceive differences in the phase spectrum. Furthermore, the perception of phase was found to be somewhat local in frequency, but there is interaction between nearby auditory bands. In addition, the phase spectrum affects the perceived timbre, and the effects are more perceivable the lower the fundamental frequency is. Based on the results, an auditory model for explaining these effects was developed. It aims to mimic the firing rate of the neurons in the cochlea. The output of the auditory model and the listening-test results were compared showing a good match.

0 INTRODUCTION

Traditionally in audio signal processing the phase spectrum of an audio signal is assumed not to play a significant role in general. The claim that the ear is "phase deaf" was already suggested by Ohm [1], and Helmholtz came to the same conclusion in his tests [2]. This assumption about the auditory system is often exploited in audio signal processing, such as in audio coding [3,4,5,6]. The phase spectrum of an audio signal is often modified by the processing, for example, due to downmixing, quantization, and, especially, decorrelation. Decorrelation techniques, which basically scramble the phase spectrum, are needed in these methods for obtaining incoherent signal components.

Hence, these techniques assume that humans do not perceive modifications in the phase spectrum. As listening tests show, this assumption holds well for most of the signals [4,6,7]. However, recently it has been noticed that with certain signals, such as applause-type [8,9] and anechoic speech signals [10], humans are sensitive to changes in the phase spectrum. In order to obtain the optimal perceptual quality, a different kind of processing is needed with these "phase-sensitive" signals. Thus, finding an objective measure that predicts how perceivable the phase modifications are with any given signal would be useful. The aim of this work is to develop such a measure.

Human perception of the phase spectrum has been studied after the work of Ohm and Helmholtz, and several studies clearly show that humans are not phase deaf [11,12, 13,14]. For example, a cosine-phase harmonic complex signal, in which all of the components start simultaneously at their maximum amplitude, is perceived differently than a random-phase signal, in which the starting phase of each component is random [14]. In addition, even small changes in the phase spectrum within auditory frequency bands have been observed to yield perceivable differences, whereas changes between the bands are reported not to be perceivable [15], or at least large changes in the phase are needed in order to have a perceivable difference [14]. Furthermore, changing the phase of even a single component of a harmonic complex signal can be perceivable [16]. The perception of the phase spectrum has also been studied in relation to many topics, such as concert hall acoustics [17,18,19], pitch perception [20], vowel identification [21], masking [22], speech processing [23], and binaural rendering [24].

This article starts by reviewing the basics of human hearing and previous studies about phase perception. Based on these, a set of formal listening tests were arranged using synthetic harmonic complex signals in order to obtain more detailed knowledge about the properties of phase perception. Differences in the phase spectrum are known to affect perception, but the significance of these differences is not completely clear. In experiments 1 and 2, the perceptual significance of phase distortion was compared to magnitude distortion. Furthermore, it is not known if the differences in perception due to the phase modifications are global or local in frequency. In order to study this, phase modifications were applied to different bandwidths in experiment 3. Experiment 4 studied how wide in frequency is the perceptual effect of changing the phase at a certain frequency. In addition, the phase spectrum has been seen to affect the timbre. In experiment 5, it is shown that the loudness of the

lowest harmonics can be controlled with the phase spectrum. Finally, in experiment 6, the frequency dependence of the phase-related effects was studied. Based on the listening-test results, an auditory model is presented, which aims to mimic the firing rate of the neurons in the cochlea. It is suggested that the output of the model can be used to predict differences in perception due to differences in the phase spectrum. This assumption is evaluated by comparing the results of the listening tests to the output of the model.

1 BACKGROUND

The basics of the human auditory system are discussed in this section and earlier studies related to phase perception are reviewed.

1.1 Human Auditory System

The sensitivity of the human hearing system to the signal phase is of interest in this study, and thus the mechanisms involved are reviewed here. The pressure signal in the ear canal causes vibrations on the eardrum, which are transmitted through the ossicles in the middle ear to the cochlea [25]. The cochlea converts the mechanical vibrations to neural pulses with hair cells, which are organized tonotopically along the basilar membrane [25]. The cells sensitive to high frequencies are located near the position where the vibration enters the membrane, and the tuning frequency decreases at further positions. The frequency selectivity of the mechanism has been found to follow the equivalent rectangular bandwidth (ERB) [26], meaning that the vibrations having frequencies inside the auditory bandwidth are not processed separately, but their joint effect is seen in the neural response.

The membrane and the mechanisms of the cells cause a frequency-dependent delay. When an impulsive sound arrives to the ear, the cells tuned to high frequencies respond earlier than the cells tuned to low frequencies due to the transmission delay in the basilar membrane and also due to mechanical factor of the system. Experimental studies suggest that the frequency-dependent group delay produced by the cochlear filtering is at least partly compensated for at a higher processing level [27]. However, there does not seem to be unambiguous data of the exact delays of the total auditory system [28].

The hair cells transmit their binary responses through the auditory nerve to the brainstem. Often the responses of individual fibers are not found to be interesting in research, and instead the firing rate inside each auditory band is considered to carry the auditory information. The firing rate can be seen as a band-pass filtered signal as the starting point [29]. In addition to that, it also carries information of the phase of the signal in the ear canal. When single sinusoids are presented, the firing rate shows a pulse for each period of the sinusoid at a temporal position corresponding to a certain value in the phase of the sinusoid. The temporal length of the pulse is about 0.5–1.0 ms, depending on the neuron type and frequency [30]. The length can be assumed to be of that order or slightly higher at all frequencies, since (a) with a sinusoid input at low frequencies the system responds very accurately to a distinct phase of the sinusoid, (b) at high frequencies the phase-locking effect is lost [30], and (c) the temporal accuracy of hearing is of the order of 1–2 ms with impulsive signals [31].

1.2 Previous Phase-Perception Studies

As mentioned in Section 0, for a long time human hearing was thought to be phase deaf. A few studies are discussed in this section that clearly show that this is not true. In addition, a few interesting effects are shown that are caused by modifying the phase spectrum.

Plomp and Steeneken studied the effect of the phase spectrum on the timbre perception with a number of formal listening tests [13]. In one of the tests, signals with an identical magnitude spectrum but with a different phase spectrum were compared. The signals were harmonic complex tones, consisting of the first ten harmonics of the tone with different phase relations, e.g.,

$$\begin{aligned} x_1(t) &= \cos(2\pi f_0 t) + \frac{1}{2}\cos(2\pi 2 f_0 t) \\ &\quad + \frac{1}{3}\cos(2\pi 3 f_0 t) + \frac{1}{4}\cos(2\pi 4 f_0 t) + \ldots \\ x_2(t) &= \sin(2\pi f_0 t) + \frac{1}{2}\cos(2\pi 2 f_0 t) \\ &\quad + \frac{1}{3}\sin(2\pi 3 f_0 t) + \frac{1}{4}\cos(2\pi 4 f_0 t) + \ldots, \end{aligned} \tag{1}$$

where f_0 is the fundamental frequency and t is time. The result was that the tones with alternating sine and cosine components exhibited a significant difference when compared to signals with only sine or cosine components. In addition, this difference in the phase spectrum was compared to the difference caused by changing the slope of the magnitude spectrum of the signals. The effect of phase on the perceived timbre was found to be quantitatively smaller than the effect of changing the slope of the magnitude spectrum by 2 dB/oct, and it is less for higher than for lower frequencies.

Patterson performed similar tests [14]. He studied the effect of two different kinds of phase modifications: local and global phase changes. The local phase changes mean changes in the phase spectrum inside the ERB bands. The local changes were caused by using alternating-phase signals (APH), which were created by shifting the starting phase of every other component by the same fixed amount D. APH signals were compared to cosine-phase signals (CPH), i.e.,

$$\begin{aligned} x_{\mathrm{CPH}}(t) &= \cos(2\pi f_0 t) + \cos(2\pi 2 f_0 t) \\ &\quad + \cos(2\pi 3 f_0 t) + \cos(2\pi 4 f_0 t) + \ldots \\ x_{\mathrm{APH}}(t) &= \cos(2\pi f_0 t) + \cos(2\pi 2 f_0 t + D) \\ &\quad + \cos(2\pi 3 f_0 t) + \cos(2\pi 4 f_0 t + D) + \ldots . \end{aligned} \tag{2}$$

This is similar to [13], but in [14] the value of D can be changed, whereas in [13] D was effectively locked to 90 degrees. The aim of the study was to find the needed D in

order to obtain a perceivable difference between the CPH and the APH signals. *D* was found to be dependent on the fundamental frequency, intensity, and the spectral location of the signal. Already 15 degrees caused a perceivable difference in certain cases, whereas more than 60 degrees was needed in some cases.

In the second set of experiments [14], global phase changes were studied, which mean that changes in the phase spectrum are minimized inside the ERB bands, while applying sufficiently large relative phase changes between the different ERB bands. This was performed by using monotonic-phase signals, where successive harmonics are progressively shifted in time. The phase is always changed in the same direction, but the amount of change is decreased using a constant deceleration. In practice, the high frequencies are delayed compared to low frequencies, or vice versa. It was found that the global phase changes are not perceivable if the total time delay across the frequency channels is less than 4–5 ms. Furthermore, discrimination was found to be largely independent of signal properties other than bandwidth.

In addition, Patterson presented an auditory model that predicts when a phase change produces an audible change in timbre [14]. The model contains a simplified model of the cochlea, consisting of an auditory filter bank and units that record the times of the larger peaks in the filter outputs. Furthermore, the times of the peaks are adjusted in time by computing the cross-correlation between different frequency bands. The patterns of these adjusted peaks on a time-frequency plot can be used to determine whether two sounds have a different timbre. The model successfully predicted the differences noticed in the listening tests presented in that article.

Moore and Glasberg studied the ability to detect a change in the relative phase of a single component in a harmonic complex tone [16]. The tone contained the first 20 harmonics, and all except one started in cosine phase. The phase of the remaining harmonic was shifted, which caused that human listeners heard a pure tone "pop out" from the complex tone. The pitch of this tone corresponded to the frequency of the phase-shifted harmonic. The aim was to find out what is the minimum required amount of phase shift in order to perceive this phase-shifted component. In some cases a phase shift of 2–4 degrees was found to be perceivable.

2 LISTENING TESTS

Section 1.2 reviewed some existing knowledge that showed that the human hearing is sensitive to changes in the phase spectrum and a few effects were presented that are caused by modifying it. These effects are studied further, and a few new effects are presented in this section. Formal listening tests are conducted in order to validate their existence and importance. Experiments 1 and 2 study the perceptual significance of phase distortion compared to magnitude distortion. Experiment 3 studies the effect of applying only partial phase modification. The bandwidth of the perceptual effect of phase modification is studied in experiments 3 and 4. The effect of phase on the perception of bass, i.e., the loudness of the lowest frequencies, is studied in experiment 5, and the frequency dependency of phase-related effects is studied in experiment 6.

2.1 Phase-Sensitive Signals

Before the actual listening tests, a number of informal listening tests were conducted in order to find out what kind of real signals are "phase sensitive," i.e., they are perceived differently if the phase spectrum is modified. A large number of signals were tested, including, for example, different instruments, speech, and ambient recordings. The informal testing was performed by scrambling the phase spectrum of the signal and comparing the phase-scrambled version of the signal to the original version. The scrambling was performed by convolving the signal with a decorrelation filter that was designed to make the phase spectrum random but affecting the magnitude spectrum and the temporal envelope minimally. Only a small part of the tested signals were found to be phase sensitive. These signals included, for example, speech, trumpet, and trombone sounds recorded in anechoic conditions. Two common characteristics with these signals is that they are essentially harmonic complex signals and the phases of the harmonics have certain fixed relations. For example, in case of voiced phonemes, the glottal pulse excites all harmonics with fixed phase relations for each cycle of the fundamental frequency. In contrast, for example in case of reverberant signals, the phase relations between the harmonics are random.

Since controlling the phase of real signals can be difficult, synthetic signals are used as test signals instead. The properties of the test signals were selected to be similar to the real signals that were found to be phase sensitive. Thus, the signal should be a harmonic complex signal, the starting phases of the harmonics should be aligned, the fundamental frequency should be around 100 Hz, and the envelope of the magnitude spectrum should be similar to the spectrum of real instrument and speech sounds. This kind of signal can be created as a sum of cosines by controlling the phase and the gain of the single components. All the test signals studied in this article were created using

$$x(t) = G \cdot \sum_{i=1}^{\infty} g_i \cdot \cos(2\pi \cdot i \cdot f_0 \cdot (t - \tau_i) + \phi_i)/i, \quad (3)$$

where G is the gain controlling the overall level of the signal, g_i is a frequency-dependent gain for controlling the magnitude spectrum, i is the sequential number of the harmonic, τ_i is a frequency-dependent delay, and ϕ_i is a frequency-dependent angle for controlling the phase spectrum. The time-domain and the frequency-domain presentations of two example signals created with this equation are shown in Fig. 1. The signals were created using MATLAB [32] with the sampling rate of 48 kHz.

The signal in the upper panel was created with the following parameter values: $f_0 = 100$ Hz, $g_i = 1$, $\phi_i = (i - 1) \cdot \pi/2$, and $\tau_i = 0$ ms. These kind of signals are called in-phase signals in this paper because the phase change between the harmonics is constant. The amount of the constant phase change and the phase of the first harmonic determine

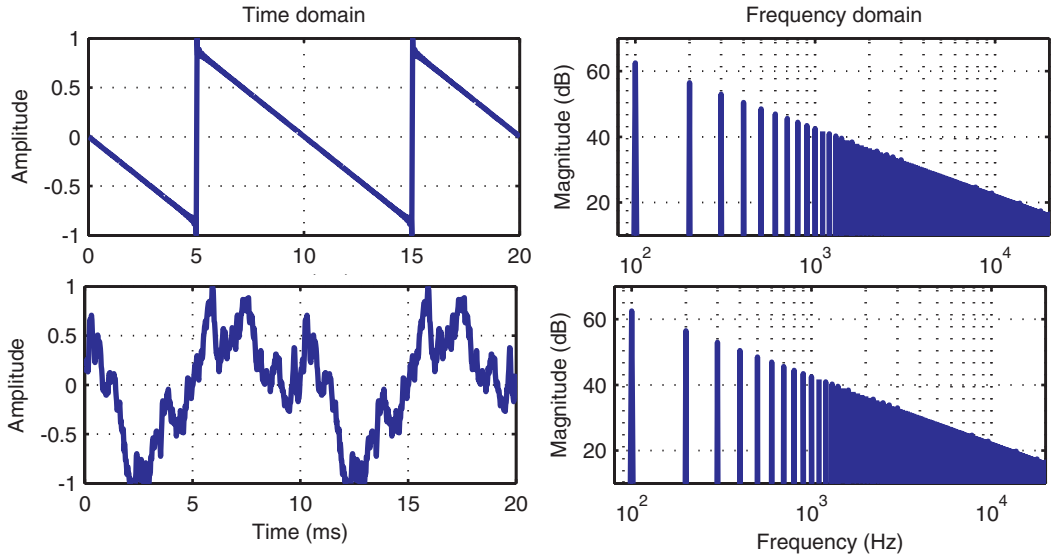

Fig. 1. Time- and frequency-domain presentations of two signals used in the listening tests.

the phase angle to which all of the harmonics "lock" into, i.e., this angle takes place at the same time instant with all harmonics once in the cycle of the fundamental frequency. With this signal the locked angle is $-\pi/2$, producing the value $\cos(-\pi/2) = 0$ at the same time instant with all harmonics (Time = 5 and 15 ms in Fig. 1), and thus, the quick rise of the sawtooth waveform. Other waveforms can also be created by using a different constant phase change, e.g., a spike train by using $\phi_i = 0$. All these waveforms are perceived in a relatively similar way [13,14], so any of them can be selected to reflect the in-phase signals.

The in-phase signal presented in Fig. 1 can be described to have a strong, low pitch, and a "buzzy" quality [33]. Perceptually, the opposite of the in-phase signals are random-phase signals, which have otherwise exactly the same parameters, but the phase is determined by $\phi_i \sim \mathcal{U}(0, 2\pi)$, where $\mathcal{U}$ stands for taking a random value from the uniform distribution (see the lower panel). The random-phase signals are perceived very differently than the in-phase signals even though the magnitude spectra are identical. They sound colored compared to the in-phase signals, they are perceived to be thinner, and the buzzy quality is absent. In addition, the distance of the auditory event is perceived to be larger.

These two signals can be considered as reference signals in this article that demonstrate significant differences due to phase changes. In addition, these signals can be thought to represent certain real-world signals. The in-phase signals can be considered to represent voiced vowels in anechoic conditions, whereas the random-phase signals represent harmonic signals without the phase-alignment property as well as all other harmonic signals in a reverberant room or after decorrelation.

2.2 Listening Test Procedure

Fourteen subjects, excluding the authors, all with earlier experience in listening tests, participated in the tests. The listening tests consisted of six *experiments*. Each experiment consisted of a certain number of *test cases*. Each test case consisted of a certain number of *signals* to be evaluated at the same time. These test cases were repeated a certain number of times producing the total number of *test sets* in each experiment, which were evaluated one after another. The signals for the different *repetitions* of the same test case were created using identical parameter values. For the signals containing randomized values, the randomizations of the different repetitions were identical or different, depending on the experiment.

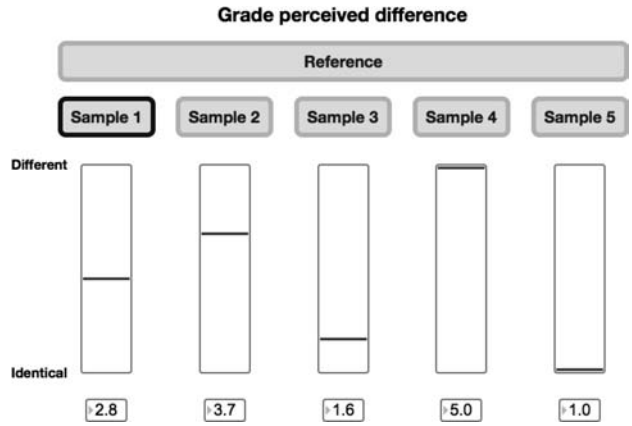

Fig. 2. Graphical user interface used in experiment 1.

The subjects used a graphical user interface to select the sample (i.e., the signal) to be played and to write in the score. See Fig. 2 for an example of the interface, which was slightly different in all experiments. The scale in the score was from 1 to 5 in steps of 0.1 The subjects listened to one sample at a time and graded different samples in a multiple-stimulus test. They were able to change between different samples freely, and the length of each sample was 2.5 seconds. All samples were time-invariant. The order of the samples was randomized for each test set, and the subjects did not know which sample they were listening to. In addition, the order of the test sets was randomized.

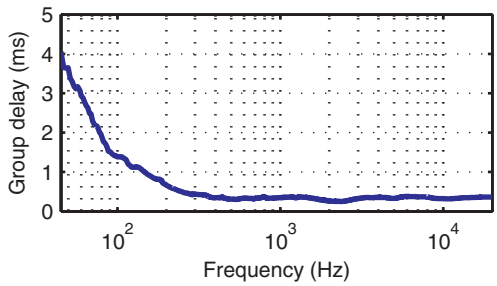

Fig. 3. Group delay of the headphones used in the listening tests, Sennheiser HD650.

Each experiment was conducted in two parts. The first part was a training session where the subjects did a short version of the actual listening test. The subjects were able to listen to different samples and to get familiarized with the user interface. This way the subjects became thoroughly familiar with the differences under study. Familiarization is recommended in [34], since it can transform some subjects with an initially low ability into experts for the purposes of the test. The subjects were given as much time for the training as they wanted. The durations of the training sessions were about 5 minutes on average. After the training session, all subjects reported that they perceived at least some differences between the samples and were ready for the actual test. The subjects had a short break between different parts of the test. The actual experiment lasted about 10 minutes, depending on the subject and the experiment. The length of a single experiment was selected to be relatively short in order to avoid listening fatigue, as the signals used were noticed to cause it easily.

The listening test was performed in a quiet listening room using Sennheiser HD650 headphones. The group delay of the headphones, shown in Fig. 3, was measured using a linear-phase microphone, B&K 4192. The delay is seen to increase at low frequencies. This kind of behavior is common with headphones as well as loudspeakers. Correcting the phase response of the headphones is possible in theory, but in practice obtaining a pure impulse out of the headphones is difficult. Furthermore, coupling between the ear and the headphones affects the needed correction. Thus, correcting the phase response was omitted. Using different headphones with different phase responses was tested informally before the test. It was noticed that the selection of the headphones affects the perception of the signals studied in Section 2.4.5, but for other experiments no differences were perceived.

2.3 Statistical Analysis

Repeated-measures analysis of variance (RM-ANOVA) was used for the statistical analysis of the results. The univariate approach was taken and the correction for the violation of sphericity was applied when required, based on Mauchly's test. The correction method applied was Greenhouse-Geisser when $\varepsilon < 0.75$ and Huynh-Feldt when $\varepsilon > 0.75$. In the following sections, all the significant main effects and interactions are presented and further discussion is performed based on the results. In experiments 2, 4, and 5, the repetitions of a same combination were pre-averaged before the RM-ANOVA was applied. In experiments 1, 3, and 6, the repetitions were not pre-averaged because the randomizations in the different repetitions were different (called *variation* in the following).

2.4 Experiments

The different experiments are discussed separately in detail in the following sections, and the results of the listening tests are presented. The audio samples used in the experiments can be downloaded from http://www.acoustics.hut.fi/go/jaes-phase.

2.4.1 Experiment 1: Comparison of Phase Distortion Effects to Magnitude Distortion Effects

The in-phase and the random-phase signals, presented in Section 2.1, were found to be perceived differently in the informal listening test. The significance of this difference compared to magnitude spectrum randomization is studied in this section with formal listening tests. A similar test was performed in [13] (see Section 1.2), but in this experiment, the magnitude spectrum manipulation was different. The aim was to use such a manipulation that caused a similar perception of coloration as the phase scrambling.

There were five different signals in the sole test case of experiment 1, as presented in Table 1. Signals 1 and 2 correspond to the in-phase and the random-phase signals in Fig. 1, respectively. Signals 3–5 have an identical phase spectrum as signal 1. The magnitude spectra of these signals were modified by multiplying the harmonics with individual gains obtained from the normal distribution with the standard deviation of 1, 2, or 4 dB, respectively. Parameter G was selected to be such that the sound pressure level of signal 1 was 65 dB with linear weighting, and G was identical for signal 2. For signals 3–5, G was selected in a way that the A-weighted levels were equal to signal 1.

The task of the subjects was to grade the perceived difference compared to the reference signal, which was identical to signal 1. The subjects were asked to use the entire scale, with the sample having the largest difference graded as 5 and at least one of the samples with 1, i.e., being identical. Hence, the results are relative. The test case was repeated three times with different variations in the randomization for both the magnitude and the phase spectrum.

Two-way RM-ANOVA was applied to the results of experiment 1. The within-subjects factors were *variation* and *signal*. RM-ANOVA revealed the following significant factors: main effects *variation*, $F(2, 26) = 12.351$, $p < 0.05$, and *signal*, $F(2.165, 28.139) = 220.429$, $p < 0.05$; and interaction *variation*signal*, $F(3.583, 46.582) = 7.428$, $p < 0.05$.

The significant interaction is further inspected. Fig. 4 shows the mean opinion score (MOS) plots with 95% confidence intervals for the interaction. The perceived difference due to phase-spectrum randomization is seen to be larger than the differences due to the magnitude-spectrum

Table 1. Values of the parameters in Eq. 3 for the signals in experiment 1.

Signal	f_0 (Hz)	g_i	ϕ_i	τ_i(ms)
1 - In phase	100	1	$(i-1)\cdot\pi/2$	0
2 - Random phase	100	1	$\sim\mathcal{U}(0,2\pi)$	0
3 - 1dB std	100	$10^{(L/20)}$, $L\sim\mathcal{N}(0,1^2)$	$(i-1)\cdot\pi/2$	0
4 - 2dB std	100	$10^{(L/20)}$, $L\sim\mathcal{N}(0,2^2)$	$(i-1)\cdot\pi/2$	0
5 - 4dB std	100	$10^{(L/20)}$, $L\sim\mathcal{N}(0,4^2)$	$(i-1)\cdot\pi/2$	0

modifications used in the experiment. The general tendency of the results is in line with the result in [13], even though they cannot be directly compared due to different manipulation of both the magnitude and the phase spectrum.

2.4.2 Experiment 2: Effect of a Shift in Phase of a Single Component

In earlier studies (e.g., [16], see Section 1.2), it has been found that changing the phase of a single harmonic can cause a perception of a sinusoid popping out of the tone if all other harmonics are in phase. As an addition to the previous studies, the perceived relative level of this sinusoid is studied in this experiment.

There were five different signals in test case 1 (see Table 2). In the case of signal 1, all the harmonics are in phase. In the case of signal 2, all the harmonics are in phase except the harmonic at 3 kHz, which is shifted by 180 degrees in order to obtain the largest possible effect. In signals 3–5 all the harmonics are in phase similarly as in signal 1, but the magnitude of the harmonic at 3 kHz is amplified by 3, 6, and 9 dB, respectively. The amplification of the harmonic component was noticed to cause a similar effect as the phase shift of the harmonic in the informal listening. The same conclusion was made in [16]. The aim of this experiment was to find out how large an amplification is needed in order to obtain an equal effect. Parameter G was selected to be such that the sound pressure level of signal 1 was 65 dB with linear weighting, and G was identical for all test signals.

The task of the subjects was to grade the perceived level of the sinusoid that pops out of the harmonic complex signal. The subjects were asked to use the entire scale such that the sample with the perception of the loudest added component is graded as 5, and the samples where the component was not perceived as 1. Hence, the results are relative. In the second test case the signals were otherwise equal to test case 1, but the values for ϕ_i were selected randomly. These phase values were identical for all signals. The same phase shift of 180 degrees was applied to signal 2. Both test cases were repeated twice, yielding four test sets altogether.

Two-way RM-ANOVA was applied to the results of experiment 2. The within-subjects factors were *test case* and *signal*. RM-ANOVA revealed the following significant factors: main effects *test case*, $F(1, 13) = 76.820$, $p < 0.05$, and *signal*, $F(1.784, 23.194) = 477.213$, $p < 0.05$; and interaction *test case*signal*, $F(2.351, 30.564) = 159.382$, $p < 0.05$.

The significant interaction is further inspected. Fig. 5 shows the mean plots with 95% confidence intervals for the interaction. It can be seen that, in case of the in-phase signal, changing the phase of a single harmonic causes a perception of an added sinusoid, which has the loudness in between the cases of the sinusoid amplified with 6 and 9 dB. In case of the random-phase signal, changing the phase does not cause a perception of an added sinusoid, which agrees with the results in [16].

2.4.3 Experiment 3: Effects of Different Amounts of Phase Modification

In Section 2.4.1 it was noticed that phase randomization can cause a significant effect. This section studies how the perception is affected if the randomization is limited

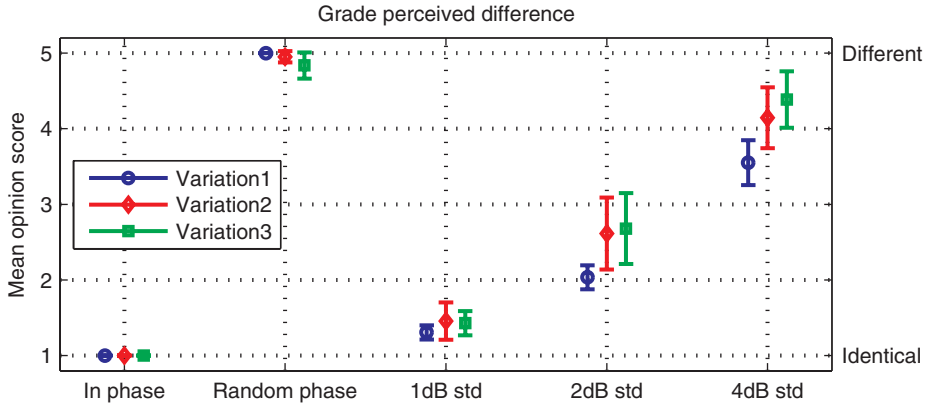

Fig. 4. Results of experiment 1: The effect of phase randomization and various magnitude-spectrum randomizations, when compared to the reference (identical to *In phase*). Means and 95% confidence intervals are shown.

Table 2. Values of the parameters in Eq. 3 for the signals in experiment 2.

Case	Signal	f_0 (Hz)	g_i	ϕ_i	τ_i(ms)
1	1 - Normal	100	1	$(i-1)\cdot\pi/2$	0
	2 - 180° shift	100	1	$\begin{cases}(i-1)\cdot\pi/2+\pi, & i=30\\(i-1)\cdot\pi/2, & i\neq 30\end{cases}$	0
	3 - 3dB boost	100	$\begin{cases}10^{(L/20)}, L=3, & i=30\\1, & i\neq 30\end{cases}$	$(i-1)\cdot\pi/2$	0
	4 - 6dB boost	100	", $L=6$	$(i-1)\cdot\pi/2$	0
	5 - 9dB boost	100	", $L=9$	$(i-1)\cdot\pi/2$	0
2	1 - Normal	100	1	$\sim\mathcal{U}(0,2\pi)$	0
	2 - 180° shift	100	1	$\begin{cases}\sim\mathcal{U}(0,2\pi)+\pi, & i=30\\\sim\mathcal{U}(0,2\pi), & i\neq 30\end{cases}$	0
	3 - 3dB boost	100	$\begin{cases}10^{(L/20)}, L=3, & i=30\\1, & i\neq 30\end{cases}$	$\sim\mathcal{U}(0,2\pi)$	0
	4 - 6dB boost	100	", $L=6$	$\sim\mathcal{U}(0,2\pi)$	0
	5 - 9dB boost	100	", $L=9$	$\sim\mathcal{U}(0,2\pi)$	0

either within a certain interval of angles or within a certain frequency interval.

In the first test case of the experiment, the phase randomization was restricted within certain angles, as described in Table 3. Similar effects can be caused by the room reverberation with different direct-to-reverberant ratios. In the case of signal 1, all the harmonics are in phase, whereas in the case of signal 5, the phases are completely randomized. For signals 2–4 the randomization is restricted to within ±20°, ±45°, and ±90°, respectively. In the second test case of the experiment, the bandwidth of the phase randomization was changed. For example, audio codecs can cause the phase spectrum to be modified only at certain frequencies. The bandwidth of the randomization varied from a single harmonic to two thirds of an octave and two octaves. Inside the selected band, the phases of the harmonics were random, whereas outside it they were aligned.

The task of the subjects was to grade the perceived difference compared to two anchor signals, which were identical to signals 1 and 5. The subjects were advised to grade a signal with the score of 1 if identical to anchor 2 and with 5 if identical to anchor 1. If the signal was equally different from both anchors, the score of 3 was to be given, otherwise the score was to be closer to the anchor that it reminded of more. The experiment was repeated with three different randomizations for both test cases, resulting in six test sets altogether. Parameter G was selected as in experiment 2.

Two-way RM-ANOVA was applied separately to the results of both test cases of experiment 3. The within-subjects factors were *variation* and *signal*. In test case 1, RM-ANOVA revealed the following significant factors: main effect *signal*, $F(1.991, 25.889) = 648.129$, $p < 0.05$; and interaction *variation*signal*, $F(8, 104) = 2.803$, $p < 0.05$. In test case 2, RM-ANOVA revealed the following significant factor: main effect *signal*, $F(4, 52) = 411.381$, $p < 0.05$.

The interactions are further inspected (although it is not statistically significant in test case 2). Fig. 6 shows the mean plots with 95% confidence intervals for the interactions. It can be seen that already partial randomization of the phase spectrum causes a difference in the perception. The more randomization is performed, the larger is the difference compared to the in-phase signal. The same tendency is seen for the band-limited randomization. Randomization in a narrow band causes a perceivable difference, and the difference is increased when the bandwidth of the randomization is increased. In informal listening, the phase scrambling of a certain band was observed to affect the

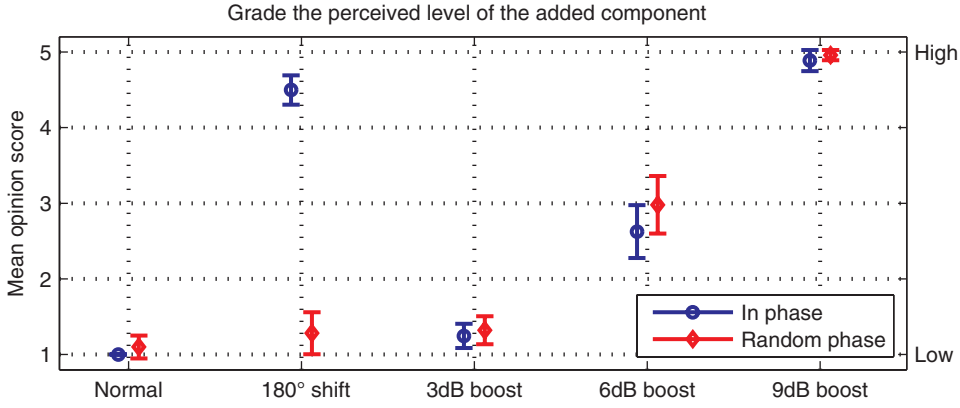

Fig. 5. Results of experiment 2: The perceived level of an added sinusoid, caused by phase shifting and various levels of amplification. Means and 95% confidence intervals are shown.

Table 3. Values of the parameters in Eq. 3 for the signals in experiment 3.

Case	Signal	f_0 (Hz)	g_i	ϕ_i	τ_i(ms)
1	1 - 0°	100	1	$(i-1)\cdot\pi/2$	0
	2 - 20°	100	1	$(i-1)\cdot\pi/2+\sim\mathcal{U}(-\pi/9,\pi/9)$	0
	3 - 45°	100	1	$(i-1)\cdot\pi/2+\sim\mathcal{U}(-\pi/4,\pi/4)$	0
	4 - 90°	100	1	$(i-1)\cdot\pi/2+\sim\mathcal{U}(-\pi/2,\pi/2)$	0
	5 - 180°	100	1	$\sim\mathcal{U}(0,2\pi)$	0
2	1 - 0	100	1	$(i-1)\cdot\pi/2$	0
	2 - Single	100	1	$\begin{cases}(i-1)\cdot\pi/2+\pi, & i=30\\(i-1)\cdot\pi/2, & i<30\vee i>30\end{cases}$	0
	3 - 2/3oct	100	1	$\begin{cases}\sim\mathcal{U}(0,2\pi), & 26\le i\le 35\\(i-1)\cdot\pi/2, & i<26\vee i>35\end{cases}$	0
	4 - 2oct	100	1	$\begin{cases}\sim\mathcal{U}(0,2\pi), & 15\le i\le 60\\(i-1)\cdot\pi/2, & i<15\vee i>60\end{cases}$	0
	5 - Broad	100	1	$\sim\mathcal{U}(0,2\pi)$	0

perception only inside and near the band. Below and above the frequency region the sound was perceived to be as buzzy as the reference signal. Hence, it is assumed that the perception of phase is somewhat local in frequency, which is studied in more detail in the following section.

2.4.4 Experiment 4: Effect of Applying Delay for High Frequencies of an In-Phase Signal

The main result of Section 2.4.2 is that a discontinuity in the phase spectrum can cause a perception of an added tone. In the next experiment, a discontinuity is created by adding a constant delay to the harmonics above a certain frequency. Below and above this frequency the harmonics are in phase. In informal listening, this kind of delay was seen to cause a perception similar to that of modifying the phase of a single harmonic, but now a narrow-band noise tone is perceived to pop out instead of a sinusoidal component. This suggests that the phase of a harmonic affects the perception of the other harmonics, i.e., there is interaction between the harmonics in our hearing system. The bandwidth of this interaction is studied in this section.

It is possible to mute the harmonics around the frequency, where the delay is applied. If the delay causes a difference in the perception even when the frequencies around the cross-over frequency are muted, that would suggest that the harmonics interfere with each other with a wider bandwidth than the bandwidth of the muted harmonics. If the delaying of the high frequencies does not cause any difference in the perception after the muting, the bandwidth of the interaction between the harmonics can be assumed to be smaller than the bandwidth of the muted harmonics.

There were three different signals in each test case (see Table 4). The harmonics in signal 1 were in phase, as were those in signal 2, except that above 3 kHz the harmonics were delayed by 5 ms. In signal 3 the phase of the harmonics was random. The delay was selected to be half of the period of the fundamental frequency, causing the largest effect according to the informal tests. There were three test cases: in test case 1, no harmonics were muted; in test case 2, the harmonics ± 1/3 octave around 3 kHz were muted (in all signals); and in test case 3, the harmonics ± 1 octave were muted (in all signals).

In each test case, the task of the subjects was to grade the perceived difference compared to the reference signal, which was identical to signal 1. The subjects were asked to use the entire scale such that the sample with the largest difference should be graded as 5 and at least one of the samples as 1, i.e., being identical. Hence, the results are

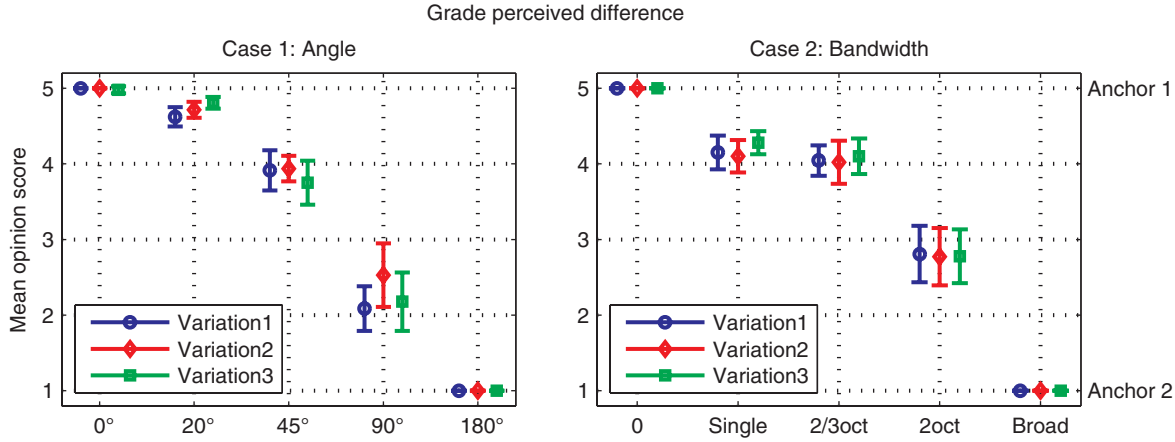

Fig. 6. Results of experiment 3: The effect of different amounts of randomization, when compared to two anchors (identical to *0°* and *180°* in case 1 and *0* and *Broad* in case 2). Means and 95% confidence intervals are shown.

Table 4. Values of the parameters in Eq. 3 for the signals in experiment 4. Different test cases are separated by horizontal lines. The signals in cases 2 and 3 are identical to the signals in case 1, except that the harmonics are muted near the cross-over frequency in all signals (i.e., signals 1, 2, and 3).

Mute	Signal	f_0 (Hz)	g_i	ϕ_i	τ_i(ms)
off	1 - In phase	100	1	$(i-1)\cdot\pi/2$	0
	2 - Time shift @ 3kHz	100	1	$(i-1)\cdot\pi/2$	$\begin{cases}0, & i \le 30\\ 5, & i > 30\end{cases}$
	3 - Random phase	100	1	$\sim\mathcal{U}(0,2\pi)$	0
2.6-3.5kHz	”	”	$\begin{cases}0, & 26 \le i \le 35\\ 1, & i < 26 \vee i > 35\end{cases}$	”	”
1.5-6kHz	”	”	$\begin{cases}0, & 15 \le i \le 60\\ 1, & i < 15 \vee i > 60\end{cases}$	”	”

relative. The experiment was repeated twice for all test cases, resulting in six test sets altogether. Parameter G was selected as in experiment 2.

Two-way RM-ANOVA was applied to the results of experiment 4. The within-subjects factors were *test case* and *signal*. RM-ANOVA revealed the following significant factors: main effects *test case*, $F(2, 26) = 85.902$, $p < 0.05$, and *signal*, $F(1.003, 13.038) = 2452.813$, $p < 0.05$; and interaction *test case*signal*, $F(4, 52) = 90.154$, $p < 0.05$.

The significant interaction is further inspected. Fig. 7 shows the mean plots with 95% confidence intervals for the interaction. It is seen that the time shift causes a clear difference compared to the in-phase signal if the signals are broadband. Muting the harmonics at one third octave below and above the cross-over frequency makes the perceived difference smaller, and no difference is perceived if the harmonics are muted one octave below and above the cross-over frequency. This would suggest that the bandwidth of the interaction between the harmonics is about one octave in each direction.

2.4.5 Experiment 5: Effect of Phase on the Perception of Bass

Previous studies indicate that the phase spectrum affects the perceived timbre (e.g., [13,14, 33]), but how the perception is affected is not thoroughly described. One interesting effect was found in the informal listening. Some signals with an identical magnitude spectrum but different phase spectrum were perceived to contain a different amount of bass, i.e., the perceived loudness of the lowest harmonics depended on the phase spectrum. This is studied further with a formal test.

There were five different signals in each test case (see Table 5). It was found in informal listening that signal 2 is perceived to contain more bass than signal 1, even though they have identical magnitude spectra, but signal 2 is the negative of signal 1, i.e., $s_2 = -s_1$. These signals were compared to signals that have phase spectra identical to signal 1, but the magnitude spectrum is amplified at low frequencies. The amplification was performed with a function that resembles a shelving filter. The shape of the function was tuned based on informal listening in order to obtain a similar timbre to signal 2. The amplification is largest at the fundamental frequency and decreases at higher frequencies, as specified in Table 5. The amplification of the fundamental frequency is 1, 2, and 4 dB for signals 3, 4, and 5, respectively. Parameter G was selected to be such that the sound pressure level of signal 1 with the fundamental frequency of 100 Hz was 65 dB with linear weighting. At other fundamental frequencies G was selected to be such that the perceived loudness was equal to the 100-Hz case, as determined in the informal listening. Usually A-weighting is used to produce equal loudness, but with these

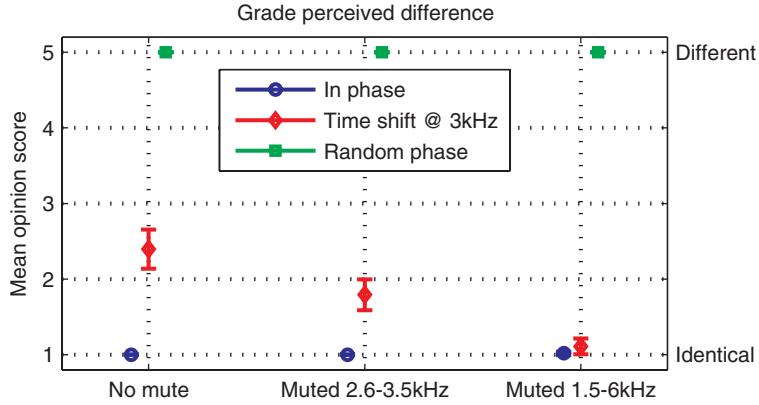

Fig. 7. Results of experiment 4: The effect of a delay at high frequencies, when compared to the reference (identical to *In phase*). The harmonics near the cross-over frequency were muted in cases 2 and 3 (in all signals). Means and 95% confidence intervals are shown.

Table 5. Values of the parameters in Eq. 3 for the signals in experiment 5. Other parameter values than f_0 are identical in all test cases.

Case	Signal	f_0 (Hz)	g_i	ϕ_i	τ_i(ms)
1	1 - Inverse	50	1	$(i-1)\cdot(-\pi/2)$	0
	2 - Normal	50	1	$(i-1)\cdot\pi/2$	0
	3 - 1dB boost	50	$1+\frac{10^{L/20}-1}{i^{0.8}}$, $L=1$	$(i-1)\cdot(-\pi/2)$	0
	4 - 2dB boost	50	", $L=2$	$(i-1)\cdot(-\pi/2)$	0
	5 - 4dB boost	50	", $L=4$	$(i-1)\cdot(-\pi/2)$	0
2	"	100	"	"	"
3	"	200	"	"	"
4	"	400	"	"	"

signals neither the linear nor the A-weighting produced perceptually equal loudness. G was identical for all test signals within each test case.

The task of the subjects was to grade the perceived level of the bass. The subjects were asked to use the entire scale with the sample with most bass to be graded as 5 and that with least bass as 1. Hence, the results are relative. The experiment was carried out for four fundamental frequencies, 50, 100, 200, and 400 Hz, and was repeated twice for each frequency. Thus, there were eight test sets in all.

In informal listening it was found that while differences in the bass were relatively easy to perceive when listening in a normal room with some background noise like air ventilation, it was difficult to hear any differences between the test signals when listening in a quiet listening room. This counterintuitive effect of the background noise is interesting by itself but goes beyond the scope of this paper and is left for future studies. Thus, white background noise with the sound pressure level of 45 dB with linear weighting was added. The noise was a continuous 1-minute sample that was looped. With the background noise present, differentiating between the samples was significantly easier, according to informal listening.

Two-way RM-ANOVA was applied to the results of experiment 5. The within-subjects factors were *fundamental frequency* and *signal*. RM-ANOVA revealed the following significant factors: main effect *signal*, $F(1.786, 23.218) = 93.168$, $p < 0.05$; and interaction *fundamental frequency*signal*, $F(3.424, 44.517) = 11.389$, $p < 0.05$.

The significant interaction is further inspected. Fig. 8 shows the mean plots with 95% confidence intervals for the interaction. It can be noticed that signals 1, 3, 4, and 5 are graded similarly in all cases and that the amplification of the low frequencies gradually increases the perceived level of bass. In the case of signal 2, the perceived level of bass depends on the frequency. At low fundamental frequencies the perceived level is relatively high, whereas at higher fundamental frequencies it is low. In addition, inspecting the results of individual subjects, it was noticed that different subjects perceived the level of bass very differently but consistently in the case of signal 2. Studying this phenomenon more thoroughly is left for future studies.

It should be noted that the results of this experiment depend on the headphones used. The group delay of the headphones and the polarity of the signal can differ between different transducers. As a result, different phase characteristics are required with certain headphones in order to obtain the perception of maximal bass and vice versa. Nevertheless, the result that our perception of the level

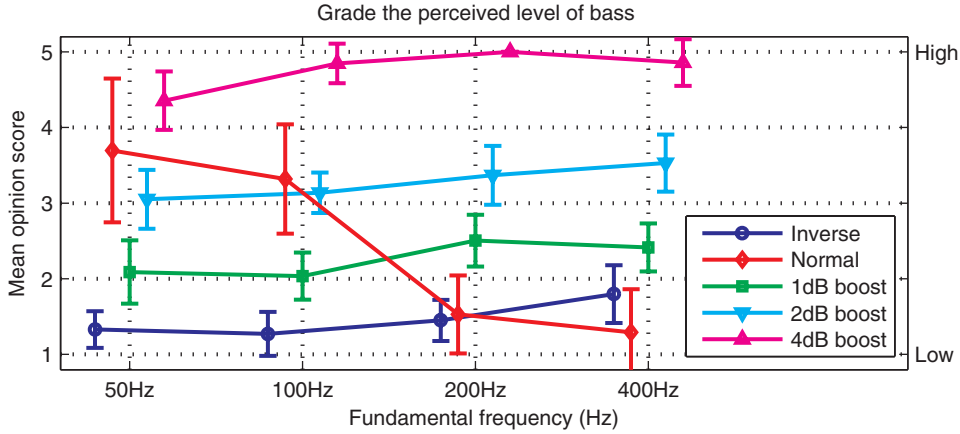

Fig. 8. Results of experiment 5: The effect of the phase-spectrum and various magnitude-spectrum manipulations on the perceived level of bass. Means and 95% confidence intervals are shown.

Table 6. Values of the parameters in Eq. 3 for the signals in experiment 6.

Signal	f_0 (Hz)	g_i	ϕ_i	τ_i(ms)
1 - In phase	50, 100, 200, 400, 800, 1600	1	$(i-1)\cdot\pi/2$	0
2 - Random phase	50, 100, 200, 400, 800, 1600	1	$\sim\mathcal{U}(0,2\pi)$	0

of the bass is dependent on the phase spectrum remains valid.

2.4.6 Experiment 6: Effect of Fundamental Frequency to Perception of Phase Modification

Earlier studies have found that many effects due to phase-spectrum modifications become smaller when the fundamental frequency is increased (e.g., [13,14,16,23], see Section 1.2). This frequency dependency is studied next. The aim is to find out how the effect of the phase spectrum behaves in relation to the fundamental frequency and to find out how high the fundamental frequency has to be so that differences in the phase spectrum cannot be perceived anymore.

In signal 1, all the harmonics were in phase, whereas in signal 2 the phases were randomized. Corresponding signal pairs were created for six fundamental frequencies ranging from 50 Hz to 1.6 kHz (see Table 6). Parameter G was selected as in experiment 5.

The task of the subjects was to grade the perceived difference compared to the reference signal, which was identical to signal 1. All six signal pairs were graded at the same time. However, each signal pair had their own reference in which to compare the test signals. The grading was performed in two phases. First, the task of the subjects was to find the hidden reference in each pair and to grade it with 1, i.e., identical. Second, the subjects were asked to compare the differences between the other signal in the pair and the reference to the differences in the other pairs. The pair with the largest difference was asked to be graded with 5. The experiment was repeated three times. Thus, there were three test sets, and each test set contained six signal pairs to be evaluated simultaneously.

Three-way RM-ANOVA was applied to the results of experiment 6. The within-subjects factors were *fundamental frequency*, *variation*, and *signal*. RM-ANOVA revealed the following significant factors: main effects *fundamental frequency*, $F(2.258, 29.359) = 105.874$, $p < 0.05$, and *signal*, $F(1, 13) = 540.661$, $p < 0.05$; and interaction *fundamental frequency*signal*, $F(2.270, 29.509) = 321.192$, $p < 0.05$.

The significant interaction is further inspected. Fig. 9 shows the mean plots with 95% confidence intervals for the interaction. The experiment has two outcomes. First, the largest difference between the in-phase and the random-phase signals is to be found at the lowest fundamental frequency, and the amount of difference decreases as the fundamental frequency is increased, which is in agreement with earlier studies [13,14,16,23]. Second, the fundamental frequency, above which differences due to phase spectrum modification cannot be perceived, is between 800 and 1600 Hz.

3 AUDITORY MODEL FOR ANALYZING PHASE PERCEPTION

A few effects in perception that can be caused by modifying the phase spectrum were presented in the previous section. In this section an auditory model is developed to describe these effects. Using the auditory model, a measure is suggested to detect whether a human can perceive a certain difference in the phase spectrum. Finally, the output of the auditory model and the suggested measure are compared to the results of the listening tests.

3.1 Auditory Model

An auditory model was developed in order to explain the differences in the perception due to phase-spectrum

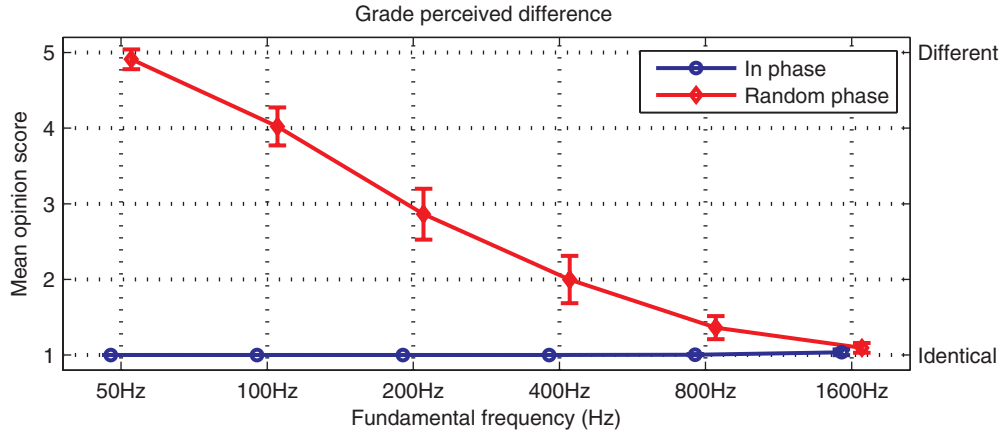

Fig. 9. Results of experiment 6: The effect of the fundamental frequency on the perception of phase randomization when compared to the reference (identical to *In phase*). Means and 95% confidence intervals are shown.

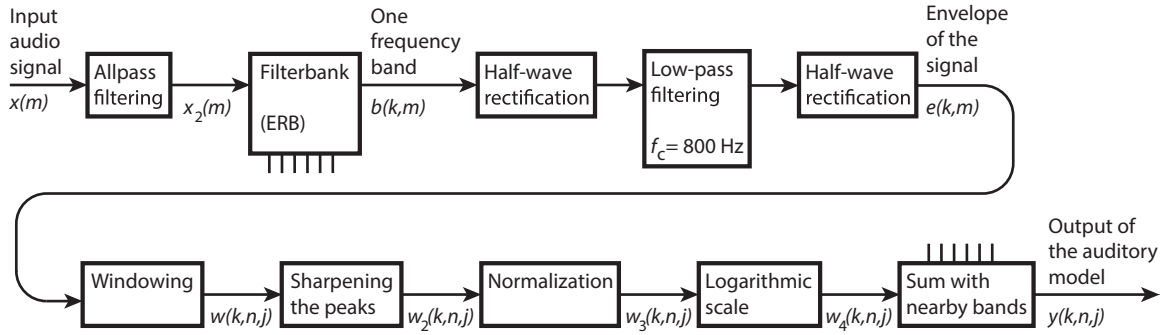

Fig. 10. Block diagram of the auditory model.

modifications. The aim of the model is to mimic the firing rate of the neurons in the cochlea. However, the aim of the model is not to be physically accurate. Instead, it should be thought of more as a tool for visualizing phase-related effects. The suggested model is based on knowledge about human hearing, especially the auditory model presented in [29], and it has been fine-tuned according to the results of the listening tests presented in Section 2.

The block diagram of the auditory model is shown in Fig. 10. The input to the model is a time-domain audio signal $x(m)$, where m is time in samples. $x(m)$ is a discrete-time version of $x(t)$ in Eq. 3. First, $x(m)$ is all-pass filtered in order to introduce a frequency-dependent delay that corresponds to the delay caused by the headphones and the human hearing. This delay adjustment is required for explaining the results of experiment 5 (see Section 3.3). However, as discussed in Sections 2.2 and 1.1, neither of the delays can be accurately measured. Thus, the delay adjustment is based on informal listening tests. The in-phase signal presented in Fig. 1 was found to be the most buzzy, the loudest, and to contain the most bass. It is assumed in this article that those effects are caused by the firing of the neurons being in sync at all frequencies with this signal. Thus, the all-pass filtering stage was designed in a way that also the output signals of the model are in sync with this signal. The filtering was realized by applying a 90-degree phase shift using the Hilbert transform. The 90-degree phase adjustment was applied to all signals in order to approximate the frequency-dependent delay of the headphones and the human hearing. A more accurate delay adjustment remains as a future research topic. Furthermore, it should be noted that the suggested delay adjustment was determined using Sennheiser HD650 (see Fig. 3). With other headphones, depending on the group delay, a different kind of adjustment might be needed.

The delay-adjusted signal $x_2(m)$ is divided in frequency using a filter bank mimicking the auditory frequency bands created according to the ERB bands (see Section 1.1). The delay of the filter bank is independent of frequency. The output of the filter bank is a time-domain signal for each frequency band, $b(k, m)$, where k is the frequency-band index.

These band-pass filtered signals are half-wave rectified, low-pass filtered, and half-wave rectified again in order to obtain the envelope of the band-pass signal without negative signal values. The properties of the low-pass filter were selected based on the listening test results presented in Section 2.4.6. The listeners were not able to distinguish changes in the phase spectrum above a fundamental frequency of 800 Hz. Thus, the temporal fine structure of the envelope signal should not contain variations faster than this. The low-pass filter was implemented using an FIR filter with the length of 100 samples and the cut-off frequency of 800 Hz. Below 800 Hz, the envelope signal is not modified, but above 800 Hz it is slowly attenuated, and at 1.6 kHz it is practically zero.

The envelope signals $e(k, m)$ are cut into 20-ms frames with a rectangular window and processed further separately within these frames. The signals in this article are time-invariant, so considering only a single frame is sufficient. For time-variant signals, a hop size of, e.g., 10 ms can be used. These windowed signals are defined as $w(k, n, j)$, where n is time in samples within the temporal frame and j is the temporal frame index.

The envelope signals can be seen as a starting point for modeling the firing rate of the neurons [29] (see Section 1.1). However, as discussed in Section 1.1, the firing rate shows at each period of a sinusoid a pulse at a temporal position corresponding a certain value in the phase of the sinusoid, and the temporal length of the pulse is somewhat independent of frequency. In the case of half-wave rectified band-pass signals, the length of the pulse is longer at low frequencies. Thus, in order to have equally wide peaks at all frequencies, the envelope signals are processed with the following function

$$\hat{w}_2(k, n, j) = w(k, n, j)^{p_k},\ p_k = 1 + 3 \cdot 0.8^{k-1}, \qquad (4)$$

and the amplitude of the highest peak within the frequency band is preserved by

$$w_2(k, n, j) = \hat{w}_2(k, n, j)\frac{\max_n(w(k, n, j))}{\max_n(\hat{w}_2(k, n, j))}. \qquad (5)$$

This signal is normalized so that the largest sample value corresponds to 1 in order to allow easier comparison of signals having different levels

$$w_3(k, n, j) = \frac{w_2(k, n, j)}{\max_{k,n}(w_2(k, n, j))}. \qquad (6)$$

In addition, since the human perception of loudness can be better described using the logarithmic scale instead of the linear [31], the level differences between the frequency bands are presented using the logarithmic scale with a

dynamic range of 60 dB

$$w_4(k,n,j) = \frac{\max(20\cdot\log_{10}(\max_n(w_3(k,n,j))) + 60, 0)\cdot w_3(k,n,j)}{60\cdot\max_n(w_3(k,n,j))}. \quad (7)$$

However, it should be noted that the output of the auditory model is presented using a linear scale within the frequency band.

In the final stage, the frequency-band signals are summed with the nearby frequency bands. The summing is performed using a Hann window, which means that the adjacent frequency bands have a large weight whereas further bands have only a small weight. The total width of the Hann window is 8 ERB bands, which roughly corresponds to 2 octaves. The summing of the neighboring frequency bands is based on the results of the listening test. In Section 2.4.4, it was noticed that the discontinuity at a certain frequency affects also the perception of the neighboring frequency bands, not only the frequency band where this discontinuity occurs. According to the test, all the harmonics one octave below and above the discontinuity have to be muted in order to avoid a difference in the perception. Thus, this article suggests that this interference between the neighboring frequency bands can be estimated by summing the output of the nearby frequency bands together, weighted in frequency with the Hann window.

Furthermore, the summing of the neighboring frequency bands is needed in order to explain the results of Section 2.4.5. At low frequencies, the ERB bands are so narrow that the adjacent harmonics are not present inside the same frequency bands. However, experiment 5 found that the perceived level of bass depends on the relative phase between the adjacent harmonics at low frequencies. By summing the neighboring frequency bands, the adjacent harmonics interfere with each other, also at low frequencies.

Signals with different phase properties are processed with this model, and the output of the model is inspected and compared to the perception of the sound. The output of the model, $y(k,n,j)$, can be used to create a time-frequency plot, which shows the neural firing rate (NFR) for each time-frequency tile, i.e., for each frequency band k at discrete time instants n (see Fig. 11 for a few example cases). In Figs. 11(a) and (b), the NFR patterns can be seen for the in-phase and the random-phase signals presented in Fig. 1, respectively. It can be seen that for the in-phase signal the NFR signal contains sharp peaks with the spacing of the cycle time of the fundamental frequency at all frequencies, and in between the peaks the NFR equals to zero. In addition, the peaks occur at the same time instant at all frequencies. In the case of the random-phase signal, there are no strong peaks, and the energy is almost equally spread in time. In addition, the weak peaks take place at different time instants. The other subfigures in Fig. 11 are discussed in Section 3.3.

3.2 Crest-Factor of Neural Firing Rate

Inspecting the NRF of the in-phase and the random-phase signals in Figs. 11(a) and (b), it can be seen that for the in-phase signal the energy is concentrated at the peaks, whereas for the random-phase signal the energy is almost equally spread in time. Thus, it is suggested that the human perception of the phase spectrum can be described using the crest factor of the neural firing rate (CFNFR), i.e., the ratio between the loudest amplitude values and the mean amplitude value. The CFNFR is computed for each frequency band from

$$c(k) = \log_{10}\left(\frac{m_{10\%}(k)}{m_{\text{all}}(k)}\right), \quad (8)$$

where $m_{10\%}(k)$ is the mean of the largest 10% of the NFR values inside the 20-ms frame and $m_{\text{all}}(k)$ is the mean of all NFR values inside the frame. This produces values between 0 and 1, where 0 means that the NFR is equally spread in time and 1 means that the high NFR is concentrated into small areas. The CFNFR values are plotted in Fig. 12 for all the signals presented in Fig. 11.

3.3 Analyzing the Signals with the Auditory Model and the Crest-Factor

The results of the formal listening tests were presented in Section 2.4, and an auditory model for explaining these results was presented in Section 3.1. This section studies how well the auditory model can explain the effects due to modifications in the phase spectrum. The plots presented in Figs. 11 and 12 are compared to the results of the listening tests.

The NFR plots of in-phase and random-phase signals in Figs. 11(a) and (b) were discussed in Section 3.1. Inspecting the corresponding CFNFR plots, Figs. 12(a) and (b), shows that CFNFR is significantly higher at all frequencies in the case of the in-phase signal. In experiment 1, a significant perceptual difference between these two signals was noticed. Thus, this difference can be described with the model.

The phase shift of 180 degrees of a single component of an in-phase signal was noticed to cause the perception of a sine tone to pop out of the complex tone in experiment 2. In the CFNFR plot (see Fig. 12(c)), this can be seen as a drop around the frequency of the shift. With random-phase signals, the phase shift of a single component does not cause significant differences in the CFNFR, since their structure is already relatively random. Thus, no difference is perceived in this case, as was observed in the experiment.

In the first part of experiment 3, gradually increasing the amount of randomization was seen to cause the perception to change from similar to the in-phase signal to closer to the random-phase signal. One of the test signals can be seen in Fig. 11(d). Comparing it to the in-phase and the random-phase signals, Fig. 11(d) shows that there still is a clear peak curve similarly as in Fig. 11(a), but there is also some energy in between the peaks. The CFNFR values in Fig. 12(d) are in between the in-phase and the random-phase signals at all frequencies. In the second part of experiment 3, the phase randomization was applied only to a narrow band. The case of a single component was discussed earlier

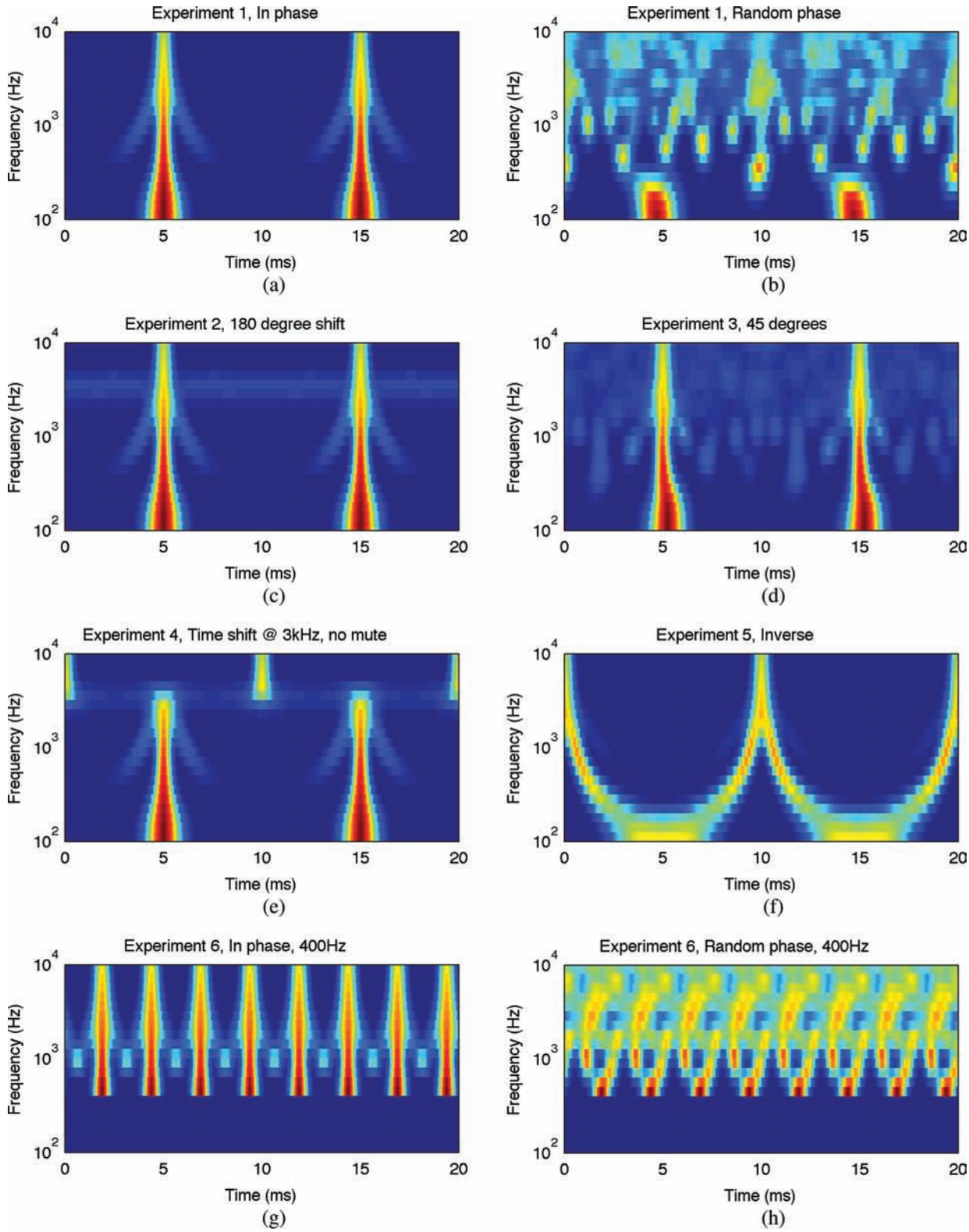

Fig. 11. Output of the auditory model for several signals employed in the listening tests, i.e., the neural firing rate (NFR) for each time-frequency tile. Red = high firing rate, blue = no firing.

(see Figs. 11(c) and Fig. 12(c)). For other bandwidths the CFNFR plots would be similar, but the drop in the CFNFR value would take place over a wider frequency range and the values would be smaller. This coincides with the perception of the signals of the informal listening tests. The wider the bandwidth of the randomization, the wider is the bandwidth of the perceived difference.

Experiment 4 indicated that applying a delay at high frequencies causes a difference in the perception. If all harmonics one octave below and above the frequency of

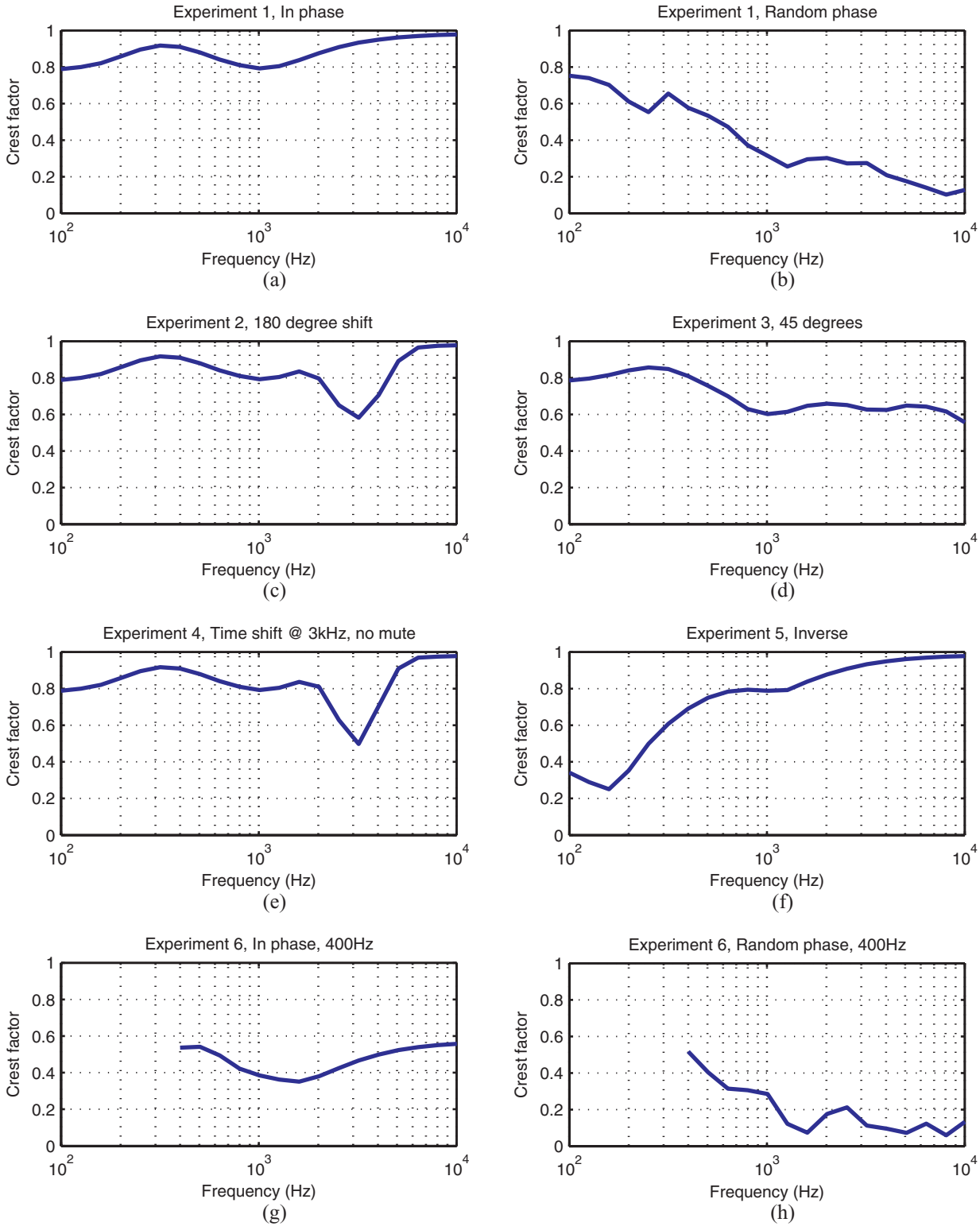

Fig. 12. Crest factor of the neural firing rate (CFNFR) corresponding to the figures in Fig. 11.

the delay are muted, no difference is perceived anymore. Comparing Figs. 12(a) and (e) shows that the delay causes a clear drop in the CFNFR plot. The drop starts one octave below the frequency of the delay and ends one octave above it. In the case where the frequencies one octave below and above the cross-over frequency are muted, the harmonics above the cross-over do not interact with the harmonics below it, and the CFNFR plots are identical, similar to the perception in the formal listening test.

In experiment 5, the perceived level of bass was observed to be affected by modifying the phase spectrum. The signal with the higher perceived level of bass is shown in Fig. 11(a) and the signal with the lower level in Fig. 11(f).

It can be seen that the CFNFR is significantly lower in Fig. 12(f) than in Fig. 12(a) at low frequencies, but at other frequencies the figures are identical. Thus, large CFNFR values at low frequencies are assumed to indicate a perception of louder bass and vice versa. According to the informal listening tests, the two signals are perceived to be identical at high frequencies. It should be noted that the CFNFR values would be identical with these signals also at low frequencies without the all-pass filtering stage in the model.

Experiment 6 showed that the effect of phase randomization is smaller the higher the fundamental frequency is. Comparing the in-phase and the random-phase signals with the fundamental frequencies of 100 Hz and 400 Hz in Figs. 12(a), (b), (g), and (h), it can be seen that the difference in the CFNFR values becomes smaller when the fundamental frequency is increased. If the fundamental frequency is larger than 800 Hz, the CFNFR values are identical regardless of the phase spectrum due to the low-pass filter stage in the model. Similarly, above 800 Hz the listeners were not able to reliably distinguish between the in-phase and the random-phase signals.

3.4 Discussion

As was discussed in the previous section, the model presented in this article can be used to detect whether there is a difference in perception due to phase spectrum modification. The suggested auditory model could, e.g., be used in audio coding to detect cases when phase spectrum modification should be avoided. The input and the output of the codec could be analyzed with the model, and if there is a difference in the model output, the phase modifications, such as decorrelation, could be omitted for these time-frequency tiles.

In addition, some assumptions about how humans perceive different sounds can be derived from the output of the model. If the CFNFR is high at low frequencies, the perceived amount of bass is high and vice versa. Correspondingly, at mid and high frequencies, high CFNFR values correspond to a perception of a buzzy signal, whereas low values are perceived to be more noise-like and colored. The frequency, where this difference in the effect begins, cannot be determined based on these listening tests. However, the informal listening tests indicate that if there are two or more harmonics inside an ERB band, the phase spectrum affects to the perceived "buzzyness," whereas if there is only one harmonic inside the band, the effect is only in the tone color or the amount of bass. Furthermore, it is assumed that if the CFNFR value of a certain frequency band differs significantly from the values of the nearby bands, the complex tone is perceived as two separate tones.

The auditory model, suggested in Section 3.1, is now compared to existing models. The only relevant auditory model known by the authors has been presented in [14] (discussed in Section 1.2). The main difference compared to the model presented in this article is that in [14] only the peaks of the band-pass signals are detected, whereas in this article the rate of the firing of the neurons is estimated, which was suggested to be useful in Section 3.2. Additionally, in the model presented in this article, the frequency bands interact with each other and the temporal alignment of the frequency bands is time-invariant. These properties are required in the model to distinguish between the different signals discussed in Sections 2.4.4 and 2.4.5.

4 CONCLUSION

Human ability to perceive differences in sounds due to the modification of the phase spectrum was studied in this article. Formal listening tests were arranged and synthetic harmonic complex signals were used as test signals. The results of the tests confirm that humans are not "phase deaf," the perceived difference due to randomization of the phase spectrum can be larger than the difference due to randomization of the magnitude spectrum with a standard deviation of 4 dB.

In addition, it seems that the mechanisms leading to phase perception are somewhat local in frequency, e.g., phase-spectrum modifications at high frequencies do not affect the perception at low frequencies. Nevertheless, the phase spectrum affects the perception of the neighboring frequencies. According to the tests, the phase of a component at a certain frequency affects the perception of frequencies about one octave lower and higher. Thus, there is interaction between nearby auditory frequency bands but the effect is not global. Furthermore, changes in the phase spectrum in both the narrow and the wide band can cause differences in the perception.

Based on informal listening tests and earlier studies, the signals for which the phase between the harmonics is aligned can be described to have a strong low pitch and a "buzzy" quality, whereas random-phase signals are perceived to be colored, thinner, and absent of the buzzy quality. According to the results of the formal listening tests, the effects of phase modification are perceived to be larger the lower the fundamental frequency is, and for signals with a fundamental frequency above 800 Hz, the differences in the phase spectrum cannot be perceived. In addition, the perceived level of bass frequencies can be affected by the phase spectrum. The difference due to phase-spectrum modification corresponds to amplification of the magnitude spectrum at low frequencies by 2–4 dB on average. However, this effect was found to be greatly dependent on the individuals.

Based on the results, an auditory model to explain these effects was developed. It aims to mimic the firing rate of the neurons in the cochlea. After comparing the output of the auditory model to the results of the listening tests, it was found that the crest factor of the neural firing rate for each frequency band can be used to explain differences in the perception due to phase-spectrum modifications. At the lowest frequencies of the tone, the high crest factor indicates a perception of loud bass, whereas at mid and

high frequencies of the tone, the high crest factor indicates a perception of a buzzy sound.

5 ACKNOWLEDGMENTS

The Fraunhofer IIS and the GETA Graduate School have supported this work. The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement n^o [240453].

6 REFERENCES

[1] G. S. Ohm, "Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen," *Ann. Phys. Chem.* (1843).

[2] H. L. F. von Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (F. Vieweg & Sohn, 1863).

[3] F. Baumgarte and C. Faller, "Binaural Cue Coding–Part I: Psychoacoustic Fundamentals and Design Principles," *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 509–519 (Nov. 2003).

[4] E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegård, "Low Complexity Parametric Stereo Coding," presented at the *116th Convention* of the Audio Engineering Society (2004 May), convention paper 6073.

[5] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 55, pp. 503–516 (2007 June).

[6] J. Herre, K. Kjörling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Rödén, W. Oomen, K. Linzmeier, and K. S. Chong, "MPEG Surround–The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding," *J. Audio Eng. Soc.*, vol. 56, pp. 932–955 (2008 Nov.).

[7] J. Vilkamo, T. Lokki, and V. Pulkki, "Directional Audio Coding: Virtual Microphone Based Synthesis and Subjective Evaluation," *J. Audio Eng. Soc.*, vol. 57, pp. 709–724 (2009 Sep).

[8] G. Hotho, S. van de Par, and J. Breebaart, "Multichannel Coding of Applause Signals," *EURASIP J. Advances in Signal Processing* (2008).

[9] M.-V. Laitinen, F. Kuech, S. Disch, and V. Pulkki, "Reproducing Applause-Type Signals with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 59, pp. 29–43 (2011 Jan./Feb.).

[10] M.-V. Laitinen and V. Pulkki, "Utilizing Instantaneous Direct-to-Reverberant Ratio in Parametric Spatial Audio Coding," presented at the *133rd Convention* of the Audio Engineering Society (2012 Oct.), convention paper 8804.

[11] M. R. Schroeder, "New Results Concerning Monaural Phase Sensitivity," *J. Acoust. Soc. Am.*, vol. 31, p. 1579 (abs.) (1959).

[12] E. de Boer, "A Note on Phase Distortion and Hearing," *Acustica*, vol. 11, pp. 182–184 (1961).

[13] R. Plomp and H. J. M. Steeneken, "Effect of Phase on the Timbre of Complex Tones," *J. Acoust. Soc. Am.*, vol. 46, p. 409 (abs.) (1969).

[14] R. D. Patterson, "A Pulse Ribbon Model of Monaural Phase Perception," *J. Acoust. Soc. Am.*, vol. 82, pp. 1560–1586 (1987 Nov.).

[15] E. de Boer, *Handbook of Sensory Physiology*, ch. "On the Residue and Auditory Pitch Perception" (Springer, 1976).

[16] B. C. J. Moore and B. R. Glasberg, "Difference Limens for Phase in Normal and Hearing-Impaired Subjects," *J. Acoust. Soc. Am.*, vol. 86, pp. 1351–1365 (1989 Oct.).

[17] D. Griesinger, "Phase Coherence as a Measure of Acoustic Quality, Part One: The Neural Mechanism," *20th International Congress on Acoustics*, Sydney, Australia (2010 Aug.).

[18] D. Griesinger, "Phase Coherence as a Measure of Acoustic Quality, Part Two: Perceiving Engagement," *20th International Congress on Acoustics*, Sydney, Australia (2010 Aug.).

[19] T. Lokki, J. Pätynen, S. Tervo, S. Siltanen, and L. Savioja, "Engaging Concert Hall Acoustics Is Made Up of Temporal Envelope Preserving Reflections," *J. Acoust. Soc. Am. Express Letters*, vol. 129, pp. EL223–EL228 (2011 June).

[20] S. Santurette and T. Dau, "The Role of Temporal Fine Structure Information for the Low Pitch of High-Frequency Complex Tones," *J. Acoust. Soc. Am.*, vol. 129, pp. 282–292, (2011 Jan.).

[21] 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.*, vol. 100, pp. 2382–2392 (1996 Oct.).

[22] A. Kohlrausch and A. Sander, "Phase Effects in Masking Related to Dispersion in the Inner Ear. II. Masking Period Patterns of Short Targets," *J. Acoust. Soc. Am.*, vol. 97, pp. 1817–1829 (1995 Mar.).

[23] H. Pobloth and W. B. Kleijn, "On Phase Perception in Speech," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Phoenix, AZ, USA (1999 March).

[24] J. Breebaart, F. Nater, and A. Kohlrausch, "Spectral and Spatial Parameter Resolution Requirements for Parametric, Filter-Bank-Based HRTF Processing," *J. Audio Eng. Soc.*, vol. 58, pp. 126–140 (2010 Mar.).

[25] H. Hudde, *Communications Acoustics*, ch. "A Functional View on the Peripheral Human Hearing Organ" (Springer, 2005).

[26] B. C. J. Moore and B. R. Glasberg, "Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns," *J. Acoust. Soc. Am.*, vol. 74, pp. 750–753 (1983 Sept.).

[27] M. Wojtczk, J. A. Beim, C. Micheyl, and A. J. Oxenham, "Perception of Across-Frequency Asynchrony and

the Role of Cochlear Delays," *J. Acoust. Soc. Am.*, vol. 131, pp. 363–377 (2012 Jan.).

[28] S. Uppenkamp, S. Fobel, and R. D. Patterson, "The Effects of Temporal Asymmetry on the Detection and Perception of Short Chirps," *Hearing Research*, vol. 158, pp. 71–83 (2001 Aug.).

[29] M. Karjalainen, "A New Auditory Model for the Evaluation of Sound Quality of Audio Systems," in *IEEE International Conference on Acoustics, Speech and Signal Processing* (1985).

[30] P. X. Joris, L. H. Carney, P. H. Smith, and T. C. Yin, "Enhancement of Neural Synchronization in the Anteroventral Cochlear Nucleus. I. Responses to Tones at the Characteristic Frequency," *J. Neurophysiol.*, vol. 71, no. 3, pp. 1022–1036 (1994).

[31] B. C. J. Moore, *An Introduction to the Psychology of Hearing* (Academic Press, 1982).

[32] "MATLAB." MathWorks, http://www.mathworks.com/products/matlab/ (June 2013).

[33] B. C. J. Moore, "Interference Effects and Phase Sensitivity in Hearing," *Phil. Trans. R. Soc. Lond.*, pp. 833–858 (2002).

[34] ITU-R BS.1116, "Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems," International Telecommunications Union, Geneva, Switzerland (1997).

THE AUTHORS

Mikko-Ville Laitinen

Sascha Disch

Ville Pulkki

Mikko-Ville Laitinen received an M.Sc. (Tech.) degree, from the Helsinki University of Technology, Finland, in 2008, majoring in acoustics and audio signal processing. In 2006, he was working for the loudspeaker company Genelec, Finland. Since 2007 he has been working as a researcher at Aalto University, Finland, where his research topics include spatial audio, multichannel reproduction, psychoacoustics, and audio codecs. Furthermore, he was a visiting researcher at Fraunhofer IIS, Germany, in 2012. Currently he is completing his doctoral degree in the field of spatial audio at Aalto University.

●

Sascha Disch received his Diplom-Ingenieur degree in electrical engineering from the Technical University Hamburg-Harburg (TUHH), Germany in 1999. From 1999 to 2007 he joined the Fraunhofer Institute for Integrated Circuits (IIS), Erlangen, Germany. At Fraunhofer, he worked in research and development in the field of perceptual audio coding and audio processing. In MPEG standardization of parametric spatial audio coding he contributed as a developer and served as a co-editor of the MPEG Surround standard. From 2007 to 2010 he was a researcher at the Laboratory of Information Technology, Leibniz University Hanover (LUH), Germany, from which he received his Dr. Ingenieur degree in 2011. During that time, he also participated in the development of the Unified Speech and Audio Coding (USAC) standard at MPEG. Since 2010, Dr. Disch is affiliated with Fraunhofer IIS again. His research interests include waveform and parametric audio signal coding, audio bandwidth extension and digital audio effects.

●

Biography for Ville Pulkki was published in the September issue of the *Journal*.