Virtual pitch integration for asynchronous harmonics

John H. Grose,[a] Joseph W. Hall III, and Emily Buss
Department of Otolaryngology/Head & Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7070

(Received 17 January 2002; revised 3 August 2002; accepted 23 August 2002)

This experiment examined the generation of virtual pitch for harmonically related tones that do not overlap in time. The interval between successive tones was systematically varied in order to gauge the integration period for virtual pitch. A pitch discrimination task was employed, and both harmonic and nonharmonic tone series were tested. The results confirmed that a virtual pitch can be generated by a series of brief, harmonically related tones that are separated in time. Robust virtual pitch information can be derived for intervals between successive 40-ms tones of up to about 45 ms, consistent with a minimum estimate of integration period of about 210 ms. Beyond intertone intervals of 45 ms, performance becomes more variable and approaches an upper limit where discrimination of tone sequences can be undertaken on the basis of the individual frequency components. The individual differences observed in this experiment suggest that the ability to derive a salient virtual pitch varies across listeners.
[DOI: 10.1121/1.1514934]

PACS numbers: 43.66.Hg, 43.66.Mk [NFV]

I. INTRODUCTION

Several lines of evidence point to the existence of an integration window for virtual pitch extraction. First, the virtual pitch extracted from harmonic complexes can be influenced by surrounding, temporally nonoverlapping, complexes (Carlyon, 1996; Micheyl and Carlyon, 1998). A parsimonious account for this finding is that an integration window centered on the signal complex includes information from the surrounding fringe complexes, and this ''overintegration'' affects the virtual pitch estimation of the signal. Second, the virtual pitch generated by a given harmonic complex can be influenced by a single tone that does not overlap the harmonic complex in time (Ciocca and Darwin, 1999). When a resolvable harmonic is missing from the harmonic complex, a separate tone can affect the pitch of the complex as long as the frequency of the tone differs by less than about 8% from that of the missing harmonic. Ciocca and Darwin showed that this influence could extend over silent intervals between the two 90-ms stimuli (tone and complex) of up to 160 ms, but failed to occur if the duration of the synchronous harmonic complex was sufficiently long (e.g., 410 ms). They interpreted this finding in terms of a pitch integration period having a duration of about 200–300 ms. A third line of evidence comes from parametric studies of fundamental frequency (f_0) discrimination that manipulate the number, duration, and frequency content of harmonic complexes. The duration of an integration window for pitch appears to depend on the harmonic content of the stimulus, being longer for unresolved harmonics than for resolved harmonics (White and Plack, 1998; Plack and White, 2001). In addition, the pitch integrator can be ''reset'' by level discontinuities in the stimuli, such as temporal gaps (Plack and White, 2000; Plack and White, 2001). A final line of evidence for a pitch integration window is the finding that virtual pitch can be extracted from a set of harmonics even when those harmonics do not overlap in time. Hall and Peters (1981) demonstrated that a sequence of three successive low-order harmonics, each 40 ms in duration and separated from each other by 10 ms, could generate a virtual pitch provided they were presented against a broadband noise background. A virtual pitch was not generated when the harmonic sequence was presented in quiet. Using a similar timing sequence, Houtsma (1984) found that some listeners could extract a virtual pitch from nonsynchronous harmonics even in quiet. However, he also found that the majority of listeners in his study could not reliably express melodic intervals using virtual pitches extracted from nonoverlapping harmonics.

The primary purpose of this study was to test the notion of a pitch integration period using a procedure similar to that used by Hall and Peters (1981). The approach was to present sequences of brief tones that either did or did not form low-order harmonic sequences, and to systematically vary the interval between successive tones. An f_0 discrimination task was used to determine whether a virtual pitch was generated by the tone sequences. Such a task can probe pitch in terms of differential scaling on a high/low dimension, but it cannot assess pitch in the musical sense of recognizing specific frequency ratios. The hypothesis to be tested here is that virtual pitch discrimination is feasible only for harmonically related tones whose intertone intervals (ITIs) allow them to fall within the pitch integration period. Once the ITI exceeds this criterion, no virtual pitch is perceived and the task can no longer be performed on the basis of virtual pitch. By the same token, a discrimination task based on the perception of virtual pitch is also not feasible for tone sequences that are not harmonically related.

[a]Author to whom correspondence should be addressed. Electronic mail: jhg@med.unc.edu

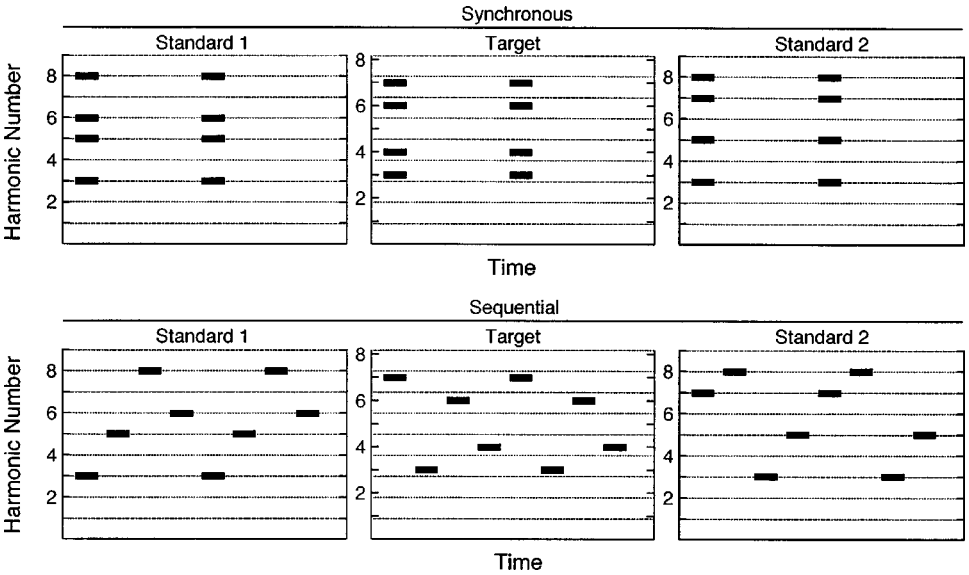

FIG. 1. Stimulus schematic showing a typical 3AFC trial for synchronous presentation of harmonic tones (upper panel) and sequential presentation of harmonic tones (lower panel). Within a trial, selection of the four tones is random for each observation interval, allowing for different tone complexes in the two standard intervals. Within an interval, the four tones are presented twice. Horizontal dashed lines across intervals are isofrequency lines for reference.

II. METHOD

A. Subjects

Five listeners with normal hearing participated in this experiment. They ranged in age from 19 to 42 years (mean $=25$ years), and all had audiometric thresholds less than 20 dB HL across the octave frequencies 250–8000 Hz (ANSI, 1996). None of the listeners had extensive musical training, although informal questioning indicated that at least three had some degree of musical experience: observers 1 and 5 played instruments, and observer 4 sang in a choir. All received a regimen of familiarization and training with the task as described below.

B. Stimuli

The stimuli for the virtual pitch discrimination task were four-component tone complexes. In the main experiment using harmonic complexes, each component was 40 ms in duration, including a 10-ms $\cos^2$ rise/fall ramp, and the individual components were presented either synchronously or sequentially. Figure 1 shows a schematic of the stimuli in each of the observation intervals in a typical three-alternative forced-choice (3AFC) trial. For both synchronous and sequential configurations, two replications of the complete stimulus were presented within each observation interval, resulting in either two complex tone bursts or a stream of eight individual tone bursts per interval. For each observation interval, the four components were selected randomly from among the third to tenth harmonics of a given f_0. Across trials, the nominal f_0 of the standard was 200 Hz, and the nominal f_0 of the signal was varied adaptively relative to this standard by a factor of 1.26. A virtual pitch region of 200 Hz was selected because this region has been extensively studied (e.g., Plomp, 1976). The direction of f_0 shift of the signal was always to a higher frequency. However, from trial to trial, the actual f_0 of the standard stimuli varied randomly by $\pm 20\%$ of 200 Hz, and the actual f_0 of the signal maintained proportionality to this actual standard f_0 within a trial. The randomization of actual f_0 from trial to trial, coupled with the random selection of harmonics from interval to interval within a trial, was implemented to undermine the cue effectiveness of monitoring individual components within the harmonic complexes to perform the task. Nevertheless, as the frequency separation between the standard and signal f_0's increased, the probability of the signal complex containing higher frequency components than the standard complex increased. Consequently, the task could well be undertaken by monitoring just the higher frequency components. To test for this limit, stimulus configurations were also constructed using logarithmically spaced tones. No virtual pitch cue would be expected for these stimuli, so discrimination performance was assumed to reflect the effectiveness of monitoring individual component tones. In these configurations, stimulus components were randomly drawn in each observation interval from a series of eight tones that were equally spaced on a log scale between the nominal frequencies of 333 and 1995 Hz. Again, from trial to trial the actual frequencies in this series varied by $\pm 20\%$.

The main independent variable was the intertone interval (ITI) between successive components when presented sequentially. This ITI was defined as the interval between the zero-voltage point at the offset of one component and the zero-voltage point at the onset of the next component. The ITIs ranged in 15-ms steps from 0 to 90 ms. The choice of ITI determined the overall duration of the eight-tone sequence in each observation interval of a 3AFC trial (Fig. 1, panel B). To maintain a complementary temporal pattern for the synchronous conditions (Fig. 1, panel A), the interval between the two complex tone bursts also varied systematically with the choice of ITI. The relation between the repetition interval for the complex tone bursts in the synchronous conditions and the ITI in the sequential conditions was $4(40+\mathrm{ITI})$. Thus, for example, an ITI of 15 ms between successive tones in the sequential condition corresponded to a repetition interval of 220 ms for the complex tone bursts in the complementary synchronous condition.

Although most of the test stimuli were composed of 40-ms tones that were presented twice within each observation interval, two other longer-duration stimulus configurations were also included. Both of these were comprised of

320-ms synchronous complex tones which were presented only once in each observation interval. In one configuration the tones were harmonically related, and in the other the tones were logarithmically spaced. The rules for random draws within an observation interval and jitter of root frequency across trials were the same as those described for the synchronous conditions above.

Each component within the four-component complex was presented at a level of 60 dB SPL. This level was chosen because pilot listening found it to be a comfortable level that generated a clear virtual pitch. The complexes were generated digitally from trial to trial at a sampling rate of 10 kHz (TDT AP2), output via a 16-bit DAC (TDT PD1), antialias filtered at 4000 Hz (Kemo VBF8), attenuated (TDT PA4), and presented monaurally to the left ear using a Sennheiser HD580 headphone. A continuous broadband background noise (0–8000 Hz) was always present at an overall level of 52 dB SPL.

C. Procedure

Prior to the experiment, each listener undertook a familiarization/training regimen. In the first phase of this regimen, the listener was presented with 320-ms harmonic complexes that alternated between harmonics 3–12 of 200 Hz and harmonics 3–12 of 225 Hz. The purpose of this phase was to familiarize the listener with virtual pitch using a perceptually salient example. The alternating sequence was discontinued by the listener once s/he felt comfortable with the perception of an alternating pitch. The second phase of the regimen was similar except that the complex with the virtual pitch of 200 Hz was comprised of harmonics 4–13, whereas the complex with the virtual pitch of 225 Hz was comprised of harmonics 2–11. Here, the complex with the lower virtual pitch contained the highest absolute frequency component. The final phase of the regimen again consisted of alternating 200-Hz and 225-Hz virtual pitch complexes, but now the 200-Hz complex was comprised of harmonics 7–10, whereas the 225-Hz complex was comprised of harmonics 3–6. In this configuration, the entire frequency content of the complex with the higher virtual pitch was lower than the frequency content of the complex with the lower virtual pitch. This three-phase regimen familiarized the listener with virtual pitch using complexes that increasingly diverged in their frequency content and number of components. None of the listeners experienced difficulty extracting an alternating virtual pitch in the three training phases, and all completed the training within 15–30 min.

In the experiment proper, an adaptive stepping rule was incorporated into the 3AFC procedure to converge on the 79.4% correct point of the psychometric function. The signal, consisting of the upward-shifted frequency set, occurred in one of the three observation intervals at random. Following three correct responses in a row, the frequency separation between the standard and signal f_0's was reduced by a factor of 1.26; following one incorrect response, the frequency separation was increased by the same factor. A threshold track was terminated after 12 reversals in direction of frequency separation, and the geometric mean of the frequency differences at the final eight reversal points was taken as the estimate of threshold for that track. For each condition, at least four threshold estimates were collected and the geometric mean of the estimates was taken as threshold for that condition.

All listeners began with the synchronous 320-ms harmonic stimulus condition. This was the condition anticipated to provide the most salient perception of virtual pitch, and the condition was repeated until systematic improvements in threshold no longer occurred. At this point data collection began and at least four further replications were undertaken. Following this condition, all listeners then received the conditions incorporating synchronous 40-ms harmonic tone bursts. Recall that in these conditions, each observation interval contained two tone bursts (see Fig. 1) which were separated in time by an interval that corresponded to the sequence repetition time associated with the complementary sequential presentation of the individual components. All listeners were tested first in the synchronous condition that had a repetition interval of 160 ms (≅0-ms ITI). Each listener was also tested in a synchronous condition corresponding to at least one other ITI; the selection of this condition varied from listener to listener. Conditions involving sequences of harmonically related tone bursts were tested next. The order of ITIs was not random; rather, each listener began with the 0-ms ITI condition and progressed on to successively longer ITIs until performance declined markedly. The final set of three conditions involved the logarithmically spaced tones. These included the 40-ms tones presented sequentially at an ITI of 0 ms, the complementary synchronous presentation of the 40-ms tones, and the synchronous 320-ms tones.

III. RESULTS

The results of the experiment are displayed in Fig. 2. Because there was some individual variability in performance, the individual results are displayed separately in the five panels. Each panel plots pitch discrimination against a measure of temporal interval appropriate for given conditions (repetition interval for the synchronous conditions, ITI for the sequential conditions). Points to the left of the abscissa break refer to the conditions involving 320-ms tones where a measure of temporal interval is not applicable. Symbols in each graph indicate stimulus type as defined in the figure legend. For each listener, the boundaries of performance are usefully described by the thresholds for the 320-ms tone conditions. For 320-ms harmonic complexes (open circles), virtual pitch discrimination thresholds ranged from about 2 Hz to about 11 Hz across listeners. Because it was assumed that the 320-ms harmonic complexes would provide the listeners with the best opportunity to derive a virtual pitch, these thresholds represent the lower limit of performance for the task. For log-spaced complexes (solid circles), discrimination thresholds ranged between about 65 Hz to about 173 Hz across listeners. The inharmonic log-spaced tones were unlikely to generate a virtual pitch and so these thresholds represent the upper bound of performance wherein the discrimination task is reduced to identifying the interval containing the highest individual frequency(ies). In each panel, horizontal dashed lines extend from these reference thresholds across the remaining conditions.

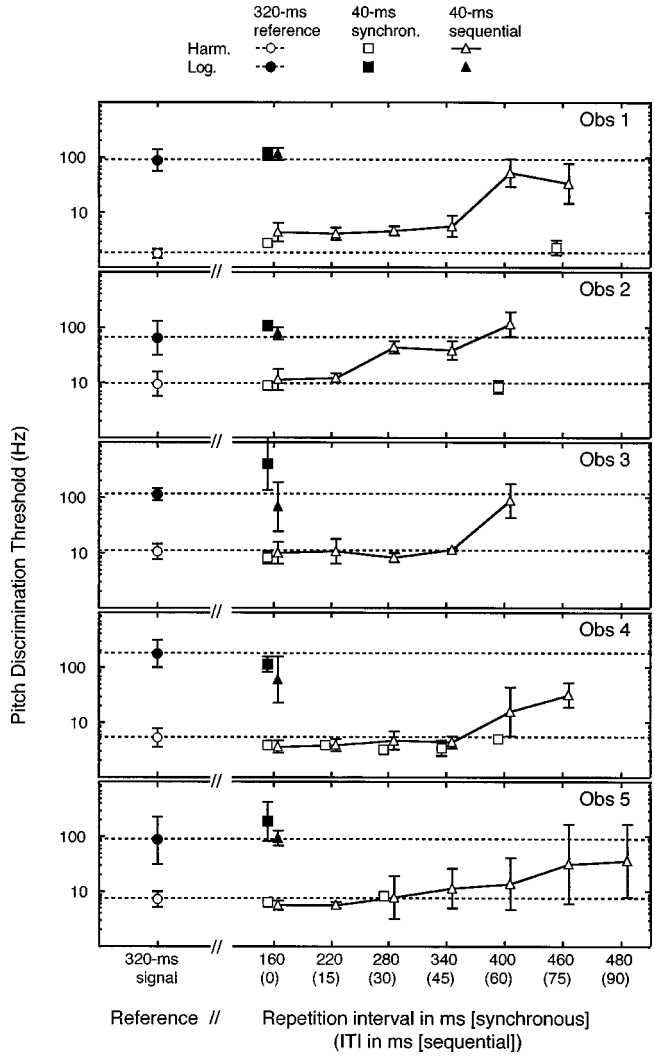

FIG. 2. Individual results plotting discrimination threshold against a measure of temporal interval. For synchronous conditions (squares) this was repetition interval; for sequential conditions (triangles) this was ITI. Each panel shows the mean results of one listener. Error bars indicate ± 1 standard deviation computed in the log-transform domain. Open circle: 320-ms synchronous harmonic complex; Solid circle: 320-ms synchronous log-spaced complex; Open squares: 40-ms synchronous harmonic complexes; Solid square: 40-ms synchronous log-spaced complex; Open triangles: 40-ms sequential harmonic tones; Solid triangle: 40-ms sequential log-spaced tones. Horizontal dashed lines in each panel are reference lines associated with the 320-ms stimulus thresholds.

The performance bounds set by the long-duration stimulus configurations provide a context in which to assess the results of the 40-ms tone burst conditions. When the brief tone bursts were logarithmically spaced, performance remained poor whether the tones were presented synchronously (solid squares) or sequentially (solid triangles). A repeated-measures analysis of variance (ANOVA) on the log-transformed thresholds from all three logarithmically spaced conditions indicated no significant difference between conditions ($F_{2,8}=3.3$; $p=0.09$). This is consistent with the assumption that a monitoring of absolute frequency content rather than a derived virtual pitch underlies performance for logarithmically spaced tones. The grand average discrimination threshold across all listeners for these three conditions was about 108 Hz.

For the 40-ms synchronous harmonic tone bursts (open squares), performance was generally on a par with the synchronous 320-ms duration harmonic stimulus (open circle) independent of the overall repetition interval of the tone burst complexes within each observation interval. It can be seen from Fig. 2 that the number of synchronous tone burst conditions and the choice of tone burst repetition intervals varied across listeners, although all received the condition having a repetition interval of 160 ms ($\cong$0-ms ITI). Nevertheless, for each listener where multiple conditions were run, thresholds remained relatively constant. To assess this apparent equivalence, log-transformed thresholds for three synchronous harmonic conditions were submitted to a repeated-measures ANOVA. These three conditions were: (1) the 320-ms condition; (2) the 40-ms condition having a repetition interval of 160 ms ($\cong$0-ms ITI); and (3) the remaining 40-ms synchronous condition yielding the *highest* threshold for an individual. (For Obs. 3, no remaining ITI conditions were tested; therefore, a univariate analysis was undertaken to accommodate the missing data point.) The analysis indicated that these thresholds did not differ across conditions ($F_{2,7}=0.24$; $p=0.79$). This suggests that the virtual pitch derived from a single 40-ms tonal complex was sufficiently salient to perform the virtual pitch discrimination task, and that increasing the duration of the stimulus complex to 320 ms, or providing a repetition of the 40-ms complex within an observation interval added no further benefit. This suggestion is reminiscent of White and Plack (1998), who found little change in d' for f_0 discrimination for resolved harmonic complexes once the signal duration exceeded 40 ms. Furthermore, they found little effect of gap duration between two brief harmonic complexes on f_0 discrimination.

The primary interest of this experiment was in the conditions where the 40-ms harmonically related tone bursts were presented sequentially, shown as triangles in Fig. 2. For all listeners, discrimination performance was at or near the lower limits for ITIs close to 0 ms, but approached the upper bounds of performance for longer ITIs. It is apparent that the ITI at which performance began to depart from the lower limit differed across listeners, although performance had clearly declined in most listeners by an ITI of 60 ms. Moreover, for some listeners the departure from baseline performance was relatively abrupt (e.g., Obs. 1), whereas for others it was more gradual (e.g., Obs. 5). To provide an efficient test of this data pattern, log-transformed thresholds from three conditions were submitted to a repeated-measures ANOVA. The three conditions were: (1) synchronous harmonic complexes corresponding to an ITI of 0 ms (repetition interval $=160$ ms); (2) sequences of harmonically related tones having an ITI of 45 ms; and (3) sequences of harmonically related tones having an ITI of 60 ms. Comparison of (1) and (2) would indicate whether performance with an ITI of 45 ms was still as good as for simultaneously presented components. Comparison of (2) and (3) would indicate whether performance had, on average, declined at an ITI of 60 ms. The results of the analysis indicated that there was a significant effect of stimulus condition across these three conditions ($F_{2,8}=18.315$; $p=0.001$). The first of the two planned comparisons described above indicated that thresholds for

the tone sequence with an ITI of 45 ms did not differ reliably from thresholds for the synchronous harmonic complex corresponding to an ITI of 0 ms ($p=0.09$). This conservative comparison indicates that pitch discrimination for harmonic tone sequences presented with ITIs of 45 ms was still at the lower bounds of performance, suggesting that the task was still being done on the basis of virtual pitch. The second of the two planned comparisons indicated that thresholds for the tone sequence with an ITI of 60 ms were significantly different from thresholds for an ITI of 45 ms ($p=0.004$). This finding indicates that, on average, the breakpoint in performance between baseline (good) performance and a significant decline in performance occurred for intertone intervals between 45 and 60 ms.

IV. DISCUSSION

The purpose of this experiment was to test the notion of an integration period for virtual pitch using a paradigm of sequentially presented harmonic tones. A parsimonious interpretation of the results is that a virtual pitch can be generated by sequences of 40-ms tones presented against a noise background for intervals between the tones of up to about 45 ms. For ITIs longer than about 60 ms, performance declines and becomes more variable, suggesting that a reliable virtual pitch is no longer being generated. The relation between these ITIs and an integration period for virtual pitch is not entirely straightforward. Because of the random selection of harmonic numbers from interval to interval within a 3AFC trial and the random order of presentation of these harmonics, the generation of a virtual pitch from just two successive components in the tone sequence would be an unreliable cue because the pitch could vary randomly across even the standard intervals. That is, two successive tones from the random sequence presented in standard interval 1 might be compatible with one virtual pitch (e.g., 600 and 800 Hz generating an f_0 of 200 Hz), whereas two successive tones from the random sequence presented in standard interval 2 might be compatible with another virtual pitch (e.g., 800 and 1600 Hz generating an f_0 of 800 Hz). Given this, it seems reasonable that any pitch integration period must cover at least three of the tones in the sequence.[1] Thus, a virtual pitch integration period must be at least 210 ms long (three 40-ms tones separated by ITIs of 45 ms). By the same reasoning, an interval of 240 ms would extend beyond the integration period (three 40-ms tones separated by ITIs of 60 ms). Integration periods in the region of 200 ms are compatible with the results of Hall and Peters (1981), whose three-tone sequences covered at interval of 140 ms, as well as those of Ciocca and Darwin (1999), whose data suggest that the pitch integration period continues for 170–250 ms following the onset of a complex tone.

Although the data of this experiment are discussed in the context of an integration period for virtual pitch, one trend in the data suggests that an alternative interpretation might deserve consideration. It is evident from Fig. 2 that for at least two of the listeners (Obs. 1 and 5) the deterioration in performance appeared to plateau below the upper bound indicated by thresholds for the logarithmically spaced tones. It was argued that this upper bound represented the point at which the task was undertaken on the basis of discrimination of individual frequencies rather than discrimination of virtual pitch. A plateau in performance below this point might suggest that some (nonsalient) virtual pitch information can be gleaned from the tone sequences, even for relatively long ITIs. One strategy with which this might be accomplished is to view the tone sequences as arpeggios where the listener internally constructs the f_0 that is the best fit to a given arpeggio. The discrimination task therefore becomes one of identifying the observation interval where the tone sequence represents an arpeggio of an f_0 that is different from the other two observation intervals. Here, the limitation in performance is the duration over which the harmonic relation between successive tones can be remembered accurately.

The ability to integrate frequency components that do not overlap in time into a single percept, in this instance virtual pitch, is concordant with other manifestations of perceptual integration. In particular, the perception of multifrequency-channel speech exhibits analogous integrative features. Greenberg and Arai (1998) have shown that the perception of speech that is filtered into multiple narrow bands can tolerate misalignments of up to 140 ms across bands. If temporal misalignments between filtered speech bands can be viewed as a special case of temporally nonoverlapping frequencies, then the resilience of speech perception to this manipulation is congruent with a process of spectral integration over time. Our own work on amplitude modulation of multiple narrow bands of speech has shown that speech perception is unaffected by the relative phase of 10-Hz modulation across bands (Buss *et al.*, 2001). In other words, speech perception is not sensitive to whether adjacent bands are modulated in phase or 2π radians out of phase. The latter case can be thought of as yielding frequency regions whose energy contents do not overlap in time.

In summary, this experiment has confirmed that a virtual pitch can be generated by a series of brief harmonically related tones that are separated over time. Robust virtual pitch information can be derived for intervals between successive 40-ms tones of up to about 45 ms, consistent with a minimum estimate of integration period of about 210 ms. Beyond ITIs of 45 ms, performance becomes more variable and approaches the upper limit of performance where discrimination of tone sequences can be undertaken on the basis of the individual frequency components. The individual differences observed in this experiment suggest that the ability to derive a salient virtual pitch varies across listeners.

ACKNOWLEDGMENTS

We thank Drs. Adrianus Houtsma and Christophe Micheyl for their helpful comments on an earlier version of this paper. This work was supported by NIH NIDCD (5 R01 DC00418-13).

[1]A simulation incorporating the random sequencing of harmonic tones used in this experiment indicated that a three-tone sequence compatible with a virtual pitch of only 200 Hz (in contrast to, for example, 200 or 600 Hz) occurred about 90% of the time.

ANSI (**1996**). ANSI S3.6-1996, ''American National Standard Specification for Audiometers'' (American National Standards Institute, New York).

Buss, E., Hall, J. W., III, and Grose, J. H. (**2002**). ''Effect of masker amplitude modulation modulation coherence for speech signals filtered into narrow bands,'' J. Acoust. Soc. Am. (in press).

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

Ciocca, V., and Darwin, C. J. (**1999**). ''The integration of nonsimultaneous frequency components into a single virtual pitch,'' J. Acoust. Soc. Am. **105**, 2421–2430.

Greenberg, S., and Arai, T. (**1998**). ''Speech intelligibility is highly tolerant of cross-channel spectral asynchrony,'' J. Acoust. Soc. Am **103** (Pt. 2), 3057.

Hall, J. W., and Peters, R. W. (**1981**). ''Pitch for nonsimultaneous successive harmonics in quiet and noise,'' J. Acoust. Soc. Am. **69**, 509–513.

Houtsma, A. J. M. (**1984**). ''Pitch salience of various complex sounds,'' Music Percept. **1**, 296–307.

Micheyl, C., and Carlyon, R. P. (**1998**). ''Effects of temporal fringes on fundamental-frequency discrimination,'' J. Acoust. Soc. Am. **104**, 3006–3018.

Plack, C. J., and White, L. J. (**2000**). ''Perceived continuity and pitch perception,'' J. Acoust. Soc. Am. **108**, 1162–1169.

Plack, C. J., and White, L. J. (**2001**). ''Temporal integration in pitch perception,'' in *Proceedings of the 12th International Symposium on Hearing. Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Mierlo, The Netherlands).

Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).

White, L. J., and Plack, C. J. (**1998**). ''Temporal processing of the pitch of complex tones,'' J. Acoust. Soc. Am. **103**, 2051–2063.