The use of temporal cues for frequency discrimination of the fundamental component in a complex tone

Nicolas Le Goff
Philips Research, Prof. Holstlaan 4, Postbox WO 02, 5656AA Eindhoven, The Netherlands, nicolas.le.goff@philips.com

Armin Kohlrausch
Philips Research, Prof. Holstlaan 4, Postbox WO 02, 5656AA Eindhoven, The Netherlands, armin.kohlrausch@philips.com

Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Experiments were performed to study the effect of mistuning the fundamental component in a harmonic complex tone. The subjects had to distinguish the complex tone with its lowest harmonic at the fundamental frequency from a complex with the lowest harmonic being shifted in frequency. Complex tones had a fundamental frequency of 100 Hz and comprised the first 12 harmonic components. They had either a flat-spectrum or were low-pass filtered with a slope of -5 dB or -10 dB/oct. Frequency discrimination thresholds for a flat-spectrum complex tone were about 2 Hz and somewhat lower than data in the literature. Applying a spectral slope leads to significantly lower discrimination thresholds. In order to test potential explanations, additional experiments as well as model simulations were performed. The influence of the overall level was tested and the thresholds were compared with those for a pure tone of 100 Hz. Simulations were conducted using a gammachirp and an excitation pattern model. The detection of the mistuning in the case of a flat-spectrum complex tone seems to be done using spectral cues, while detection for a low-pass filtered complex appears to be based on temporal cues within auditory filters with center frequencies clearly above 100 Hz.

1 Introduction

The experiments described in this contribution were originally motivated by the perceptual evaluation of a new loudspeaker system for efficient reproduction of low frequencies [1]. As part of the design, listening experiments were conducted in order to investigate the perceptual sensitivity at low frequencies. The detectability of the mistuning of the fundamental frequency of a complex tone was therefore studied. Complex tones had a fundamental frequency of 100 Hz and consisted of the first twelve harmonics of equal amplitude. Experiments were also run with low-pass filtered complex tones, and an unexpected sensitivity was found in this particular case. Two hypotheses regarding this phenomenon were formulated and these are tested by the experiments reported in this contribution. The hypothesis were focused on the nature of the cues used in order to perform the detection of the mistuning. The additional experiments allowed to distinguish between the hypotheses. These hypotheses were further evaluated by model simulations.

2 General method

The objective of the experiments was to determine the amount by which the fundamental component in a complex tone can be mistuned before this mistuning can be detected. The measurements are divided into two parts and the results will be separately presented.

2.1 Stimuli

All stimuli were generated at a sampling frequency of 44.1 kHz, using Matlab running on a PC, and played out via an external D/A converter, which also served as headphone amplifier. The components were generated in zero phase. As was previously shown [4], the direction of mistuning does not influence the results, therefore the frequency of the mistuned fundamental component was always changed upwards. Each stimulus was windowed in the time domain with a Hanning-window function, thus they had a steady-state duration of 360 ms and started and ended with a ramp of 30 ms each. The stimuli were diotically presented to the subjects by means of Beyerdynamic DT 990 headphones. Ten different stimuli divided in two series were used. An overview of all the stimuli is presented in Table 1.

2.2 Procedure

Thresholds were determined using a 3-interval 3-alternative forced-choice procedure with an adaptive-frequency adjustment. At the start of each run the amount of mistuning was chosen as small as possible while the detection could still be very clearly done. As shown in Table 1 this amount differed across stimuli. On each trial the subject was presented with three successive complex tones, separated by 100 ms of silence. The mistuned complex, i.e. the target, occurred with equal probability in one of the three observation intervals. Imme-
Table 1: Summary of the characteristics of the stimuli

<table>
<thead>
<tr>
<th>Stimulus number</th>
<th>Slope dB/Oct</th>
<th>Overall level dB</th>
<th>Initial mistuning Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>70</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>-5</td>
<td>70</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>-10</td>
<td>70</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>pure tone</td>
<td>60</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>60</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>-5</td>
<td>60</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>-10</td>
<td>60</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>50</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>-5</td>
<td>50</td>
<td>3.5</td>
</tr>
<tr>
<td>10</td>
<td>-10</td>
<td>50</td>
<td>3.5</td>
</tr>
</tbody>
</table>

diate feedback was provided by displaying “wrong” or “correct” after each answer. The amount of mistuning depended on the subject’s responses using a two-down one-up procedure. Using this procedure, the frequency difference leading to 71% correct discrimination was estimated. There were three repetitions per subject and per condition. The experiments were run in a measurement room fulfilling the requirements specified by the recommendation ITU-RBS 1116 regarding sound insulation and background noise. Five subjects participated in the experiments, they had normal hearing and were experienced in listening experiments.

3 Experiment I

In this first experiment, four stimuli were used, three complex tones and a pure tone.

3.1 Flat-spectrum complex tone: stimulus 1

As a starting point, a simple complex tone consisting of the first twelve harmonics of equal amplitude (60 dB) was chosen (stimulus 1). The fundamental component had a frequency of 100 Hz and the overall level was equal to about 70 dB SPL.

3.2 Sloped-spectrum complex tone: stimuli 2 and 3

In connection with the application context, two extra stimuli with sloped spectrum and the same overall level as stimulus 1 were used. Stimulus 2 and stimulus 3 were low-pass version of stimulus 1, with a slope of respectively -5 dB/oct. and -10 dB/oct.

3.3 Pure tone: stimulus 4

A condition was also run using the same procedure, with a pure tone signal at a frequency of 100 Hz with a level of 60 dB SPL.

3.4 Results

The results from the five subjects are similar and Fig. 1 shows that the average thresholds were 2.4 Hz for the flat-spectrum complex tone (stimulus 1), 0.4 Hz for the -5 dB/oct. low-pass filtered complex tone (stimulus 2) and 0.25 Hz for the -10 dB/oct. low-pass filtered complex tone (stimulus 3). As a first remark, one can say that the threshold for flat-spectrum complex tone (stimulus 1) is somewhat in line with similar results published by Moore et al. [6]. A second remark is that the thresholds measured for the sloped-spectrum complex tone were respectively six and ten times lower than that of the flat-spectrum complex tone, which is unexpectedly low. At this point two hypotheses are considered in order to describe the higher sensitivity for the sloped-spectrum complex tone stimuli.

The first hypothesis is to consider that applying a spectral slope on a complex tone makes the fundamental frequency stand out more from the remaining harmonics. By using changes in the excitation pattern, i.e. spectral cues, the discrimination threshold should decrease with slope steepness and should be lowest for the fundamental component presented in isolation. Now looking at the result of the pure tone test one can see that the average detection threshold for stimulus 4 was 1.7 Hz. This threshold is lower than that of the flat-spectrum stimulus, but higher than that of the sloped-spectrum stimuli. If the main effect of applying a slope would have been an increase of the usability of the spectral cues, the pure tone case would have had the lowest thresholds, which is obviously not the case. In other words, one can say that the lower average thresholds of the sloped-spectrum complex
tones (stimulus 2 and stimulus 3), compared to the pure tone (stimulus 4), prove that subjects use extra information present in the spectral-slope complex tone compared to a pure tone. This additional information could consist of temporal cues in the envelope of the stimulus. Therefore, a second hypothesis is that not only spectral cues are used, but also temporal cues play a role in the detection process of the mistuning. In order to test this second hypothesis, experiment II was run.

4 Experiment II

Following the raise of the second hypothesis, new conditions were designed.

4.1 Level attenuated complex tones: stimuli 5 to 10

In connection with stimuli 1, 2 and 3, a pair of extra stimuli at lower overall level was designed for each of them. Stimuli 4 and 7 are the 60 and 50 dB versions of stimulus 1, stimulus 5 and 8 are the 60 and 50 dB versions of stimulus 2 and stimuli 6 and 9 are the 60 and 50 dB version of stimulus 3 (see Table 1).

4.2 Results

As shown in Fig. 2 decreasing the overall level of the complex tones decreased the detection threshold for the flat-spectrum complex tone decreasing from 2.4 Hz (stimulus 1) to 2.1 Hz (stimulus 5) and to 2 Hz (stimulus 8). However, it had the opposite effect on the complex tones with a sloped-spectrum. For the -5 dB/oct. filtered complex tones, the threshold increased from 0.4 Hz (stimulus 2) to 0.9 Hz (stimulus 6) and to 1.45 Hz (stimulus 9). Likewise, for the -10 dB/oct. filtered complex tones, the threshold increases from 0.25 Hz (stimulus 3) to 0.5 Hz (stimulus 7) and to 0.8 Hz (stimulus 10). A different approach, already suggested in [3], is now considered. It states that temporal cues could also be involved in the detection process of the mistuning. By applying a slope on a complex tone, the relative contribution of the fundamental component in higher filters is made stronger and thus increases the usability of temporal interactions within each individual filter. Furthermore this observation is in line with earlier observations made in [4]. The first observation was that the mistuning detection is at least partially based on the use of temporal cues in the auditory filters, and the second observation was that the cues used for the detection of the mistuning are at least partially placed two or three octaves higher than the fundamental component.

Now looking at the results for the level-attenuated stimuli in Fig. 2 one can notice that decreasing the overall level of the complex tones has a different effect on the measurement thresholds depending on whether the complex tone had a flat or a sloped-spectrum. In the case of a flat-spectrum, decreasing the overall level of the stimuli, causes the threshold to decrease (stimuli 1, 5 and 8). In the case of the two different sloped-spectrums, -5 dB/oct. and -10 dB/oct., decreasing the overall level causes a raise of the detection thresholds (stimuli 2,6,9 and stimuli 3,7,10). Taking a look back at psychoacoustic models, it is well known that the width of the auditory filters depends on the amplitude of the signal traveling along the basilar membrane, a stronger signal causing wider auditory filters and vice versa. Consequently, the reduction of the overall level of the stimuli, makes the width of the auditory filter smaller. Since the temporal information is caused by the interaction of adjacent harmonics within one filter, narrower auditory filters would contain less temporal information. Therefore, one would expect a raise of the detection thresholds for conditions in which subjects use temporal cues, which is what one can observe on the basis of the present results. Furthermore, narrower filters increase the frequency selectivity. Therefore, the detection of a phenomenon like a mistuning, by means of spectral cues, would be easier with lower stimuli levels. Once again, this is what one can actually observe looking at the present results. This indicates that subjects predominantly use spectral cues to detect the mistuning in the case of flat-spectrum stimuli and that subjects predominantly use temporal cues to detect the mistuning in the case of sloped-spectrum stimuli.

5 Simulations

The previous conclusions were evaluated by using an excitation pattern model as well as a gammachirp model. As a first step, the level of usability of spectral cues for different stimuli was evaluated by using an
The excitation pattern model. The excitation patterns for a flat-spectrum complex tone (stimulus 1), a -10 dB/oct. sloped-spectrum complex tone (stimulus 3) and a pure tone (stimulus 4) were compared. The simulations were run using an implementation by Rao et al. [8] of the excitation pattern model by Glasberg and Moore [7]. The model was run using the stimuli used for the experiments and the result is shown in Fig. 3. One should keep in mind that the overall level of stimuli 1 and 3 was kept constant, therefore the level of the lowest components of stimulus 3 was higher than for stimulus 1. For the clarity of the representation, the level of the stimulus 3 is attenuated so that the patterns match at 100 Hz. The representation shows that a variation of the 100 Hz component of the three stimuli has a different effect on the three excitation patterns. The excitation pattern of the pure tone (continuous) is the most affected, because the whole pattern is shifted horizontally. The complex tone with a -10 dB/oct. slope (dotted) is less modified and the excitation pattern of the flat-spectrum (dashed) complex tone is the least modified. This decrease of the effect is due to the spectral masking caused by the presence of the higher harmonics in the complex tones. This observation gives a different conclusion regarding the use of spectral cues than the experimental data. Effectively, the detection thresholds for these three stimuli show that, in the case of the sloped-spectrum stimulus, the detection cannot be only based on spectral cues. Therefore, it was decided to run simulations with a gammachirp filter bank in order to evaluate the potential contributions of temporal cues.

Simulations were run using a gammachirp filter bank as well as an outer and middle-ear filtering as specified by [2]. The simulations were run using a model by Irino [5]. Fifty channels between 100 and 1000 Hz were used for the computation. Figure 4 allows to compare the envelopes of the output signals of the auditory filter centered at 400 Hz for stimuli 1 (left panel) and 3 (right panel). The continuous lines show the envelope for the harmonic stimulus, the dashed line represent the envelope for a stimulus with a mistuning of 6 Hz. For both stimuli, a mistuning of 6 Hz is larger than their respective detection thresholds. Even in this extreme conditions, one can easily see that the two curves for the sloped-spectrum stimulus (right) are more different from each other than those for the flat-spectrum stimulus (left). The 50-ms section was even chosen in order to see a maximum difference between the two curves of the flat-spectrum stimulus (left). Therefore the hypothesis that the detection of the mistuning for the sloped-complex is based on the detection of the variation in the temporal envelope of the signals in the auditory filters clearly above the fundamental component is qualitatively confirmed.

In order to evaluate this hypothesis, a new representation of the output signals of the gammachirp filter bank was computed. The idea is that the subjects are capable of storing and comparing an internal representation of the target and reference signals during the 3-AFC paradigm. The differences in the waveform between the target and the reference are therefore the cue for detection. Thus it was decided to plot the difference between the corresponding outputs for the harmonic and the mistuned stimulus. As before, it is actually the envelope of the signals that is represented on a logarithmic scale. This time the model was used for making a comparison abased on a filter centered on 350 Hz for stimulus 1 and 3 and for two different mistunings of 6 and 0.3 Hz. The results are shown in Fig. 5. One can see that, for a given mistuning, the difference of the
logarithm of the envelopes of the output of the model for an auditory filter centered at 350 Hz, is 5 to 9 dB higher for the sloped-spectrum complex tone (bottom) than for the flat-spectrum stimulus (top), for both values of mistuning. This observation supports the hypothesis that for a sloped-spectrum complex tone, the detection of the mistuning would be much easier by using temporal cues, and more specifically that subjects store and compare an internal representation of the envelopes of the mistuned stimulus and the harmonic stimulus in auditory filters above the fundamental frequency.

6 Conclusion

The presented experiments and observations based on model simulation results show that the detection of the mistuning of the fundamental component in a complex tone is a process that has a different nature depending on the characteristics of the complex tone itself. Based on a strict spectral approach, the excitation pattern model suggests that a mistuning of the fundamental component in the sloped-spectrum complex tone is more easily detected on the basis of a frequency cue than for a flat-spectrum complex tone. However, this does not mean the detection of such a mistuning is not based on spectral cues for the flat-spectrum complex tone. Furthermore, experiments show that subjects are actually using additional cues for the detection in the case of the sloped-spectrum complex tones. Results from the simulations from a gammachirp filterbank also show that the usability of the temporal cues within each auditory filter a few octaves above the fundamental frequency is much better for a sloped-spectrum complex tone compared to a flat-spectrum complex tone. These results, combined with previous observations regarding the use of temporal cues in conditions with the sloped-complex tones [4] suggest that in the case of a sloped-spectrum complex tone the detection is mainly based on the use of temporal cues. Regarding the detection in a flat-spectrum complex tone, the differences of temporal variations in individual auditory filters between harmonic and mistuned signals are not large enough to allow the use of temporal cues, suggesting that the detection is most likely based on spectral cues.

References