The integration of nonsimultaneous frequency components into a single virtual pitch

V. Ciocca[a)]
Department of Speech and Hearing Sciences, University of Hong Kong, Hong Kong

C. J. Darwin
Experimental Psychology, University of Sussex, Brighton BN1 9QG, United Kingdom

The integration of nonsimultaneous frequency components into a single virtual pitch was investigated by using a pitch matching task in which a mistuned 4th harmonic (mistuned component) produced pitch shifts in a harmonic series (12 equal-amplitude harmonics of a 155-Hz $F0$). In experiment 1, the mistuned component could either be simultaneous, stop as the target started (pre-target component), or start as the target stopped (post-target component). Pitch shifts produced by the pre-target components were significantly smaller than those obtained with simultaneous components; in the post-target condition, the size of pitch shifts did not decrease relative to the simultaneous condition. In experiment 2, a silent gap of 20, 40, 80, or 160 ms was introduced between the nonsimultaneous components and the target sound. In the pre-target condition, pitch shifts were reduced to zero for silent gaps of 80 ms or longer; by contrast, a gap of 160 ms was required to eliminate pitch shifts in the post-target condition. The third experiment tested the hypothesis that, when post-target components were presented, the processing of the pitch of the target tone started at the onset of the target, and ended at the gap duration at which pitch shifts decreased to zero. This hypothesis was confirmed by the finding that pitch shifts could not be observed when the target tone had a duration of 410 ms. Taken together, the results of these experiments show that nonsimultaneous components that occur after the onset of the target sound make a larger contribution to the virtual pitch of the target, and over a longer period, than components that precede the onset of the target sound. © *1999 Acoustical Society of America.* [S0001-4966(99)02604-1]

PACS numbers: 43.66.Ba, 43.66.Hg, 43.66.Mk [JWH]

INTRODUCTION

Several studies have demonstrated that virtual (or residue) pitch is mainly determined by the frequency of the low-numbered, resolved partials of complex tones (Moore *et al.*, 1985; Plomp, 1967; Ritsma, 1967). The results of these studies have led to the formulation of pitch perception theories which state that virtual pitch is calculated mainly on the basis of the frequencies of the resolved, low-frequency components (see, e.g., Goldstein, 1973; Terhardt *et al.*, 1982a,b). Most studies on the perception of the virtual pitch of complex sounds have employed frequency components that had simultaneous onsets and offsets (for recent reviews, see Houtsma, 1995; Hartmann, 1996). By contrast, the partials of many natural sounds start and stop at different times and do not have the same patterns of amplitude changes over time (see, e.g., Risset and Wessel, 1982). Since changes in the spectra of natural sounds may occur within periods of a few tens of milliseconds, it is likely that pitch perception processes calculate virtual pitch by integrating frequency information over time rather than by taking a snapshot of the frequency content of a complex tone at one particular moment in time.

Two types of experiment have demonstrated that pitch perception processes are sensitive to the context in which a complex sound is presented: (i) Darwin and Ciocca (1992) showed that the pitch shifts produced by mistuning the 4th harmonic (mistuned component) of a harmonic series were reduced if the mistuned component started 80 ms or more before the other harmonics; (ii) Darwin *et al.* (1995) demonstrated that a reduction in the contribution of the mistuned component to the pitch of a complex (target) sound could be obtained when a sequence of four tones, which were identical to the mistuned component, preceded the target. The effects of onset asynchrony and sequential grouping investigated by these studies showed that preceding context can be used to exclude frequency components from the calculation of virtual pitch. The current paper asks whether the virtual pitch of a harmonic series (target tone) can be affected by frequency components that stop before its onset, or start after its offset.

Hall and Peters (1981) demonstrated the temporally integrative nature of pitch perception by performing experiments on the pitch of a sequence of three 40-ms sinusoids, separated by 10-ms gaps of silence; the total duration of each sequence was 140 ms. The frequency of each sinusoid corresponded to either the 3rd, 4th, and 5th harmonic of a 200-Hz fundamental frequency (sequence 200), or the 4th, 5th, and 6th harmonic of a 180-Hz $F0$ (sequence 180). The order of the harmonics within each sequence was always from the lowest to the highest numbered. The pitch of these sequences was measured by using a discrimination task and a

a)Electronic mail: vciocca@hkusua.hku.hk

pitch matching task, and by presenting the sequences either in quiet (high S/N ratio) or in the presence of a background noise (low S/N ratio). The latter condition was included because Houtsma and Goldstein (1971; cited in Hall and Peters, 1981) had previously reported that two nonsimultaneous harmonics presented in quiet did not give rise to the perception of a single virtual pitch. Hall and Peters found that subjects heard the sequences as having a single virtual pitch, corresponding to the fundamental frequency of the partials, in the low S/N condition. In quiet, subjects heard a spectral pitch based on the frequency of the lowest and first (in order of presentation) harmonic of each sequence. These results confirmed Hall and Peter's hypothesis that pitch mechanisms must integrate frequency information over time in order to compute virtual pitch. Moreover, these findings suggest that the period over which information about a single virtual pitch is integrated (called ''pitch integration period,'' hereafter) could be as long as 140 ms.

The present study extended Hall and Peters' findings by using the pitch matching paradigm developed by Moore *et al.* (1985), and employed in more recent studies (Darwin and Ciocca, 1992; Ciocca and Darwin, 1993; Darwin *et al.*, 1994). Darwin and his colleagues measured the pitch shifts of a harmonic series (target sound) produced by mistuning the 4th harmonic (mistuned component) by various amounts. This procedure gives an estimate of the contribution of a single component to the pitch of a complex sound. The main goal of the current paper was to explore temporal order effects by measuring the pitch shifts produced by a nonsimultaneous mistuned component which could either precede or follow the presentation of the target. An additional purpose of the study was to investigate whether pitch shifts occur when a silent gap of variable duration is introduced between the target sound and a nonsimultaneous component, and whether pitch shifts occur at different silent gap durations for preceding and following components. Although a study by Rakowski and Hirsh (1980) measured the effects of the frequency of a sine wave (leading tone) on the pitch of another (target) sinusoid as a function of the duration of a silent interval which separated the two tones, to the best of our knowledge the effects of a nonsimultaneous frequency component on the virtual pitch of a complex sound have not been previously investigated.

I. EXPERIMENT 1

In this experiment, pitch shifts produced by a mistuned component which was simultaneous with the target were compared with those obtained with two types of nonsimultaneous mistuned components: (i) mistuned components which preceded the target [pre-target components; see Fig. 1(a)], and (ii) mistuned components which started as the target stopped [post-target components; see Fig. 1(b)]. There was no temporal overlap between the target and the nonsimultaneous mistuned components.

A. Method

On each trial, the subjects adjusted the pitch of a matching harmonic series to the pitch of a target complex. The
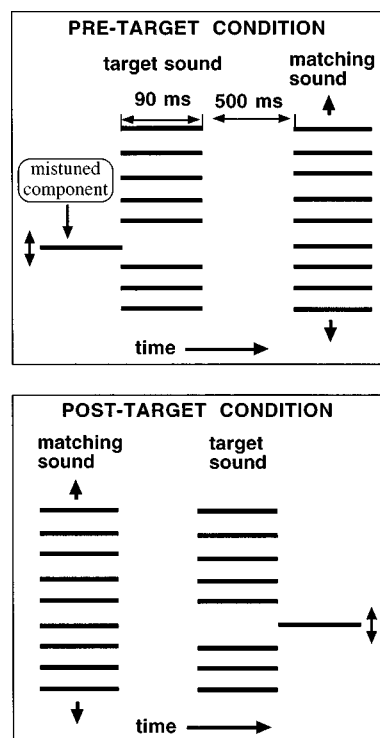

FIG. 1. Stimulus configuration for the pre-target and post-target conditions of experiment 1.

target complex contained harmonics 1–3 and 5–12 of 155 Hz, together with a possibly mistuned 4th harmonic (mistuned component). The mistuned component had a frequency of either 570, 600, 610, 620, 630, 640, or 670 Hz. These frequencies corresponded to mistunings of −8.0%, −3.2%, −1.6%, 0, +1.6%, +3.2%, and +8.0%, respectively. The matching sound consisted of the first 12 harmonics of a fundamental frequency which subjects could adjust between 151 and 159 Hz (155±4 Hz). The mistuned component, the target tone, and the matching tone had a total duration of 90 ms, including 5-ms rise/fall raised-cosine ramps. The duration of the stimuli was the same as the duration used in previous pitch matching experiments (Darwin and Ciocca, 1992; Ciocca and Darwin, 1993), and was selected because it was found to give reliable pitch shift estimates while being short enough for the purpose of the current study (see also experiment 3, Sec. III).

Different matching tasks, administered in separate sessions, were used to measure pitch shifts for pre- and post-target components. In the pre-target task, the onset-to-onset time between the mistuned component and the target was either 0 (simultaneous components) or 90 ms [pre-target components; see Fig. 1(a)], and the matching tone followed the target sound. In the post-target task, mistuned components could either start at the same time as the target (simultaneous components) or start 90 ms after the onset of the target [i.e., they target as the target stopped; see Fig. 1(b)]; the order of presentation of the target and matching tones was the opposite to that of the pre-target condition (i.e., the matching tone always preceded the target tone). The change in the order of the stimuli for performing the matching task with post-target components was introduced to ensure that

the mistuned component affected only the pitch of the target tone. If the same matching task (i.e., if the matching tone followed the target tone) had been used to measure pitch shifts for post-target components, the latter would have intervened between the target and the matching tones. In this case, mistuned components might have affected the pitch of both the matching and the target tones, or could have acted as distractors, thereby making the matching task generally more difficult. Informal observations indicated that adjusting the pitch of the matching tone to match the pitch of a subsequent target tone proved to be slightly more difficult than matching the pitch of the tones in the standard order of presentation. However, the increase in the variability of pitch matches was not very large, and listeners learned to perform the post-target matching task as accurately as the pre-target task over the course of an experimental session (see Sec. I B). In both tasks, a 500-ms silent interval separated target and matching tones.

All components were presented to the left ear at the same amplitude (corresponding to 58 dB, sound-pressure level, for a 1000-Hz tone). All components started at sine (0) phase. Sounds were presented through Sennheiser HD414 headphones in a double-walled IAC booth. At the beginning of each trial, the fundamental frequency of the matching tone was chosen at random from the permitted range (155±4 Hz). Subjects adjusted the frequency of the matching sound by moving a roller-ball up or down (Darwin and Ciocca, 1992). Subjects were allowed to perform as many adjustments as they required in order to achieve a satisfactory match. Each target sound was matched five times in a quasi-random order within each experimental session; one session took 1–2 h to complete, including rest breaks. Within a trial, the mistuned components, especially a nonsimultaneous one, could often be heard out as an individual pure tone. Subjects were instructed to ignore the mistuned component and to focus their attention on matching the pitches of the two complex sounds.

To summarize, the experiment consisted of 21 conditions: 7 frequencies (570, 600, 610, 620, 630, 640, or 670 Hz) by 3 types of mistuned component (simultaneous, pre-target, or post-target). The two nonsimultaneous conditions were run in separate sessions since they required different matching tasks. In each session, the simultaneous and the nonsimultaneous conditions were repeated five times. Pre- and post-target sessions were run on separate days, and their order was counterbalanced across subjects such that half the subject participated in the pre-target session first, and vice versa.

Sounds were synthesized in real-time at 44.1 kHz using custom software (Russell and Darwin, 1991) written for the 56001 processor of the Digidesign Audiomedia II board, and output through that board's 16-bit DACs and anti-aliasing filters. The board was attached to an Apple MacIIcx computer which controlled the experiment.

Eight university students, six of whom were musically trained, participated in the experiment. All had participated in previous pitch matching experiments.
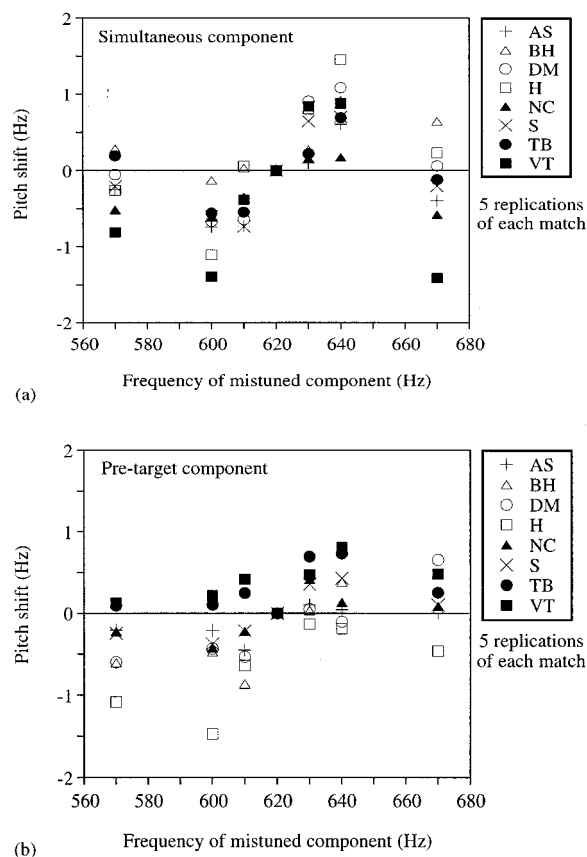

FIG. 2. Individual pitch shifts, averaged over five repetitions, for the pre-target task of experiment 1. Pitch shifts for the simultaneous components are shown in (a); shifts for the pre-target mistuned components are displayed in (b).

B. Results and discussion

The effect of the mistuning of the 4th harmonic on the pitch of the target tone was determined by measuring the difference between the pitch matches for each amount of mistuning relative to the condition in which the 4th harmonic was not mistuned. The negative and positive pitch shifts for simultaneous and pre-target components in the pre-target task are displayed in Fig. 2(a) and (b), respectively. Pitch shifts for individual listeners were calculated as the difference between the average of pitch matches for each −ve/+ve mistuning and the average of the pitch matches for the in-tune, 620-Hz condition (0-Hz point on the y-axis). The pattern of pitch shifts for the simultaneous components replicates the pattern obtained with similar stimuli in previous pitch matching experiments; pitch shifts were largest for mistunings of ±20 Hz, and were on average close to 0 Hz for mistunings of ±50 Hz (see, for example, Darwin and Ciocca, 1992). The pitch shifts for pre-target components were generally smaller than those produced by simultaneous components, but there were individual differences. Subjects showed a large reduction in the size of pitch shifts either for negative mistunings (subjects VT, NC, and TB), or for positive mistunings of pre-target components (subjects DM, BH, and H) relative to the simultaneous component condition. The remaining subjects (AS and S) showed a reduction of pre-target shifts for both negative and positive mistunings [see Fig. 2(b)].
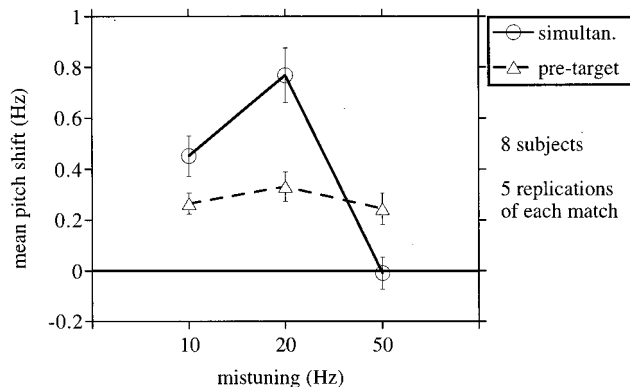
FIG. 3. Mean pitch shifts with error bars (standard errors of the mean) for the pre-target task of experiment 1. Mean shifts for the pre-target mistuned component are displayed as open triangles, and those for the simultaneous component as open circles.

The mean pitch shifts for the pre-target task, averaged across eight subjects, are displayed in Fig. 3. Mean pitch shifts were calculated as half the difference between matches for each amount of positive and negative mistuning for each subject (Moore *et al.*, 1985). A two-way ANOVA with repeated measures was applied to the individual mean pitch shifts obtained in the pre-target sessions. The factors were the ''mistuning'' of the mistuned component (10, 20, or 50 Hz), and the ''type of mistuned component'' (simultaneous or pre-target). The main effect of ''mistuning'' was statistically significant [$F(2,14) = 19.07$, $p < 0.0001$]. In the simultaneous condition, mean pitch shifts were largest at mistunings of 20 Hz (3.2% of the harmonic frequency), and were virtually absent at mistunings of 50 Hz. Shifts for both the 10- and 20-Hz mistunings were significantly higher than for mistunings of 50 Hz (Tukey HSD test, $p < 0.01$). The effects of mistuning with simultaneous components is consistent with previous findings which showed that pitch shifts are at a maximum for mistunings of about 3%, and then decrease to zero with mistunings of 8% (Darwin and Ciocca, 1992; Moore *et al.*, 1985). In the pre-target condition, only the 20-Hz mistuning produced shifts which were significantly different from those of the 50-Hz, simultaneous condition (Tukey HSD test, $p < 0.05$). Pitch shifts for the three pre-target mistunings did not differ from each other (Tukey HSD test, $p > 0.05$). The difference in the effect of mistuning between the simultaneous and the pre-target conditions resulted in a significant ''mistuning'' by ''type of mistuned component'' interaction [$F(2,14) = 14.44$, $p < 0.0005$]. The main effect of ''type of mistuned component'' showed that overall, pitch shifts were smaller in the pre-target than in the simultaneous conditions [$F(1,7) = 5.59$, $p < 0.05$]. Pitch shifts for the simultaneous components were significantly larger than those for pre-target components only for mistunings of 20 Hz (Tukey HSD test, $p < 0.01$); for mistunings of 10 and 50 Hz, pitch shifts were not significantly different (Tukey HSD test, $p > 0.05$).

The negative and positive pitch shifts for the simultaneous and the post-target components in the post-target task are displayed in Fig. 4(a) and (b), respectively. The overall pattern and size of pitch shifts for the post-target and for the
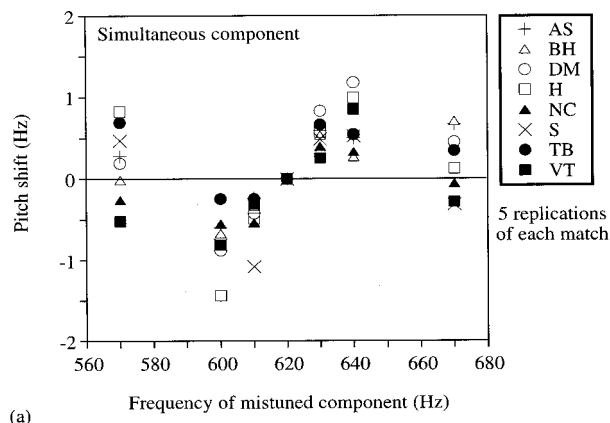
(a)
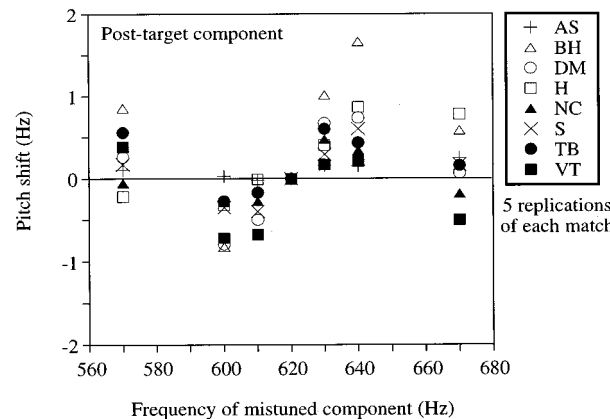
(b)

FIG. 4. Individual pitch shifts, averaged over five repetitions for the post-target task of experiment 1. Pitch shifts for the simultaneous components are shown in (a); shifts for the post-target mistuned components are displayed in (b).

simultaneous components is similar. For most subjects (six out of eight) a reduction in the size of the pitch shifts occurred either for positive or for negative mistunings of 10–20 Hz in the post-target condition [Fig. 4(b)] relative to the simultaneous condition [Fig. 4(a)]; for two listeners, BH and AS, the patterns of pitch shifts for post-target components differed from that of the other subjects. Listener AS, who showed virtually no pitch shifts in the post-target condition, also showed very small shifts in the pre-target condition. This listener was perhaps able to make a reliable estimate of the pitch of the target very quickly, possibly by focusing only on the first few tens of milliseconds of the target sound. Unlike AS, listener BH perceived much larger pitch shifts in the post-target than in the simultaneous condition for positive mistunings.

A two-way ANOVA with repeated measures was applied to the mean pitch shifts (averages of −ve and +ve pitch shifts for each amount of mistuning) for the post-target session (see Fig. 5). The factors were ''mistuning'' (10, 20, or 50 Hz) and ''type of mistuned component'' (simultaneous or post-target). The main effect of ''mistuning'' was statistically significant [$F(2,14) = 36.25$, $p < 0.0001$]. The pattern of pitch shifts as a function of mistuning is very similar in the simultaneous and the post-target conditions: Pitch shifts were largest at 20-Hz mistunings and decreased to about zero at mistunings of 50 Hz. The effect of ''type of mistuned
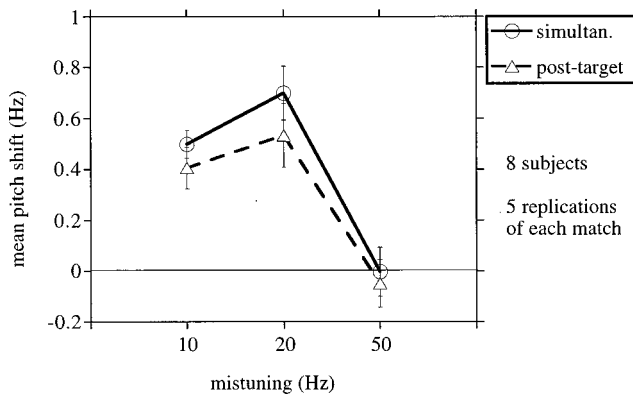
FIG. 5. Mean pitch shifts with error bars (standard errors of the mean) for the post-target task of experiment 1. Mean shifts for the post-target mistuned component are displayed as open triangles, and those for the simultaneous component as open circles.

component'' was not statistically significant [$F(1,7)=2.96$, $p>0.05$]. The difference between simultaneous and post-target components was not statistically significant for any amount of mistuning (Tukey HSD tests). The ''mistuning'' by ''type of mistuned component'' interaction was not statistically significant [$F(2,14)=0.15$, $p>0.05$].

In order to perform a direct comparison of the effects of a preceding and a following nonsimultaneous component on the pitch of the target sound, a two-way ANOVA with repeated measures was applied to the mean pitch shifts for the pre- and post-target conditions. The factors were the ''mistuning'' (10, 20, or 50 Hz) and the ''type of mistuned component'' (pre-target or post-target). Pre- and post-target mistuned components had different effects on the mean pitch shifts as a function of the amount of mistuning, as shown by a statistically significant ''type'' by ''mistuning'' interaction [$F(2,14)=5.62$, $p<0.05$]. The main effect of ''mistuning'' was also statistically significant [$F(2,14)=11.93$, $p<0.001$]; the main effect of ''type'' was not statistically significant [$F(1,7)=0.21$, $p>0.05$].

To summarize, this experiment showed that a mistuned component which is nonsimultaneous can produce pitch shifts in a target complex. This finding extends the results of Hall and Peters' (1981) study which found that listeners integrated a sequence of three brief sinusoids into a single virtual pitch when these stimuli were presented in noise (low S/N ratio condition). While Hall and Peters reported a lack of virtual pitch perception in the quiet condition of their study, the integration of the nonsimultaneous component into the virtual pitch of the target tone was observed in quiet in the present experiment. This difference could be due to the nature of the stimuli: When only three nonsimultaneous harmonics are used, low S/N ratios may be necessary to be able to perceive a single virtual pitch. On the other hand, the virtual pitch of the target complex could be perceived in quiet, and the current task measured whether virtual pitch was affected by the presence of a nonsimultaneous component (instead of measuring whether a virtual pitch could be perceived at all). Second, this experiment showed that pitch shifts in a target complex sound were smaller when the mistuned component preceded the target than when the target and the mistuned component were presented simultaneously.

By contrast, the post-target and simultaneous conditions produced virtually identical pitch shifts. The fact that the contribution of a nonsimultaneous component to the pitch of the target complex is asymmetrical with respect to the order of presentation of the stimuli might indicate that pitch perception mechanisms calculate virtual pitch in a ''forward'' fashion. That is, pitch processes integrate acoustic information which follows the onset of the complex into a single virtual pitch; acoustic energy which precedes the onset of a harmonic complex would be less strongly integrated into the pitch of the complex. This asymmetry may be explained by the fact that the onset of a harmonic complex causes pitch perception mechanisms to search for components which could ''fit'' into the pitch of a complex sound. This search could be based on a ''harmonic sieve'' principle (Scheffers, 1983). A post-target component would therefore contribute to the pitch of the target complex as long as its frequency is not too far from the harmonic frequency. When the mistuned component precedes the target complex (pre-target condition), the calculation of the virtual pitch of the target sound would not begin until immediately after the offset of the component; therefore, the mistuned component would be less likely to be included in the pitch of the harmonic complex.

II. EXPERIMENT 2

If the pitch processor calculates virtual pitch in a forward fashion then the pitch integration period should be relatively long following the onset of the harmonic complex (post-target components), but it should be short for information that precedes the onset of the complex sound (pre-target components). This prediction was tested in the second experiment by varying the duration of a silent gap which separated pre-target and post-target mistuned components from the target tone.

A. Method

The pre- and post-target tasks were run in separate experimental sessions as in the first experiment. In both tasks, the duration of the silent interval which separated the target from the nonsimultaneous mistuned component was varied and could be either 0, 20, 40, 80, or 160 ms. In the pre-target task, the interval separated the offset of the mistuned component from the onset of the target; in the post-target task, the silent interval separated the offset of the target complex and the onset of the mistuned component. A condition in which the mistuned component and the target tone were simultaneous was also included in both experimental session. Experiment 1 showed that mistunings of ±3% (600- and 640-Hz mistuned components) produced the largest pitch shifts in the target sound, in agreement with the findings of previous pitch matching studies (see, e.g., Darwin and Ciocca, 1992). Therefore, only three mistuned components (600, 620, and 640 Hz) were included in this experiment.

The stimuli were played through Sennheiser HD 250 Linear earphones in a single walled IAC booth which was located in a sound-insulated room. The software/hardware
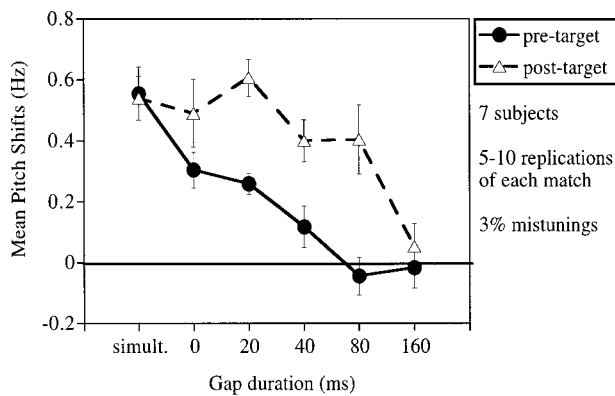
FIG. 6. Mean pitch shifts with error bars (standard errors of the mean) for the pre- and post-target tasks of experiment 2. Mean shifts for mistunings of ±3% are displayed for the simultaneous component and for each interval duration of the pre-target (filled symbols) and post-target tasks (open symbols).

setup was the same as in the previous experiment, except that a PowerMac 7100/70AV was used to run the experiment instead of a MacIIcx.

Seven listeners, one of whom was the first author, took part in the experiment; three of the listeners were musically trained. All listeners had participated in previous pitch matching experiments, but none had taken part in experiment 1. Subjects completed between five and ten matches for each experimental condition, depending on their pitch matching accuracy after five matches per condition. The criterion for stopping after five matches was a standard error smaller than 0.4 Hz in all conditions. The experiment was run in separate blocks during which subjects completed 2–3 matches for each condition; 2–4 blocks were run by each listener, with each block taking 1–2 h to complete.

B. Results

Mean pitch shifts for the pre- and post-target sessions, averaged across listeners, are displayed in Fig. 6. A two-way ANOVA with repeated measures was applied to the mean pitch shifts for each subject in each experimental condition. The "temporal overlap" between the mistuned component and the target sound (simultaneous, 0-, 20-, 40-, 80-, or 160-ms gap), and the "type of mistuned component" (pre-target versus post-target) were the experimental factors. Mean pitch shifts were larger for post-target than for pre-target components (main effect of "type") $[F(1,6) = 12.26, p < 0.05]$. The overall pattern of shifts indicates that pitch shifts decreased as the silent gap duration increased, as shown by a statistically significant main effect of "temporal overlap" $[F(5,30) = 16.18, p < 0.0001]$.

For pre-target components, the 0-ms condition produced lower mean pitch shifts than the simultaneous component (planned comparison, $p < 0.02$); this finding replicates the results obtained in the 20-Hz mistuning condition of experiment 1. Pitch shifts were effectively eliminated when a silent gap of 80 ms separated the offset of the mistuned component and the onset of the target; this condition was significantly different from the 20-ms condition (planned comparison, $p < 0.005$). The 40- and 80-ms conditions were not signifi-

cantly different from each other (planned comparison, $p > 0.05$). The results of the post-target task replicated those of experiment 1, which showed that the pitch shifts for the synchronous and the 0-ms delay conditions were virtually identical for mistunings of 20 Hz. Mean pitch shifts dropped to zero in the 160-ms gap condition, which was significantly different from the 0-ms condition (planned comparison, $p < 0.001$). The 40- and 80-ms conditions were not significantly different from the 0-ms condition (planned comparison, $p > 0.05$).

As the delay between the nonsimultaneous component and the target was increased, mean pitch shifts decreased at different rates for pre-target and for post-target components ("temporal overlap" by "type" interaction) $[F(5,30) = 3.11, p < 0.05]$. The mean pitch shifts for pre- and post-target components were significantly different in the 20-ms, 40-ms, and 80-ms conditions (planned comparisons, $p < 0.01$). The pre- versus post-target difference in pitch shifts for the 0-ms condition just failed to reach statistical significance (planned comparison, $p = 0.07$).

To summarize, pitch shifts were significantly reduced when the silent interval between the pre-target component and the target complex was 20 ms or longer; when a silent gap of 80 ms (or longer) was introduced between the offset of the component and the onset of the target, pitch shifts were virtually eliminated. For post-target components, pitch shifts were identical to those produced by a simultaneous component for delays of up to 40 ms; pitch shifts were reduced to zero only for the longest delay (160 ms). These results support the previous suggestion that the pitch integration period extends for a longer interval following than preceding the onset of a complex sound.

III. EXPERIMENT 3

The results of the previous experiments have been considered to be indicative of the duration of the "pitch integration period" (i.e., the temporal window within which pitch processes integrate acoustic information into a single virtual pitch) with respect to the onset of the target. This interpretation is based on the assumption that, in the current task, the formation of virtual pitch (i) is triggered by the onset of the target, (ii) includes a relatively short interval prior to the onset of the target, and (iii) ends once a reliable estimate of virtual pitch has been achieved following the onset of the target. If this assumption is valid, then pitch shifts obtained as a function of silent gap duration would represent an estimate of the pitch integration period. However, an alternative interpretation of the results is that pitch processes might not cease the calculation of virtual pitch as soon as a reliable estimate of the pitch of the target has been made; instead, pitch processes may keep revising their estimate as long as information is received. Consider, for example, the post-target stimuli: it is possible that listeners use only the last portion of the target in conjunction with the nonsimultaneous component for performing the pitch matching task. If this were the case, the inclusion of the duration of the target into the estimate of the duration of the pitch integration period would not be warranted. A simple test of these hypotheses
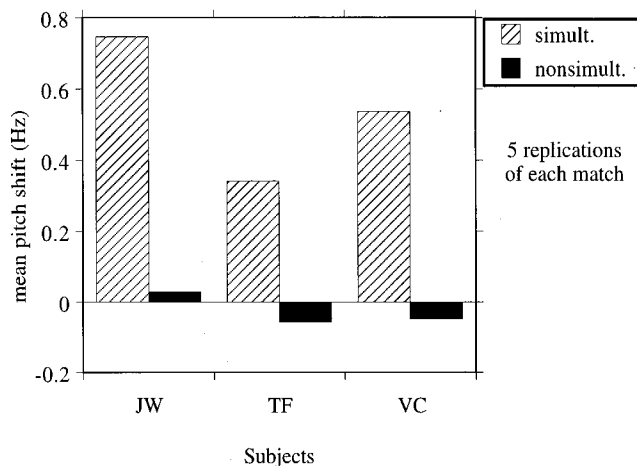
FIG. 7. Mean pitch shifts for the stimuli of experiment 3. Mean shifts for mistunings of $\pm 3\%$ are displayed for the simultaneous and the post-target conditions.

can be carried out by using stimulus durations in excess of the longest estimated duration of the pitch integration period (i.e., about 200 ms for post-target stimuli).

This experiment employed the post-target task only, since post-target components produced pitch shifts for longer gap durations than pre-target components in the previous experiments. The stimuli had a fixed duration of 410 ms; this duration was the same as that of the stimuli used in previous pitch matching tasks in which pitch shifts were measured for simultaneous mistuned components (Darwin and Ciocca, 1992). Only the post-target condition with no silent gap between the target and the nonsimultaneous component (0-ms delay) was tested. If subjects perceive the pitch of the target sound by integrating acoustic information from the onset of the target until a reliable pitch estimate has been achieved then nonsimultaneous components should not produce pitch shifts in this experiment, since the pitch of the target would be calculated well before the offset of the target. By contrast, if the alternative hypothesis is correct then we would expect to observe pitch shifts for 410-ms target sounds followed by nonsimultaneous components of the same duration.

A. Method

The stimuli were identical to those used in the post-target session of experiment 1, except that: (i) all the stimuli had a fixed duration of 410 ms instead of 90 ms, and (ii) only mistunings of ± 20 Hz were used. Therefore, a total of six experimental conditions were employed in this experiment: two types of mistuned component (simultaneous versus post-target) by three frequencies (600, 620, and 640 Hz). Subjects completed five matches for each condition.

The apparatus was identical to that employed in experiment 2.

Three listeners, one of whom was the first author, took part in the experiment. All listeners had taken part in previous pitch matching experiments, and two had taken part in experiment 2; one of the listeners was musically trained. The experiment took about 30 min to complete.

B. Results

The pitch shifts for $-$ve and $+$ve mistunings, averaged across five trials, are displayed in Fig. 7 for each subject. This figure shows that the nonsimultaneous mistuned components produced no pitch shifts, while the simultaneous components did. Since the pattern of pitch shifts for simultaneous and nonsimultaneous components was identical for the three subjects, it was not considered necessary to perform any statistical analysis on the data. The results of this experiment fully support the idea that listeners begin calculation of the virtual pitch at the onset of the target tone, and complete such calculation as soon as a reliable pitch estimate has been achieved. This process is completed before the offset of the stimuli used in the present experiment, which suggests that the pitch integration period for the current stimuli is shorter than 410 ms. This conclusion supports the interpretation of the nonsimultaneous pitch shifts obtained in experiment 2 in terms of estimated duration of the pitch integration period.

IV. GENERAL DISCUSSION

A. Summary

The present study investigated the integration of a nonsimultaneous mistuned component into the virtual pitch of a harmonic complex. The first experiment showed that, for brief target tones: (i) the pitch shifts produced by a post-target mistuned component were virtually identical to the shifts produced by a simultaneous mistuned component; (ii) pre-target components produced significantly smaller pitch shifts than simultaneous components. The findings of experiment 2 indicated that pitch shifts occurred at longer gap durations for post-target components (up to 80 ms) than for pre-target components (up to 40 ms). The third experiment demonstrated that the current pitch matching procedure can be employed for obtaining reliable estimates of the duration of the pitch integration period with respect to the onset of the target sound.

B. Matching virtual or spectral pitches?

The interpretation of the results of this study in terms of the contribution of a nonsimultaneous component to the virtual pitch of the target tone relies on the assumption that subjects were matching the virtual pitch of the target and matching tones. It has been suggested that in pitch discrimination or matching experiments, subjects could perform these tasks by comparing the spectral pitches of corresponding frequency components instead of comparing the virtual pitches of the complex tones (Faulkner, 1985). In principle, the matching of the spectral pitch of the mistuned component with the corresponding 4th harmonic of the (fully harmonic) matching tone might well have occurred in the present study, particularly when the component was nonsimultaneous and could be easily heard as having a distinct spectral pitch. However, there are several reasons for rejecting the idea that listeners were matching spectral pitches of individual components in the present experiments. First, Moore (1987) demonstrated that pitch shifts produced by (simultaneous) mistuned components can be measured even when the

matching tone does not contain the harmonic that corresponded to the mistuned component. Second, if listeners matched the pitch of the mistuned component, pitch shifts should have been larger than those normally observed in pitch matching studies. For example, in the present experiments pitch shifts were about ±1 Hz, on average, when the 4th harmonic was mistuned by ±20 Hz. If subjects had adjusted the matching tone so that its 4th harmonic matched the pitch of the mistuned component, then pitch shifts should have been much larger than those observed in the present study. Moreover, the matching of the spectral pitch of mistuned components would have produced a monotonic increase in the size of pitch shifts with increasing mistuning,[1] which was not the case. Third, matching the spectral pitch of the mistuned component should have been easier, thereby producing larger pitch shifts, in the nonsimultaneous than in the simultaneous conditions. Finally, matching the pitch of the mistuned component should have been easier in the pre-target conditions, in which the mistuned component was the first sound within a trial, than in the post-target conditions, in which the first sound within a trial was the matching tone having a randomly selected fundamental frequency. Contrary to these predictions, pitch shifts were never larger in the nonsimultaneous than in the simultaneous conditions, and were usually larger in the post- than in the pre-target condition. These considerations support the idea that, in the present study, subjects matched the virtual pitches of the target and matching tones rather than matching the spectral pitches of individual components.

C. Temporal integration and pitch perception

Although further research is needed to obtain a precise estimate of the duration and of the window shape of the pitch integration period, a tentative model of temporal pitch integration can be formulated from the present data by assuming that: (i) the temporal integration window is rectangular; (ii) the calculation of virtual pitch is triggered by the onset of the target; (iii) the pitch integration period includes a relatively short interval preceding, and a longer interval following, the onset of the target; (iv) the endpoint of the pitch integration period occurs within the interval spanning from the longest gap duration at which pitch shifts were observed to the gap duration at which pitch shifts were eliminated. While (i) is assumed for simplicity, criteria (ii), (iii), and (iv) can be justified on the basis of the results of experiments 2 and 3. It might be argued that criterion (ii) applies to post-target but not to pre-target conditions, because subjects might have started the calculation of the virtual pitch of the target as soon as they heard a pre-target component at the beginning of a trial (in spite of instructions to ignore pre-target components and focus on the matching of the pitch of target and matching sounds). However, this hypothesis does not explain the decrease in pitch shifts with increasing delays in the pre-target conditions of experiment 2, nor is it clear how the pitch of the target could be calculated on the basis of a single mistuned component. It is more likely that pitch estimation is initiated at the onset of the target, but that the pitch integration period includes an interval prior to the onset of the target, as specified by criterion (iii). This criterion is consistent

with the assumptions that the auditory input is first stored into a ''preperceptual auditory image,'' and that auditory processes such as those involved in pitch perception operate on the information stored in the preperceptual auditory image (Massaro, 1972).

On the basis of these criteria, the duration of the pitch integration period can be estimated to be 40–80 ms prior to target onset, and 170–250 ms following the onset of the target sound (the latter duration also includes the target tone duration of 90 ms). Robinson and Patterson (1995) investigated the identification of the pitch chroma of vowels as a function of the number of cycles, and found that performance reached an asymptote at 16–32 cycles, which corresponds to durations of 122–244 ms and 92–184 ms for $F0$'s of 131 and 174 Hz, respectively. Their results are consistent with the results of experiment 3, which suggest that pitch perception processes integrate acoustic information into a single virtual pitch until a reliable estimate has been generated. The current estimates are also in agreement with the results of studies which investigated fundamental frequency discrimination (Carlyon, 1996; Plack and Carlyon, 1995). For example, Carlyon (1996) demonstrated that listeners perceptually integrate portions of a masker which occurred before and after a 200-ms target complex into the target's virtual pitch over a period of 100 ms or longer (''overintegration''). Carlyon found that overintegration occurred when the target was composed of unresolved, but not of resolved, harmonics. However, it is possible that overintegration could be observed with resolved harmonics for target durations shorter than 200 ms. Evidence in support of this prediction is found in two recent studies. Plack and Carlyon (1995) reported that the $F0$ discrimination of the fundamental frequency of two 50-ms complex tones deteriorated when duration of the gap between the tones was reduced from 500 ms to 50 ms, even when the two tones consisted of resolved harmonics. Micheyl and Carlyon (1998) found that $F0$ discrimination of 100-ms target tones composed by resolved components was degraded by the presence of complex tones which preceded and followed the targets. Further evidence for the fact that nonsimultaneous resolved components can be integrated into a single pitch comes from studies which employed pure tone stimuli (Rakowski and Hirsh, 1980; Kelly and Watson, 1986).

The current estimate of the duration of the pitch integration period is consistent with previous suggestions that the auditory system integrates acoustic energy over a period of few hundreds of milliseconds for the perception of loudness (''loudness summation;'' Zwislocki, 1969), backward and forward masking (Wilson and Carhart, 1971), and amplitude modulation detection (Sheft and Yost, 1990). The present results are also compatible with temporal integration models based on shorter time constants (few milliseconds), such as the ''multiple look'' model proposed by Viemeister and Wakefield (1991). This model states that the auditory system takes samples or ''looks'' of the incoming acoustic signal about every 3 ms. These samples are stored in memory and can be selected for further processing. The present data could be explained by such model if it is assumed that pitch processes can combine information across frequency channels

from samples which occurred over periods up to about 200 ms, despite the temporal separation between the nonsimultaneous component and the other components of the target tone (see, for example, White and Carlyon, 1997, for evidence in support of this kind of temporal integration).

The present findings suggest that pitch processes calculate virtual pitch by giving a higher weighting to acoustic information which follows the onset of a complex sound than to information which precedes it. This temporal asymmetry complements the finding that onset asynchrony is more effective than offset asynchrony in preventing an asynchronous harmonic from contributing to the timbre of a vowel (Darwin, 1984). This asymmetry is also consistent with the results of previous studies which showed that the final portion of a signal is given a higher weighting than initial information for the perception of the pitch of frequency glides (Nabelek *et al.*, 1970) and ramped versus damped sinusoids (Patterson, 1994a,b). A larger degradation in performance with trailing maskers (which followed the presentation of target tones) than with leading maskers was observed in order discrimination (Kelly and Watson, 1986), and pitch identification tasks (''backward recognition masking;'' Massaro, 1975). However, the effects of trailing maskers in those studies were not frequency specific, and were interpreted in terms of the interference of the maskers with the stored representation of (brief) target sounds. Therefore, it is not clear that this interference effect played a major role in the pitch matching paradigm used in the present study, which measured the integration of pre- and post-target components into the virtual pitch of the target.

D. Implications for models of pitch perception

The current findings, which imply that pitch processes integrate acoustic energy over a period of few hundred milliseconds for calculating virtual pitch, have implications for models of pitch perception. First, these findings suggest that phase information is not likely to be an important factor for the integration of a frequency component into a single virtual pitch, since mistuned components which followed the offset of the target by up to 80 ms of silence produced relatively large pitch shifts. It is interesting to note that, unlike pitch perception, the detection of the mistuning of a harmonic was found to be affected by its phase (Hartmann, 1988). Second, these results are difficult to reconcile with models of pitch perception that estimate virtual pitch by combining autocorrelation functions of the activity of individual nerve fibers (see recent models by Meddis and Hewitt, 1991a,b; Yost *et al.*, 1996; Meddis and O'Mard, 1997). These models calculate virtual pitch by summing the output of autocorrelation functions across frequency channels; the summary autocorrelation function thus obtained (''summary autocorrelogram'') is then used to estimate virtual pitch, which corresponds to the period of the highest peak in the summary autocorrelogram. The models require that the energy of frequency components which is used to produce a summary autocorrelogram occur within a ''time constant,'' which is typically a period of few milliseconds (see also a related model by de Cheveigne, 1998). The results reported here, as well as those of the studies reviewed above, have demonstrated that the integration of nonsimultaneous acoustic energy for pitch perception can occur over periods of 100 ms or longer. In order to account for these findings, an autocorrelation model would have to be able to store autocorrelation functions for each frequency channel and for periods of at least 100 ms, and then apply some temporal smoothing before the summary autocorrelogram is calculated.[1] The integration of nonsimultaneous energy over relatively long durations for the perception of virtual pitch could be perhaps more easily implemented by models which compute virtual pitch from the frequency of resolved components (Goldstein, 1973; Terhardt *et al.*, 1982a). These models could be modified to specify differential weightings of frequency components for the calculation of pitch, depending on the components' time of occurrence with respect to the onset of complex sounds.

[1]We are grateful to Robert Carlyon for this suggestion.

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. (**1993**). ''Effects of onset asynchrony on pitch perception: Adaptation or grouping?,'' J. Acoust. Soc. Am. **93**, 2870–2878.

Darwin, C. J. (**1984**). ''Perceiving vowels in the presence of another sound: constraints on formant perception,'' J. Acoust. Soc. Am. **76**, 1636–1647.

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. **9**, 3381–3390.

Darwin, C. J., Ciocca, V., and Sandell, G. J. (**1994**). ''Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic,'' J. Acoust. Soc. Am. **95**, 2631–2636.

Darwin, C. J., Hukin, R. W., and Al-Khatib, B. Y. (**1995**). ''Grouping in pitch perception: Evidence for sequential constraints,'' J. Acoust. Soc. Am. **98**, 880–885.

de Cheveigne, A. (**1998**). ''Cancellation model of pitch perception,'' J. Acoust. Soc. Am. **103**, 1261–1271.

Faulkner, A. (**1985**). ''Pitch discrimination of harmonic complex signals: Residue pitch or multiple component discriminations,'' J. Acoust. Soc. Am. **78**, 1993–2004.

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

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

Hartmann, W. M. (**1988**). ''Pitch perception and the segregation and integration of auditory entities,'' in *Auditory Function—Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowen (Wiley, New York), pp. 623–645.

Hartmann, W. M. (**1996**). ''Pitch, periodicity, and auditory organization,'' J. Acoust. Soc. Am. **100**, 3491–3502.

Houtsma, A. J. M. (**1995**). ''Pitch perception,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 267–295.

Houtsma, A. J. M., and Goldstein, J. L. (**1971**). *Perception of musical intervals: Evidence for the central origin of the pitch of complex* (Tech. Re. 484): MIT, Cambridge.

Kelly, W. J., and Watson, C. S. (**1986**). ''Stimulus-based limitations on the discrimination between different temporal orders of tones,'' J. Acoust. Soc. Am. **79**, 1934–1938.

Massaro, D. W. (**1972**). ''Stimulus information vs processing time in auditory pattern recognition,'' Percept. Psychophys. **12**, 50–56.

Massaro, D. W. (**1975**). ''Backward recognition masking,'' J. Acoust. Soc. Am. **58**, 1059–1065.

Meddis, R., and Hewitt, M. (**1991a**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and Hewitt, M. (**1991b**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity,'' J. Acoust. Soc. Am. **89**, 2883–2894.

Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.

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

Moore, B. C. J. (**1987**). ''The perception of inharmonic complex tones,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 180–189.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). ''Relative dominance of individual partials in determining the pitch of complex tones,'' J. Acoust. Soc. Am. **77**, 1853–1860.

Nabelek, I. V., Nabelek, A. K., and Hirsh, I. J. (**1970**). ''Pitch of tone bursts of changing frequency,'' J. Acoust. Soc. Am. **48**, 536–553.

Patterson, R. D. (**1994a**). ''The sound of a sinusoid: spectral models,'' J. Acoust. Soc. Am. **96**, 1409–1418.

Patterson, R. D. (**1994b**). ''The sound of a sinusoid: time-domain models,'' J. Acoust. Soc. Am. **96**, 1419–1428.

Plack, C. J., and Carlyon, R. P. (**1995**). ''Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics,'' J. Acoust. Soc. Am. **98**, 1355–1364.

Plomp, R. (**1967**). ''Pitch of complex tones,'' J. Acoust. Soc. Am. **41**, 1526–1533.

Rakowski, A., and Hirsh, I. J. (**1980**). ''Post-stimulatory pitch shifts for pure tones,'' J. Acoust. Soc. Am. **68**, 467–474.

Risset, J. C., and Wessel, D. L. (**1982**). ''Exploration of timbre by analysis and synthesis,'' in *The Psychology of Music*, edited by D. Deutsch (Academic, New York), pp. 25–58.

Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.

Robinson, K., and Patterson, R. D. (**1995**). ''The stimulus duration required to identify vowels, their octave, and their pitch chroma,'' J. Acoust. Soc. Am. **98**, 1858–1865.

Russell, P., and Darwin, C. J. (**1991**). ''Real-time synthesis of complex sounds on a Mac II with 56001 DSP chip,'' Br. J. Audiol. **25**, 59–60.

Scheffers, M. T. (**1983**). ''Sifting vowels: Auditory pitch analysis and sound segregation,'' Unpublished Ph.D. thesis, Groningen University, The Netherlands.

Sheft, S., and Yost, W. A. (**1990**). ''Temporal integration in amplitude modulation detection,'' J. Acoust. Soc. Am. **88**, 796–805.

Terhardt, E., Stoll, G., and Seewann, M. (**1982a**). ''Algorithm for extraction of pitch salience from complex tonal signals,'' J. Acoust. Soc. Am. **71**, 679–688.

Terhardt, E., Stoll, G., and Seewann, M. (**1982b**). ''Pitch of complex signals according to virtual pitch theory,'' J. Acoust. Soc. Am. **71**, 671–678.

Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.

White, L. J., and Carlyon, R. P. (**1997**). ''Detection of signals having expected and unexpected temporal structures,'' Hearing Res. **112**, 141–146.

Wilson, R. H., and Carhart, R. (**1971**). ''Forward and backward masking: Interactions and additivity,'' J. Acoust. Soc. Am. **49**, 1254–1263.

Yost, W. A., Patterson, R., and Sheft, S. (**1996**). ''A time domain description for the pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **99**, 1066–1078.

Zwislocki, J. J. (**1969**). ''Temporal summation of loudness: An Analysis,'' J. Acoust. Soc. Am. **46**, 431–441.