Published in final edited form as:

J Acoust Soc Am. 2009 February ; 125(2): 1059–1066. doi:10.1121/1.3056568.

Further examination of pitch discrimination interference between complex tones containing resolved harmonics

Hedwig E. Gockel^{a)} and Robert P. Carlyon

MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK

Christopher J. Plack

School of Psychological Sciences, University of Manchester, Manchester M13 9PL, UK

Abstract

Pitch discrimination interference (PDI) is an impairment in fundamental frequency (F0) discrimination between two sequentially presented complex (target) tones produced by another complex tone (the interferer) that is filtered into a remote spectral frequency region. Micheyl and Oxenham (2007) reported a modest PDI for target tones and interferers both containing resolved harmonics when the F0 difference between the two target tones (ΔF_0) was small. When the interferer was in a lower spectral region than the target, a much larger PDI was observed when ΔF_0 was large (14%–20%), and, under these conditions, performance in the presence of an interferer was *worse* than at smaller ΔF_0 s. The present study replicated the occurrence of PDI for complex tones containing resolved harmonics for small ΔF_0 s. In contrast to Micheyl and Oxenham's findings, performance in the presence of an interferer always increased monotonically with increasing ΔF_0 . However, when the interferer was in a lower spectral region than the target (and not vice versa), some subjects needed verbal instructions or modified stimuli to choose the correct cue, indicating an asymmetry in spontaneous obviousness of the correct listening cue across conditions.

I. INTRODUCTION

Many of the periodic sounds that we encounter in everyday life are broadband. Most current models of pitch perception assume that the derivation of pitch (which, for most harmonic complex tones, corresponds to the fundamental frequency, F0) is based on the combination of information from different frequency regions (Terhardt, 1974; Goldstein, 1973; Moore, 1982; Meddis and Hewitt, 1991; Meddis and O'Mard, 1997). This is consistent with empirical evidence showing that information can be integrated across different frequency regions (Moore *et al.*, 1984; Gockel *et al.*, 2007). However, it has also commonly been assumed that people can listen selectively to specific frequency regions in order to extract the F0s of multiple sources present at the same time. For example, the ability to identify two vowels presented simultaneously is better when the vowels have different F0s than when they have the same F0 (Scheffers, 1983). Although the improvements seen with very small F0 differences (ΔF_0 s) may be due to within-channel effects (Culling and Darwin, 1994), substantial benefits may be obtained at larger ΔF_0 s by comparing information from different parts of the spectrum. Models of this effect assume that F0 information can be extracted independently from different parts of the basilar membrane, and are based on the fact that the two vowels have formants at different frequencies, so that the spectrum is dominated at

^{a)}Corresponding author: Hedwig Gockel, Tel.: +44 1223 355294 ext 703. Fax: +44 1223 359062, Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk.

PACS: 43.66.Hg, 43.66.Fe, 43.66.Ba

different frequencies by the two different vowels (Assmann and Summerfield, 1990; 1994; Meddis and Hewitt, 1992; de Cheveigné, 1993).

Recently, Gockel *et al.* (2004; 2005a) showed that the ability of listeners to extract the pitch of a harmonic complex can be impaired in the presence of another harmonic complex (the interferer) even when the two tones occupy different frequency regions. This effect was called pitch discrimination interference (PDI) and its existence indicates that listeners have difficulties in independently extracting F0 information from different parts of the spectrum. In these studies, a two-interval two-alternative forced-choice task (2I-2AFC) was used to measure sensitivity (d') to the F0 difference between two target harmonic complexes in the presence of an interferer whose F0 was constant across the two intervals and that was filtered into a remote frequency region. The target contained harmonics that were unresolved by the peripheral auditory system, while the interferer contained resolved harmonics and produced a more salient pitch (see e.g., Shackleton and Carlyon, 1994). The degree to which F0 discrimination of the target was impaired depended on the similarity of the interferer's F0 and the nominal F0 of the target; the more similar the F0s the worse was performance. The studies also showed that PDI was larger when the interferer contained resolved harmonics than when it contained only unresolved harmonics, suggesting that the relative pitch salience of target and interferer was important. Furthermore, introducing a 400-ms onset and 200-ms offset asynchrony between target and interferer reduced but did not abolish PDI. Even a continuously presented interferer reduced sensitivity to the F0 changes of the target by a small but significant amount.

As mentioned above, in all of these experiments the target complex contained only unresolved harmonics and so its pitch was not very salient. The difference in F0 between the target complexes in the two intervals (ΔF_0), was chosen so as to give good but not ceiling performance levels, and so was relatively large (mostly between 3.5% and 7.1%). Micheyl and Oxenham (2007) tested whether PDI would occur if both the interferer *and* the target contained resolved harmonics. Each complex was bandpass filtered (8th-order Butterworth filter) with fixed 3-dB cutoff frequencies at either 125-625 Hz (LOW region) or 1375-1875 Hz (MID region). The nominal F0 of the target complex was 250 Hz. The interferer F0 was constant across the two intervals of a trial and had an F0 mid-way between the F0s of the target tones. The authors found that, for small ΔF_0 s (0.875%-1.75%), PDI was observed for targets containing resolved harmonics, irrespective of whether the target was filtered into the LOW and the interferer into the MID region, or vice versa. For larger values of ΔF_0 , no significant PDI was observed when the target was in the LOW and the interferer in the MID region. When the target was in the MID and the interferer was in the LOW region, significantly greater PDI was observed for *large* ΔF_0 s (14%-20%) than for small ΔF_0 s. In the absence of the interferer, performance increased monotonically with increasing ΔF_0 (being at ceiling at ΔF_0 s of 14-20%), but in the presence of the interferer, performance first increased and then, at ΔF_0 s of 14-20%, decreased markedly to nearly chance level. For an even larger ΔF_0 of 40%, performance recovered but was still somewhat below that observed without an interferer. This non-monotonic behaviour is contrary to our own data obtained in pilot experiments using large ΔF_0 values and comparable stimuli, and therefore seemed surprising.

Micheyl and Oxenham (2007) suggested two possible explanations for the non-monotonicity. One is that, in the presence of the LOW-region interferer, subjects listened analytically to the MID-region target tones, i.e., to the frequencies of specific harmonics, rather than listening synthetically to the F0 of the target complexes. The frequency ratio between consecutive audible harmonics of the target complex with a nominal F0 of 250 Hz filtered into the MID region (1375-1875 Hz) was about 1.17. This means that, for ΔF_0 s of 14% or 20% between the two target tones, the frequencies of harmonics n and $n+1$ in the

target complex with the higher F0 were close to the frequencies of harmonics $n+1$ and $n+2$ in the target complex with the lower F0. If subjects listened analytically to shifts in the frequency of an individual harmonic of the target complex across the two intervals, they likely would have compared the “wrong” (non-corresponding) harmonics in the two target complexes. This would lead to systematic errors, because for F0s larger than half the difference between consecutive harmonics, the frequencies of harmonics in the higher-F0 target were *lower* than the frequencies of those harmonics in the lower-F0 target complex that were closest to them in frequency. This could explain the poor performance at F0s of 14%-20% when the target and interferer were filtered into the MID and LOW regions, respectively [for details of the argument see Micheyl and Oxenham (2007)].

The other explanation suggested by Micheyl and Oxenham was that the pitch of the complex in the LOW region was more salient because it contained lower-numbered and more dominant harmonics and thus suppressed the perception of the pitch of the complex in the MID region. For F0s of 14-20%, where performance was worst, the F0 difference between the *interferer* and target would be 7-10%. Micheyl and Oxenham (2007) suggested that a “harmonic sieve” tuned to the F0 of the dominant interferer might be applied and that the harmonics of the targets in the MID region would be suppressed as they “fall between the teeth of the harmonic comb”. At smaller F0s the target harmonics fell less precisely between the “teeth”, and so this effect would be reduced.

The objectives of the present study were to investigate: (i) whether we would replicate the non-monotonic pattern of results observed by Micheyl and Oxenham (2007) under conditions slightly changed so as to encourage synthetic listening (Experiment 1) and, after failing to do so, under identical conditions (Experiment 2); (ii) which (if any) of their explanations was more likely to be correct, by increasing the overall number of subjects tested under identical conditions (Experiment 2), in order to preclude differences in analytic versus synthetic listening biases across listeners as explanation.

II. EXPERIMENT 1: FILTER REGION VARYING WITH F0

A. Rationale

It is well known that, for complex tones with only a few harmonics excluding the fundamental, many listeners perceive the pitches of individual part-tones rather than the residue pitch (Smoorenburg, 1970; Laguitton *et al.*, 1998; Patel and Balaban, 2001; Schneider *et al.*, 2005). We hypothesized that, in Micheyl and Oxenham’s (2007) study, the fixed filter cutoffs (1375 and 1875 Hz) and the large roving range for F0 (6 semitones) might have encouraged subjects to listen analytically to the target in the MID region, i.e., to the individual harmonics, rather than listening synthetically to the overall low pitch. Considering that the F0 was sometimes as high as 300 Hz, a fixed filter with a bandwidth of 500 Hz did not pass many harmonics and this might well have encouraged subjects to listen to individual harmonics moving in and out of the fixed filter passband when the F0 was changed. To test this hypothesis, we slightly modified the conditions to encourage synthetic listening and also to make them more similar to those used in previous studies on PDI (Gockel *et al.*, 2004; 2005a). There were three changes. First, the mean F0 was varied across trials over a range of $\pm 10\%$ rather than over the 6-semitone range (approximately 41%) used by Micheyl and Oxenham (2007). Second, the cutoff frequencies for the two filter regions were not fixed at absolute frequency values, as in Micheyl and Oxenham (2007), but varied together with, and by the same proportion as, the mean F0 across trials. Third, the spectrum level of a continuous background of pink noise was set to 15 dB (re: 20 μ Pa) at 1 kHz compared to the 12 dB value used by Micheyl and Oxenham (2007).

If Micheyl and Oxenham's (2007) first explanation was correct *and* the present modifications indeed encouraged synthetic listening, one would expect to observe a monotonic relationship between F_0 and performance in the presence of an interferer, irrespective of whether the target and interferer were in the MID and LOW regions, respectively, or vice versa. If this were the case, it would also rule out the "suppression" theory suggested by Micheyl and Oxenham (2007). If Micheyl and Oxenham's (2007) first explanation was correct *and* the present modifications were *unsuccessful* in encouraging synthetic listening, or alternatively if their second explanation was correct, then one would expect to observe a similar pattern of results to them.

B. Methods

1. Stimuli—In a 2I-2AFC task, subjects had to discriminate between the F_0 s of two sequentially presented target complex tones with a nominal F_0 of 250 Hz and a fixed difference, F_0 , between the F_0 s of the two tones within a block. In one, randomly chosen, interval the target complex had an F_0 equal to $F_0 - F_0/2$, while in the other interval its F_0 was $F_0 + F_0/2$. The values of F_0 were 0.225, 0.45, 0.9, 1.8, 3.6, 7, 14 and 20 % of the F_0 . The target was presented either alone, or with a synchronously gated harmonic complex (the interferer) with an F_0 that was identical in the two intervals of a trial, and which equalled the arithmetic mean of the F_0 s of the target in the two intervals. Thus, the interferer's F_0 was never identical to the F_0 of the simultaneously presented target complex. Each harmonic complex was bandpass filtered (slopes of 48 dB/octave) into one of two frequency regions. The nominal cutoff frequencies (3-dB down points) were 125 and 625 Hz for the LOW region and 1375 and 1875 Hz for the MID region. These frequency regions were chosen following Carlyon and Shackleton (1994) and, according to e.g. Shackleton and Carlyon (1994), the MID-region complex would contain at least some (the 6th and 7th) resolved harmonics (Plomp, 1964; Plomp and Mimpen, 1968; Moore and Ohgushi, 1993; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003). When an interferer was presented with the target, the two complex tones were always filtered into different frequency regions.

The level per component in the passband was always 45 dB SPL and components were added in sine phase. The nominal stimulus duration was 400 ms, including 5-ms raised-cosine onset and offset ramps. The silent interval between the two stimuli within a trial was 500 ms. Tones were generated and bandpass filtered digitally; bandpass filtering was achieved with a linear-phase FIR filter (order 16000) implemented in MATLAB with flat passband and linear slopes on a logarithmic frequency scale of 48 dB/oct. They were played out using a 16-bit digital-to-analog converter (CED 1401 plus), with a mean sampling rate of 40 kHz. The actual sampling rate was varied randomly over the range $\pm 10\%$ between trials (this also produced a slight variation in duration and in the filter cutoff frequencies). This was done to discourage subjects from using a long-term memory representation of the F_0 of the stimuli, and to encourage them to compare the F_0 of the target sounds across the two observation intervals within each trial. Stimuli were passed through an antialiasing filter (Kemo 21C30) with a cutoff frequency of 17.2 kHz (slope of 96 dB/oct), and presented monaurally to the subject's left ear, using Sennheiser HD250 headphones. A continuous background of pink noise with spectrum level of 15 dB (re: 20 μ Pa) at 1 kHz was presented in order to mask possible distortion products and to prevent subjects from relying on possible within-channel cues arising from the interaction of components at the outputs of auditory filters having center frequencies mid-way between the two spectral regions. Calculation of excitation patterns (following Moore *et al.*, 1997) for the pink noise and for the two complexes together showed that the excitation level of the pink noise at the output of an auditory filter mid-way between the two regions was about 20 dB above the excitation level of the primary components and thus would mask any interactions between the target

and interferer. Subjects were seated individually in an IAC double-walled sound-attenuating booth.

2. Procedure—Subjects had to indicate the interval in which the target sound had the higher F0. They were requested to focus attention on the target and to ignore the interferer as much as possible when it was present. Each interval was marked by a light, and visual feedback was provided after each trial. The value of F0 and the absence or presence of the interferer was fixed within a block of 105 trials. The first five trials were considered as “warm-up” trials and results from those were discarded. For a given condition, usually a block of “target alone” trials was run first, so that subjects were familiar with the timbre of the target. This was followed by two blocks with the interferer present, and another block with the target alone (following an ABBA design), before a new condition was tested. Conditions were run in a pseudo-random order (allowing for the ABBA design within a given value of F0). The frequency regions into which the target and the interferer were filtered were swapped between the first and second halves of a session. The total duration of a session was about 2 h, including rest times. For all subjects, except one (see below), at least 400 trials were run per condition. This corresponded to about 6 sessions.

3. Subjects—Five subjects participated, one of whom was the first author. They ranged in age from 19 to 44 years, and their absolute thresholds at octave frequencies between 250 and 8000 Hz were within 15 dB of the ISO (2004) standard. All of the subjects had participated in earlier studies on PDI, and all except one had some musical training. Four subjects participated in all conditions, while the fifth was not run for conditions with the smallest and two largest F0s (0.225, 14 and 20 %), due to time constraints. For this fifth subject, 100 trials were collected in the target alone conditions with F0s of 7, 3.6 and 1.8%, and in the target plus interferer conditions with F0s of 7 and 3.6% of the F0, for which his performance was at ceiling (100 % correct). In the remaining conditions at least 200 trials were collected.

C. Results and Discussion

Figure 1 shows the mean d' values and the corresponding standard errors averaged across subjects as a function of F0. Open and solid symbols show results in the absence and presence of an interferer, respectively. When the target was in the LOW region and the interferer in the MID region (Figure 1a, top), performance increased monotonically with increasing F0, until d' reached a very high asymptotic value¹, both in the absence and in the presence of the interferer. Some PDI was observed when performance was above chance, at F0 values of 0.45, 0.9, and 1.8 %. No PDI was observed at F0s of 3.6 % and larger. Planned contrasts (one-tailed paired-samples t-tests) indicated that the impairment in the presence of the interferer was significant at F0 values of 0.45 % [$t=3.98$, $p<0.01$], 0.9 % [$t=3.33$, $p<0.05$] and 1.8 % [$t=2.53$, $p<0.05$]. This finding is in good agreement with Micheyl and Oxenham's (2007) data. When the target was in the LOW and the interferer in the MID region, they too reported no significant PDI for F0 values of 3.5 % and larger.

When the frequency regions of target and interferer were exchanged, so that the target was in the MID region and the interferer in the LOW region (Figure 1b, bottom), the same pattern of results was observed. Performance increased monotonically with increasing F0, both in the absence and in the presence of the interferer. Again, significant PDI was observed when performance was above chance, at F0 values of 0.45 % [$t=4.63$, $p<0.01$], 0.9 % [$t=5.05$, $p<0.01$] and 1.8 % [$t=3.16$, $p<0.05$]. No significant PDI was observed for

¹In occasional cases where performance was correct in 100% of the trials, and thus the d' value could not be determined, the performance level was adjusted downwards to a value of 99.75%. This resulted in a d' value of 3.97.

F0s of 3.6 % and more. For small values of F0, the present findings agree well with those of Micheyl and Oxenham (2007), who reported significant PDI for values of F0 smaller than 3.5% as long as performance in the absence of an interferer was clearly above chance (so that floor effects were avoided). For larger values of F0, i.e., 7 % and more, the present results contrast markedly with those of Micheyl and Oxenham (2007). They reported increased PDI for those F0 values, and near-chance performance for F0s of 14 and 20 % in the presence of the LOW-region interferer.

To summarize, for F0s smaller than 7 %, the present findings replicate those of Micheyl and Oxenham (2007); when both target and interferer had resolved components, significant PDI was observed when F0 was 1.8 % or less (and performance for the target alone was above chance level), independent of whether the target was in the LOW and the interferer in the MID region, or vice versa. In contrast, for F0s larger than 3.5 %, the results of the two studies differed markedly. Here, no significant PDI was observed, independent of the frequency regions occupied by target and interferer, while Micheyl and Oxenham (2007) reported the largest amount of PDI for F0s of 14 and 20 % and a non-monotonic relationship between F0 and the sensitivity to this difference in F0 when the target and the interferer were filtered into the MID and LOW regions, respectively.

It is possible that, in the present study, some PDI would have been observed for F0s larger than 7% if performance in the absence of the interferer had been below ceiling. However, ceiling effects seem to be an unlikely explanation for the difference in results between the two studies given (i) the size of the PDI effects observed by Micheyl and Oxenham when the target was in the MID region and (ii) the fact that performance for the target alone, for F0s larger than 3.5% was close to ceiling in their study as well.

Note that brief testing using two subjects showed that, for F0=14 %, the level of the interferer in the LOW region could be increased by up to 25 dB SPL without much effect on performance for F0 discrimination of the target in the MID region. Thus, slight differences in the transfer functions of the headphones used would not explain the different findings of the two studies.

The current findings rule out the “suppression” theory suggested by Micheyl and Oxenham (2007) and seem to favour their first explanation (both described in the Introduction) for the observed non-monotonic relationship between F0 and performance in their data, in terms of analytical listening. As described in the Rationale above, the present experiment used stimuli that were slightly modified from Micheyl and Oxenham’s, in order to enhance synthetic listening. It is conceivable that these modifications led to the observed monotonic relationship between F0 and the sensitivity to this difference in F0 even when the target and the interferer were filtered into the MID and LOW regions, respectively. Alternatively, the difference across studies might have resulted from the choice of subjects. The second experiment tested these two alternatives, by more closely replicating the conditions of Micheyl and Oxenham (2007) with four new and two of the current subjects.

III. EXPERIMENT 2: FILTER REGION FIXED

A. Rationale

The results of the first experiment replicated Micheyl and Oxenham’s finding of a moderate PDI for small F0s, but here performance increased monotonically with increasing F0 in the presence of an interferer, irrespective of the frequency region. The objective of the second experiment was to investigate whether stimuli identical to the ones used by Micheyl and Oxenham (2007) would induce analytical listening, and yield similar results to theirs.

Listeners vary naturally in their ability to hear the residue pitch of complex tones containing only a few harmonics, when the fundamental component is absent. Although unlikely, the following scenario can be imagined: by chance, none of the five listeners who took part in Micheyl and Oxenham's Experiment 1, where the target and the interferer were filtered into the MID and the LOW region, respectively, listened synthetically when the interferer was present. Additionally, by chance, all of our five subjects happened to be synthetic listeners. This scenario could lead to the observed differences in the results between the two studies. To decrease the chance of this alternative explanation being applicable, four new subjects participated in experiment two, in addition to two subjects from the first experiment. If differences between subjects' analytical listening tendency were the reason for the different findings across the two studies, having four new subjects would increase the chance of observing a non-monotonic pattern of results for at least some of the subjects.

B. Methods

The stimuli and procedure were the same as for Experiment 1 with the following exceptions: (1) The mean F0 was varied across trials over a range of ± 3 semitones (approximately 41% total range); (2) The cutoff frequencies for the two filter regions were fixed at absolute frequency values of 125-625 Hz for the LOW and 1375-1875 Hz for the MID region; (3) The spectrum level of the continuous background of pink noise was set to 12 dB (re: 20 μ Pa) at 1 kHz. For a given condition, a block of "target alone" trials was always run first, followed by a block of the same condition with the interferer present. These conditions replicate those of Micheyl and Oxenham (2007).

Only the larger values of F0 were tested, for which the results differed across the two studies. When the target was in the MID region, the values of F0 were 3.6, 7, 14 and 20 % of the F0 for all subjects, plus 1.8 % for one subject (see below). When the target was in the LOW region, values of F0 were 3.6, 7, and 14 % of the F0 for all subjects.

The stimuli were played with a fixed sampling rate of 20 kHz and passed through an antialiasing filter (Kemo 21C30) with a cutoff frequency of 8.6 kHz (slope of 96 dB/oct). The roving of the mean F0 was achieved by generating the appropriate stimuli digitally rather than by randomizing the sampling rate. They were bandpass filtered using pairs of cascaded hardware filters (one high pass, one low pass, Kemo 21C30 series, each with slopes of 48 dB/oct).

Six subjects participated, one of whom was the first author. They ranged in age from 25 to 47 years, and their absolute thresholds at octave frequencies between 250 and 8000 Hz were within 15 dB of the ISO (2004) standard. Two of the subjects had also participated in Experiment 1 of this study, and all except two had some musical training. Stimuli were delivered to the left ear for five subjects and to the right ear for the sixth subject.

All subjects received some practice, until performance remained stable across blocks of trials. The amount of practice needed differed across subjects (see below) and was between four and 10 h for the four subjects who were new to the task. In the experiment proper, at least 400 trials were run for each condition and subject.

C. Results and Discussion

The solid lines in figure 2 show the mean d' values and the corresponding standard errors across subjects as a function of F0. Open and solid symbols show results in the absence and presence of an interferer, respectively. When the target was in the LOW region and the interferer in the MID region (Figure 2a, top), performance was at or close to ceiling for all subjects and conditions. A repeated measures two-way analysis of variance (ANOVA) was carried out with factors presence of interferer and F0, using the mean d' value for each

subject and condition as input. The ANOVA revealed no significant main effect and no significant interaction. This result was expected and replicates the finding of Experiment 1 in the present study and of Experiment 2 in Micheyl and Oxenham (2007).

When the target was in the MID region and the interferer in the LOW region (Figure 2b, bottom), one of the new subjects performed generally worse than the other five subjects. Her data are plotted using dashed lines, separately from the mean data for the other subjects. While the performance of the other five subjects was at or close to ceiling for all conditions (as in Experiment 1), the sixth subject generally had problems with the MID-region target, in spite of extensive training over 10 h. It is important to note that (i) this subject had some musical training, and (ii) performance in the presence of the interferer for F_0 of 14 and 20 % was not worse than for the smaller values of F_0 . Thus, in spite of her below-ceiling performance, the pattern of results was monotonic. The results of a repeated measures two-way ANOVA, performed on the data for the target in the MID region, showed no significant main effect and no significant interaction, independent of whether data from the odd performing sixth subject were included in the analysis or not.

Overall, the pattern of results for Experiment 2 was very similar to that for Experiment 1, and again contrasts with Micheyl and Oxenham's (2007) findings for large F_0 s when the target and the interferer were in the MID and LOW regions, respectively. This finding indicates that the differences between the stimuli used in Experiment 1, on the one hand, and in Experiment 2 and by Micheyl and Oxenham on the other hand, were unlikely to be the cause of the difference between the present results and those of Micheyl and Oxenham. In other words, it seems unlikely that the exact stimulus conditions chosen by Micheyl and Oxenham strongly encouraged analytical listening, and thus led to the breakdown in performance for F_0 s of 14 and 20 % when the target and the interferer were in the MID and LOW regions, respectively.

It is also very unlikely that, by chance, all nine subjects in the present study were synthetic listeners *and* in Micheyl and Oxenham's study all five subjects were analytic listeners. It is not possible to calculate exactly the probability for running one experiment with five subjects to find that by chance all five of them are analytical listeners *and* then run another experiment with nine subjects to find that by chance all nine of them are synthetic listeners, because the exact distribution of analytic versus synthetic listeners in the general population for the present stimuli is not known. However, to estimate the upper boundary for the probability of the occurrence of these two combined events, we assume that the probability that a randomly drawn subject will listen to the present (and Micheyl and Oxenham's) stimuli in an analytical way is 0.357 ($=5/14$), i.e., we assume that, for the present stimuli, this is the proportion of analytical listeners in the population, as this value will maximize the probability of the combined event occurring. With this assumed value, the upper boundary for the probability of the combined event is calculated as $0.000109 [= 0.357^5 * (1-0.357)^9]$. In other words, the probability that chance differences between the subjects of the two studies are responsible for the observed differences in the results is less than 0.0002. Thus, the current findings seem to rule out Micheyl and Oxenham's first explanation for their non-monotonic pattern of results in terms of analytical listening as well as their second explanation, the "suppression theory".

Some insight into a possible reason for their non-monotonic pattern of results comes from the following observations during the initial training phase of the experiment. Of the four subjects who were new to this task, three (including the "different" worse-performing subject) needed some specific clarification about the cue they were supposed to listen to when the target was in the MID region and the interferer was in the LOW region. In the present study, listeners' training always started with large values of F_0 of 20 and 14 % of

the F0 for both the target alone and target-plus-interferer conditions. When the target was in the LOW region, no subject had difficulties in picking up the correct cue, i.e. after two or three blocks of trials, subjects' performance was at or close to ceiling for those large values of F0, even in the presence of the interferer. However, when the target was in the MID region and the interferer in the LOW region, two subjects showed performance close to chance in the presence of the interferer after two blocks, in spite of reaching ceiling in the target-alone condition. For one subject (one without formal musical training), it was sufficient to stress and explain again that the target sound to which attention should be directed was the one with the harsher timbre, that sounded less full and smooth than the other sound. This subject's performance was at or close to ceiling for the large values of F0 after five blocks (500 trials), even in the presence of the interferer. A second subject replied to the verbal descriptions with: "...but there is no second tone". For this subject, the stimulus duration was increased to 1 s, and at the start of a short practice block of 11 trials, the level of the target sound was increased above that of the interferer by about 30 dB. Over the course of this practice block, the level of the target in the presence of the interferer was reduced by 3 dB after each trial until it reached its "normal" level, which was equal to that of the interferer. This led to an "Aha effect" of recognition on the part of the subject. Two more of these short practice blocks were run with the normal 400-ms duration of the stimuli and slowly reducing level, before normal practice continued. After these three short special practice / cueing blocks, the subject performed at ceiling for large F0s. For a third subject (the "different" worse performer) the verbal description and special cueing did not have much of an impact; she consistently performed more poorly than the other subjects when the target was in the MID region, in both the presence and the absence of the interferer. Apart from her, generally, once a subject knew "what to listen for", when questioned they always reported hearing two tones, and this percept seemed to be irreversible.

These observations indicate that the "correct" listening cue was not always picked up spontaneously. Specifically, when the target and the interferer were in the LOW and MID regions, respectively, all subjects spontaneously listened to the correct cue. In contrast, when the target and the interferer were in the MID and LOW regions, respectively, the correct cue was spontaneously obvious for only some of the subjects, while others needed instructions or special cueing. This indicates an asymmetry in the salience of the correct cue, which might be related to the fact that the dominance region for a complex with F0 of 250 Hz is closer to the LOW than to the MID region (Plomp, 1967; Ritsma, 1967; Moore *et al.*, 1985; Gockel *et al.*, 2005b) i.e., the LOW region complex had a somewhat more salient pitch than the MID region complex. Alternatively, the asymmetry could be related to the fact that sounds in music (and every day life) usually have more energy in the lower harmonics than in the higher harmonics. This could have biased subjects to pay more attention to the lower-frequency range than to the higher-frequency range of the present stimuli.

Either way, the present findings indicate that the non-monotonic pattern of results observed by Micheyl and Oxenham (2007) was not due to a hard-wired suppression of the perception of the pitch of the complex in the MID region due to the presence of the complex in the LOW region. The present findings also show that it is unlikely that the difference between the results of Micheyl and Oxenham and the present study was due to analytical listening, caused either by their specific stimuli or by individual differences between subjects across studies. Although we did not observe non-monotonic performance for any subject or condition, we believe that a possible explanation for the results reported by Micheyl and Oxenham may lie in our finding that, when the target and interferer were in the MID and LOW regions, respectively, some subjects (two out of nine) needed special cueing during training in order to attract their attention to the correct cue. Micheyl and Oxenham did not report using any such cueing.

IV. SUMMARY AND CONCLUSIONS

F0 discrimination between two sequentially presented 400-ms harmonic target complexes was measured. Their nominal F0 was 250 Hz. The target tones were presented either alone or simultaneously with another harmonic complex, the interferer. The interferer had a constant F0 in the two intervals of the 2I-2AFC task which corresponded to the arithmetic mean of the targets' F0s. The arithmetic mean of the target F0 together with the interferer's F0 varied randomly across trials over a range of $\pm 10\%$ in Experiment 1 and over a ± 3 semitone range (approximately 41% total range) in Experiment 2. The target and the interferer were always bandpass filtered into separate and remote frequency regions: the LOW (125 - 625 Hz) and the MID region (1375-1875 Hz). In Experiment 1, the filter cutoffs varied proportionally with the mean F0 across trials. In Experiment 2, the filter cutoffs were fixed. Both complexes contained some resolved harmonics and thus had a salient pitch. The experimental findings were as follows:

1. Performance in the absence and in the presence of the interferer was a monotonic function of F0, the frequency difference between the two target tones. This was true regardless of whether the target was in the LOW region and the interferer in the MID region, or vice versa.
2. Some PDI (reduced performance in the presence of the remote interferer) was observed for small values of F0 (1.8 % or less). The pattern of PDI was the same whether the target and the interferer were filtered into the LOW and HIGH regions, respectively, or vice versa.
3. In both experiments, no significant PDI was found for values of F0 equal to or larger than 3.6%, regardless of whether the target was in the LOW region and the interferer in the MID region, or vice versa. Specifically, in Experiment 2, where stimuli were identical to those used by Micheyl and Oxenham (2007), no near-chance performance was observed when the target was in the MID region and the interferer in the LOW region for F0 values of 14 and 20%.
4. In Experiment 2, for large values of F0, some of the subjects who were new to the task needed some special training when the target and the interferer were in the MID and LOW regions, respectively, but not when the target and the interferer were in the LOW and MID regions, respectively.

The finding of some PDI for small values of F0 (1.8 % or less) between two complex tones containing resolved harmonics replicates Micheyl and Oxenham's (2007) results. The finding of a monotonic relationship between sensitivity and F0 when the target was in the MID region and the interferer in the LOW region contrasts with the results of Micheyl and Oxenham (2007), even in experiment 2, which replicated their conditions closely. However, the observations during the training phase indicate that, for some subjects, there is an asymmetry in the initial salience of the correct listening cue across conditions. Although we never observed the non-monotonic performance reported by Micheyl and Oxenham, our results do clearly show that cueing subjects to listen to the correct part of a sound can have a strong effect on performance.

Acknowledgments

This work was supported by EPSRC Grant EP/D501571/1. We thank Brian Moore, Alain de Cheveigné, Christophe Micheyl, an anonymous reviewer and Ruth Litovsky for helpful comments on a previous version of this manuscript.

REFERENCES

- Assmann PF, Summerfield AQ. Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *J. Acoust. Soc. Am.* 1990; 88:680–697. [PubMed: 2212292]
- Assmann PF, Summerfield Q. The contribution of waveform interactions to the perception of concurrent vowels. *J. Acoust. Soc. Am.* 1994; 95:471–484. [PubMed: 8120258]
- Bernstein JG, Oxenham AJ. Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? *J. Acoust. Soc. Am.* 2003; 113:3323–3334. [PubMed: 12822804]
- Carlyon RP, Shackleton TM. Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? *J. Acoust. Soc. Am.* 1994; 95:3541–3554.
- Culling JF, Darwin CJ. Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating. *J. Acoust. Soc. Am.* 1994; 95:1559–1569. [PubMed: 8176059]
- de Cheveigné A. Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing. *J. Acoust. Soc. Am.* 1993; 93:3271–3290.
- Gockel H, Carlyon RP, Moore BCJ. Pitch discrimination interference: The role of pitch pulse asynchrony. *J. Acoust. Soc. Am.* 2005a; 117:3860–3866. [PubMed: 16018488]
- Gockel H, Carlyon RP, Plack CJ. Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms. *J. Acoust. Soc. Am.* 2004; 116:1092–1104. [PubMed: 15376675]
- Gockel H, Carlyon RP, Plack CJ. Dominance region for pitch: Effects of duration and dichotic presentation. *J. Acoust. Soc. Am.* 2005b; 117:1326–1336. [PubMed: 15807021]
- Gockel HE, Moore BCJ, Carlyon RP, Plack CJ. Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency. *J. Acoust. Soc. Am.* 2007; 121:373–382. [PubMed: 17297792]
- Goldstein JL. An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* 1973; 54:1496–1516. [PubMed: 4780803]
- Acoustics - Reference zero for the calibration of audiometric equipment - Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones. International Organization for Standardization; Geneva: 2004. ISO 389-8
- Laguitton V, Demany L, Semal C, Liegeois-Chauvel C. Pitch perception: A difference between right- and left-handed listeners. *Neuropsychologia.* 1998; 36:201–207. [PubMed: 9622185]
- Meddis R, Hewitt M. Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery. I: Pitch identification. *J. Acoust. Soc. Am.* 1991; 89:2866–2882.
- Meddis R, Hewitt M. Modeling the identification of concurrent vowels with different fundamental frequencies. *J. Acoust. Soc. Am.* 1992; 91:233–245. [PubMed: 1737874]
- Meddis R, O'Mard L. A unitary model of pitch perception. *J. Acoust. Soc. Am.* 1997; 102:1811–1820. [PubMed: 9301058]
- Micheyl C, Oxenham AJ. Across-frequency pitch discrimination interference between complex tones containing resolved harmonics. *J. Acoust. Soc. Am.* 2007; 121:1621–1631. [PubMed: 17407899]
- Moore, BCJ. *An Introduction to the Psychology of Hearing*. 2nd Ed.. Academic Press; London: 1982.
- Moore BCJ, Glasberg BR, Baer T. A model for the prediction of thresholds, loudness and partial loudness. *J. Audio Eng. Soc.* 1997; 45:224–240.
- Moore BCJ, Glasberg BR, Peters RW. Relative dominance of individual partials in determining the pitch of complex tones. *J. Acoust. Soc. Am.* 1985; 77:1853–1860.
- Moore BCJ, Glasberg BR, Shailer MJ. Frequency and intensity difference limens for harmonics within complex tones. *J. Acoust. Soc. Am.* 1984; 75:550–561. [PubMed: 6699293]
- Moore BCJ, Ohgushi K. Audibility of partials in inharmonic complex tones. *J. Acoust. Soc. Am.* 1993; 93:452–461. [PubMed: 8423261]
- Patel AD, Balaban E. Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nature Neuroscience.* 2001; 4:839–844.
- Plomp R. The ear as a frequency analyzer. *J. Acoust. Soc. Am.* 1964; 36:1628–1636.
- Plomp R. Pitch of complex tones. *J. Acoust. Soc. Am.* 1967; 41:1526–1533. [PubMed: 6075560]

- Plomp R, Mimpfen AM. The ear as a frequency analyzer II. *J. Acoust. Soc. Am.* 1968; 43:764–767. [PubMed: 5645825]
- Ritsma RJ. Frequencies dominant in the perception of the pitch of complex sounds. *J. Acoust. Soc. Am.* 1967; 42:191–198. [PubMed: 6052077]
- Scheffers, MTM. Sifting vowels: auditory pitch analysis and sound segregation. Groningen University; The Netherlands: 1983. Ph.D. Thesis
- Schneider P, Sluming V, Roberts N, Scherg M, Goebel R, Specht HJ, Dosch HG, Bleeck S, Stippich C, Rupp A. Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience.* 2005; 8:1241–1247.
- Shackleton TM, Carlyon RP. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* 1994; 95:3529–3540. [PubMed: 8046144]
- Smootenburg GF. Pitch perception of two-frequency stimuli. *J. Acoust. Soc. Am.* 1970; 48:924–942. [PubMed: 5480388]
- Terhardt E. On the perception of periodic sound fluctuations (roughness). *Acustica.* 1974; 30:201–213.

FIG. 1.

Mean performance (d') in an F0-discrimination task, and the associated standard errors (across subjects) obtained in experiment 1. The nominal F0 of the target tones was 250 Hz. Open and solid symbols show performance in the absence and presence of an additional synchronously presented complex tone (the interferer). Figure 1a (top) shows performance when the target was filtered into a LOW frequency region from 125 to 625 Hz and the interferer was filtered into a MID frequency region from 1375 to 1875 Hz. Figure 1b is as figure 1a, but with the target in the MID region and the interferer in the LOW region.

FIG. 2.

Mean performance (d') in an F0-discrimination task, and the associated standard errors (across subjects) obtained in experiment 2. Otherwise as Figure 1. Figure 2b (bottom) shows mean results and corresponding standard errors across five subjects (circles connected by solid lines), and separately the results of one “different” worse-performing subject (circles connected by dashed lines).