Thresholds for hearing mistuning of the fundamental component in a complex sound

Nicolas Le Goff, Ronald M. Aarts, Armin Kohlrausch

Philips Research Laboratories WO02
Prof. Holstlaan 4
5656 AA Eindhoven
The Netherlands
{nicolas.le.goff,ronald.m.aarts,armin.kohlrausch}@philips.com

Abstract

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. Thresholds were determined for fundamental frequencies of 60 Hz and 100 Hz. The complexes had either a flat spectrum, or components were generated with a spectral slope of -5 dB/octave or -10 dB/octave. Conditions with the second or both the second and third harmonic omitted from the complex were also included. In additional conditions, the complex was slowly amplitude-modulated and/or presented with a simultaneous distracting sound. Results without modulation and distracter are comparable to data in the literature. Presenting the sound with a spectral slope strongly lowers thresholds. Adding a distracter or applying amplitude modulation both lead to higher thresholds.

1. Introduction

Sound reproduction with small loudspeakers at low frequencies with a reasonable efficiency is very difficult, since the efficiency is proportional to the square of the product cone area and force factor (Bl) while it is inversely proportional to the moving mass [2]. At the resonance frequency, however, the efficiency can be much higher. A new optimal design for such a loudspeaker has been devised—which is beyond the scope of this present paper—which has a much higher efficiency than current bass drivers. This is realized at the expense of a decreased efficiency for frequencies beyond the resonance frequency. Therefore a special frequency mapping was applied to overcome this problem. To study the perceptual consequences of this mapping technique, psychophysical tests were conducted, which will be the topic of the present paper. The stimuli will have an increasing complexity, where the simplest ones are designed to allow comparison with results known from the literature. The most complex ones are similar to some types of pop music. In this way the stimuli mimic properties of sounds which might be reproduced via this special driver.

2. Method

2.1. Stimuli

The goal of the experiments was to determine the amount by which the fundamental component of a harmonic complex can be modified in frequency before this mistuning becomes detectable. Fifteen different stimuli were used, which had fundamental frequencies of either 100 Hz or 60 Hz. MATLAB was used to generate the digital stimuli with a sampling frequency of 44.1 kHz. The level of all complex tones was adjusted to 70.3±0.5 dB. In all stimuli the components within the complex tone were added in zero phase. The stimuli had onsets and offsets ramps of 30 ms duration, with envelopes shaped according to a Hanning window. The steady-state duration was 340 ms. Stimuli were presented diotically through both earphones of a Beyerdynamics DT 990 headset. The frequency response of the headphones, as measured in a Bruel & Kjær artificial ear type 4153, was flat within ±4 dB over the range 60–2000 Hz. The results of the experiment are presented separately for the fundamental frequency set to 100 Hz and 60 Hz. Table 1 provides a summary of the stimulus parameters.

2.1.1. Group 1: f_0=100 Hz

This group comprises stimuli 1 to 6. All stimuli had a fundamental frequency of 100 Hz. Stimulus numbers 1,2,3 and 6 consisted of the first 12 components, while stimuli 4 and 5 contained the first 6 and 18 components, respectively. Stimuli 1 and 6 had a flat spectrum. The complex tones in the stimuli 2,4 and 5 had a spectral slope of -5 dB/octave, while the stimulus 3 had a spectral slope of -10 dB/octave.

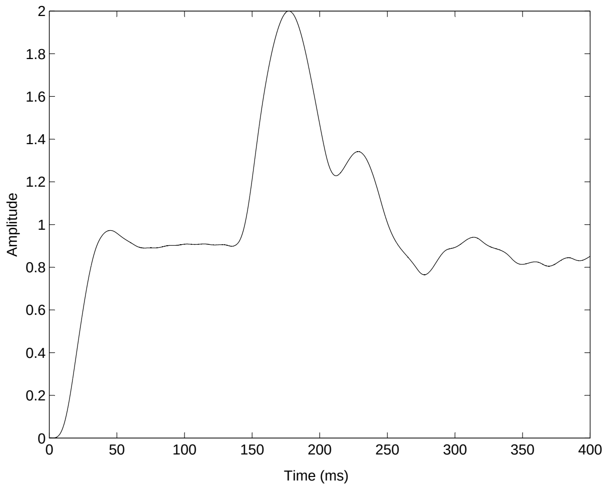

Figure 1: *Envelope of the complex tone in stimuli 10 and 15.*

2.1.2. Group 2: f_0=60 Hz

This group comprises stimuli 7 to 15. All stimuli had a fundamental frequency of 60 Hz, and a spectral slope of -5 dB/octave was applied to the complex tones. Stimulus 7 consisted of the first 12 harmonics. In all other stimuli, the second harmonic, with a frequency of 120 Hz, was removed. In stimulus 9 the third harmonic with a frequency of 180 Hz was also removed. Stimulus 10 was modified by an amplitude modulation shown in Fig. 1. The modulator was designed by envelope extraction of a beat in a pop music song in order to reproduce amplitude variations occurring in such music in the low frequency region. In stimulus 11 a Gaussian noise, bandlimited between 750 Hz and 1250 Hz, with an overall level of 55 dB SPL was added. In stimulus 12 a speech-like sound was added as a distracter (see Fig. 3 for the stimulus spectrum). In, respectively, stimuli 13 and 14, a 750 Hz and 500 Hz high-pass filtered version of the voice was added. Stimulus 15 had both the full frequency range version of the voice distracter and the modulator shown in Fig. 1 was applied to the complex tone.

2.2. Procedure

Thresholds were determined using a three-alternative forced-choice (3AFC) procedure in combination with an adaptive frequency adjustment. In the discrimination task, three observation intervals were presented during each trial. Two intervals contained the reference stimulus, while one interval contained the target stimulus. The silent interval between observation intervals was 100 ms. For all stimulus conditions, with the exception of stimulus 6, the frequency of the fundamental component was shifted upwards. The subjects' task was to indicate which interval contained the target. The procedure was designed such that the first stimuli were very easily discriminable.

Table 1: *Parameters of the stimuli. N:noise, V: Voice, HP7, HP5: high pass filtered voice at 750 Hz and 500 Hz respectively*

Stim. #	# of Comp.	Slope *dB/Oct.*	Lacking Comp.#	Mod.	Dist.
1	12	0	-	-	-
2	12	-5	-	-	-
3	12	-10	-	-	-
4	6	-5	-	-	-
5	18	-5	-	-	-
6	12	0	-	-	-
7	12	-5	-	-	-
8	12	-5	2	-	-
9	12	-5	2,3	-	-
10	12	-5	2	Yes	-
11	12	-5	2	-	N
12	12	-5	2	-	V
13	12	-5	2	-	HP7
14	12	-5	2	-	HP5
15	12	-5	2	Yes	V

In this way the subject could both concentrate on a specific frequency range and focus on the cues (s)he had to follow. Frequency shifts were controlled with a 2-down 1-up adaptive tracking procedure, meaning that the frequency of the fundamental component was shifted closer to its harmonic value after two correct answers while only one wrong answer was required to push it further from its harmonic value. At the start of each run, frequencies were adjusted with 0.3 Hz steps. Step sizes were halved after each second reversal until a minimum step-size of 0.02 Hz was reached. Afterwards 8 other reversals were measured and the median of these last 8 reversals was used as the measured threshold of that run. Immediate feedback regarding the correctness of the response was provided and the next trial started immediately after the answer. For each condition three thresholds were measured per subject.

2.3. Subjects

All five subjects were familiar with psychoacoustical experiments and had normal hearing.

3. Experimental results

As mentioned before, the experiment is divided in two groups. The data are presented separately for each group and are computed by averaging across the results of the five subjects and the three repetitions per subject.

3.1. Results from Group 1

In this first part, the complex tones had a fundamental frequency of f_0=100 Hz. The results for stimulus numbers

1 to 6 are shown in Fig. 2 by the first 6 bars. The thresholds for detecting the frequency shift are expressed as a percentage with respect to the frequency of the fundamental component, i.e. 100 Hz in this group. In this case the results agree with the thresholds expressed in Hz.
The results show the following:

- Influence of spectral slope (stimuli 1, 2 and 3)
  The measured thresholds are highest for the complex with a flat spectrum and amount to a bit more than 2%. Applying a slope on the amplitudes of the complex tone noticeably decreases the threshold of detection by factors of 3 and 5 in the case of slopes of -5 and -10 dB/octave, respectively.

- Influence of the number of components (stimuli 2, 4 and 5)
  A higher number of components in the complex tone makes the detectability of the mistuning more difficult and vice versa. By going from 6 (L3) to 18 components (L4), thresholds increase from 0.42 to 0.82 %.

- Direction of frequency change (stimuli 1 and 6)
  The direction of mistuning the frequency does not have a significant influence on the threshold values, especially considering that the standard deviation for data corresponding to stimuli 1 and 6 is, respectively, 0.9 % and 0.8 %.

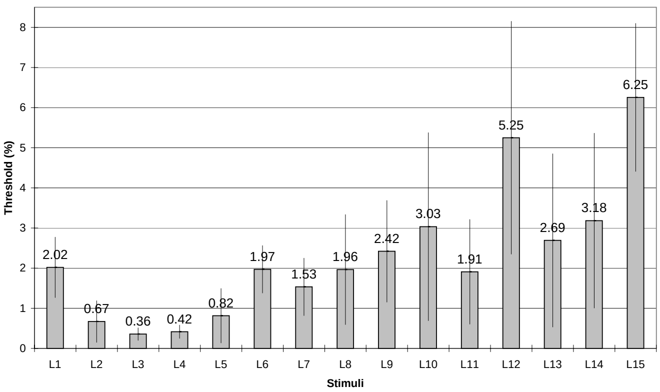

Figure 2: *Average thresholds values and ± standard deviation for all subjects and three repetitions for each stimulus.*

3.2. Results from Group 2

The results for this group of stimuli are represented by bars 7 to 15 in Fig. 2. As in the previous figure the threshold for detectability of mistuning is expressed as a percentage with respect to the frequency of the fundamental component, i.e. 60 Hz in this group. The results reveal the following trends:

- Influence of the fundamental frequency (stimuli 7 and 2)
  Decreasing the fundamental frequency from 100 Hz to 60 Hz causes the (relative) thresholds to increase by more than a factor of 2, from 0.67% to 1.53%.

- Influence of leaving out individual harmonics (stimuli 7, 8 and 9)
  This comparison clearly shows that the threshold of detection increases when the first individual harmonics above the fundamentalare removed. This effect is shown for removing one (L8) and two (L9) components.

- Influence of an irregular amplitude modulation (stimuli 8 and 10)
  Adding the amplitude modulator shown in Fig. 1 to the stimulus 8 increases the threshold by about 50 %.

- Influence of an additional noise distracter (stimuli 8 and 11)
  Adding the Gaussian noise does not have any significant effect, especially taking into account the standard deviation of the data for stimuli 8 and 11 of 0.8 % and 0.7 %, respectively.

- Influence of a speech-like distracter (stimuli 8, 12–14)
  Adding a speech-like sound to stimulus 8 creates a significant difficulty. The threshold increases by a factor of 2.7 and reaches 5.25% for the stimulus 12. One has to take into account that in this case the voice has quite some spectral overlap with the complex tone as shown in Fig.3. By removing the low-frequency part of the distracter (L13 and L14), the thresholds become smaller, but stay clearly above the value obtained without this distracter (L8).

- Combined influence of distracter and modulator (stimuli 8,10, 12 and 15)
  Stimulus 15 represents the combination of the effect of the voice present in stimulus 12 and the amplitude modulator present in stimulus 10. All in all these two distracters causes a raise of the threshold by a factor of 3.2 leading to a value of 6.25%.

4. Discussion

In the discussion we will first compare the data, mainly from group 1, with results from the literature, and will then discuss the outcomes from group 2 in the light of the intended loudspeaker application.

As said before, the condition 1 was chosen in line with one of the conditions measured in the paper by Moore et al. [1]. His threshold for mistuning the fundamental of a 100 Hz complex was about 5 % which is more

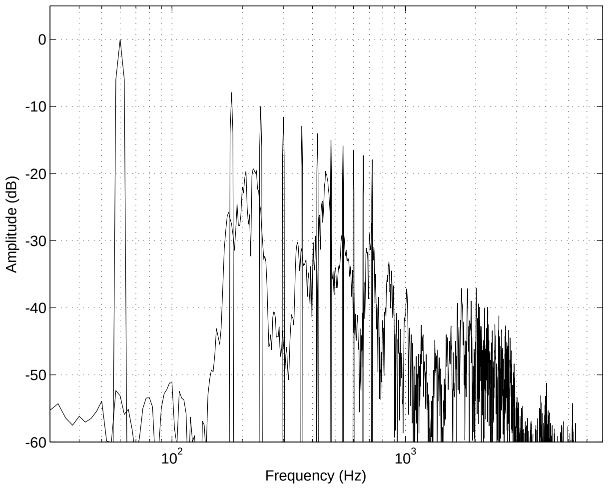

Figure 3: *Spectrum of the complex tone and the speech-like sound in stimulus 12.*

than a factor 2 higher than the value obtained in condition L1. One reason for this lower value might be a small experimental difference: while Moore et al. presented their stimuli to one ear only, presentation was diotic in the present experiments. It is known from the literature that for tasks like pure-tone frequency discrimination, there is a monaural-diotic difference, leading to lower thresholds in the diotic condition (cf. [3, 4]). It can be stated that both these thresholds values for mistuning the fundamental in a complex tone are higher than pure-tone frequency just noticeable differences (jnds) which are about 1 % at 100 Hz [5]. The trend that thresholds increase for a lower frequency of the fundamental is in line with observations for both pure tones and harmonic complexes.

Introducing a slope to the complex strongly reduces thresholds. One possibility is that such a slope makes the fundamental to stand out spectrally from the remaining harmonics which could make the use of spectral cues more effective. Another possibility is that, with a steeper slope, the temporal interactions in filters tuned to harmonics above the fundamental increases, because the low-frequency attenuation of these filters is counteracted by the spectral slope in the stimulus. In this way, the relative contribution of the fundamental to the filtered output waveform is stronger and mistuning could be detected more easily based on the change in the temporal waveform at the output of individual auditory filters. There are a number of observations which favor the temporal hypothesis over the spectral one. First of all, the discrimination thresholds for a complex with a 5-dB/octave slope are much lower than pure-tone frequency jnds from the study by Formby [5]. For 100 Hz, the values are 0.67 % vs. 1.2 %, respectively, and at 60 Hz, the corresponding values are 1.5 % and 4.5 %. Secondly, leaving out the harmonics directly above the fundamental makes it to stand out even more spectrally, which should lead to lower thresholds according to the spectral hypothesis. On the other hand, with fewer harmonics, the strength of the temporal interaction in higher filters is reduced. The observation that thresholds *increase* when leaving out harmonics (conditions 8 and 9) again supports the explanation based on temporal effects.

The stimuli in group 2 showed that with increasing temporal and spectral complexity the ability to detect the mistuning of the fundamental decreases. The effect of an irregular modulator could be that it becomes harder to use subtle changes in temporal patterns which lead to the low thresholds for the isolated and unmodulated complex. The strong influence of a distracting voice is probably based both on direct masking if there is strong spectral overlap (stimulus 12) and on a distracting effect (stimuli 12–14), if there is temporal and spectral structure present in the distracter (compare the negligible effect of the noise distracter). All these results support the notion that mistuning becomes difficult to detect once the target complex is embedded in a spectrally and temporally *rich* sound context, as it is typical for applications in modern multimedia reproduction devices.

5. References

[1] Moore, B.C.J., Peters, R.W., and Glasberg, B.R. ”Thresholds for the detection of inharmonicity in complex tones”, J. Acoust. Soc. Amer., vol. 77, 1861–1867, 1985.

[2] Vanderkooy, J., Boers, P.M., and Aarts, R.M. ”Direct-Radiator Loudspeaker Systems with High Bl”, J. Audio Eng. Soc., vol. 51, 625–634, 2003.

[3] Langhans, A., and Kohlrausch, A. ”Differences in auditory performance between monaural and diotic conditions. I: Masked thresholds in frozen noise”, J. Acoust. Soc. Am., vol. 91, 3456– 3470, 1992.

[4] Jesteadt, W., and Wier, C.C. ”Comparison of monaural and binaural discrimination of intensity and frequency”, J. Acoust. Soc. Am., vol. 61, 1599–1603, 1977.

[5] Formby, C. ”Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners”, J. Acoust. Soc. Amer., vol. 78, 70–77, 1985.