Detection and $F0$ discrimination of harmonic complex tones in the presence of competing tones or noise

Christophe Micheyl,[a] Joshua G. W. Bernstein, and Andrew J. Oxenham[b]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

(Received 5 October 2005; revised 26 May 2006; accepted 3 June 2006)

Normal-hearing listeners' ability to "hear out" the pitch of a target harmonic complex tone (HCT) was tested with simultaneous HCT or noise maskers, all bandpass-filtered into the same spectral region (1200–3600 Hz). Target-to-masker ratios (TMRs) necessary to discriminate fixed fundamental-frequency ($F0$) differences were measured for target $F0$s between 100 and 400 Hz. At high $F0$s (400 Hz), asynchronous gating of masker and signal, presenting the masker in a different $F0$ range, and reducing the $F0$ rove of the masker, all resulted in improved performance. At the low $F0$s (100 Hz), none of these manipulations improved performance significantly. The findings are generally consistent with the idea that the ability to segregate sounds based on cues such as $F0$ differences and onset/offset asynchronies can be strongly limited by peripheral harmonic resolvability. However, some cases were observed where perceptual segregation appeared possible, even when no peripherally resolved harmonics were present in the mixture of target and masker. A final experiment, comparing TMRs necessary for detection and $F0$ discrimination, showed that $F0$ discrimination of the target was possible with noise maskers at only a few decibels above detection threshold, whereas similar performance with HCT maskers was only possible 15–25 dB above detection threshold. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2221396]

PACS number(s): 43.66.Dc, 43.66.Fe, 43.66.Hg [JHG] Pages: 1493–1505

I. INTRODUCTION

In everyday life, several sound sources are often simultaneously present in the environment. An essential task for the auditory system is to analyze these complex acoustic mixtures in order to detect, identify, and track sounds of interest amid other sounds (Bregman, 1990). For humans and many animal species, the target sounds often fall under the category of harmonic complex tones (HCTs). This encompasses voiced speech sounds and animal vocalizations, the sounds produced by most musical instruments, and many artificially produced alarm signals. Thus, determining how HCTs are "heard out" in the presence of various other types of interfering sounds is an important step toward a better understanding of auditory perception in everyday situations. Besides its theoretical importance, this type of research could have interesting applications in the design of artificial auditory scene analysis systems for automatic speech recognition, automated musical transcription, or perceptual coders for audio compression. Similarly, understanding how the normal auditory system processes HCTs in the presence of other sounds may improve our relatively limited understanding of the listening difficulties experienced by hearing-impaired listeners and cochlear-implant users in environments where multiple sound sources are simultaneously present.

Our understanding of listeners' abilities to hear out and identify HCTs in the presence of competing sounds stems primarily from studies involving either vowel identification or fundamental-frequency ($F0$) identification or discrimination. In so-called double- or concurrent-vowel experiments, two (usually synthetic) vowels are presented simultaneously to a listener, whose task is to identify them. The results of such experiments have demonstrated an important role of differences in fundamental frequency ($\Delta F0$) between the two vowels in promoting correct identification (e.g., Scheffers, 1983; Summerfield and Assmann, 1991; Culling and Darwin, 1993; de Cheveigné *et al.*, 1995). Various models have been proposed to explain this effect (Parsons, 1976; Weintraub, 1987; Stubbs and Summerfield, 1988; Assmann and Summerfield, 1990; Meddis and Hewitt, 1992; de Cheveigné *et al.*, 1995; de Cheveigné, 1997; Cariani, 2001). These models, which operate in the spectral or temporal domain, usually involve as a first stage the estimation of at least one of the two $F0$s present, and as a second stage the use of that $F0$ information to either select or suppress groups of harmonics. An assumption of this class of model is that the improvement in concurrent-vowel identification with increasing $\Delta F0$ is related to $F0$-based perceptual segregation. However, it has been suggested (Culling and Darwin, 1994) that $F0$-based segregation may only work for $\Delta F0$s larger than those at which the improvement in concurrent-vowel identification performance typically plateaus (i.e., about 1 semitone, or 6%). At smaller $\Delta F0$s, identification could be mediated by beats between adjacent harmonics in the experimental steady-state sounds, which may allow "glimpsing" of the individual vowels at different times [Assmann and Summerfield, 1994; Culling and Darwin, 1994; however, see de

[a]Author to whom correspondence should be addressed. Electronic mail: cmicheyl@mit.edu
[b]Present address: Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455. Electronic mail: oxenham@umn.edu

Cheveigné (1999b) for a critique of this interpretation]. Furthermore, it has been shown that listeners can identify above chance even pairs of vowels that have the same $F0$ (e.g., Scheffers, 1983; Zwicker, 1984; Assmann and Summerfield, 1989; Qin and Oxenham, 2005). Thus, good performance in a concurrent-vowel experiment does not necessarily imply $F0$-based perceptual segregation.

Segregation of concurrent HCTs has also been studied using $F0$ identification (Beerends and Houtsma, 1989) and $F0$ discrimination (Carlyon 1996a; 1997). Beerends and Houtsma (1989) found that the ability of listeners to correctly identify the pitches of two simultaneous two-component complexes was strongly impaired only when none of the four components was sufficiently separated in frequency from the others to be individually resolved by the peripheral auditory system. This finding suggests that the ability to exploit $\Delta F0$s in order to hear out concurrent HCTs may be constrained by the interference of spectral components within the auditory periphery, due to limited cochlear frequency resolution. Using rather different techniques, Carlyon (1996a; 1997) also concluded that peripheral resolvability played an important role in determining listeners' ability to hear out one HCT in the presence of another. The first study (Carlyon, 1996a) measured listeners' performance in a sequential $F0$ discrimination task between two consecutive target HCTs in the absence and presence of a simultaneous masker HCT, bandpass-filtered into the same spectral region, with the same $F0$ in the two observation intervals. In the condition where the target and masker both consisted primarily of resolved harmonics, performance was only moderately affected by the masker. In the condition where the target and masker only contained unresolved harmonics, listeners were still able to perform the task, but reported not being able to hear two sounds; instead, they heard the target and masker as a single "crackly" sound. Carlyon concluded that performance in that condition was probably based on the discrimination of global changes in the pitch evoked by the target-plus-masker mixture, rather than the pitch of the target alone. Specifically, the rate of envelope peaks of the combined masker and target increased with increases in the target $F0$, giving a so-called "mean-rate cue." This conclusion was supported by a later study (Carlyon, 1997), which showed chance performance in a similar task when the mean-rate cue was neutralized by using pseudorandom pulse trains generated in such a way that, when mixed together, they contained the same number of pulses.

The present study was designed to gain additional information regarding the perceptual segregation of concurrent harmonic complexes occupying overlapping spectral regions, and the role of peripheral frequency resolution in this ability. Similar to Carlyon (1996a, 1997), we used an $F0$-discrimination paradigm, with two important differences. First, instead of using a fixed target-to-masker ratio (TMR) of 0 dB [as in Carlyon (1996a)], we measured the TMR necessary to detect a fixed $\Delta F0$ between the targets. A fixed $\Delta F0$, which was suprathreshold in the absence of the masker, provided us with an operational definition for listeners' ability to "hear out" the target and enabled us to test this ability over a range of TMRs. Second, the $F0$ of the masker was roved across intervals, to discourage listeners from using any composite pitch of the target and masker, deriving from a "mean-rate" cue. Varying degrees of peripheral resolvability of the harmonics were obtained by using $F0$s ranging from below 100 Hz to above 400 Hz and bandpass filtering the stimuli between 1200 and 3600 Hz. For example, with the 100-Hz target $F0$, the spectral components of the target were unresolved from each other, and thus remained unresolved after the addition of the masker; with the 200-Hz target, some components were resolved but may have been unresolved after being added to the masker (depending on the TMR and masker $F0$); at the 400-Hz $F0$, the target components were well resolved from each other and some of them could still remain resolved after the masker was added. Finally, in addition to using harmonic maskers, we also used noise maskers. Besides the fact that noise constitutes another common source of interference in real-life environments, the results of the noise-masker conditions provide control or reference data, against which to compare the effects of HCT maskers.

A preliminary experiment measured $F0$ discrimination thresholds for isolated HCTs. The main experiment (experiment 2) measured the TMR required to discriminate the $F0$ of a HCT in the presence of harmonic or noise maskers. Experiment 3 repeated experiment 2, but without roving the masker $F0$ across the two intervals of each trial, to investigate the possible roles of masker distraction and of composite (signal-plus-masker) percepts when they provide reliable information. Experiment 4 investigated the effect of introducing additional segregation cues between the masker and target, based on onset and offset asynchronies. The final experiment measured detection thresholds for the target HCTs in both harmonic and noise maskers, in order to determine the relationship between the TMR necessary for the detection of a target and the TMR necessary for pitch discrimination judgments.

II. GENERAL METHODS

A. Stimuli

In all experiments, the target stimuli were 500-ms (total duration) HCTs gated on and off with 20-ms raised-cosine ramps. They were bandpass filtered between 1200 and 3600 Hz using digital trapezoidal filters (on a log frequency scale) with slopes of 48 dB/oct. The filtering was done in the spectral domain by individually adjusting the amplitude of each sinusoidal component and then summing the components together. Depending on the experiment, the target HCTs were either presented alone (experiment 1) or accompanied by simultaneous maskers (experiments 2–5). When present, the maskers were usually 500 ms in duration (including 20-ms raised-cosine ramps) and were usually gated on and off together with the targets (with the exception of experiment 4). The maskers were either HCTs or Gaussian noise, bandpass filtered in the same way as the targets. When the masker was a HCT, depending on the condition being tested, its nominal (i.e., average) $F0$ was either equal to ($C0$ condition), 7 semitones lower than ($C-7$ condition), or 7 semitones higher than ($C+7$ condition) the nominal target

$F0$. In most experiments, the masker $F0$ was roved across trials and intervals. The details of the roving are provided within the descriptions of each experiment.

The target HCTs had nominal $F0$s of 100, 200, or 400 Hz; actual target $F0$s were roved across trials over a 6-semitone range (±3 semitones), independent of the masker $F0$. Within the selected passband (1200–3600 Hz), these nominal $F0$s result in different degrees of peripheral resolvability of the harmonics. While the exact limit between resolved and unresolved harmonics can vary somewhat depending on how the measurements are made and how resolvability is defined, it is generally believed that the first six to ten harmonics are peripherally resolved (Plomp, 1964; Bernstein and Oxenham, 2003). With the stimulus passband used here (1200–3600 Hz), target HCTs having a nominal $F0$ of 100 Hz contained only harmonics higher than the 10th, i.e., all peripherally unresolved; target HCTs with a nominal $F0$ of 200 Hz contained harmonics between the 5th and the 21st, thus including some resolved and some unresolved components; and target HCTs with a nominal $F0$ of 400 Hz contained only harmonics less than about the 10th, i.e., all resolved. Of course, this analysis only holds for each HCT in isolation; the addition of a masker HCT reduced the degree to which individual harmonics were resolved in ways that depended on the TMR as well as on the specific masker and target $F0$s; we will return to this question when discussing the results.

The starting phases of the target and masker components were drawn randomly and independently from a uniform distribution spanning 0°–360° on each presentation. This was done in order to avoid providing listeners with consistent cues associated with a specific choice of phase relationship for the stimulus components (e.g., de Cheveigné, 1999b).

All stimuli were presented in lowpass-filtered (1200-Hz cutoff) pink noise. The purpose of this background noise was to prevent listeners from detecting distortion products generated at lower harmonic frequencies, which could have confounded the interpretation of the results. To ensure that distortion products would always be masked, the 1/3rd-octave band level of the lowpass noise was adjusted on each trial to be equal to the highest level per component of the target or masker. The background noise was turned on 500 ms before the onset of the first target on a trial and off 500 ms after the offset of the second target on the same trial. It had a total duration of 2.5 s, including 20-ms raised-cosine ramps.

B. Procedure

In all experiments, thresholds were measured using a two-interval, two-alternative forced-choice (2I-2AFC) procedure with an adaptive three-down one-up rule, which tracked the 79.4%-correct point on the psychometric function (Levitt, 1971). Each trial consisted of two 500-ms observation intervals separated by 500 ms. The intervals were marked by "lights" on a virtual response box, displayed on a computer screen. In all experiments except the last, the two intervals contained target HCTs that differed in $F0$ by an amount, $\Delta F0$, which was either varied adaptively (experiment 1) or kept constant (experiments 2–4) within each experimental run. The interval containing the target with the higher $F0$ was selected randomly with a 0.5 probability, and the listener's task was to indicate whether the higher-$F0$ target occurred in the first or second interval. Responses were entered via a computer keyboard (key "1" for interval 1, key "2" for interval 2). Visual feedback was provided following each trial. The procedure and task used in the final experiment are described separately under Sec. VII.

When a masker was present, its overall [root-mean-square (rms)] level was either kept fixed at 56 dB SPL (experiments 2–4) or roved across observation intervals between 51 and 61 dB SPL (experiment 5). The rms level of the target was always set relative to the (actual) rms level of the masker. In all experiments except the first, the TMR was the variable in the adaptive tracking procedure. After three consecutive correct responses, the TMR was decreased; after each incorrect response, it was increased. At the beginning of a block of trials, the step size by which the level of the target was increased or decreased was set to 6 dB; it was reduced to 4 dB after the first reversal in the direction of tracking from increasing to decreasing, and to 2 dB after the second such reversal. Thereafter, the step size was kept fixed at 2 dB. The procedure stopped after a total of six reversals with the 2-dB step size. The threshold was defined as the average TMR at the last six reversal points. Each condition was repeated at least three times for each listener.

C. Apparatus

The stimuli were generated digitally and played out via a soundcard (LynxStudio LynxOne) with 24-bit resolution and a sampling frequency of 32 kHz. The stimuli were then passed to a headphone buffer (TDT HB6) before being presented to the listener via the left earpiece of Sennheiser HD 580 headphones. Subjects were seated in a double-walled sound-attenuating chamber.

D. Listeners

Four listeners (one female, ages between 22 and 28 yr) took part in the study. They all had pure-tone hearing thresholds less than 20 dB HL at octave frequencies between 250 and 8000 Hz. All had some musical education and had several years of experience playing an instrument.

III. EXPERIMENT 1. $F0$ DISCRIMINATION THRESHOLDS

A. Rationale

The aim of this preliminary experiment was to measure $F0$ difference limens (DL$F0$s) for the target HCTs at each of the three nominal target $F0$s tested in the absence of any interference. This enabled the individual adjustment of the (fixed) $\Delta F0$ between the two targets in the other experiments, such that it corresponded to a constant proportion of each listener's DL$F0$.

B. Methods

DL$F0$s for target complexes with a nominal $F0$ of 100, 200, or 400 Hz were measured using a 2I-2AFC paradigm

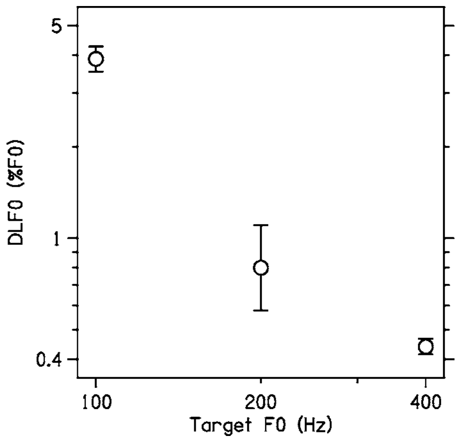

FIG. 1. Mean $F0$ discrimination thresholds or DL$F0$s for HCTs with nominal $F0$s of 100, 200, and 400 Hz (experiment 1). Thresholds are expressed in percent of the nominal $F0$, and plotted on a logarithmic scale. The error bars represent standard errors of the mean across listeners.

with an adaptive three-down one-up procedure. At the start of each trial block, the $\Delta F0$ between the two targets was set to 20% of the actual reference $F0$, which was drawn randomly on each trial from a 6-semitone range centered on the nominal $F0$ as explained in Sec. II. The $\Delta F0$ was then adaptively varied by a factor of 4 for the first two reversals, 2 for the next two reversals, and $\sqrt{2}$ for the last six reversals. The DL$F0$ was computed as the geometric mean of $\Delta F0$ at the last six reversal points. Each listener completed at least three such runs at each nominal target $F0$. The resulting DL$F0$s were geometrically averaged to produce a single DL$F0$ estimate per listener per condition. The overall level of the target was kept constant at 56 dB SPL. As in all other experiments, the background lowpass noise was present to mask distortion products but, unlike in subsequent experiments, there was no masker in the same spectral region as the target.

C. Results

The mean DL$F0$s, averaged (geometrically) across listeners are shown in Fig. 1. They were largest at the 100-Hz $F0$ (around 4%), and smallest at the 400-Hz $F0$ (around 0.45%). This is consistent with data from other studies using comparable stimulus conditions. In particular, the finding of substantially larger DL$F0$s in the 100-Hz $F0$ condition (for which the passband of the target HCT contained only unresolved harmonics) than in the 200- and 400-Hz $F0$ conditions (for which the target contained resolved harmonics), is consistent with earlier results (e.g., Hoekstra, 1979; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003; 2005).

IV. EXPERIMENT 2. *F*0 DISCRIMINATION OF HARMONIC TARGETS IN SIMULTANEOUS HARMONIC OR NOISE MASKERS

A. Methods

The target stimuli were similar to those used in experiment 1. However, in contrast to experiment 1, the $\Delta F0$ between the two consecutive targets was kept constant throughout each run. Depending on the condition, the $\Delta F0$ for each listener was set to 2 or 4 times the DL$F0$ in quiet, as measured individually in experiment 1. The target was always accompanied by a HCT or noise masker, filtered into the same spectral region, and the TMR was varied adaptively. The overall masker level was fixed at 56 dB SPL. The masker $F0$ was roved over a 6-semitone range across trials. Further roving was introduced between the two intervals within each trial, to discourage listeners from basing their judgments on any global percepts evoked by the target-plus-masker mixtures. The within-trial roving was set to 3 times the $\Delta F0$ between the two targets, corresponding to 6 or 12 times each listener's DL$F0$.

B. Results and discussion

Figure 2 shows the threshold TMRs averaged across all four listeners for the different testing conditions. These data were analyzed using a repeated-measures analysis of variance with target $F0$ (100, 200, or 400 Hz), masker type ($C-7$, $C0$, $C+7$, or noise), and $\Delta F0$ (2 or 4 times the listener's DL$F0$) as within-listener factors. A general finding, which was observed in this and all subsequent experiments of the same type, is that TMRs were lower at the larger $\Delta F0$ ($4\times$DL$F0$) than at the smaller $\Delta F0$ ($2\times$DL$F0$) [$F(1,3)=41.31$, $p=0.008$]. This can be explained rather simply by considering that increasing the $\Delta F0$ between the two targets made it easier for listeners to detect which was the higher in pitch, so that listeners could tolerate lower TMRs while still

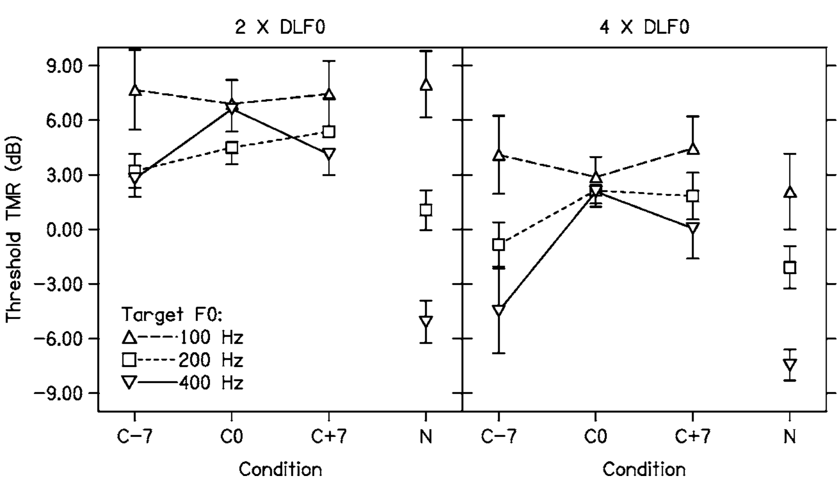

FIG. 2. TMRs required for 79.4%-correct target $F0$ discrimination in the presence of harmonic or noise maskers (experiment 2). Left and right panels show data using $\Delta F0$s between the targets in each trial corresponding to two and four times each listener's DL$F0$, respectively. In this and subsequent figures, the masker type is indicated under the horizontal axis: "*N*" stands for noise and "*C*-7," "*C*0," and "+*C*7" stand for a complex tone where the $F0$ range of the masker was centered 7 semitones below, equal to, or 7 semitones above that of the target, respectively. Symbols represent target complex nominal $F0$s: downward-pointing triangles for 100 Hz, circles for 200 Hz, and upward-pointing triangles for 400 Hz. In this and subsequent figures, the error bars represent standard errors of the mean across listeners. In cases of overlap, error bars on one side of the symbol are omitted for clarity.

achieving about 79% correct responses in the discrimination task. The same general trends were usually observed at both values of $\Delta F0$ in all experiments, so that there were no significant interactions between this and the other experimental factors. The effects of target $F0$ and masker type are addressed below.

1. Effect of average target *F*0

Overall, the threshold TMRs improved significantly as the target's nominal $F0$ increased from 100 to 400 Hz [main effect: $F(1.20,3.60)=10.55$, $p=0.035$; linear contrast: $F(1,3)=1.887$, $p=0.041$].[1] This effect was particularly large in the noise-masker conditions (plotted on the right in each panel of Fig. 2), where the mean threshold TMR decreased by more than 10 dB as the nominal target $F0$ increased from 100 to 400 Hz [$F(1.40,4.21)=20.04$, $p=0.0008$]. With harmonic maskers the reduction in threshold TMR with increasing target $F0$ was generally not as dramatic, and was only present when the average masker $F0$ was different from that of the target $F0$ [$F(1.31,3.92)=8.54$, $p=0.018$].

The reason why TMRs in noise vary so dramatically with target $F0$ is not completely clear, although it is consistent with earlier studies (Bilsen, 1973; Hoekstra, 1979). Bilsen (1973) showed that a 6% change in the $F0$ of a two- or three-component HCT was more easily masked by broadband noise for high than for low harmonic numbers. Hoekstra (1979, pp. 42–43) measured rate discrimination thresholds for bandpass-filtered periodic pulse trains in a noise masker as a function of TMR, with pulse rate (or $F0$) as a parameter. His results indicate that as TMR decreases, thresholds increase much more dramatically for low-$F0$ than for high-$F0$ pulse trains in a given spectral region. Envelope pitch information carried by unresolved harmonics may be more susceptible to noise than the spectral or fine-structure information carried by resolved harmonics, and this might explain the very high thresholds in the 100-Hz $F0$ condition. For resolved harmonics, such as those in the 200- and 400-Hz $F0$ conditions, if discrimination could be based on the frequencies of individual harmonics, then the improvement with increasing $F0$ might be due to the higher level per component in the higher $F0$ conditions: every doubling in $F0$ leads to a halving of the number of components within the passband and, for a constant overall level, a concomitant doubling (3-dB increase) in the intensity of each component. The extent to which these different explanations can account for the results remains to be explored.

In the presence of a HCT masker whose $F0$ was different from that of the target, thresholds decreased with increasing $F0$. This is probably due to the concomitant increase in average spacing between adjacent spectral components of the composite stimulus. This presumably resulted in greater peripheral separation of the target and masker components on average, so that the likelihood of there being some peripheral auditory filters with a relatively large TMR at the output increased. In this case, therefore, the improvement in TMR with increasing masker and target $F0$ may relate to an increase in the audibility of the target. The question of target audibility is addressed further in experiment 5.

2. Effect of masker type

The effect of masker type was dependent on the target $F0$ range. For the lowest average target $F0$ (100 Hz), there was no effect of masker type: TMRs at threshold were approximately the same whether the masker was a noise or a HCT, and were roughly constant across the three different relative masker $F0$ ranges. At the 200-Hz target $F0$, some emerging trends for TMRs to vary with masker type and masker $F0$ were apparent; harmonic maskers appeared on average to produce higher threshold TMRs than noise maskers, and harmonic maskers with an $F0$ below that of the target appeared to produce lower threshold TMRs than maskers with an $F0$ equal to, or a higher than, the target $F0$. However, these trends failed to reach significance. In contrast, for the highest average target $F0$ (400 Hz), mean TMRs at threshold were strongly dependent on masker type, varying over a 14-dB range. Threshold TMRs for the 400-Hz targets were generally lower in noise than in HCT maskers [$F(1,3)=284.77, p<0.001$]. This may reflect a pitch-interference effect, whereby the pitch of the masker interferes with the ability of listeners to discriminate the pitch of the target (Gockel *et al.*, 2004). Consistent with this interpretation, threshold TMRs were highest when the average target and masker $F0$s were the same, possibly reflecting the difficulty in distinguishing the masker from the target when they are both drawn from the same range of $F0$s.

3. Implications of positive and negative TMRs with HCT maskers

Threshold TMRs of 0 dB or less imply that the pitch of both the target and HCT masker may have been heard by the listeners. This is because the masker level was either the same as or higher than the target level and therefore was likely to have had at least as clear a pitch as the target (assuming that the pitch salience of the masker and target are roughly equal when played in isolation). The threshold TMR was negative for the 400-Hz target $F0$ in the $4\times \mathrm{DL}F0$ condition when the masker and target $F0$s were different, and approached zero for one of the 200-Hz target $F0$ conditions.

Positive TMRs are less straightforward to interpret. It may be that the target essentially masks the pitch of the masker at high TMRs. This seems particularly likely for the 100-Hz $F0$ target, where threshold TMRs were rarely below +3 dB. If the masker $F0$ was not perceived in the 100-Hz $F0$ target conditions, this would explain why changes in the masker $F0$ range had no effect, and why the TMRs were roughly the same for both noise and HCT maskers. The inaudibility of the masker $F0$ for the 100-Hz $F0$ target condition would be consistent with Carlyon's (1996a) observation that when listeners were presented with a target and masker containing in their passband only unresolved harmonics having approximately the same level, listeners could not perceptually separate the two, and heard the mixture as a fused sound.

Because our technique involved an adaptive tracking threshold procedure, and always started with a high TMR, it is also possible that a positive TMR threshold reflects the minimum level difference at which it was possible to distinguish the target and masker based on loudness differences. In other words, the psychometric function may be nonmonotonic, with a minimum reached when the target and masker are at similar levels. In that case, it is possible that performance might again improve at negative TMRs, when listeners could focus on the quieter of the two sounds present (for an example from the speech perception literature, see Brungart, 2001). However, the most likely situation for a level difference to be effective would be in cases where the $F0$ range of the masker and target are the same. In the other cases, it appears unlikely that a level difference would provide a more salient cue than the large pitch-range differences already present between the masker and target.

In summary, the positive TMRs observed in most of the cases tested in experiment 2 suggest that it is usually not possible to hear out the pitch of a target in the presence of a synchronously gated, higher-level harmonic masker in the same spectral region, even when substantial target-masker $F0$ differences exist and when the target and masker themselves contain mostly resolved harmonics when presented in isolation.

V. EXPERIMENT 3. INFLUENCE OF WITHIN-TRIAL RANDOMIZATION OF THE MASKER *F*0

A. Rationale

In the previous experiment, the masker $F0$ was randomized between the two observation intervals on each trial. This was done to encourage listeners to base their judgments on the $F0$ of the target, rather than on some overall sensation derived from the mixture of the masker plus the target, such as Carlyon's (1996a) mean-rate cue. Although this approach helps to ensure that responses were based on the pitch of the target alone, it has the disadvantage that performance might be limited by the potentially distracting variations in the masker $F0$. In other words, performance in experiment 2 may have been limited by confusions between the target and masker, even when the listeners were able to perceptually separate the two simultaneous sounds. The observed tendency for TMRs to improve with increased target-masker $F0$ differences is consistent with this idea.

In this experiment, we kept the masker $F0$ constant across the two observation intervals to test two hypotheses related to the observed threshold TMR measurements of experiment 2. The first hypothesis was that the threshold TMRs were limited by masker distraction or target-masker confusion in the 400-Hz (and possibly 200-Hz) target $F0$ conditions, where the threshold TMRs measured in experiment 2 were usually the lowest, indicating that the masker $F0$ may have been heard and was sometimes louder than the target. If so, then a constant masker $F0$ should lead to an improvement (a decrease) in the threshold TMRs by decreasing the possible influence of masker distraction or target-masker confusions. The second hypothesis was that mean-rate cues could play a role in determining thresholds with complex consisting of only unresolved harmonics (Carlyon, 1996a; 1997). If so, then introducing a useful mean-rate cue by keeping the masker $F0$ constant should lead to an improvement in performance with the 100-Hz $F0$ target.

B. Methods

The stimuli and procedure for this experiment were the same as those for experiment 2, except that the $F0$ of the masker was the same in both intervals of a given trial. The masker $F0$ was still roved across trials, as in the previous experiment. Four threshold estimates were obtained in each condition for each listener. The mean threshold TMRs shown below were computed as the mean of these threshold estimates across repetitions and listeners.

C. Results and discussion

The average threshold TMRs for this experiment are shown in the top two panels of Fig. 3. The results displayed the same general trends as those of experiment 2. In particular, thresholds improved with increasing target $F0$ $[F(1.77,5.29)=28.66, p=0.002]$, and at the highest nominal target $F0$ tested (400 Hz), there was some advantage to having a difference in average $F0$ between target and masker $[F(1,3)=16.20, p=0.028]$. To facilitate comparisons between these results and those of experiment 2, threshold TMR differences (obtained by subtracting the thresholds measured in experiment 2 from those measured in corresponding conditions of the current experiment) are plotted in the lower two panels of Fig. 3. Negative values indicate an improvement in TMRs with the elimination of the within-trial rove of the masker $F0$.

The comparison reveals that keeping the masker $F0$ constant across observation intervals resulted in substantial improvements in thresholds at 400 Hz $[F(10,3)=296.09, p<0.001]$, but not at 100 or 200 Hz ($p>0.05$). This outcome is not predicted under the hypothesis that the lack of within-trial randomization of the masker $F0$ would allow listeners to take advantage of Carlyon's (1996a) mean-rate cue, because it is an envelope-based cue that should have an effect mainly at low $F0$s, where the harmonics of the target were not peripherally resolved.[2] It may be that our use of random-rather than constant sine- or cosine-phase harmonics, resulting on average in a weaker representation of the fluctuations in the stimulus temporal envelope, reduced the usefulness of the mean-rate cue. Instead, the outcome is consistent with our earlier observation that the threshold TMRs measured at the 100-Hz $F0$ in experiment 2 were usually above 0 dB, indicating that the masker was less intense than the target. Under such conditions, the target was presumably more salient than the masker, making it perhaps less necessary for the listeners to actively ignore the irrelevant changes in the masker $F0$. These results suggest that the masker $F0$ was playing little or no role for both the 100- and 200-Hz target $F0$ conditions, making it even less likely that the positive TMRs measured in experiment 2 reflect segregation based on simple loudness differences and nonmonotic psychometric functions as discussed in Sec. IV B 3, or target-masker confusions. Instead, they confirm our interpretation that, in that experiment, the 100- and 200-Hz target pitch could not usually be heard out when the target was less intense than the masker.

In contrast, in the 400-Hz nominal target $F0$ conditions, the threshold TMRs were usually lower, and sometimes negative, making it necessary for listeners to actively ignore

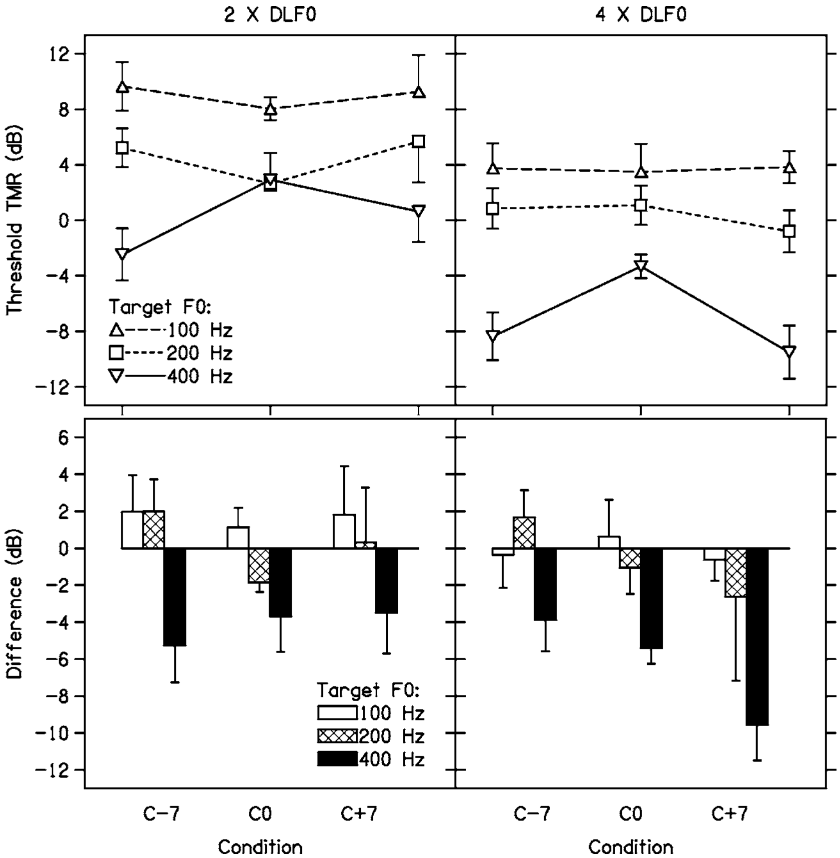

FIG. 3. Upper panels: Threshold TMRs required for 79.4%-correct $F0$ discrimination of harmonic targets in the presence of harmonic maskers having the same $F0$ in the two observation intervals within a trial. Lower panels: Differences in TMRs between experiment 3 (masker $F0$ roved across intervals) and experiment 2 (masker $F0$ held constant across intervals). Negative values indicate an improvement in TMR with the elimination of across-interval $F0$ roving in experiment 3. Left and right panels show data using $\Delta F0$s between the targets in each trial corresponding to two and four times each listener's DL$F0$ in quiet, respectively.

the irrelevant variations in masker $F0$. From that point of view, the finding that removing the within-trial variation in masker $F0$ improved thresholds at the 400-Hz $F0$ suggests that, at that nominal target $F0$, performance in experiment 2 was limited by target-masker confusions and an inability to ignore the irrelevant variation in the pitch of the masker, even though listeners may have heard the masker and target as two separate objects.

VI. EXPERIMENT 4: INFLUENCE OF ONSET-OFFSET ASYNCHRONIES

A. Rationale

Onset and offset asynchronies are usually regarded as a powerful cue for concurrent sound segregation. Components that start and end at the same time tend to be grouped together by the auditory system, while components that start or end at different times tend to be perceived as separate objects. There are many illustrations of this in the psychoacoustical literature (for reviews, see Bregman, 1990; Darwin and Carlyon, 1995). An important question is whether onset-offset asynchronies can help listeners to segregate concurrent HCTs.

Some evidence that they do can be found in two earlier studies in which listeners had to identify a synthetic vowel masked either by another vowel (Akeroyd and Summerfield, 2000) or by a flat-spectrum HCT (Demany and Semal, 1990). In the asynchronous condition, the masker started approximately 500 ms before and ended either simultaneously with (Demany and Semal, 1990) or 100 ms after (Akeroyd and Summerfield, 2000) the target. The asynchrony was found to have a beneficial effect on identification accuracy (Akeroyd and Summerfield, 2000) or masked thresholds (Demany and Semal, 1990).

On the other hand, Carlyon (1996a, 1996b) found no benefit of target-masker asynchronies in experiments which, like the present ones, required listeners to discriminate the $F0$ of two consecutive target HCTs in the presence of another HCT. In fact, when the target contained only unresolved harmonics in its passband, an asynchronous masker not only failed to help listeners, but even had a detrimental influence on performance. Carlyon (1996b) interpreted this as a sequential interference effect in pitch discrimination, whereby the leading (or trailing) portion of the masker had a negative influence on the processing of the $F0$ of the subsequent (or preceding) target.

The present experiment further tested the influence of onset-offset asynchronies on the segregation of concurrent HCTs but with two important methodological differences from the Carlyon (1996a, 1996b) studies. First, measuring threshold TMR enabled us to investigate the hypothesis that asynchronies could provide a useful segregation cue for unresolved HCTs for sufficiently large TMRs where the target $F0$ can be heard out. Second, this experiment used a target-masker asynchrony of 500 ms (instead of the 150-ms asynchrony in the Carlyon studies), testing the hypothesis that the lack of a benefit observed by Carlyon was due to an insufficiently long onset cue, as might be predicted from the results of Darwin and Ciocca (1992), who found that for one component to cease contributing to the overall pitch of a complex, the asynchrony had to exceed 300 ms.

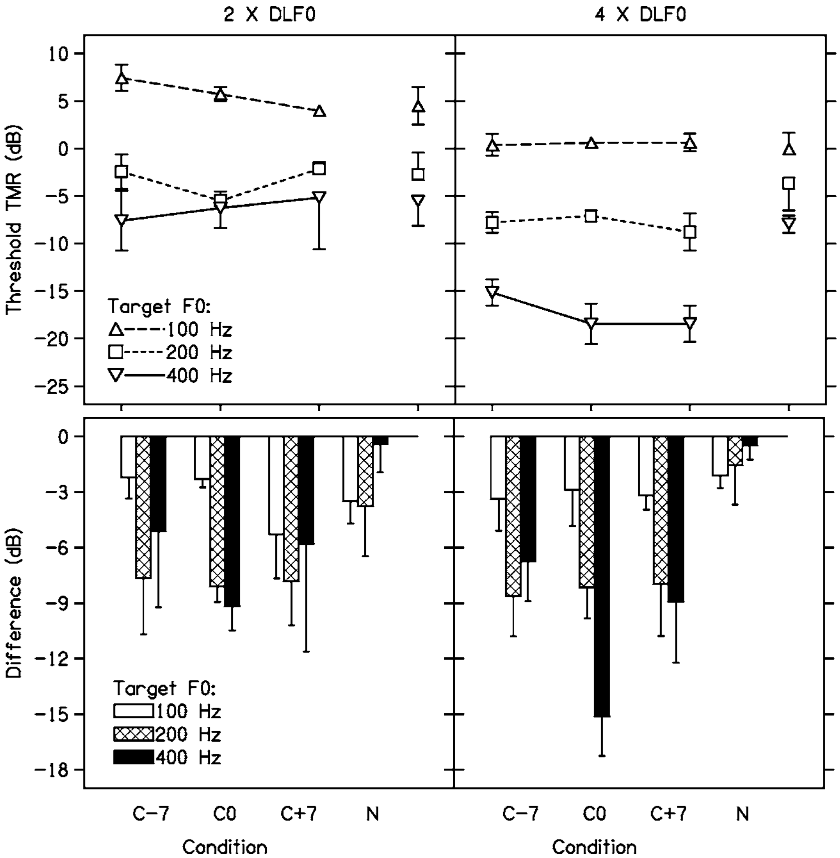

FIG. 4. Upper panels: TMRs corresponding to 79.4%-correct $F0$ discrimination of harmonic targets in the presence of asynchronous maskers (experiment 4). Lower panels: Differences in TMRs between experiment 4 (asynchronous maskers) and those measured in experiment 3 (synchronous harmonic maskers) or 2 (synchronous noise maskers). The average differences shown in these panels were computed by subtracting the synchronous TMRs measured in experiments 2 or 3 from the asynchronous TMRs measured in experiment 4. Negative TMRs indicate the benefit provided by target-masker asynchronies. Left and right panels show data using $\Delta F0$s between the targets in each trial corresponding to two and four times each listener's DL$F0$ in quiet, respectively.

B. Methods

This experiment was similar to experiment 3, except that the masker was turned on 500 ms before the onset of the first target and it continued uninterrupted until 500 ms after the offset of the second target. Thus, the masker duration was 2.5 s. The masker $F0$ remained constant throughout the 2.5 s but, as in the previous two experiments, it was roved over 6 semitones across trials. Because the masker $F0$ remained constant within each trial, the threshold TMRs were compared to the results of experiment 3, and not to the results of experiment 2, in which the masker $F0$ was also roved across intervals.

C. Results and discussion

The upper panels in Fig. 4 show the threshold TMRs measured in the current experiment. The lower panels show the difference between these TMRs and those measured in experiment 3 (synchronous maskers, fixed masker $F0$ across observation intervals) or, for the noise maskers, those measured in experiment 2. The differences were computed by subtracting the threshold TMRs measured in experiment 3 (or 2, for the noise maskers) from those measured in the current experiment, so that negative values reflect an improvement (decrease) in thresholds with the introduction of the onset-offset asynchrony. The differences were on average all negative, suggesting that when target-masker asynchronies had an effect, it was beneficial. For the noise maskers, the difference between the TMRs measured in the synchronous condition (experiment 2) and the asynchronous condition (this experiment) failed to reach statistical significance. The lack of effect of asynchronous gating with the noise maskers is probably due to the fact that these maskers differed markedly in timbre from the targets, which already provided a strong cue for perceptual segregation between the target and the masker, and made it unlikely that listeners would confuse target and masker.

For HCT maskers, the difference in TMR between the synchronous condition (experiment 3) and the asynchronous one (this experiment) proved significant at both 200 Hz [$F(1,3)=19.50, p=0.022$] and 400 Hz [$F(1,3)=11.88, p=0.041$], but just failed to reach significance at 100 Hz [$F(1,3)=8.15, p=0.065$]. These results are in contrast with those of Carlyon (1996a, 1996b), who found no benefit of target-masker asynchronies on target-$F0$ discrimination performance, even when the target and masker contained resolved harmonics in their passbands, at least prior to being mixed. This suggests that the relatively short (150-ms) asynchrony in the Carlyon studies may have been insufficient to promote the perceptual segregation of concurrent HCTs, at least for the purpose of subsequently discriminating their pitch. The studies of Demany and Semal (1990) and Akeroyd and Summerfield (2000), which found a beneficial effect of target-masker asynchrony on the identification of target vowels, also used onset asynchronies of about 500 ms. The current finding that target-masker asynchronies helped significantly at the 200 and 400 Hz nominal $F0$s indicates that two HCTs containing resolved components in their passband can be perceptually segregated, and that the segregation can be enhanced through asynchronous gating.

Our results also differ from those of Carlyon (1996a, 1996b) in that we never found the asynchronous gating to have a detrimental effect in the case where the target and masker harmonics contained only unresolved harmonics. This apparent discrepancy in outcomes could be related to the fact that the TMR in Carlyon's study was fixed at 0 dB, whereas in ours the TMR could be positive. At positive TMRs, the masker is lower in level than the target, and so may interfere less with the target, irrespective of whether it is gated synchronously or asynchronously with the target.

The finding that target-masker asynchronies did not significantly help listeners when the nominal target $F0$ was low (100 Hz) agrees with Carlyon's (1996a) observations in suggesting that it may not be possible to perceptually segregate two simultaneous complexes containing only unresolved components, even when large gating asynchronies exist between them.

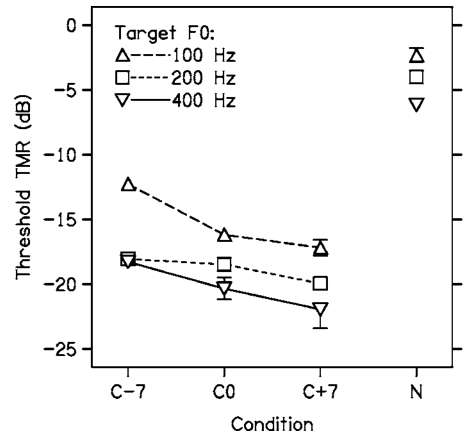

FIG. 5. Masked detection thresholds for target HCTs with nominal $F0$s of 100, 200, and 400 Hz, in the presence of harmonic or noise maskers (experiment 5).

VII. EXPERIMENT 5. DETECTION OF HARMONIC TARGETS IN NOISE AND HARMONIC MASKERS

A. Rationale

Experiments 2–4 measured the TMR necessary to discriminate a given difference in the $F0$ of the target. Some of the differences between conditions in this study may be related to differences in the audibility of the target. This final experiment was devoted to measuring the detectability of the target in the conditions of experiment 2. Thus, whereas experiment 2 required listeners to hear out the pitch of the target, here listeners were only required to detect its presence. An interesting question is whether and how the detection thresholds relate to the TMR levels required for $F0$ discrimination. One possibility is that the target pitch can be discriminated with the required level of accuracy as soon as the target is detected, in which case the target detection thresholds should be similar to the TMRs measured in the previous experiments. A more likely scenario is that the target must be set some level above its detection threshold in order for its pitch to be accurately perceived. In this case, the interesting questions are how far above its detection threshold must a target HCT be for its pitch to be correctly discriminated, and whether that level is the same or different as the masker or target $F0$ is varied, or as the masker changes from noise to a HCT.

B. Methods

Detection thresholds were measured using a 2I-2AFC procedure. In contrast to all the earlier experiments, the target was presented in only one interval, which listeners had to identify. The interval containing the target was selected randomly with a 0.5 probability prior to each trial. Another difference with the previous experiments is that the level of the masker was roved over a 10 dB range (−5 to +5 dB) around the nominal level across observation intervals as well as across trials, in order to discourage listeners from basing their judgments on differences in global loudness between the stimuli. In all other respects, the targets and maskers in this experiment had identical characteristics to those used in experiment 2.

C. Results and discussion

1. Masked detection thresholds

The masked target detection thresholds are shown in Fig. 5. Detection thresholds were considerably higher (i.e., poorer) in the presence of the noise masker than in the presence of a harmonic masker [planned comparison: $F(1,3) = 2618.24$, $p < 0.0005$]. This result can be explained by considering that when a masker is periodic, listeners can easily detect the disruption in temporal (envelope or fine structure) regularity caused by the addition of a signal with a different periodicity, in what is typically heard as fluctuations or as roughness. In contrast, when the masker is aperiodic, this cue is generally not available (Hellman, 1972; Moore *et al.*, 1998; Treurniet and Boucher, 2001b; 2001a; Gockel *et al.*, 2002; 2003). This effect may be captured by de Cheveigné's (1993, 1999a) harmonic cancellation model, which can be understood as a mechanism for detecting a disruption in temporal regularity.

Thresholds usually decreased (i.e., improved) as the nominal $F0$ of the target increased from 100 to 200 and then 400 Hz. This effect was found for both harmonic maskers [$F(1,3)=48.37, p=0.006$] and noise maskers [$F(1,3) = 117.56, p=0.002$]. The decrease in thresholds with increasing $F0$ may reflect the increasing level per component of the target HCT for a given overall stimulus level. To the extent that listeners are not able to integrate spectral information efficiently, thresholds may be dependent on the level of each component within a complex, rather than the overall level of the complex (e.g., van den Brink and Houtgast, 1990). For the $C0$, $C+7$, and N masker conditions, the decrease in threshold TMR is roughly equal to the 3 dB per $F0$ doubling that would be expected if this were the case. Another possibility is that the complexes sound more tone-like and less noise-like at higher $F0$s. Gockel *et al.* (2002) found a similar pattern of results to those reported here when the HCTs were in random phase, but not when they were in sine phase, which generally produces a more salient pitch.

The improvement in target detection thresholds with increasing $F0$ in the HCT masker conditions may be more an effect of masker $F0$ than of target $F0$. This is suggested by

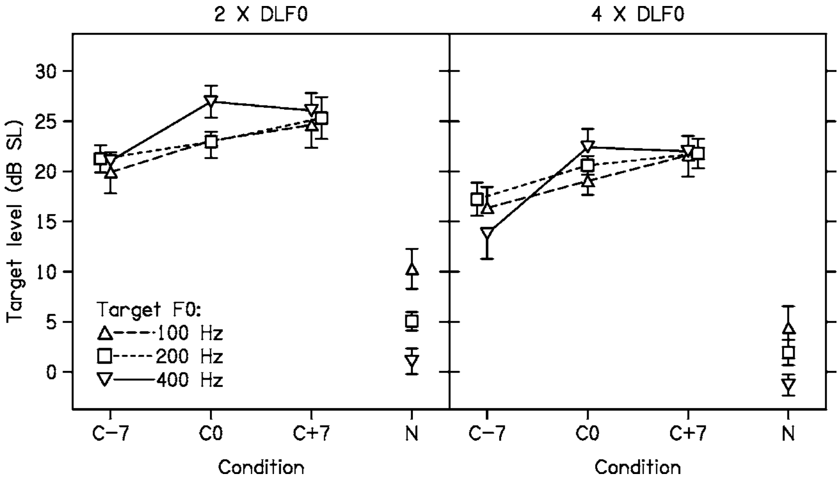

FIG. 6. Target levels above detection threshold (dB SL) required for 79.4%-correct target $F0$ discrimination in the presence of harmonic or noise maskers, showing the level required, relative to the detection threshold, to discriminate the changes in target $F0$ with an accuracy of 79.4% correct. Each "SL" value was computed by subtracting the detection threshold measured in experiment 2 (Fig. 2) from the TMR measured in experiment 5 (Fig. 5).

the following two observations: First, in the HCT masker conditions, thresholds decreased (improved) consistently with increases in the nominal $F0$ of the masker, relative to that of the target; thresholds in the $C-7$ condition were higher than in the $C0$ condition [$F(1,3)=13.93, p=0.034$], which in turn were higher than in the $C+7$ condition [$F(1,3)=20.54, p=0.020$]. Second, a comparison of conditions with very similar masker $F0$s but different target $F0$s suggests only a small and inconsistent effect of target $F0$ in the presence of a harmonic masker.[3] Thus, in contrast to thresholds in noise, thresholds in the presence of harmonic maskers do not seem to be strongly dependent on the target $F0$ (at least within the range tested here), but instead depend more on the masker $F0$. The reason for the dependence of thresholds on masker $F0$ may relate to the spectral spacing of masker components. As the $F0$ increases, the spectral gaps between masker components also increase, thereby increasing the possibility for detecting the target in spectral regions between successive masker components.

2. Levels above masked threshold required for accurate target *F*0 discrimination

In order to investigate how the TMRs required for correct target $F0$ discrimination were related to the TMRs required for correct target detection, we subtracted the threshold TMRs for detection measured in this experiment from the TMRs measured for target $F0$ discrimination in experiment 2. The resulting values are referred to as target sensation levels (SL). The SL indicates the level required, relative to the detection threshold, to discriminate the given changes in target $F0$ with an accuracy of 79.4% correct. Presenting the results in this way allows a determination of the role of audibility in determining listeners' abilities to hear out the $F0$ of the target complex. If differences in threshold TMR across conditions resulted from differences in audibility alone, then these conditions would yield similar threshold TMRs when expressed in SL terms.

Target SLs are shown in Fig. 6. The format of this figure is similar to that of Fig. 2, with the left- and right-hand panels showing the results obtained with target $\Delta F0$s set to two and four times the listener's DL$F0$, respectively. The important difference between the current figure and Fig. 2 is that here, the threshold TMRs in ordinate are expressed in dB SL, rather than relative to the masker level.

Several observations are apparent in the SL data plotted in Fig. 6. First, with the noise masker, the target SLs decreased significantly as the target $F0$ increased from 100 to 200–400 Hz [linear contrast: $F(1,3)=12.81$, $p=0.037$]. This confirms our earlier statement that, in the noise masking conditions of experiment 2, the improvement in threshold TMRs with increasing target $F0$ (shown in Fig. 2) could not be entirely explained in terms of target audibility. For the harmonic maskers, the threshold TMRs in dB SL (shown in Fig. 6) were not found to vary significantly with the target $F0$. Thus, the improvements in threshold TMRs with increasing target and masker $F0$ observed in the harmonic-masker conditions of experiment 2 (Fig. 2) were paralleled by similar improvements in detection thresholds (Fig. 5). The improvements in discrimination and detection in the presence of HCT maskers may depend on the separation of the spectral components of the target and masker, which increased with $F0$.

The target SLs measured using harmonic maskers were generally much higher than those measured with noise maskers [contrast analysis: $F(1,3)=656.35$, $p<0.001$]. Specifically, with harmonic maskers, listeners typically needed target levels between 15 and 25 dB above the detection threshold in order to correctly discriminate the changes in target $F0$. This suggests that, in HCT-masker conditions, the ($F0$) discriminability of the targets was not determined primarily by their detectability. In contrast, with noise maskers, listeners could correctly discriminate changes in the $F0$s of targets whose level was only a few decibels above their detection threshold. This is consistent with the idea that the detection of a HCT in noise may be mediated by the detection of pitch (Hafter and Saberi, 2001). A similar conclusion, for the detection of frequency modulation on a harmonic complex, was reached by Carlyon and Stubbs (1989).

VIII. GENERAL DISCUSSION

A. On the possible influence of target-masker confusion

We have generally discussed the results in terms of the ability of listeners to perceptually segregate the target from the masker. However, another logical possibility is that the target and masker are perceptually segregated (i.e., produce two separate percepts with different pitches), but that thresholds reflect a confusion on the part of the subject as to which pitch belongs to the target and which to the masker. This appears to be unlikely in the case of the 100-Hz $F0$ targets: in all cases, manipulations that should have reduced confusion, as well as improved segregation, such as keeping the masker $F0$ constant or introducing asynchronous gating, had no significant effect on threshold.

Confusion was most likely to play a role when the target $F0$ was 400 Hz and masker and target levels and $F0$s were most similar. For instance, in Fig. 2, it can be seen that the highest thresholds (poorest performance) in the 400-Hz target conditions were observed when the target and masker occupied the same $F0$ range ($C0$ conditions). On the other hand, keeping the $F0$ constant across trials within each interval (Fig. 3) did not selectively improve performance in the $C0$ condition, as might have been expected if the results in Fig. 2 were dominated by confusion effects. Thus, it appears that confusion did not play a major role in determining thresholds.

B. The role of peripheral resolvability in concurrent sound segregation

When presented in isolation, the targets in the present study comprised resolved harmonics in the $F0$ ranges of 200 and 400 Hz, but only unresolved harmonics in the 100-Hz $F0$ range. The experiments showed that manipulations designed to enhance the perceptual segregation of the masker and target improved performance for the 200- and 400-Hz targets, but not for the 100-Hz targets. This is probably because, in the case of the 100-Hz target, only one pitch was heard at a time, and the target was only discriminable when it was perceptually more salient than the masker. It might therefore be tempting to conclude that resolved harmonics are crucial for hearing out one harmonic sound in the presence of another. However, it is important to consider not only the resolvability of the target components in isolation, but also their resolvability once they are added to the masker.

The resolvability of components within a mixture depends on several factors, including the actual $F0$s of the target and masker, their relative and absolute levels and what criterion is used in defining resolvability. To provide some initial guidance, we assume the case of equal-amplitude harmonics in both the target and masker. Under such circumstances, resolvability is traditionally estimated either in terms of the minimum spacing between components (e.g., Plomp, 1964) or in terms of the number of components falling within the bandwidth of auditory filters with center frequencies within the stimulus passband (e.g., Shackleton and Carlyon, 1994). Here, we used a technique consistent with both of these approaches. We assume that a component is resolved if no other component falls within the 10-dB bandwidth of the auditory filter centered at the component's frequency. Note that this is equivalent to requiring a minimum spacing between a component and its nearest neighbor in order for that component to be deemed resolved; in the present case, that minimum distance is equal to half the 10-dB bandwidth of a model auditory filter. The 10-dB bandwidth equals 1.8 times the equivalent rectangular bandwidth (ERB) defined as: $21.4 \log_{10}(4.37\text{CF}+1)$, where CF is the filter center frequency, in kilohertz (Glasberg and Moore, 1990). Thus, we consider that a target component is resolved if its nearest neighbor is at least 0.9 ERB away from it one either side. Since 0.9 ERB is almost equivalent to a critical band, our approach is broadly consistent with the results of Plomp (1964), who concluded that "the ear is able to distinguish a simple tone in a complex sound if the frequency distance to the adjacent tones exceeds the critical bandwidth." Our approach for estimating resolvability also agrees with the common view, inspired by early as well as recent findings (e.g., Bernstein and Oxenham, 2003), that harmonics below the 10th are usually resolved. Indeed, 0.9 ERB is close to about 10%, which is roughly the distance between the 10th harmonic in a single harmonic complex and its nearest neighbors, and is also similar to the minimum distance between audible partials, as established by Plomp (1964).

Using these criteria, it is easy to show that in the 100-Hz target $F0$ condition, no target component was resolved within the stimulus passband (1200–3000 Hz) prior to the addition of the masker, from which it obviously follows that no target component was resolved after the masker was added. For the 200- and 400-Hz target $F0$ conditions, the situation is less straightforward, and computer simulations were performed in order to determine on what proportion of the trials, on average, at least one component of the target was resolved. The simulations used the same stimulus parameters as in the experiments, including the relative positions and roving range of the target and masker $F0$s. The results revealed that in conditions where the nominal $F0$ of the target was 200 Hz, the addition of the masker resulted in none of the target components being resolved. For the 400-Hz nominal target $F0$, the proportion of trials on which at least one target component was resolved in the mixture was less than 1% for the $C-7$ condition, around 6% for the $C0$ condition, and around 30% for the $C+7$ condition. Thus, for the case in which the harmonics of the target and masker are of equal amplitude, only the components of the 400-Hz $F0$ target were partially resolved when added to the masker.

This simple analysis, in combination with our results showing good performance in many conditions with a 200-Hz $F0$ target suggest the possibility that a sound may be successfully segregated from another even if none of its components is resolved within the mixture. This seems to contradict the findings of Beerends and Houtsma (1986; 1989), who found that at least one component from each two-tone complex had to be resolved for the two pitches to be successfully identified. However, their complexes consisted of only two components, which generally produce a relatively weak pitch percept, even in isolation. Nevertheless, it is premature to draw strong conclusions based on this simple analysis. More insight might be gained from comprehensive computational modeling, incorporating both relative and absolute level effects; this is beyond the scope of the current study.

C. Implications for hearing impairment and artificial auditory-scene analysis

Multiple harmonic sources are common in everyday listening environments, such as multitalker conversations and music, and present particular challenges to hearing-impaired listeners and cochlear-implant users. The current results, showing that higher TMRs are required for the successful separation of concurrent HCTs when the spectral components are not well separated at the auditory periphery, contribute to further establishing the link between the loss of peripheral frequency resolution that is often associated with cochlear damage and the difficulties in complex acoustic environments that are often reported by hearing-impaired listeners. At the same time, our finding of negative threshold TMRs in conditions where no resolved harmonic were present (e.g., the 400-Hz nominal target $F0$, $C-7$ condition in experiment 2, or the 200-Hz nominal target $F0$ conditions in experiment 4) suggest that hearing out sounds may not rely solely on their spectral resolution within a mixture.

A second area where the present findings may have interesting applications is the design of artificial auditory scene analysis systems. In recent years, mechanisms have been proposed for the separation of simultaneous HCTs, which operate based on cues present in the temporal pattern of activity from individual peripheral channels (Cariani, 2001; de Cheveigné, 2003), leading to the possibility that segregation can be achieved without spectrally resolved components. While the present results provide some evidence that concurrent HCTs can indeed be separated with some degree of success even when no resolved harmonics are present, they also indicate that factors that promote increased peripheral separation between spectral components usually promote perceptual separation. Furthermore, the observation that in conditions in which the peripheral separation of the spectral components is poorest (e.g., at the 100-Hz nominal target $F0$), positive TMRs are usually required for successful extraction of the target $F0$, suggests that some degree of spectral analysis may be required for the successful separation of simultaneous HCTs by human listeners. This is consistent with de Cheveigné's (2001, 2005) proposal that peripheral filtering may enhance the target-to-masker ratio in some channels, up to a point where temporal source-segregation mechanisms can operate effectively.

IX. SUMMARY

The ability to detect and discriminate $F0$ differences in HCT targets was measured in normal-hearing listeners in the presence of spectrally overlapping HCT or noise maskers. The following conclusions were drawn:

(1) The TMR required for listeners to be able to accurately discriminate changes in the $F0$ of a target HCT varied depending on the nature of the masker. With noise maskers, the TMR for $F0$ discrimination was generally within 10 dB of the TMR required for detection. In contrast, with a harmonic masker, the threshold TMRs for $F0$ discrimination were typically between 15 and 25 dB higher than the threshold TMR for detection. Thus, detection thresholds are probably not a valid measure of listeners' ability to perceptually segregate concurrent HCTs.

(2) When the target comprised only unresolved harmonics (100-Hz $F0$ range), TMRs were always above 0 dB for the discrimination tasks, suggesting that listeners were only able to perform the task when the target $F0$ dominated the percept, and were not able to hear the pitches of both the target and the masker simultaneously, in line with earlier studies (Beerends and Houtsma, 1986; Carlyon, 1996a).

(3) When the target contained resolved harmonics (200- and 400-Hz $F0$ range), TMRs were reduced (improved) by manipulations designed to assist in perceptually segregating the masker and target, such as difference in $F0$ range and onset/offset asynchronies. No such changes were observed with the 100-Hz $F0$ target, again suggesting that listeners' ability to hear out the pitches of harmonic complexes presented simultaneously in the same spectral region is limited by the peripheral resolvability of their components.

(4) At low or negative TMRs, even the 200- and 400-Hz targets probably contained only unresolved harmonics when combined with the masker. Therefore, harmonics may not need to remain spectrally resolved after being mixed with another source in order for the pitches of two simultaneous HCTs to be heard. However, further empirical and computational modeling studies will be required to provide a more stringent test of this preliminary conclusion.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIDCD Grant No. R01 DC 05216). The authors are grateful to R. P. Carlyon, A. de Cheveigné, and J. H. Grose, for helpful suggestions on an earlier version of this manuscript.

[1]Noninteger degrees of freedom in the reported F statistics reflect the use of the Greenhouse-Geisser correction whenever the sphericity assumption was not met.

[2]Carlyon *et al.* (2002) showed that the pitch evoked by a mixture of two HCTs comprised of unresolved harmonics corresponded to that of the HCT with the higher $F0$, and appeared to be related to the mean first-order interval between pitch pulses in the mixture. If listeners' judgments in the present experiment were based on this mean-rate pitch, thresholds should have been lower in the condition where the target $F0$ was higher than the masker's, since in that situation, according to Carlyon *et al.*'s (2002) results, the perceived $F0$ probably corresponded to that of the target. The fact that this prediction was not borne out in the present data suggests that the listeners did not rely on the mean-rate cue here.

[3]The relevant comparisons in Fig. 5 are the 100-Hz/C+7-semitones condition with the 200-Hz/C-7-semitones condition, and the 200-Hz/C+7-semitones condition with the 400-Hz/C-7-semitones condition. Both pairs of conditions have masker $F0$ ranges within a semitone of each other, but target $F0$s one octave apart. The differences between the conditions in each pair are small (2 dB or less) and in opposite directions.

Akeroyd, M. A., and Summerfield, Q. (**2000**). "Integration of monaural and binaural evidence of vowel formants," J. Acoust. Soc. Am. **107**, 3394–3406.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P. F., and Summerfield, Q. (**1989**). "Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency," J. Acoust. Soc. Am. **85**, 327–338.

Assmann, P. F., and Summerfield, Q. (**1994**). “The contribution of waveform interactions to the perception of concurrent vowels,” J. Acoust. Soc. Am. **95**, 471–484.

Beerends, J. G., and Houtsma, A. J. M. (**1986**). “Pitch identification of simultaneous dichotic two-tone complexes,” J. Acoust. Soc. Am. **80**, 1048–1055.

Beerends, J. G., and Houtsma, A. J. M. (**1989**). “Pitch identification of simultaneous diotic and dichotic two-tone complexes,” J. Acoust. Soc. Am. **85**, 813–819.

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.

Bernstein, J. G. W., and Oxenham, A. J. (**2005**). “An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination,” J. Acoust. Soc. Am. **117**, 3816–3831.

Bilsen, F. A. (**1973**). “On the influence of the number and phase of harmonics on the perceptibility of the pitch of complex signals,” Acustica **28**, 60–65.

Bregman, A. S. (**1990**). *Auditory Scene Analysis; The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).

Brungart, D. S. (**2001**). “Informational and energetic masking effects in the perception of two simultaneous talkers,” J. Acoust. Soc. Am. **109**, 1101–1109.

Cariani, P. (**2001**). “Neural timing nets,” Neural Networks **14**, 737–753.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). “Temporal pitch mechanisms in acoustic and electric hearing,” J. Acoust. Soc. Am. **112**, 621–633.

Carlyon, R. P. (**1996a**). “Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker,” J. Acoust. Soc. Am. **99**, 517–524.

Carlyon, R. P. (**1996b**). “Masker asynchrony impairs the fundamental-frequency discrimination of unresolved harmonics,” J. Acoust. Soc. Am. **99**, 525–533.

Carlyon, R. P. (**1997**). “The effect of two temporal cues on pitch judgments,” J. Acoust. Soc. Am. **102**, 1097–1105.

Carlyon, R. P., and Shackleton, T. M. (**1994**). “The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,” J. Acoust. Soc. Am. **95**, 3529–3540.

Carlyon, R. P., and Stubbs, R. J. (**1989**). “Detecting single-cycle frequency modulation imposed on sinusoidal, harmonic, and inharmonic carriers,” J. Acoust. Soc. Am. **85**, 2563–2574.

Culling, J. F., and Darwin, C. J. (**1993**). “Perceptual separation of simultaneous vowels: Within and across-formant grouping by *F*0,” J. Acoust. Soc. Am. **93**, 3454–3467.

Culling, J. F., and Darwin, C. J. (**1994**). “Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating,” J. Acoust. Soc. Am. **95**, 1559–1569.

Darwin, C. J., and Carlyon, R. P. (**1995**). “Auditory organization and the formation of perceptual streams,” in *Handbook of Perception and Cognition, Volume 6. Hearing*, edited by B. C. J. Moore (Academic, San Diego).

Darwin, C. J., and Ciocca, V. (**1992**). “Grouping in pitch perception: effects of onset asynchrony and ear of presentation of a mistuned component,” J. Acoust. Soc. Am. **91**, 3381–3390.

de Cheveigné, A. (**1993**). “Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing,” J. Acoust. Soc. Am. **93**, 3271–3290.

de Cheveigné, A. (**1997**). “Concurrent vowel identification. III. A neural model of harmonic interference cancellation,” J. Acoust. Soc. Am. **101**, 2857–2865.

de Cheveigné, A. (**1999a**). “Pitch shifts of mistuned partials: a time-domain model,” J. Acoust. Soc. Am. **106**, 887–897.

de Cheveigné, A. (**1999b**). “Waveform interactions and the segregation of concurrent vowels,” J. Acoust. Soc. Am. **106**, 2959–2972.

de Cheveigné, A. (**2001**). “The auditory system as a separation machine,” in *Physiological and Psychophysical Bases of Auditory Function*, edited by J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven Shaker Publishing BV, Maastricht, The Netherlands, pp. 453–460.

de Cheveigné, A. (**2003**). “Scene analysis without spectral analysis?,” J. Acoust. Soc. Am. **113**, 2231(A).

de Cheveigné, A. (**2005**). “Separable representations for cocktail party processing,” Forum Acusticum (FA2005), Budapest, pp. 1527–1531.

de Cheveigné, A., McAdams, S., Laroche, J., and Rosenberg, M. (**1995**). “Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement,” J. Acoust. Soc. Am. **97**, 3736–3748.

Demany, L., and Semal, C. (**1990**). “The effect of vibrato on the recognition of masked vowels,” Percept. Psychophys. **48**, 436–444.

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

Gockel, H., Carlyon, R. P., and Plack, C. J. (**2004**). “Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms,” J. Acoust. Soc. Am. **116**, 1092–1104.

Gockel, H., Moore, B. C., and Patterson, R. D. (**2002**). “Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression,” J. Acoust. Soc. Am. **111**, 2759–2770.

Gockel, H., Moore, B. C., and Patterson, R. D. (**2003**). “Asymmetry of masking between complex tones and noise: Partial loudness,” J. Acoust. Soc. Am. **114**, 349–360.

Hafter, E. R., and Saberi, K. (**2001**). “A level of stimulus representation model for auditory detection and attention,” J. Acoust. Soc. Am. **110**, 1489–1497.

Hellman, R. P. (**1972**). “Asymmetry of masking between noise and tone,” Percept. Psychophys. **11**, 241–246.

Hoekstra, A. (**1979**). “Frequency discrimination and frequency analysis in hearing,” Ph.D. thesis, Institute of Audiology, University Hospital, Groningen, Netherlands.

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.

Levitt, H. (**1971**). “Transformed up-down methods in psychoacoustics, “J. Acoust. Soc. Am. **49**, 467–477.

Meddis, R., and Hewitt, M. (**1992**). “Modeling the identification of concurrent vowels with different fundamental frequencies,” J. Acoust. Soc. Am. **91**, 233–245.

Moore, B. C., Alcántara, J. I., and Dau, T. (**1998**). “Masking patterns for sinusoidal and narrow-band noise maskers,” J. Acoust. Soc. Am. **104**, 1023–1038.

Parsons, T. W. (**1976**). “Separation of speech from interfering speech by means of harmonic selection,” J. Acoust. Soc. Am. **60**, 911–918.

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

Qin, M. K., and Oxenham, A. J. (**2005**). “Effects of envelope-vocoder processing on *F*0 discrimination and concurrent-vowel identification,” Ear Hear. **26**, 451–460.

Scheffers, M. T. M. (**1983**). “Sifting vowels: Auditory pitch analysis and sound segregation,” Ph.D. thesis, Groningen University, The Netherlands.

Stubbs, R. J., and Summerfield, Q. (**1988**). “Evaluation of two voice-separation algorithms using normally-hearing and hearing impaired listeners,” J. Acoust. Soc. Am. **84**, 1236–1249.

Summerfield, Q., and Assmann, P. F. (**1991**). “Perception of concurrent vowels: Effects of pitch-pulse asynchrony and harmonic misalignment,” J. Acoust. Soc. Am. **89**, 1364–1377.

Treurniet, W. C., and Boucher, D. R. (**2001a**). “Masked threshold difference due to masker harmonicity: Uniform modulation rates of auditory filter outputs versus periodicity in waveform fine structure,” J. Acoust. Soc. Am. **110**, 1267–1270.

Treurniet, W. C., and Boucher, D. R. (**2001b**). “A masking level difference due to harmonicity,” J. Acoust. Soc. Am. **109**, 306–320.

van den Brink, W. A. C., and Houtgast, T. (**1990**). “Spectro-temporal integration in signal detection,” J. Acoust. Soc. Am. **88**, 1703–1711.

Weintraub, M. (**1987**). “Sound separation and auditory perceptual organisation,” in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Martinus Nijhoff, Dordrecht).

Zwicker, U. T. (**1984**). “Auditory recognition of diotic and dichotic vowel pairs,” Speech Commun. **3**, 265–277.