The Journal of the Acoustical Society of America

The effect of musicianship, contralateral noise, and ear of presentation on the detection of changes in temporal fine structure

--Manuscript Draft--

Manuscript Number:	JASA-03996R3
Full Title:	The effect of musicianship, contralateral noise, and ear of presentation on the detection of changes in temporal fine structure
Short Title:	Temporal fine structure and musicianship
Article Type:	Regular Article
Corresponding Author:	Brian Cecil Joseph Moore University of Cambridge Cambridge, Cambridgeshire UNITED KINGDOM
First Author:	Emelia Tarnowska
Order of Authors:	Emelia Tarnowska
	Andrzej Wicher
	Brian Cecil Joseph Moore
Section/Category:	Psychological and Physiological Acoustics
Keywords:	musical training; temporal fine structure; pitch; contralateral stimulation
Abstract:	Musicians are better than non-musicians at discriminating changes in the fundamental frequency (F0) of harmonic complex tones. Such discrimination may be based on place cues derived from lower resolved harmonics, envelope cues derived from high harmonics, and temporal fine-structure (TFS) cues derived from both low and high harmonics. The present study compared the ability of highly trained violinists and non-musicians to discriminate changes in complex sounds that differed primarily in their TFS. The task was to discriminate harmonic (H) and frequency-shifted inharmonic (I) tones that were bandpass filtered such that the components were largely or completely unresolved. The effect of contralateral noise and ear of presentation was also investigated. It was hypothesized that contralateral noise would activate the efferent system, helping to preserve the neural representation of envelope fluctuations in the H and I stimuli, thereby improving their discrimination. Violinists were significantly better than non-musicians at discriminating the H and I tones. However, contralateral noise and ear of presentation had no effect. It is concluded that, compared to non-musicians, violinists have a superior ability to discriminate complex sounds based on their TFS, and that this ability is unaffected by contralateral stimulation or ear of presentation.

The effect of musicianship, contralateral noise, and ear of presentation on the detection of changes in temporal fine structure

Emilia Tarnowska and Andrzej Wicher

Department of Psychoacoustics and Room Acoustics, Institute of Acoustics, Faculty of Physics, Adam Mickiewicz University, Poznań, Umultowska 85, 61-614 Poland

Brian C.J. Moore[a]

Department of Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

[a]Author to whom correspondence should be addressed. Electronic mail: bcjm@cam.ac.uk

PACS numbers: 43.66.fe, 43.66.Hg, 43.66.Pn

Abstract

Musicians are better than non-musicians at discriminating changes in the fundamental frequency (F0) of harmonic complex tones. Such discrimination may be based on place cues derived from low resolved harmonics, envelope cues derived from high harmonics, and temporal fine-structure (TFS) cues derived from both low and high harmonics. The present study compared the ability of highly trained violinists and non-musicians to discriminate changes in complex sounds that differed primarily in their TFS. The task was to discriminate harmonic (H) and frequency-shifted inharmonic (I) tones that were bandpass filtered such that the components were largely or completely unresolved. The effect of contralateral noise and ear of presentation was also investigated. It was hypothesized that contralateral noise would activate the efferent system, helping to preserve the neural representation of envelope fluctuations in the H and I stimuli, thereby improving their discrimination. Violinists were significantly better than non-musicians at discriminating the H and I tones. However, contralateral noise and ear of presentation had no effect. It is concluded that, compared to non-musicians, violinists have a superior ability to discriminate complex sounds based on their TFS, and that this ability is unaffected by contralateral stimulation or ear of presentation.

I. Introduction

The perception of pitch and discrimination of the fundamental frequency (F0) of complex tones may be based on several mechanisms. For tones containing low harmonics, the auditory system may extract information about the frequencies of individual resolved harmonics from place cues (the distribution of excitation along the cochlea) and/or temporal fine-structure (TFS) cues (phase locking) and the pitch may be derived from these frequency estimates (de Boer, 1956; Thurlow, 1963; Goldstein, 1973; Wightman, 1973). For complex tones containing only very high harmonics, the pitch may be extracted from the envelope repetition rate of the waveform on the basilar membrane resulting from the interference of several harmonics (Hoekstra and Ritsma, 1977; Moore and Rosen, 1979; Houtsma and Smurzynski, 1990). For tones with intermediate harmonics, the pitch may be extracted from the time intervals between peaks in the TFS close to adjacent envelope maxima (Schouten, 1940; Schouten *et al.*, 1962; Moore and Moore, 2003). Several studies have shown that musicians perform better than non-musicians in pitch-related tasks, including F0 discrimination (Kishon-Rabin *et al.*, 2001; Micheyl *et al.*, 2006). Furthermore, musicians perform better than non-musicians for complex tones containing both resolved harmonics and high unresolved harmonics (Bianchi *et al.*, 2016; Bianchi *et al.*, 2017). However, it is not clear whether musicians are better than non-musicians in using TFS cues for F0 discrimination. One goal of the present study was to compare the ability of musicians and non-musicians to discriminate complex tones based primarily on TFS cues. A second goal was to assess the effect of applying a noise stimulus to the ear opposite to that receiving the test tones, so as to activate the efferent system. A third goal was to assess possible effects associated with the ear of presentation of the stimuli.

There is considerable research showing that musically trained subjects perform better than non-musicians on a variety of auditory tasks, and especially pitch-related tasks. For example, compared to non-musicians, musicians have smaller thresholds for the frequency discrimination of pure tones (Kishon-Rabin *et al.*, 2001), the F0 discrimination of harmonic complex tones (Micheyl *et al.*, 2006), and the detection of mistuning of a single component in an otherwise harmonic tone (Zendel and Alain, 2009). Musicians are also better than non-musicians at "hearing out" partials in complex tones (Soderquist, 1970; Fine and Moore, 1993). However, it is unclear whether the superior performance of musicians in pitch-related tasks results from a better ability to use place information, envelope information, TFS information, or some combination of these three.

The TFS1 test (Moore and Sek, 2009) is intended to assess the ability to process TFS information in complex tones. In this test, subjects are required to discriminate harmonic complex tones (H) and frequency-shifted inharmonic tones (I), in which each component is shifted upwards by the same amount in Hertz (Δf). The H and I tones have the same envelope repetition rate (equal to the F0 of the H tones), but they differ in their TFS. The phases of the components are chosen randomly for every tone, which means that the envelope shape fluctuates randomly from one tone to the next, so that the envelope shape does not provide a cue for discriminating the H and I tones. Stimuli are made up of many components and are then passed through a fixed bandpass filter (with center frequency F_c) centered on the higher components, so as to make excitation-pattern cues minimal.

The rationale behind the TFS1 test is illustrated in Fig. 1, which shows waveforms of H and I tones at the output of a simulated auditory filter centered at 1000 Hz. The F0 of the H tones was 100 Hz and F_c was 1000 Hz. The waveforms and envelope shapes differ for the two H tones shown in the top panels, because the component phases were chosen randomly for each stimulus. The perceived pitch can be predicted assuming that: (1) most nerve spikes are synchronized to the largest peaks in the TFS at the output of the auditory filter, and these occur close to the envelope peaks, as illustrated by the vertical lines in Fig. 1 (Javel, 1980); (2) the pitch corresponds to the most prominent time intervals between nerve spikes (excluding the very short intervals corresponding to immediately adjacent peaks in the TFS); (3) these most prominent intervals correspond to the intervals between peaks in the TFS close to adjacent envelope peaks, as illustrated by the arrows in Fig. 1. For the two H tones (top), the most prominent time interval is 10 ms (1/F0). When the harmonics are shifted by 50 Hz (bottom left), the most prominent time interval is 9.5 ms, while when the shift is 25 Hz (bottom right) the most prominent interval is 9.75 ms. In all cases, the perceived pitch corresponds approximately to the reciprocal of the most prominent interval (Schouten *et al.*, 1962; Moore and Moore, 2003). Discrimination of the H and I tones when the components are unresolved is thought to depend on differences in the inter-spike intervals produced by the H and I tones.

One concern with the TFS1 test is that performance may be based on differences in spectrum of the H and I tones, which would be reflected by differences in their excitation patterns. To illustrate this, spectra were calculated for ten samples of the H and I tones (together with ten different samples of the TEN) and the spectra were averaged. The spectra were then converted to excitation patterns calculated using the method described by Moore *et al*. (1997). The “diffuse-field” presentation option was used, since the headphones used in our experiments have approximately a diffuse-field response. The averaging was done to smooth random irregularities in the excitation patterns produced by the TEN, which are often large compared with the differences between the excitation patterns for the H and I tones (Jackson and Moore, 2014). Figure 2 shows excitation patterns for H and I tones with F_c = 2000 Hz and Δf /F0 = 0.5 (the frequency shift leading to the greatest difference between the H and I tones). The value of F0 was F_c/10 (left) and F_c/20 (right). The lower panels show the differences in excitation level between the H and I tones. When F_c = 10F0, the maximum difference in excitation level was 1.4 dB, which might just be detectable (Buus and Florentine, 1995) (note however, that the value of Δf /F0 at threshold for such a condition is usually much less than 0.5, so the differences in excitation level at threshold would be much smaller). For F_c = 20F0, the maximum difference in excitation level was 0.5 dB, which would be below the threshold for detecting a change in level in a limited frequency region (Buus and Florentine, 1995). Furthermore, for F_c = 20F0 there is no clear pattern of ripples in the excitation-pattern differences. These analyses suggest that excitation-pattern cues might just be sufficient for discrimination of the H and I tones when F_c = 10F0, but they are very unlikely to be sufficient for F_c = 20F0. For a review of other evidence indicating that performance of the TFS1 test is not based on the use of excitation-pattern cues when F_C/F0 is above about 10, see Moore (2019).

Mishra *et al*. (2015) and Jain *et al.* (2016) used the TFS1 test to address the question of whether Indian (Carnatic) musicians are better than non-musicians at discriminating changes in the TFS of complex sounds. Mishra *et al*. (2015) reported that adult musicians performed better than adult non-musicians on the TFS1 task, suggesting a superior ability of the former to process TFS information. Similar results were obtained by Jain *et al.* (2016) in a comparison of musically trained and untrained children, aged 8-10 years. However, in these studies, the passband of the filter had a width equal to F0, and the passband was centered at 9F0. This meant that the lowest audible harmonic in the H tone was the seventh or the eighth. Harmonics seven and eight are often regarded as being on the boundary between clearly resolved and clearly unresolved (Plomp, 1964; Bernstein and Oxenham, 2003; Moore and Gockel, 2011). Hence, it is possible that in the studies of Mishra *et al*. (2015) and Jain *et al.* (2016) the superior ability of the musicians to discriminate the H and I tones did not reflect greater sensitivity to TFS but rather reflected a superior ability to resolve the components. It has been reported that musicians have sharper auditory filters than non-musicians for a center frequency of 4 kHz (Bidelman *et al.*, 2014), but a recent study failed to find any effect of musicianship on the sharpness of auditory filters centered at 4 kHz, as measured using three methods (Moore *et al.*, 2019), and other studies have found no effect of musicianship on the sharpness of the auditory filter for lower center frequencies (Fine and Moore, 1993; Oxenham *et al.*, 2003). Nevertheless, as noted earlier, musicians have been shown to be better than non-musicians in “hearing out” individual partials in complex tones (Soderquist, 1970; Fine and Moore, 1993).

In the present study, we compared the performance of musicians and non-musicians on the TFS1 task using a bandpass filter centered at 10F0, for which the lowest components might have been partly resolved, and using a bandpass filter centered at 20F0, for which the components would have been completely unresolved. If musicians perform better than non-musicians even when the filter passband is centered at 20F0, this would strongly support the idea that musicians have a superior ability to process TFS information.

A second aim of this study was to explore the effect of contralateral stimulation (CS) on performance for the TFS1 task. The discrimination and detection of auditory stimuli presented to one ear can be affected by presentation of a non-informative stimulus to the other ear (Guinan, 2006; Perrot and Collet, 2014; Guinan, 2018). This effect is thought to be mediated by activation of the medial olivocochlear (MOC) efferent system (Collet *et al.*, 1990; Guinan, 2006; 2018). CS can lead to the suppression of otoacoustic emissions (OAE) in the ear contralateral to the CS and can change the characteristics of psychophysical tuning curves (Vinay and Moore, 2008; Wicher, 2013; Wicher and Moore, 2014; Bidelman *et al.*, 2016). Perrot and Collet (2014) reviewed the possible functions of the efferent system, including protection against acoustic trauma (Maison and Liberman, 2000) and improved hearing in noise (Micheyl and Collet, 1996). They also reviewed studies comparing the effects of CS for musicians and non-musicians. The outcomes were mixed, but there was at least some evidence for greater activation of the MOC system by CS for musicians than for non-musicians, as was also found by Bidelman *et al*. (2017).

Recently, Carney (2018) has proposed that the main role of the efferent system is to preserve the neural representation of envelope fluctuations in different frequency regions, by regulating cochlear gain so as to avoid neural saturation effects. This is relevant to performance of the TFS1 task. Although the envelope repetition rate and shape do not provide a cue for discriminating the H and I tones, detection of the difference between the H and I tones depends upon the presence of distinct envelope peaks, as illustrated in Fig. 1. Hence, preservation of the representation of the envelope shape in the auditory nervous system is important. If Carney's (2018) hypothesis is correct, then stronger activation of the efferent system might be associated with better preservation of envelope fluctuations in the auditory system and, hence, better performance of the TFS1 task. Furthermore, this effect might be stronger for musicians than for non-musicians. On the other hand, CS usually has the effect of slightly reducing the sharpness of the auditory filters in the contralateral ear (Vinay and Moore, 2008; Wicher, 2013; Wicher and

Moore, 2014; Bidelman *et al*., 2016), and this might impair performance when the bandpass filter is centered at 10F0, if performance in that case depends on the partial resolution of components.

A third aim of this study was to assess the effect of ear of presentation. It is widely believed that speech stimuli are processed primarily in the left cerebral hemisphere (leading to a right-ear advantage) and non-speech stimuli, including musical sounds, are processed primarily in the right cerebral hemisphere (leading to a left-ear advantage) (Broadbent and Gregory, 1964; Kimura, 1964), although cerebral dominance for musical and speech sounds appears to depend on the specific task that is used (Brancucci *et al*., 2005; 2008). There is also evidence that the extent of cerebral asymmetry for musical sounds differs for musicians and for non-musicians (Schlaug *et al*., 1995). We therefore assessed whether performance on the TFS1 task was better for stimuli presented to the left ear than for stimuli presented to the right ear, and whether there was any difference between musicians and non-musicians in the degree of asymmetry.

In most previous studies of the effects of musicianship on performance in pitch-related tasks, the musicians played a variety of types of musical instruments or were singers. It seems plausible that pitch discrimination skills would be greater for musicians whose instruments require precise pitch judgments and fine motor control to achieve the correct note (e.g. violinists) than for musicians who play instruments with pre-set discrete pitches (e.g. pianists). To maximize the likelihood of finding differences between musicians and non-musicians, in the present study the former all played instruments requiring precise pitch judgments and motor control to achieve the correct note; all played the violin and/or viola.

In summary, the aims of this study were: (1) To compare the performance of musicians on the TFS1 task under conditions where the components were marginally resolved and where they were completely unresolved; (2) To assess the effect of CS on performance of the TFS1 task and to compare that effect for musicians and non-musicians; (3) To assess the effect of ear of presentation on performance of the TFS1 task and to compare that effect for musicians and non-musicians. The musician group was selected to have a high likelihood of superior pitch-related skills based on extensive experience playing the violin and/or viola.

I. MATERIAL AND METHODS

The TFS1 test was conducted using a bandpass filter centered at 10F0 (experiment 1) or 20F0 (experiment 2). Ten musicians (M) and ten non-musicians (NM) were tested in each experiment. Subjects in group M were the same for the two experiments. Two subjects in the group NM differed across experiments.

A. Selection of subjects

Subjects in group M were students of the Ignacy Paderewski Music Academy in Poznań, who played the violin and/or viola. Nine were female. They began formal musical education no later than the age of 8 years (on average 6.5 years, standard deviation, SD = 0.6 years), and continued education and/or work as professional musicians, playing on average 5 hours per day. The average duration of musical training was 15.4 years (SD = 1.1 years). Subjects in group NM did not play any instrument (7 subjects) or played as amateurs not more than 2 hours per week (1 piano and 2 guitar players, for both experiment 1 and experiment 2). Seven were female in both experiments. If they played, their musical education was not formal, it started not earlier than 16 years of age, and it lasted no longer than 3 years. The average age was 22 years (SD = 0.7 years) for group M, and 25 years (SD = 1.7 years) for group NM in both experiments 1 and 2.

Audiometric thresholds were measured using an Interacoustics (Middlefart, Germany) AC40 clinical audiometer with Telephonics (Huntington, NY) TDH 39P headphones, using the recommended method in Poland, which is the same as the method recommended by the British Society of Audiology (2011). All subjects were selected to have audiometric thresholds better than 20 dB HL over the frequency range 500 to 4000 Hz. Audiometric thresholds averaged over the range 125 to 8000 Hz were 9.1 dB HL (standard deviation, SD = 6.9 dB) for group M. Audiometric thresholds for group NM were 5.8 dB HL (SD = 5.9 dB) for experiment 1 and 6.1

dB HL (SD = 6.0 dB) for experiment 2. The audiometric thresholds did not differ significantly across the M and NM groups for either experiment. As a check that cochlear outer hair-cell function was normal, distortion product otoacoustic emissions (DPOAEs) were measured over the frequency range 1000 to 4000 Hz using an Interacoustics Titan system. The signal-to-noise ratio was greater than 6 dB for all subjects, indicating normal outer hair cell function (Robinette and Glattke, 2007). The Titan system was also used to measure tympanograms. All subjects had type A tympanograms, indicating normal middle-ear function. No subjects reported any history of auditory processing disorder or other disorders that might affect auditory processing (e.g. dyslexia). Subjects were paid for their participation.

B. The TFS1 test

The TFS1 test was conducted using the method described by Moore and Sek (2009) and the software described by Sek and Moore (2012). A two-interval, two-alternative forced-choice (2AFC) task was used. Subjects were required to discriminate an H tone with fundamental frequency F0 from a tone in which all the components were shifted upwards by Δf Hz, resulting in an I tone. Each interval contained four successive 200-ms tones (including 20-ms onset and offset ramps), separated by 100 ms. One interval contained four H tones, giving the pattern HHHH. The other interval contained alternating H and I tones, giving the pattern HIHI. The subjects were instructed to choose the interval in which they heard a fluctuation in pitch.

A two-down one-up adaptive procedure was used and visual feedback was given after each trial, via the computer screen. After two successive correct responses, the value of Δf was divided by a factor, k. After one incorrect response, the value of Δf was multiplied by k. Before the first turn point, k was set to 1.25^3. Between the first and second turn points, k was 1.25^2, and beyond the second turn point, k was equal to 1.25. An adaptive track ended after eight turn points. The threshold, corresponding to 71% correct responses, was calculated as the geometric mean of the values of Δf at the last six turn points.

The maximum value of Δf was set to 0.5F0 Hz; this corresponds to the value at which the

H and I tones are most different. If the limit was reached five times during a run, the adaptive procedure ended and the percentage correct was measured for forty further trials with Δf fixed at 0.5F0. We refer to this as the constant-stimulus procedure.

One modification to the test was made. In the "standard" version of the test, the component phases are chosen randomly for every tone. This can result in perceptual differences between tones with the same magnitude spectrum (e.g. two I tones with the same value of Δf), because of differences in envelope shape. Here, the component phases were chosen randomly and independently for the first and second tones in each interval, but the component phases were the same for the first and third tones and for the second and fourth tones. This was done so that, in the interval with alternating H and I tones, the two H tones would sound similar to one another and the two I tones would sound similar to one another, thus making the task slightly easier. This was considered desirable, since the task is very difficult when the bandpass filter is centered on very high components, as it was in experiment 2.

To reduce cues due to differences in the excitation patterns of the H and I tones, the stimuli were passed through a bandpass filter. This filter was centered at 10F0 for experiment 1 and 20F0 for experiment 2. The filter had a central flat region with a width equal to 3F0. The skirts of the filter fell off at a rate of 30 dB/octave. The filter minimized differences in the spectral envelopes and excitation patterns of the harmonic and inharmonic tones, as illustrated in Fig. 2. In experiment 1, the value of F0 was either 200 Hz or 400 Hz and the bandpass filter was centered at 2000 or 4000 Hz, respectively. In order to keep the frequency regions the same in experiment 2, the value of F0 was either 100 Hz or 200 Hz, so that the bandpass filter was again centered at 2000 or 4000 Hz, respectively. The center frequencies of 2000 and 4000 Hz were chosen to be within the range where phase locking occurs (Palmer and Russell, 1986). The overall level of the tones, after bandpass filtering, was set to 45 dB SPL. This level was chosen to be sufficiently low that the efferent system would be only weakly activated (Guinan, 2006).

A threshold equalizing noise (TEN) (Moore *et al.*, 2000) extending from 50 to 11,050 Hz was used to mask combination tones and to limit the audibility of components falling on the skirts of the bandpass filter. The TEN started 300 ms before the first tone burst and ended 300 ms after the last tone burst. The TEN level was specified as the level in a 1-ERB_N wide band centered at 1000 Hz, where ERB_N stands for the average value of equivalent rectangular bandwidth of the auditory filter at moderate sound levels for listeners with normal hearing (Glasberg and Moore, 1990). The level of the TEN was set 15 dB below the overall level of the complex tone. The TEN level was about 9 dB below the level of each component in the passband, and should have been sufficient to mask components falling on the filter skirts and combination tones whose level was 9 dB or more below the level of each component in the passband. In practice, this meant that components down to the 8th might have been just audible when the passband was centered at 10F0 and components down to the 16th might have been just audible when the passband was centered at 20F0.

The TFS1-test stimuli were presented monaurally. Both the left ear and right ear of each subject were tested. Thresholds were measured in the absence and in the presence of CS. The CS was a pink noise with a frequency range from 20 to 20000 Hz and an overall level of 60 dB SPL. A pink noise at this level significantly reduces the level of DPOAEs in the opposite ear, confirming that it is effective in activating the efferent system (Wicher, 2013; Wicher and Moore, 2014). This gave eight conditions (2 F0 values × 2 test ears × 2 conditions corresponding to the presence and absence of CS). Three threshold estimates were obtained for each condition and the final threshold was taken as the geometric mean of the three estimates.

The order of testing the conditions for experiment 1 was: F0 = 400 Hz without CS; F0 = 400 Hz with CS; F0 = 200 Hz without CS; F0 = 200 Hz with CS. The order for experiment 2 was: F0 = 200 Hz without CS; F0 = 200 Hz with CS; F0 = 100 Hz without CS; F0 = 100 Hz with CS. For each combination of F0 and presence/absence of CS, the order of testing the two ears was random.

Stimuli were generated using a Dell (Round Rock, TX) Inspiron 7000 series PC with a Conxant SmartAudio (Newport Beach, CA) sound card and presented via Sennheiser (Wedemark, Germany) HD600 headphones. The equipment was calibrated with an Ono Sokki

(Yokohama, Japan) FFT Analyzer type CF-5210, a Bruel & Kjær (Nærum, Denmark) type 4152 artificial ear, and an SVAN (Warsaw, Poland) 945A sound-level meter. All testing was conducted in sound-proof booths.

II. RESULTS

A. Experiment 1

In experiment 1, for which the filter passband was centered at 10F0, the adaptive procedure was completed by all subjects in both groups. The mean thresholds are shown in Fig. 3. Thresholds were expressed as the value of Δf at threshold, Δf_{thresh}, divided by F0, to facilitate comparison across the two F0s. The SD of the thresholds across repeated runs for a given condition was approximately proportional to the geometric mean threshold for that condition. Hence, statistical analyses were based on the logarithms of the thresholds, expressed as Δf_{thresh}/F0. The log thresholds were analyzed using a mixed-model analysis of variance (ANOVA). Within-subject factors were F0 (200 or 400 Hz), ear (left, L or right, R), and presence/absence of CS. The between-subjects factor was group (M or NM). The effect of group was significant [$F(1,18) = 7.22$, $p = 0.015$], group M having lower thresholds than group NM. The effect of F0 was significant [$F(1,18) = 9.27$, $p = 0.007$], the relative threshold being lower for F0 = 400 Hz than for F0 = 200 Hz. This is consistent with earlier work using the TFS1 test and similar tests (Moore *et al.*, 2006a; Moore and Sek, 2009; Jackson and Moore, 2014). There was no significant effect of test ear, and no significant effect of CS. There were no significant interactions.

B. Experiment 2

In experiment 2, for which the filter passband was centered at 20F0, the adaptive procedure often terminated and was switched automatically to the constant-stimulus procedure, because the value of Δf reached the limit of 0.5F0. This happened in 21% of the runs for group M and in 64% of the runs for group NM. For runs that switched to the constant-stimulus procedure, scores for group NM were often in the range that would be expected by chance guessing (Miller, 1996). The greater difficulty of the TFS1 task when the bandpass filter was centered on very high harmonics was expected from previous research (Moore *et al.*, 2006a; Moore and Sek, 2009; Jackson and Moore, 2014). The following procedure was adopted to transform the results obtained using the constant-stimulus procedure to make them comparable to the threshold values obtained using the adaptive procedure. Scores from the constant-stimulus procedure were converted to values of the detectability index, $d'_{obtained}$, using standard tables (Hacker and Ratcliff, 1979). The value of d′ calculated for 40 2AFC trials can reach 0.5 with a probability ≈ 0.05 when the subject is randomly guessing (Miller, 1996). To prevent excessively high estimates of "threshold" when performance was close to chance, values of $d'_{obtained} < 0.5$ were set to 0.5. Based on the assumption that d′ is proportional to Δf, the values of $d'_{obtained}$ were then converted to the value of Δf, $\Delta f_{extrapolated}$, that would be required to give a d′ value of 0.78, the value tracked by the adaptive procedure, using the following equation:

$$\Delta f_{extrapolated} = (0.78/d'_{obtained}) \times 0.5F0 \quad \text{(Eq. 1)}$$

It should be noted that this procedure often resulted in values of $\Delta f_{extrapolated}$ that were above 0.5F0 (with a maximum of 0.78F0). Such thresholds are not meaningful, since the largest difference between the H and I tones occurs when $\Delta f = 0.5F0$. However, it is the case that performance worsens monotonically with increasing $\Delta f_{extrapolated}$. The procedure was merely used to allow all thresholds to be transformed to the same scale.

Following the procedure defined by Eq. 1, so that all scores were expressed either as Δf_{thresh} or as $\Delta f_{extrapolated}$, the results were analyzed in the same way as for experiment 1. The mean thresholds are shown in Fig. 4. A mixed-model ANOVA was conducted with the same factors as for experiment 1. The effect of group was significant [$F(1,18) = 18.95$, $p < 0.001$], group M having lower thresholds than group NM. This indicates that musicians are better at processing TFS information than non-musicians. The effect of F0 was not significant. There was no significant effect of test ear, and no significant effect of CS. There was a significant interaction between CS and ear [$F(1, 18) = 8.19$, $p = 0.01$], and a significant interaction between

CS, ear, and F0 [$F(1, 17) = 16.52$, $p < 0.001$]. However, these interactions each accounted for 2% or less of the variance in the thresholds.

C. Comparison of results for experiments 1 and 2

As neither experiment showed a significant effect of ear of presentation or presence/absence of CS, the data were averaged across these factors to facilitate comparison of the results for the two experiments. Figure 5 shows geometric mean thresholds for each frequency region (F_c = 2000 or 4000 Hz), each degree of resolvability (bandpass filter centered at 10F0 or 20F0), and each group. Thresholds for both groups were lower by a factor of about 10 when the bandpass filter was centered at 10F0 than when it was centered at 20F0. On average, thresholds were higher for the non-musician group than for the musician group by a factor of about 1.5, regardless of whether the bandpass filter was centered at 10F0 or at 20F0. However, the factor in the latter case is an underestimate of the difference between the two groups because the value of $\Delta f_{\mathrm{extrapolated}}$ was limited to 0.78F0 much more often for the non-musicians than for the musicians. When the bandpass filter was centered at 20F0, the mean thresholds for the non-musicians were consistently above 0.5F0, indicating a very poor or no ability to perform the task, whereas the thresholds for the musicians were consistently below 0.5F0, indicating above-chance performance for most subjects. It is likely that the difference between musicians and non-musicians becomes very marked when TFS cues are very weak.

III. DISCUSSION

In experiment 1, the filter passband was centered at 10F0 and the passband width was 3 F0, so the 9th harmonic of the H tones fell at the lower edge of the passband. This is comparable to the conditions of Mishra *et al*. (2015) and Jain *et al.* (2016), who used a filter passband centered at 9F0 and a passband width of F0. The effect of musicianship reported by Mishra *et al*. for an F0 of 222 Hz was similar to that found by us for an F0 of 200 Hz. However, our musicians' thresholds overall were lower (better) than theirs, and also lower than the thresholds reported for comparable conditions by Moore and Sek (2009), although they were only slightly lower than those reported by Jackson and Moore (2014) for subjects with a moderate amount of musical training. The relatively low thresholds in our study might reflect the fact that we used a modified version of the TFS1 task in which the component phases were the same for the first and third tones and for the second and fourth tones in each interval. This had the effect of eliminating perceptual differences between the two H tones in the target interval and the two I tones in the target interval. In the "standard" version of the TFS1 test, such perceptual differences can be caused by differences in envelope shape between the two H tones and between the two I tones in the target interval, which might have a distracting effect. A possible disadvantage of our modified version of the test is that for the interval containing the HHHH sequence, the first and third tones had the same envelope shape and the second and fourth tones also had the same envelope shape, introducing an ABAB pattern that might have provided a false cue. However, the fact that performance was better with our modified version of the test than with the standard version suggests that the false cue had little or no deleterious effect.

A limitation of experiment 1, and of the studies of Mishra *et al*. (2015) and Jain *et al.* (2016), is that the lowest audible harmonics in the H tone were probably the 7th or 8th. These might have been partially resolved (Bernstein and Oxenham, 2003; Moore and Gockel, 2011). The advantage of musical training revealed in these cases might reflect a superior ability of musicians to hear out partials in complex tones (Soderquist, 1970; Fine and Moore, 1993) rather than a superior ability to process TFS cues.

In our experiment 2, the filter passband was centered at 20F0, which meant that the lowest audible components were completely unresolved. As expected from previous work, the task was much more difficult in this case (Moore *et al.*, 2006b; Moore *et al.*, 2009; Moore and Sek, 2009; Jackson and Moore, 2014). The adaptive procedure was switched automatically to the constant-stimulus procedure for 21% of the runs for group M and 64% of the runs for group NM. The method that we used for transforming the data from the runs using the constant-stimulus procedure limited the extrapolated threshold, $\Delta f_{\text{extrapolated}}$, to 0.78F0. This limit was applied more often for group NM than for group M. Despite this, a clear and significant advantage of musical training was observed. Mean thresholds, expressed as Δf_{thresh}/F0, were about 0.38F0 for group M and 0.56F0 for group NM. It is possible that performance when the filter passband was centered at 20F0 was based on the excitation pattern differences illustrated in Fig. 2. However, this possibility seems unlikely given the very small sizes of the differences and given that the background TEN would have produced substantial random ripples in the excitation patterns (Jackson and Moore, 2014). The most plausible interpretation of the results is that musically trained subjects are better at using TFS information than non-musicians.

One possible reason why performance of the TFS1 task worsens when the filter passband is centered on the higher harmonics can be illustrated using Fig. 1. That figure shows the output of a simulated auditory filter centered at 1000 Hz for H and I tones with a nominal F0 = 100 Hz. It is assumed that the H and I tones with Δf = 0.25F0 (bottom right) can be discriminated if the inter-peak interval of 10 ms for the H tone can be distinguished from the inter-peak interval of 9.75 ms for the I tone. This corresponds to a Weber fraction, $\Delta t/t$, of (10 – 9.75)/10 = 0.025. If the stimuli were bandpass filtered around 20F0, then for the same frequency shift of the I tone (Δf = 0.25F0), the most prominent inter-peak interval for the I tone would be 9.875 ms, and the Weber fraction would be 0.125/10 = 0.0125. If the Weber fraction at threshold corresponds to a fixed value, then performance would be expected to worsen progressively as the filter center frequency increases.

In fact, the worsening in performance with increasing filter center frequency was greater than would be predicted assuming that the Weber fraction for time-interval discrimination is constant. For example, for a filter centered at 2000 Hz and for group M, the threshold was about 0.047 for F0 = 200 Hz (corresponding to a Weber fraction of 0.00235/5 = 0.0047) while the threshold was about 0.4 for F0 = 100 Hz (corresponding to a Weber faction of 0.2/10 = 0.02). This may be explained by the increasing ambiguity of the time intervals to be discriminated as F_c increases for a fixed F0. For both the H and I tones, there are several candidate time intervals between peaks in the TFS close to adjacent envelope maxima, as illustrated in Fig. 1. The number of TFS peaks whose amplitude is within, say, 20% of the amplitude of the largest TFS peak increases with increasing F_c. When F_c is high relative to F0, it becomes increasing unclear what time intervals evoked by the H and I tones should be compared. For example, for F0 = 100 Hz, F_c = 2000 Hz, and $\Delta f/F0 = 0.4$, the most prominent candidate intervals would be 8.5, 9.0, 9.5, 10, 10.5, 11, and 11.5 ms for the H tone and 8.3, 8.8, 9.3, 9.8, 10.3, 10.8, 11.3, and 11.8 ms for the I tone. This is illustrated for an H tone in Fig. 6. Both the H and I tones would have a highly ambiguous pitch and this probably makes the task more difficult.

For a fixed ratio of F_c to F0, the Weber fraction at threshold may correspond approximately to a fixed value. This can explain why, when the bandpass filter was centered on the 20th harmonic, performance was not worse when the filter was centered at 4000 Hz than when it was centered at 2000 Hz, despite the fact that phase locking is likely to be weaker at 4000 than at 2000 Hz (Verschooten *et al.*, 2018). To illustrate this, assume that at 4000 Hz (with F0 = 200 Hz) the threshold, $\Delta f/F0$ is 0.4. The relevant intervals to be discriminated in this case would be 5 ms and 4.9 ms (the Weber fraction is 0.1/5 = 0.02). At 2000 Hz (with F0 = 100 Hz), the relevant intervals to be discriminated would be 10 ms and 9.8 ms (the Weber fraction is 0.2/10 = 0.02). According to this interpretation, performance is limited mainly by the central processes involved in interspike-interval discrimination, rather than by the precision of peripheral phase locking, at least for center frequencies up to 4000 Hz.

Our data for experiment 2 showed better performance than would be expected from previous work. For example, Jackson and Moore (2014) reported performance that was close to chance for a group of subjects with a moderate amount of musical training when F0 was 100 or 200 Hz and the lowest component within the passband was the 16th. The difference across studies may again reflect the fact that we used a modified version of the TFS1 test, with the same selection of component phases for the first and third tones and the second and fourth tones within each interval.

The reasons why musicians are better than non-musicians at processing TFS information remain unclear. The effect might reflect better neural encoding of TFS cues for musicians, greater proficiency of musicians in using the available neural cues, or a combination of the two. Supporting the concept of superior neural encoding, it has been reported that the synchronization of brainstem responses to pitch-evoking stimuli, as measured by the frequency-following response, FFR, is stronger for musicians than for non-musicians (Bidelman *et al.*, 2011). Also, thresholds for detecting changes in the frequency of a low-frequency (660-Hz) pure tone, which are thought to depend on the use of TFS information (Moore, 1973; Moore and Ernst, 2012), are correlated with a measure of the synchronization strength of the FFR (Marmel *et al.*, 2013). On the other hand, musicians perform better than non-musicians on a great variety of tasks, including tasks that are not related to pitch perception. For example, musicians show superior performance for gap detection (Zendel and Alain, 2012) and temporal-interval discrimination (Banai *et al.*, 2012). This is consistent with the idea that musicians have generally greater proficiency in making use of the available neural information, as well as having enhanced neural coding (Banai *et al.*, 2012), perhaps because of enhanced auditory attention (Strait *et al.*, 2010; Bianchi *et al.*, 2016). It is also possible that musicians were better than non-musicians at ignoring the false cue mentioned above, but, as stated earlier, the better performance with the modified version than with the standard version of the TFS1 test suggests that the negative influence of the false cue was very small.

The results of both experiments showed no effect of CS. We had suggested that CS would activate the efferent system, helping to preserve the neural representation of envelope fluctuations in the stimuli (Carney, 2018) and hence improving performance. The failure to find an effect of CS might have been related to the relatively low presentation level of our stimuli (45 dB SPL). Neural saturation is modest at such a level, occurring only for the most sensitive neurons (Liberman, 1978; Sachs and Young, 1979), so the envelope fluctuations in the TFS1-test stimuli were probably well preserved in the auditory nerve, even without activation of the efferent system. Another possibility is that the efferent system was sufficiently activated by the test stimuli themselves, so that any activation achieved by the (more intense) CS was not necessary for good performance to be achieved. However, this seems unlikely given the low level of the test stimuli.

The results for both experiments showed no overall effect of the ear of presentation of the test stimuli. Mishra *et al*. (2015) also reported no significant effect of ear of presentation. This may indicate that there is no ear dominance in the discrimination of pitch based on changes in TFS. Alternatively, ear dominance may exist, but it may only show up under conditions where there are competing stimuli at the two ears. Experiments demonstrating a right-ear advantage for speech have often been conducted using such competing stimuli (Broadbent and Gregory, 1964).

IV. SUMMARY AND CONCLUSIONS

The ability to discriminate harmonic from frequency-shifted tones was compared for highly trained musicians (violin and/or viola players) and non-musicians under conditions where the lowest components in the tones might have been partially resolved (experiment 1) and where all components were completely unresolved (experiment 2). The effects of CS and ear of presentation were also assessed. The task was a modified version of the TFS1 task, in which the component phases were chosen randomly and independently for the first and second of the four tones within each interval, but the component phases were the same for the first and third tones and for the second and fourth tones. This eliminated distracting effects of differences in timbre between the two H tones and the two I tones within each target interval that would otherwise have occurred.

The musicians performed better than the non-musicians in both experiments, confirming that musicians have a superior ability to use TFS information. There was no effect of ear of presentation, suggesting either no effect of laterality in the processing of TFS cues or that laterality is only revealed when there are competing stimuli at the two ears. There was also no effect of CS.

ACKNOWLEDGMENTS

We thank Dr hab. Maciej Karpiński, Dr hab. Krzysztof Sowiński, and Professor Andrzej Łapa for organizational help, support, and consultation, Prof. Aleksander Sęk for helpful comments, modifications to the TFS1 software, and help with statistics, Kuba Polakowski for help in data collection, all the subjects for their patience and involvement, and the Associate Editor Joshua Bernstein and three reviewers for helpful comments on an earlier version of this paper.

Banai, K., Fisher, S., and Ganot, R. (**2012**). "The effects of context and musical training on auditory temporal-interval discrimination," Hear. Res. **284**, 59-66.

Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number?," J. Acoust. Soc. Am. **113**, 3323-3334.

Bianchi, F., Hjortkjaer, J., Santurette, S., Zatorre, R. J., Siebner, H. R., and Dau, T. (**2017**). "Subcortical and cortical correlates of pitch discrimination: Evidence for two levels of neuroplasticity in musicians," Neuroimage **163**, 398-412.

Bianchi, F., Santurette, S., Wendt, D., and Dau, T. (**2016**). "Pitch discrimination in musicians and non-musicians: Effects of harmonic resolvability and processing effort," J. Assoc. Res. Otolaryngol. **17**, 69-79.

Bidelman, G. M., Krishnan, A., and Gandour, J. T. (**2011**). "Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch," Eur. J. Neurosci. **33**, 530-538.

Bidelman, G. M., Nelms, C., and Bhagat, S. P. (**2016**). "Musical experience sharpens human cochlear tuning," Hear. Res. **335**, 40-46.

Bidelman, G. M., Schneider, A. D., Heitzmann, V. R., and Bhagat, S. P. (**2017**). "Musicianship enhances ipsilateral and contralateral efferent gain control to the cochlea," Hear. Res. **344**, 275-283.

Bidelman, G. M., Schug, J. M., Jennings, S. G., and Bhagat, S. P. (**2014**). "Psychophysical auditory filter estimates reveal sharper cochlear tuning in musicians," J. Acoust. Soc. Am. **136**, EL33-39.

Brancucci, A., Babiloni, C., Rossini, P. M., and Romani, G. L. (**2005**). "Right hemisphere specialization for intensity discrimination of musical and speech sounds," Neuropsychologia **43**, 1916-1923.

Brancucci, A., D'Anselmo, A., Martello, F., and Tommasi, L. (**2008**). "Left hemisphere specialization for duration discrimination of musical and speech sounds," Neuropsychologia **46**, 2013-2019.

British Society of Audiology (**2011**). *Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking: Recommended procedure* (British Society of Audiology, Reading, UK).

Broadbent, D. E., and Gregory, M. (**1964**). "Accuracy of recognition for speech presented to the right and left ears," Q. J. Exp. Psychol. **16**, 359-360.

Buus, S., and Florentine, M. (**1995**). "Sensitivity to excitation-level differences within a fixed number of channels as a function of level and frequency," in *Advances in Hearing Research*, edited by G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, and H. Oekinghaus (World Scientific, Singapore), pp. 401-412.

Carney, L. H. (**2018**). "Supra-threshold hearing and fluctuation profiles: implications for sensorineural and hidden hearing loss," J. Assoc. Res. Otolaryngol. **19**, 331-352.

Collet, L., Kemp, D. T., Veuillet, E., Duclaux, R., Moulin, A., and Morgon, A. (**1990**). "Effect of contralateral auditory stimuli on active cochlear micro-mechanical properties in human subjects," Hear. Res. **43**, 251-261.

de Boer, E. (**1956**). "Pitch of inharmonic signals," Nature **178**, 535-536.

Fine, P. A., and Moore, B. C. J. (**1993**). "Frequency analysis and musical ability," Music Percept. **11**, 39-53.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103-138.

Goldstein, J. L. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496-1516.

Guinan, J. J., Jr. (**2006**). "Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans," Ear Hear. **27**, 589-607.

Guinan, J. J., Jr. (**2018**). "Olivocochlear efferents: Their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses," Hear. Res. **362**, 38-47.

Hacker, M. J., and Ratcliff, R. (**1979**). "A revised table of d' for M-alternative forced choice," Percept. Psychophys. **26**, 168-170.

Hoekstra, A., and Ritsma, R. J. (**1977**). "Perceptive hearing loss and frequency selectivity," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans, and J. P. Wilson (Academic, London, England), pp. 263-271.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304-310.

Jackson, H. M., and Moore, B. C. J. (**2014**). "The role of excitation-pattern and temporal-fine-structure cues in the discrimination of harmonic and frequency-shifted complex tones," J. Acoust. Soc. Am. **135**, 1356-1570.

Jain, S., Sahana, K., and Vinutha, M. M. (**2016**). "Effect of musical training on perception of temporal fine structure cues in children," in *Fontiers of Research on Speech and Music (FRSM-2016)*, Barapida, Orissa, India, 79-82.

Javel, E. (**1980**). "Coding of AM tones in the chinchilla auditory nerve: Implications for the pitch of complex tones," J. Acoust. Soc. Am. **68**, 133-146.

Kimura, D. (**1964**). "Left-right differences in the perception of melodies," Q. J. Exp. Psychol. **16**, 355-358.

Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (**2001**). "Pitch discrimination: are professional musicians better than non-musicians?," J. Basic Clin. Physiol. Pharmacol. **12**, 125-143.

Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**, 442-455.

Maison, S. F., and Liberman, M. C. (**2000**). "Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength," J. Neurosci. **20**, 4701-4707.

Marmel, F., Linley, D., Carlyon, R. P., Gockel, H. E., Hopkins, K., and Plack, C. J. (**2013**). "Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently," J. Assoc. Res. Otolaryngol. **14**, 757-766.

Micheyl, C., and Collet, L. (**1996**). "Involvement of the olivocochlear bundle in the detection of tones in noise," J. Acoust. Soc. Am. **99**, 1604-1610.

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (**2006**). "Influence of musical and psychoacoustical training on pitch discrimination," Hear. Res. **219**, 36-47.

Miller, J. (**1996**). "The sampling distribution of d'," Percept. Psychophys. **58**, 65-72.

Mishra, S. K., Panda, M. R., and Raj, S. (**2015**). "Influence of musical training on sensitivity to temporal fine structure," Int. J. Audiol. **54**, 220-226.

Moore, B. C. J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610-619.

Moore, B. C. J. (**2019**). "The roles of temporal envelope and fine structure information in auditory perception," Acoust. Sci. Tech. **40**, 61-83.

Moore, B. C. J., and Ernst, S. M. (**2012**). "Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code," J. Acoust. Soc. Am. **132**, 1542-1547.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc. **45**, 224-240.

Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (**2006a**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480-490.

Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (**2006b**). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure information," Hear. Res. **222**, 16-27.

Moore, B. C. J., and Gockel, H. (**2011**). "Resolvability of components in complex tones and implications for theories of pitch perception," Hear. Res. **276**, 88-97.

Moore, B. C. J., Hopkins, K., and Cuthbertson, S. J. (**2009**). "Discrimination of complex tones with unresolved components using temporal fine structure information," J. Acoust. Soc. Am. **125**, 3214-3222.

Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcántara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205-224.

Moore, B. C. J., and Rosen, S. M. (**1979**). "Tune recognition with reduced pitch and interval information," Q. J. Exp. Psychol. **31**, 229-240.

Moore, B. C. J., and Sek, A. (**2009**). "Development of a fast method for determining sensitivity to temporal fine structure," Int. J. Audiol. **48**, 161-171.

Moore, B. C. J., Wan, J., Varathanathan, A., Naddell, S., and Baer, T. (**2019**). "No effect of musical training on frequency selectivity estimated using three methods," Trends Hear. **23**, 1-9.

Moore, G. A., and Moore, B. C. J. (**2003**). "Perception of the low pitch of frequency-shifted complexes," J. Acoust. Soc. Am. **113**, 977-985.

Oxenham, A. J., Fligor, B. J., Mason, C. R., and Kidd, G., Jr. (**2003**). "Informational masking and musical training," J. Acoust. Soc. Am. **114**, 1543-1549.

Palmer, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells," Hear. Res. **24**, 1-15.

Perrot, X., and Collet, L. (**2014**). "Function and plasticity of the medial olivocochlear system in musicians: a review," Hear. Res. **308**, 27-40.

Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628-1636.

Robinette, M. S., and Glattke, T. J. (**2007**). *Otoacoustic Emissions: Clinical Applications* (Thieme, New York), pp. 436.

Sachs, M. B., and Young, E. D. (**1979**). "Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate," J. Acoust. Soc. Am. **66**, 470-479.

Schlaug, G., Jancke, L., Huang, Y., and Steinmetz, H. (**1995**). "In vivo evidence of structural brain asymmetry in musicians," Science **267**, 699-701.

Schouten, J. F. (**1940**). "The residue and the mechanism of hearing," Proc. Kon. Ned. Akad. Wetenschap. **43**, 991-999.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). "Pitch of the residue," J. Acoust. Soc. Am. **34**, 1418-1424.

Sek, A., and Moore, B. C. J. (**2012**). "Implementation of two tests for measuring sensitivity to temporal fine structure," Int. J. Audiol. **51**, 58-63.

Soderquist, D. R. (**1970**). "Frequency analysis and the critical band," Psychon. Sci. **21**, 117-119.

Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (**2010**). "Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance," Hear. Res. **261**, 22-29.

Thurlow, W. R. (**1963**). "Perception of low auditory pitch: a multicue mediation theory," Psychol. Rev. **70**, 461-470.

Verschooten, E., Desloovere, C., and Joris, P. X. (**2018**). "High-resolution frequency tuning but not temporal coding in the human cochlea," PLoS biology **16**, e2005164.

Vinay, and Moore, B. C. J. (**2008**). "Effects of activation of the efferent system on psychophysical tuning curves as a function of signal frequency," Hear. Res. **240**, 93-101.

Wicher, A. (**2013**). "The effect of contralateral signal on distortion product otoacoustic emissions and psychophysical tuning curves at 1 and 2 kHz," Acta Phys. Polon. A **123**, 1001-1006.

Wicher, A., and Moore, B. C. J. (**2014**). "Effect of broadband and narrowband contralateral noise on psychophysical tuning curves and otoacoustic emissions," J. Acoust. Soc. Am. **135**, 2931-2941.

Wightman, F. L. (**1973**). "The pattern-transformation model of pitch," J. Acoust. Soc. Am. **54**, 407-416.

Zendel, B. R., and Alain, C. (**2009**). "Concurrent sound segregation is enhanced in musicians," J. Cog. Neurosci. **21**, 1488-1498.

Zendel, B. R., and Alain, C. (**2012**). "Musicians experience less age-related decline in central auditory processing," Psychol. Aging **27**, 410-417.

Figure captions

FIG. 1. Segments of waveforms of harmonic (H) tones (top) and inharmonic (I) tones (bottom) at the output of a simulated auditory filter centered at 1000 Hz, for a nominal F0 of 100 Hz. The H and I tones have the same envelope repetition rate but differ in the time intervals between peaks in the TFS close to adjacent envelope maxima, as indicated by the arrows.

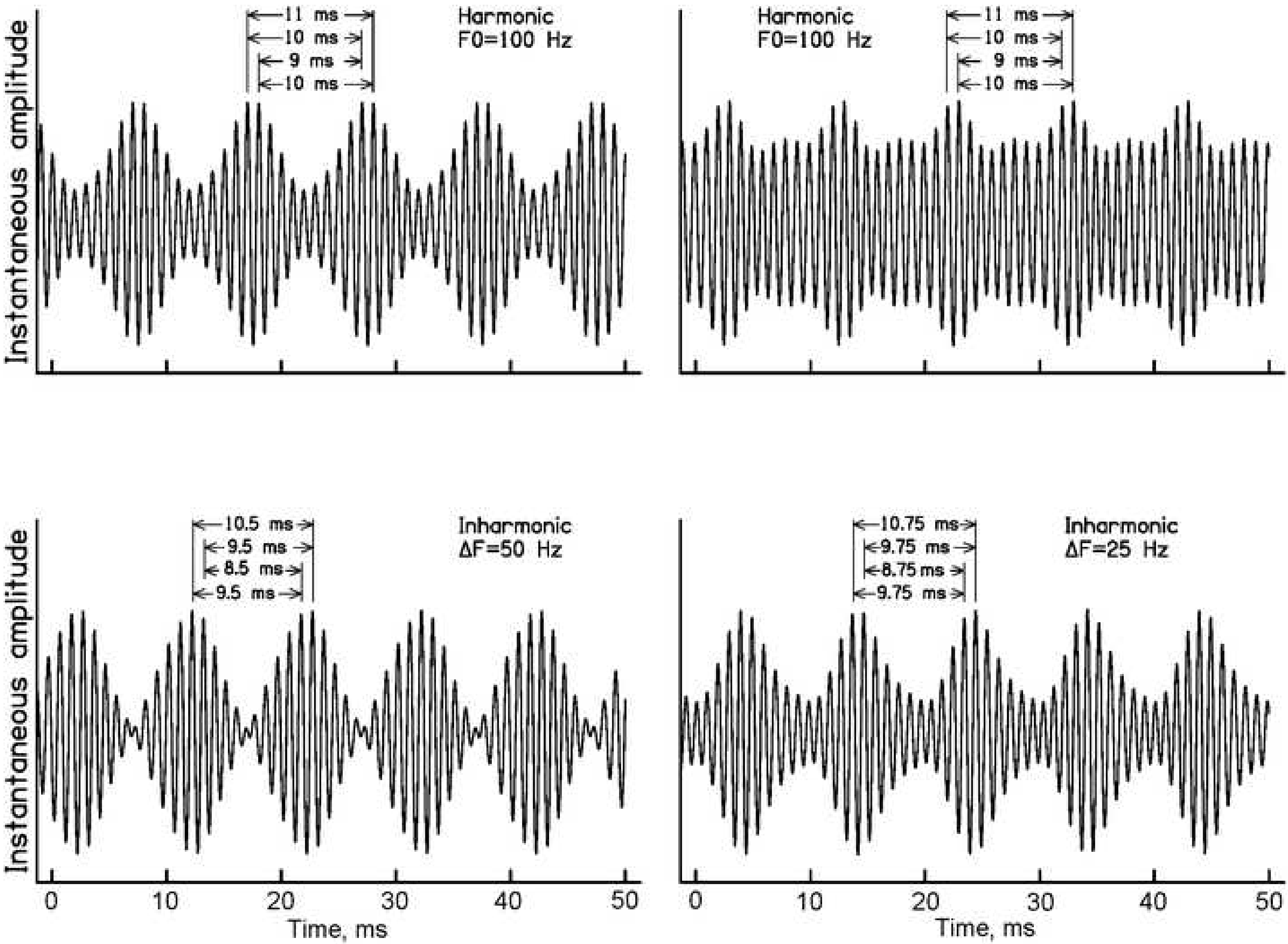
FIG. 2. Excitation patterns (top) and excitation pattern differences (bottom) for H tones (black lines) and I tones (gray lines) with Δf/F0 = 0.5. The bandpass filter was centered at 2000 Hz and F0 was 200 Hz (left) and 100 Hz (right).

FIG. 3. Geometric mean thresholds, expressed as Δf/F0, for experiment 1 for the two groups (M, shaded bars, and NM, open bars), the two ears of presentation of the test stimuli (L and R), the two F0s, and the two presentation modes (CS off and on). The bandpass filter was centered at 10F0. Error bars show ±1 standard error.

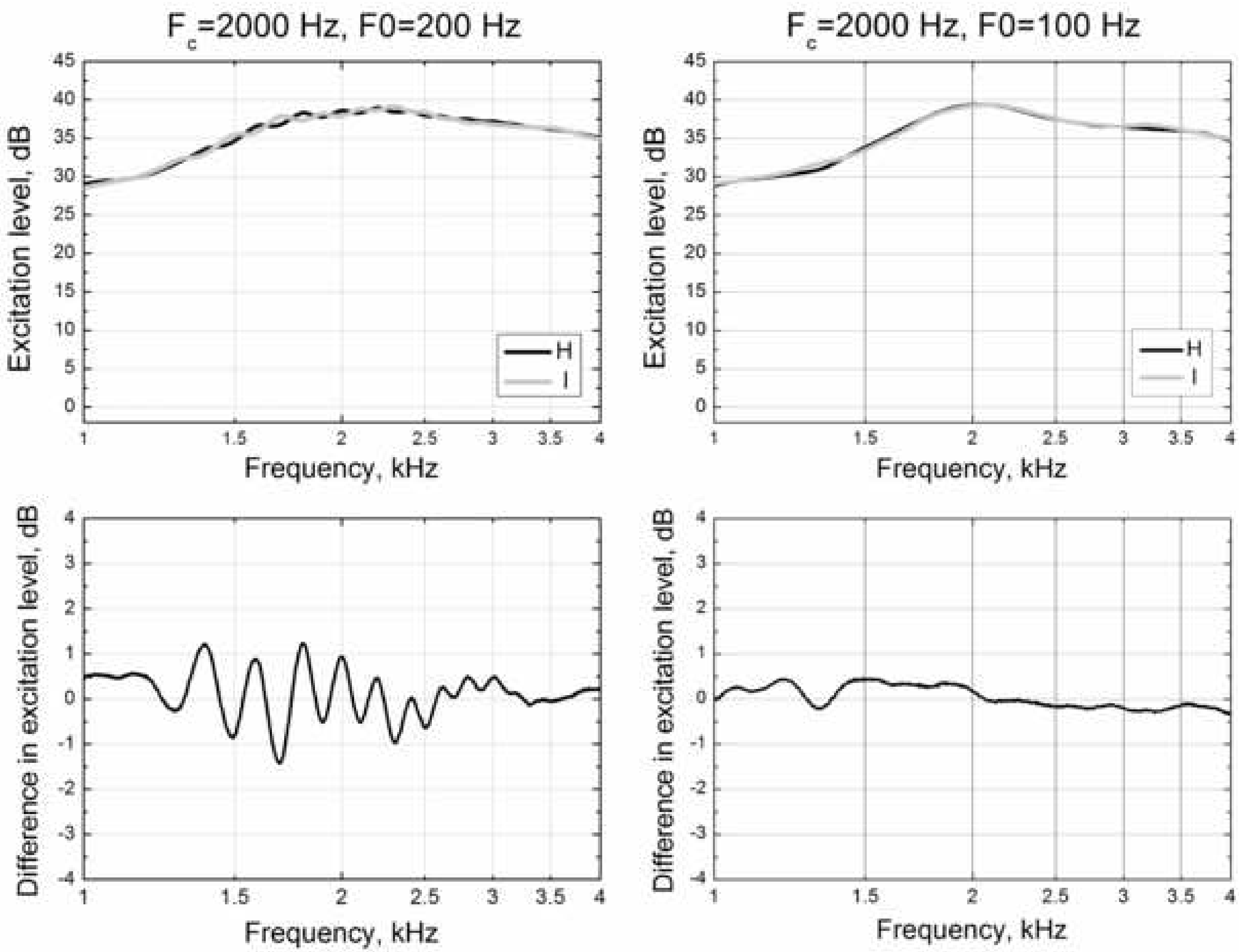
FIG. 4. As Fig. 3 but for experiment 2, for which the bandpass filter was centered at 20F0.

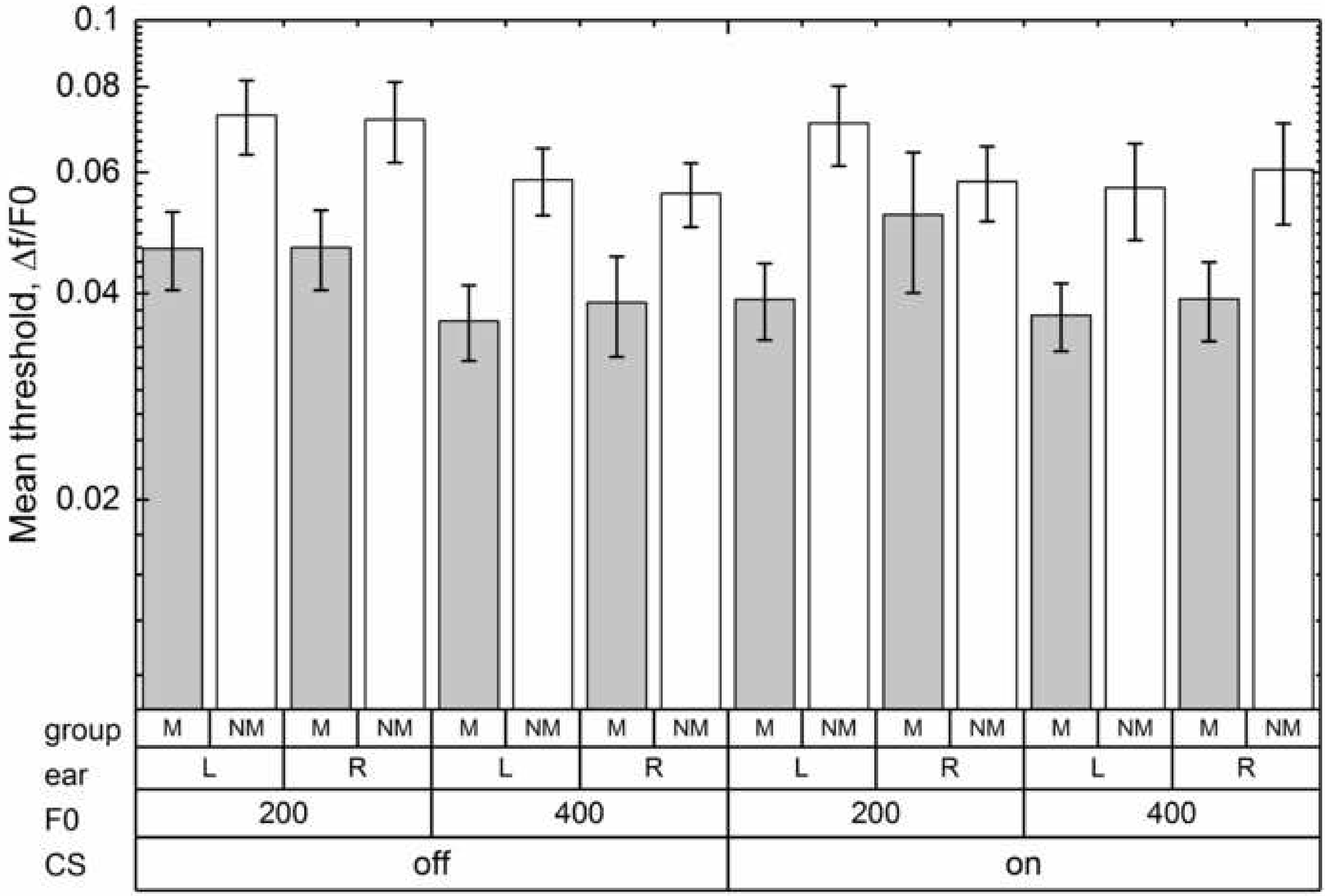
FIG. 5. Comparison of geometric mean thresholds for experiments 1 and 2, after averaging across ear of presentation and presence/absence of CS. The center frequency of the passband was 2000 or 4000 Hz and this corresponded to either 10F0 (experiment 1) or 20F0 (experiment 2).

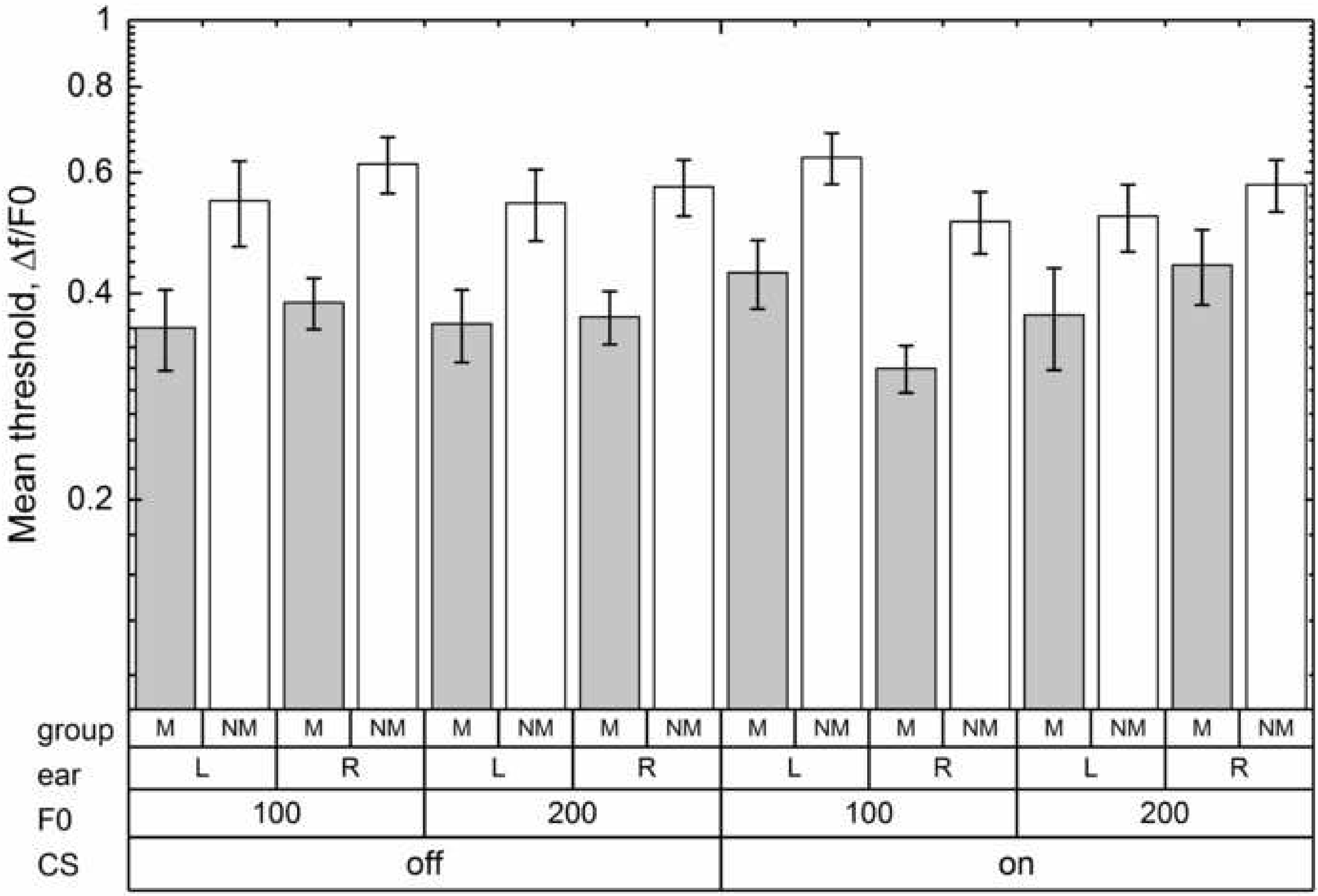
FIG. 6. Segment of the waveform of an H tone with F0 = 100 Hz and F_c = 2000 Hz at the output of a simulated auditory filter centered at 2000 Hz. The vertical lines indicate the positions of TFS peaks with amplitude within 20% of the amplitude of the largest TFS peak, and the numbers within arrows show the time intervals between those peaks, in ms.

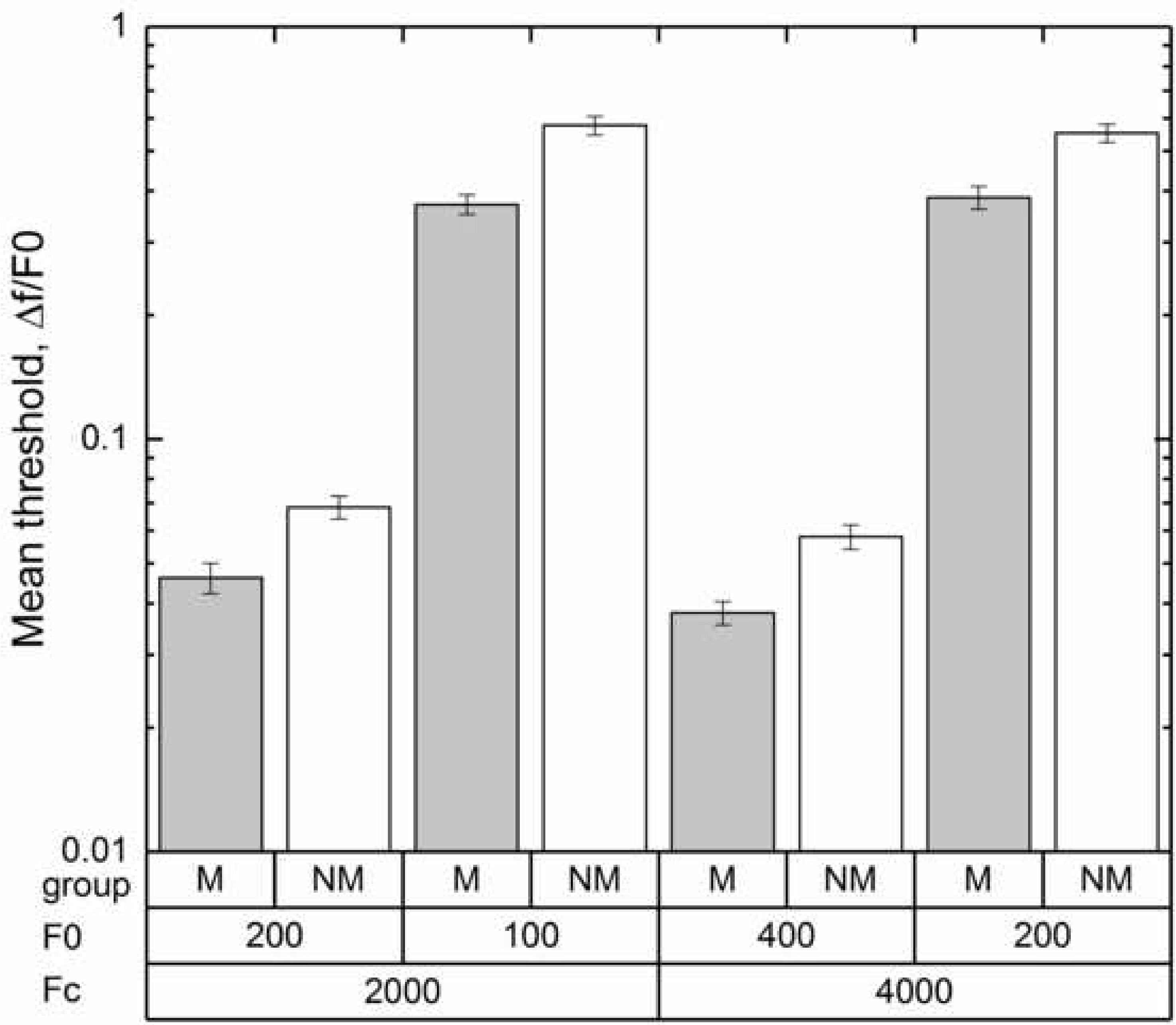


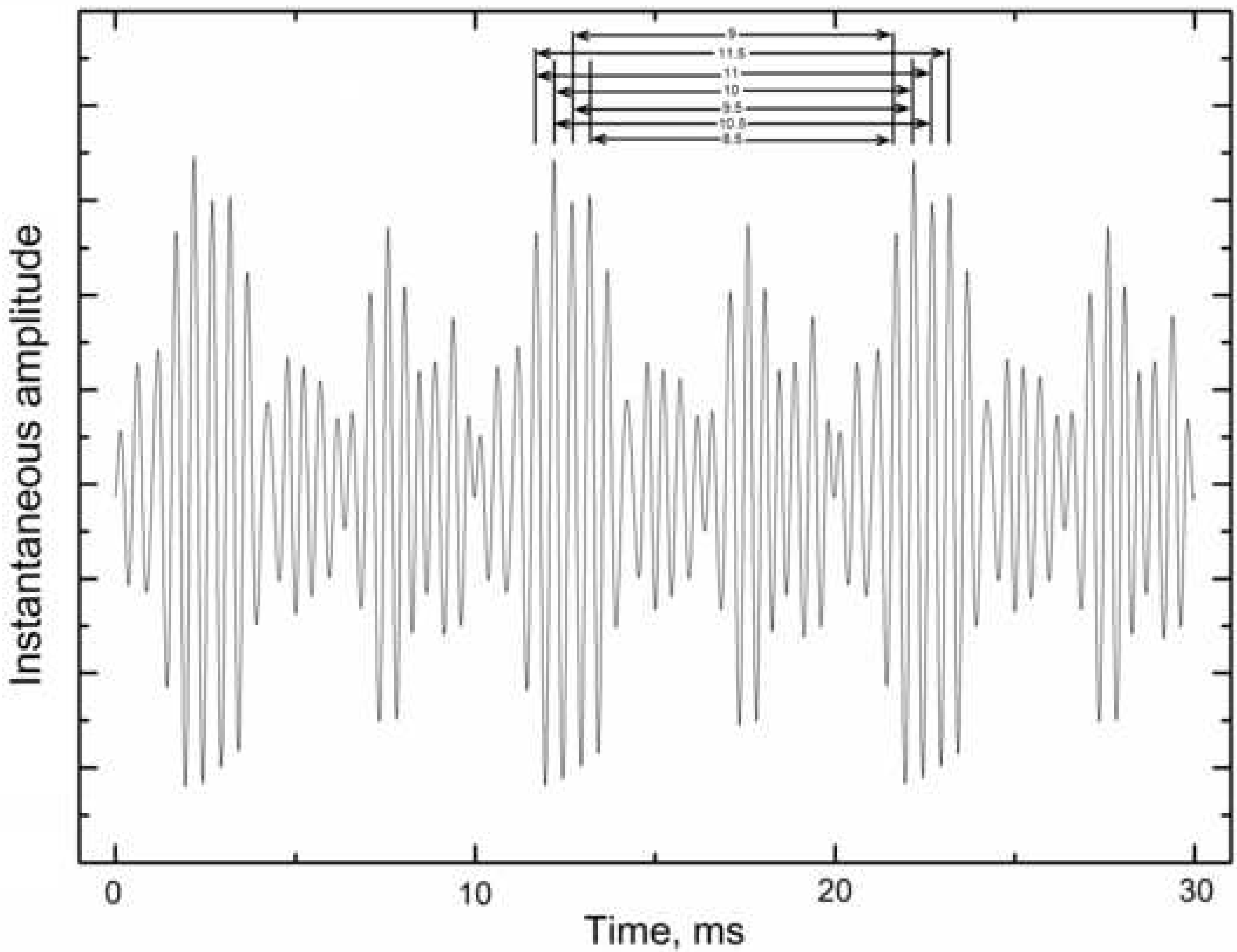
Instantaneous amplitude
11 ms
10 ms
9 ms
10 ms
Harmonic
F0=100 Hz
11 ms
10 ms
9 ms
10 ms
Harmonic
F0=100 Hz
10.5 ms
9.5 ms
8.5 ms
9.5 ms
Inharmonic
ΔF=50 Hz
10.75 ms
9.75 ms
8.75 ms
9.75 ms
Inharmonic
ΔF=25 Hz
0
10
20
30
40
50
Time, ms


F_c=2000 Hz, F0=200 Hz
F_c=2000 Hz, F0=100 Hz
Excitation level, dB
Frequency, kHz
H
I
Difference in excitation level, dB
Frequency, kHz


0.1
0.08
0.06
0.04
0.02
Mean threshold, Δf/F0
group
ear
F0
CS
M
NM
L
R
200
400
off
on


Mean threshold, Δf/F0
1
0.8
0.6
0.4
0.2
group
ear
F0
CS
M
NM
L
R
100
200
off
on


1
0.1
0.01
Mean threshold, Δf/F0
group
M
NM
M
NM
M
NM
M
NM
F0
200
100
400
200
Fc
2000
4000


9
11.5
11
10
9.5
10.8
8.5
Instantaneous amplitude
0
10
20
30
Time, ms