Frequency Discrimination in a Simultaneous Two-Tone Signal

P. Kleczkowski* and M. Pluta

AGH University of Science and Technology, Faculty of Mechanical Engineering and Robotics
Department of Mechanics and Vibroacoustics, al. A. Mickiewicza 30, 30-059 Krakow, Poland

Discrimination of just noticeable change in pitch has been widely investigated. In contrast, works on discrimination of two simultaneous pitches produced by two tones have been scarce. In this work two methods were used to find the minimum frequency interval in a two-tone at which listeners perceive two distinct pitches. In the first, listeners tuned manually the frequency interval between components in a two-tone. In the second, the frequency interval between tones in a two-tone was switched between five different fixed values, and a pitch of a probe tone was compared with pitches in a two-tone. Both methods were used in two different groups of listeners, musicians and non-musicians. The first method yielded substantially narrower intervals. Both methods revealed that musicians were able to discriminate pitches at narrower intervals than non-musicians. The raw results of tests were evaluated by two approaches, one was based on fitting of the psychometric function and the other on statistical tests.

PACS: 43.66.−x

1. Introduction

The pitch of two-tones has been widely investigated. Two of the many examples are [1] where recognition of pitch of pairs of neighbouring partials was investigated in the context of simple melodies and [2] where residue pitch of two-component complexes was investigated with the conclusion that the nonlinearity of the hearing system did not influence the perception of the residual pitch.

The simultaneous discrimination of pitch of two concurrent sounds is one of the fundamental properties of the ear, and probably the fundamental determinant of the ear's frequency resolution. Despite that, the bibliography of this topic is scarce.

Thurlow and Bernstein [3] used two simultaneous sinusoidal tones of the same amplitude, further referred to as a two-tone. They examined two subjects only, of which the first reported the threshold frequency interval of between 5% (at 4 kHz) and 10% of the standard frequency, and the second reported values three to ten times larger. Both dichotic and monaural presentations were investigated, with thresholds higher by a couple of percent in the latter case.

Plomp [4] in reviewing earlier works noticed that the estimates of resolution bandwidths ranged from less than 10 Hz to more than 500 Hz. He attributed that to the vagueness of the criterion of judgements used. In Ref. [3] the perception of two distinct pitches was evaluated. Plomp and co-workers [4, 5] investigated frequency resolution of the ear by examining the ability to hear out one particular harmonic in a harmonic tone complex. They presented a complex tone followed by a tone with the frequency matching one of the harmonics in a complex, or followed by a tone with a frequency halfway between them. Their findings are well known: five to seven

lowest harmonics could be heard in that experiment. The resolvability of harmonics in inharmonic complexes was investigated by Soderquist for four musicians and four non-musicians [6] and the musicians were found to detect pitches at about 10% narrower intervals.

In some works ([7] and references therein, [8]) so--called complementary pairs of two-tones were investigated, i.e. pairs of two-tones with slightly differing amplitudes, where the lower tone was slightly more intense in one two-tone and in the other the higher tone was more intense. The resolvability of pitches between these pairs can be considered a different class of pitch resolvability and the authors argued that the temporal fine structure of two-tones might be involved.

To the best knowledge of the authors, none of the earlier works analysed frequency discrimination in two-tones in different groups of listeners. In this work, two groups of subjects were examined: non-musicians with some audio or musical experience as well as experience in psychoacoustic tests, and active musicians with no experience in psychoacoustic tests. Two methods have been applied and their results were compared. The first was similar to that used in [3], but longer and more precisely controlled stimuli were used. The stimuli were longer to make them more natural for musicians. The second was similar to the method of [4, 5], but the stimuli were more precisely controlled and the effect of the sequence of stimuli was analysed.

2. Experiment 1

A subject listened to two simultaneous and continuous tones of the same amplitudes. Their frequencies were separated by a small interval. One of the tones had a constant frequency of a standard value (250, 1000, and 4000 Hz, respectively). The subject was asked to manually increase the other frequency until he/she clearly

* corresponding author; e-mail: `kleczkow@agh.edu.pl`

heard two separate pitches. The adjustment was performed by pressing a computer key in 1 Hz steps. At the beginning of the experiment both tones had identical phase and the change of the frequency of the other tone was upwards. The subject could shift that frequency upwards or downwards (but not below the standard frequency). The subject did not see the computer screen so she/he did not know what the frequency difference between the tones was.

The "Siggen" software [9] and the Asus 1201NL netbook were used to generate the sounds. They were converted to an analogue signal by the M-Audio Fast Track Pro USB audio interface and reproduced diotically by Beyerdynamic DT770 Pro closed headphones at the sound level of 40 dB(A), in order to minimise combination tones. The listeners were seated in a sound insulated room of a recording studio. Background noise level was 28 dB(A).

The experiment was participated by 19 students of the major in acoustical engineering at the AGH University of Science and Technology (denoted AE in the results) and 15 students of the Faculty of Instruments of the Academy of Music in Cracow (denoted AM in the results). The test took 2 to 5 min for each participant.

3. Experiment 2

In this experiment, the setup and the test conditions were the same as in experiment 1.

The experimental paradigm was similar to that used by Plomp [4, 5]. Sound samples were organized in trials. Each trial consisted of two time intervals, and each time interval contained a pair of stimuli. The first element of the pair was a two-tone, separated by a fixed interval in frequency. The standard frequencies and intervals are shown in Table I.

TABLE I

Standard frequencies and the corresponding frequency intervals in two-tones used in experiment 2.

Standard frequency [Hz]	Intervals [Hz]				
250	4	8	16	32	64
1000	16	32	64	128	256
4000	64	128	256	512	1024

The second element was a single tone burst of the same duration and equal amplitude. The experiment consisted in comparing the pairs within the trial, which differed in the frequency of the second element of the pair. Its frequency varied according to the test condition: it could be either equal to the lower frequency in the two-tone (further referred to as condition L), equal to the higher frequency in the two-tone (condition H), or equal to the arithmetic mean of these frequencies (condition M). In each trial, the pair containing the single tone burst of

condition L or H was compared with the pair containing the tone of condition M. The experimental task was to choose the interval (pair) that contained the single tone of the frequency equal to one of the frequencies in the two-tone, i.e. the tone of condition L or H but not M.

Four arrangements of conditions within the trial were used. All of them pertain to the second element in the time interval (pair), as the first element in each condition was the two-tone: H followed by M (further referred to as HM), L followed by M (LM), M followed by H (MH) and M followed by L (ML). These arrangements are shown in Fig. 1.
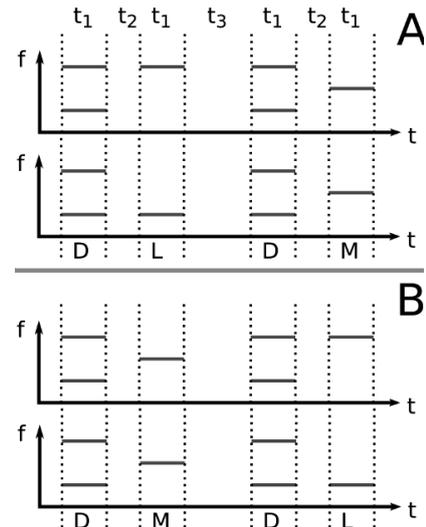

Fig. 1. Four arrangements of conditions within the trial. D (for "double") — the two-tone, H — the tone following the two-tone has the frequency of the higher tone, M — the tone following the two-tone has the frequency equal to $(f_H - f_L)/2$, L — the tone following has the frequency of the lower tone. Part (B) contains the sequence of the following tones reversed relative to part (A).

The duration of each of the tone bursts was 1.25 s, including linearly ramped attack and decay, each lasting 0.1 s. The time structure of the trial is shown in Fig. 2. The time intervals were separated by 1 s, while the bursts within the interval were separated by 0.1 s.
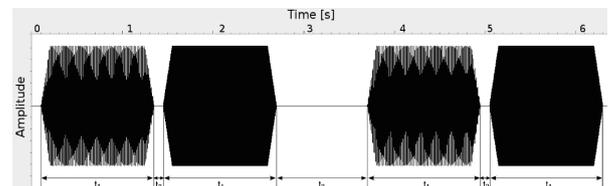

Fig. 2. Time structure of the stimuli within the trial. t_1 — 1.25 s, t_2 — 0.1 s, attack/decay — 0.1 s, t_3 — 1 s.

A custom software for running psychoacoustic tests developed in C programming language by the second author

(MP) was used. In this software, the subject activates a trial at any time by pressing a "Play" button in the screen. The test was run according to the two alternative forced choice (2AFC) paradigm. The screen contained buttons for two alternative responses. Each trial was presented once, the time to answer was unlimited. Feedback was not provided. The subject was informed about the progress of the test.

Three standard frequencies were investigated, with five different frequency intervals (Table I). For each combination of frequencies four arrangements of Fig. 1 were used as trials and each was repeated three times in the test. Thus, the total number of trials presented to the subject was $3 \times 5 \times 4 \times 3 = 180$. All stimuli were generated offline by another custom software written in C and were stored in as "wave" files in computers used for running the tests. The test software mentioned above controlled the playback of these files. The test was divided in three sessions, each devoted to one standard frequency. The 60 trials in each session were reproduced in random order, identical for each of the participants. The sessions were separated by short breaks.

4. Results

4.1. Experiment 1

The raw results were the intervals in Hz claimed by individual subjects as minimum to raise the percept of two pitches. The results were analysed for each of the two groups of participants (AE and AM). All means of threshold intervals were lower in the AM group and it was statistically significant at $p < 0.05$. The results of experiment 1 are shown in Table II.

TABLE II

Results of experiment 1. Groups of subjects: AM — Academy of Music, AE — acoustical engineering. μ — mean, s — standard deviation.

Standard frequency [Hz]	AM $\mu \pm s$ [Hz]	IA $\mu \pm s$ [Hz]	AM median [Hz]	IA median [Hz]	AM+IA median [Hz]
250	7.73 ± 4.17	31.21 ± 17.72	7	26	12.5
1000	22.4 ± 21.31	54.95 ± 28.83	14	60	32.5
4000	26.93 ± 39.47	58.63 ± 45.23	12	51	24

For the AM group (musicians) the distributions of the results significantly departed from normal, as shown by both skewness and kurtosis parameters exceeding 2. For the AE group (non-musicians) only the distribution at standard frequency of 4 kHz significantly departed from normal. According to the F-test the variances showed no significant differences at 1 kHz and 4 kHz, but a substantial difference at 250 Hz (the ratio of 18). The variance was lower in the results of the AM group for all standard frequencies.

Both the Kolmogorov–Smirnov and Wilcoxon non-parametric tests confirmed statistical significance of the differences of means between the AM and AE groups at $p < 0.05$ significance level. Also the t-test, although not quite reliable in this case, was performed for the 1 kHz and 4 kHz conditions and confirmed statistical significance of differences in means.

4.2. Experiment 2

The raw results were percentages of correct responses in each experimental condition. In total 120 percent correct results were obtained: 2 groups $\times$3 standard frequencies $\times$5 intervals $\times$4 sequences. As expected, the results improved as the size of the interval was increased.

Figure 3 presents percent correct results as a function of the interval, for AM and AE groups, separately for each of the standard frequencies. The fits shown in Fig. 3 are described in Sect. 4.2.1.

The number of frequency intervals investigated in experiment 2 was five, with equal distances in the logarithmic scale. Therefore, in order to obtain a more accurate value of the threshold of discrimination of two pitches an interpolation procedure was used. It was based on two different methods of evaluation of results: fitting of the psychometric function and a statistical test. The objective was to compare the performance of both approaches.

4.2.1. Evaluation of results with the psychometric function

The objective of the experiment was to find the average threshold of perception in each group for each standard frequency (six thresholds). The psychometric functions were obtained from 5 data points with the Weibull sigmoidal function fit given by [10]:

$$y = A + (1 - A - B)\left[1 - \exp\left(-\left(\frac{x}{C}\right)^D\right)\right], \qquad (1)$$

where A denotes the distance of the lower bound from 0, B denotes the distance of the upper bound from 1, C
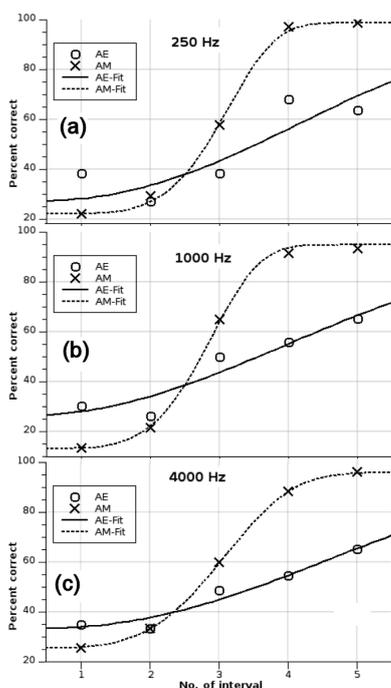
Fig. 3. Percent correct results as a function of the interval, for AM and AE groups, separately for each of the standard frequencies: (a) standard frequency (s.f.): 250 Hz, (b) s.f.: 1000 Hz, (c) s.f.: 4000 Hz. Results are fitted by the psychometric function. Above parts: standard frequencies. No. of the interval $= j + \log_2(f)$ (see (1)).

and D are the parameters of the curve returned by the Levenberg–Marquardt least squares fitting procedure used [11], x denotes the magnitude of a stimulus. In our case: $x = j + \log_2(f)$, where j is a constant as in Fig. 3, f denotes frequency.
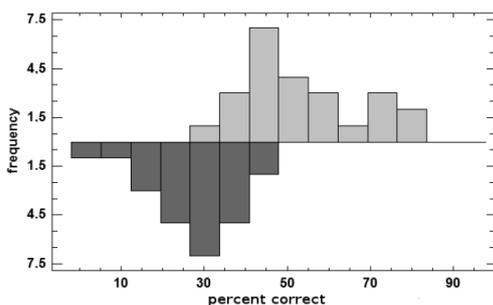

Fig. 4. The comparison of frequency histograms: the distribution of results in the middle-sized interval, shifted downwards by 3.4% (above the horizontal axis) and the distribution of results in the narrowest interval (below the horizontal axis).

In the psychometric function applied to our experiment the abscissa is the logarithmic scale of frequency separations, so that the corresponding scale of subsequent numbers of frequency intervals is linear.

Due to the bias (see discussion) in the nature of the experiment 2 in all conditions the lower bounds of percent correct results were below 50%. The significance of these results was tested by the following procedure. The results for the middle-sized interval in all conditions, with an average of 53.4% were taken as the reference distribution, as a percent correct result close to 50% is normally expected as a lower bound in 2AFC experiments. This distribution was shifted downwards by 3.4% to obtain a distribution centred at 50%. Then it was compared by the t-tests with the distributions obtained from all results at the narrowest frequency interval and the second narrowest interval (both had percent correct averages below 50%). In both cases the results turned out highly significant ($p < 10^{-6}$). The two distributions for the narrowest intervals were also compared between themselves and this test showed no significant difference of their averages. Figure 4 presents the comparison of frequency histograms of the reference distribution with the distribution of results in the narrowest interval.

Because of these results the lower bounds A in (1) were assumed as those obtained as percent correct results in each of the three standard frequencies in the AM group, respectively: 22.2% at 250 Hz, 13.3% at 1 kHz, and 25.6% at 4 kHz. In the AE group the percent correct results were non-monotonic in two narrowest intervals, therefore the lower bound A was chosen as the lower in these two intervals producing 27.2% at 250 Hz, 26.4% at 1 kHz, and 33.3% at 4 kHz.

TABLE III

Threshold values of frequency separation for pitch discrimination obtained by both methods. AE — non-musicians, AM — musicians, TPF — threshold from psychometric function, TBD — threshold from binomial distribution.

Standard frequency [Hz]	Group of listeners	TPF [Hz]	TBD [Hz]
250	AE	40.4	20.65
250	AM	16.4	12.08
1000	AE	174.6	55.22
1000	AM	55.9	36.82
4000	AE	922.2	249, 66
4000	AM	261.8	209.60

The upper bound was assumed equal to 95%, as is often the case because of the lapse rate [12, 13] unless it was measured as higher. The fits obtained are shown in Fig. 3. The thresholds of perception were obtained as the values of abscissa corresponding to percent correct values given by $(u - l)/2$, where u denotes the upper bound and l denotes the lower bound, and are given in Table III.

4.2.2. Evaluation of results with the binomial distribution

The binomial distribution was used to determine the unidirectional cumulative probability

$$P(k \geq k_0) = \sum_{k=k_0}^{N} \frac{N!}{k!(N-k)!} p^k q^{N-k}, \qquad (2)$$

where k denotes a number of correct answers out of N opportunities, k_0 denotes a threshold value of correct answers to obtain an assumed cumulative probability, N denotes the no. of participants (15 in AM group, 23 in AE group) times the no. of repetitions of trials (3), producing respectively $N_{AM} = 45$ for AM and $N_{AE} = 69$ for AE, p denotes probability of the correct answer and was different in all conditions: in AM — 0.222 at 250 Hz, 0.133 at 1 kHz, 0.256 at 4 kHz, in AE: 0.272 at 250 Hz, 0.264 at 1 kHz, 0.333 at 4 kHz, $q = 1 - p$. A significance level of 0.01 was assumed.

Thus six values of k_0 were obtained. They were converted to threshold percent correct values by calculating k_0/N.

In order to make these results more comparable with those from Sect. 3 a simple linear interpolation procedure was performed for each of the six conditions

$$r = r_L + \frac{T-L}{H-L}, \qquad (3)$$

where r_L denotes the number of a frequency interval (see Fig. 3 and Table I), characterised by the percent correct result just below the one obtained above, r denotes the interpolated value of abscissa on r_L axis, T denotes the percent correct value corresponding to k_0 in (2), L denotes percent correct result in a given condition just below T, H denotes percent correct result just above T.

Finally, the r values were converted to frequencies according to

$$f = 2^{r+(2j-1)}, \qquad (4)$$

where j denotes the number of the standard frequency: No. 1: 250 Hz, No. 2: 1 kHz, No. 3: 4 kHz.

The thresholds are given in Table III.

5. Discussion and conclusions

The participants of the AM group attained significantly better frequency discrimination than the AE group. The narrowest interval of experiment 2 was the exception, as AE group performed better. This was confirmed by the paired data t-test, at $p < 0.05$. The reason for this anomaly is unknown.

In relative numbers, the average discrimination thresholds in experiment 1 were 8.3% of the standard frequency at 250 Hz, 4% at 1 kHz and 1.1% at 4 kHz. The medians pointed to even lower thresholds. These results indicate similar resolution as given in [3–5] for 250 Hz, but nearly twice better at 1 kHz and surprisingly better (nearly an order of magnitude) at 4 kHz. A relatively large group of subjects (34 together) and precision of contemporary equipment support our results. However, the results of experiment 2 do not. There seem to be an important difference between the methods of both experiments which is yet to be investigated, especially following the findings in [7].

Ten out of 15 listeners declared to hear distinct pitches at the interval of 8 Hz or less. These responses seem questionable, as beatings are the dominant percept at these intervals. These results were not removed however, in order to stay consistent with the rule of examining the feeling of distinct pitches in musicians. Despite that, the variance in the AM group was lower at all standard frequencies, indicating higher consistency in the performance of this group.

The results of experiment 2 are closer to earlier reports. Their evaluation depends on the percent correct value assumed as a threshold. According to both evaluation methods used, the discrepancy between the results of experiments 1 and 2 are large and requires further investigation.

The results for the narrowest interval where no discrimination occurs are lower than 50%. This was caused by the fact that when no discrimination occurs, the pitch of a two-tone is perceived as the average of the two frequencies [7]. Thus the listeners tended to point to M rather than to H or L time intervals, which produced this bias in the results.

The statistical analysis of the results of experiment 1 indicated that musicians can discriminate simultaneous pitches at narrower intervals than non-musicians. This has been confirmed in a very convincing way by the analysis of experiment 2.

The binomial distribution was demonstrated to yield considerably lower thresholds of perception. This can be interpreted as follows. The binomial distribution reveals the threshold above which the results were not obtained by chance, i.e. resulted from an audible effect. However, this indicates a just noticeable change of percept, which in this case is distant from the true perception of two distinct pitches. The psychometric function indicates the threshold at which listeners become to really perceive two pitches. On the other hand, the psychometric function is less reliable due to the limitations of the model and errors in a fitting procedure. This accounts for a much higher discrepancy in the results for the AE than for the AM group. The difference was bigger for the AE group, since their psychometric functions had considerably milder slopes, leading to lower consistency in the data. An important conclusion is that the psychometric function is suitable for evaluation of the results of well trained subjects, while the binomial distribution may be more suitable for untrained listeners.

References

[1] A.J.M. Houtsma, L.A. Goldstein, *J. Acoust. Soc. Am.* **51**, 520 (1972).

[2] A. Mielczarek, A. Dobrucki, *Proc. Forum Acust.*, Budapest 2005, p. 1629.

[3] W.R. Thurlow, S. Bernstein, *J. Acoust. Soc. Am.* **29**, 515 (1957).

[4] R. Plomp, *J. Acoust. Soc. Am.* **36**, 1628 (1964).

[5] R. Plomp, A.M. Mimpen, *J. Acoust. Soc. Am.* **43**, 764 (1968).

[6] R. Plomp, *Aspects of Tone Sensation: A Psychophysical Study*, Academic Press, London 1976.

[7] L.L. Feth, H. O'Malley, *J. Acoust. Soc. Am.* **62**, 940 (1977).

[8] K.H. Arehart, P.F. Rosengard, *J. Acoust. Soc. Am.* **105**, 2454 (1999).

[9] Siggen, `http://www.comp.leeds.ac.uk/jj/linux/siggen.html`, access: 18.10.11.

[10] A.R. Jennings, Ph.D. Thesis, University of Newcastle Upon Tyne, 2005.

[11] W.H. Press, S. Teukolsky, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed., Cambridge University Press, Cambridge 1992.

[12] F.A. Wichmann, N.J. Hill, *Perception Psychophys.* **63**, 1293 (2001).

[13] H. Dai, C. Micheyl, *J. Acoust. Soc. Am.* **130**, 263 (2011).