Grouping in pitch perception: Evidence for sequential constraints

C. J. Darwin,[a] R. W. Hukin, and Batul Y. Al-Khatib
University of Sussex, Brighton BN1 9QG, England

(Received 7 October 1994; revised 28 November 1994; accepted 9 February 1995)

Evidence is presented that sequential auditory grouping constraints apply to the perception of pitch. Experiment 1 shows that the pitch changes produced by mistuning the fourth harmonic of a 90-ms 12-harmonic 155-Hz fundamental complex tone are substantially reduced when the complex is preceded by four 90-ms tones at the same frequency as the mistuned component. Both the pitch changes and their reduction by the tonal sequence precursor remain when the mistuned component and the precursor are presented contralateral to the remaining components. Experiment 2 shows that reducing the level of the same mistuned component reduces the size of the pitch change, but only if the mistuned component is presented ipsilaterally. To the extent that adaptation can be equated with a physical reduction in level, this result provides further evidence against peripheral adaptation playing a significant role in the auditory grouping of harmonics in pitch perception. © *1995 Acoustical Society of America.*

PACS numbers: 43.66.Hg, 43.66.Mk

INTRODUCTION

Although the problem of how we perceive the pitch of a complex tone has been extensively studied for the case where the listener hears a single complex (Goldstein, 1973; Houtsma, 1984; Meddis and Hewitt, 1991; Carlyon and Shackleton, 1994), the naturally more common case of pitch perception when there are multiple periodic sound sources present has only received attention relatively recently (Beerends and Houtsma, 1986; Beerends, 1989; Beerends and Houtsma, 1989; Darwin and Ciocca, 1992; Assmann and Paschall, 1994).

Widely accepted models of pitch perception such as Goldstein's implicitly assume that all of the (resolved) frequency components present at a particular time are relevant to the calculation of the pitch of a complex tone. When the signal is noisy or when more than one periodic sound is present at a time, this unwarranted assumption can lead to serious errors in pitch estimation. A practical implementation of Goldstein's theory for measuring pitch in speech discards harmonics that are implausibly far in frequency from a harmonic of a particular fundamental—the "harmonic sieve" (Duifhuis *et al.*, 1982).

The tolerance of a possible perceptual harmonic sieve has been estimated in experiments where the effect of mistuning a low-numbered harmonic on the pitch of a complex has been measured. They show that the contribution that a mistuned harmonic makes to the pitch of a complex decreases as it is progressively mistuned according to a Gaussian function with a standard deviation of about 3% (Moore *et al.*, 1985a; Darwin, 1992; Darwin *et al.*, 1994).

Factors other than harmonicity also influence whether a frequency component will be included in the calculation of the pitch of a complex tone. A slightly mistuned harmonic that starts at a different time from the remainder will make less of a contribution than if it had started simultaneously (Darwin and Ciocca, 1992; Ciocca and Darwin, 1993), and one that is frequency modulated coherently with the remainder will make more of a contribution than one that is mistuned by the same amount in an unmodulated sound (Darwin *et al.*, 1994). However, presenting the mistuned component contralateral to the remaining components, or giving all components coherent amplitude modulation, have rather little effect on the mistuned component's contribution to pitch (Darwin and Ciocca, 1992; Darwin *et al.*, 1994).

All the properties discussed so far—mistuning, onset-time, common modulation, lateral position—are examples of cues that influence primitive grouping mechanisms (Bregman, 1990) concerned with *simultaneous* perceptual organization. The main purpose of the experiment described in the present paper is to test whether pitch perception is also subject to auditory grouping mechanisms that are concerned with *sequential* organization. A secondary aim was to investigate whether sequential grouping cues were more effective when combined with differences in lateral position. The fact that a particular auditory grouping cue is effective in one perceptual paradigm (such as vowel identification) does not guarantee that it will also be effective in another paradigm such as pitch perception.

A potent way to perceptually remove a single frequency component from a complex using sequential organization is to embed the complex in a short sequence of tones at the same frequency. The single frequency component of the complex then is grouped as an additional tone of the sequence and may thereby be removed from the complex. This technique has been used (Darwin *et al.*, 1989; Hukin and Darwin, in press) to perceptually remove energy at 500 Hz from the calculation of vowel quality.

The experiment reported here asks whether a similar manipulation is effective at removing a single frequency component from a complex for the purpose of calculating its pitch. It uses the pitch-matching paradigm introduced by Moore *et al.* (1985a) and used previously in this laboratory to investigate auditory grouping in pitch perception.

---

[a] Electronic mail: cjd@epvax.sussex.ac.uk

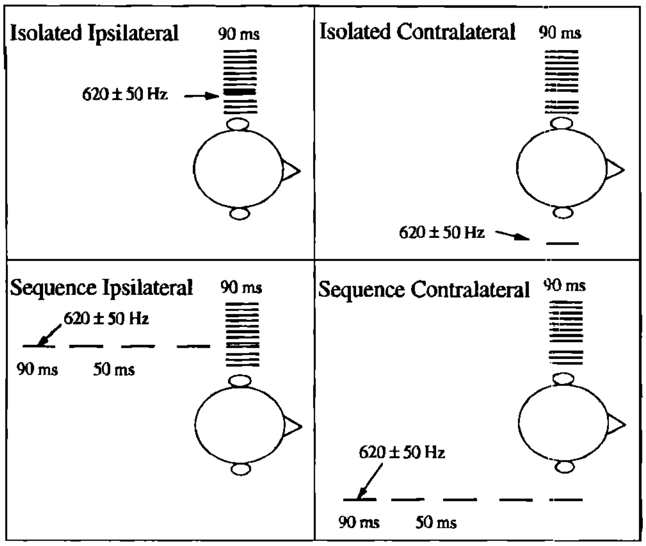

FIG. 1. Stimulus configuration in experiment 1 for the target complex tone and the mistuned component presented either ipsilaterally or contralaterally in isolated or sequence conditions.

I. EXPERIMENT 1

If a preceding tonal sequence is able to perceptually remove a mistuned harmonic from a complex so that it makes less of a contribution to the pitch of that complex, then we would expect to find that the changes in pitch of the complex produced by the mistuned component will be reduced by preceding the complex with a tone sequence at the frequency of the mistuned component.

A. Method

The experiment had two major conditions that were presented in separate blocks of trials. Figure 1 is a schematic of the stimuli. In the "isolated" condition subjects heard a single presentation of two 90-ms complex tones separated by 500 ms each time they pressed the mouse button. The first was the target (which had a mistuned component), the other was adjustable. In the "sequence" condition, the target tone was preceded by a sequence of four 90-ms pure tones (separated by 50 ms of silence) at the same frequency and amplitude as the mistuned component. The subject's task was to match the pitch of the adjustable tone to that of the target by making manual adjustments to a rollerball between button presses. Subjects had as long as they wished to complete the experiment and could make as many adjustments as they wished in each trial. They indicated their satisfaction with a match by a key press. Within each block, the mistuned (and sequence) components were presented randomly either ipsilaterally or contralaterally to the rest of the complex, which always went to the left ear.

The target tone consisted of harmonics 1–3 and 5–12 of a 155-Hz fundamental together with a possibly mistuned 4th harmonic. The frequency of the mistuned component was either 570, 600, 610, 620 (harmonic frequency), 630, 640, or 670 Hz corresponding to mistunings of −8%, −3.2%, −1.6%, 0%, +1.6%, +3.2%, and +8%. The adjustable tone consisted of the first 12 harmonics of a 155±4-Hz fundamental.

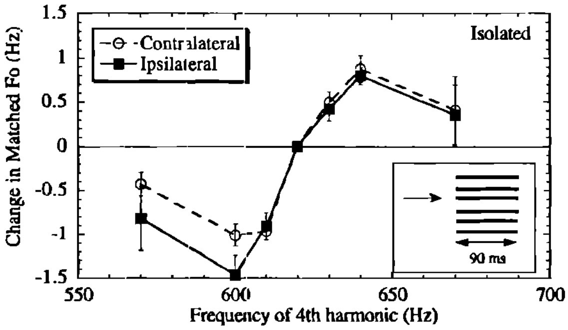

FIG. 2. Average normalized pitch matches (with standard errors) to a 12-harmonic complex with a fundamental frequency 155 Hz, when the 4th harmonic was mistuned. The mistuned component was presented either contralaterally or ipsilaterally.

At the start of every trial the adjustable tone was allocated a fundamental frequency value at random from within the limits 155 Hz±4 Hz. This value was then adjusted by subjects using a rollerball whose movement was visually represented by a cursor on a screen. If a subject's adjustments caused the value to overshoot the aforementioned preset limits a warning sound was heard and the adjustable tone was reset at random and within the limits. The setting of such limits was intended as a guide to subjects, most of whom had no experience of pitch matching.

There were 14 conditions in each block: 7 mistunings by 2 possible sides of presentation (ipsilateral or contralateral). For each condition 5 matches were made, making a total of 70 trials for each block. The experiment consisted of two such blocks each taking about 40 min to complete. Four subjects took the "isolated" condition first and three the "sequence" condition first; subjects had several days' break between conditions. Seven paid subjects participated in the experiment, including the authors, all of whom had had musical training and reported themselves as having normal hearing.

Sounds were synthesized in real time at 44.1 kHz using custom software written for the 56001 processor the Digidesign Audiomedia board and output through that board's 16-bit DACs and antialiasing filters. The board was attached to an Apple MacII computer which controlled the experiment. Sounds were presented to subjects through Sennheiser HD 414 headphones in a double walled IAC booth. All the partials of the complex tones had the same amplitude (corresponding to 58 dB SPL for a 1000-Hz tone) and started with sine phase. Stimulus levels were calibrated using a B&K type 2610 measuring amplifier with a type 4153 artificial ear and a 4140 sound level calibrator.

B. Results

In order to remove individual constant matching biases from the data, each subject's pitch matches were expressed as deviations from the fundamental frequency match when the 4th harmonic was in tune (620 Hz). Figure 2 shows this

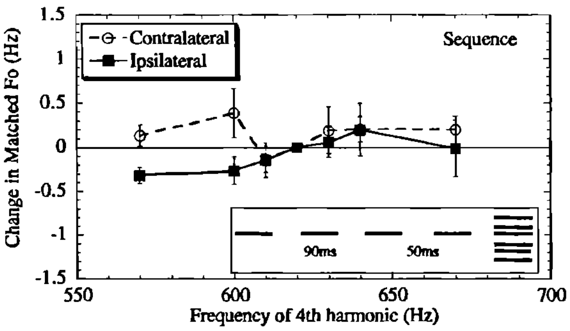

FIG. 3. Average normalized pitch matches (with standard errors) to a 12-harmonic tone with fundamental frequency 155 Hz with a mistuned 4th harmonic preceded by a captor tone sequence presented either contralaterally or ipsilaterally.

average normalized match across subjects for each mistuning in the isolated conditions; Fig. 3 shows the average normalized matches for the sequence conditions.

The general pattern of the results in the isolated condition in Fig. 2 replicates previous experiments, with the maximum pitch change of about ±1 Hz occurring at about ±3% mistuning. Figure 3 shows the predicted reduction in the size of these pitch changes, when the complex is preceded by the tone sequence.

Analysis of variance of the data from all conditions showed significant interaction between the isolated/sequence factor and level of mistuning ($F_{6,36}=18.8$; $p=0.0001$). This interaction reflects the fact that the pitch shifts in the sequence condition are smaller than those in the isolated condition. There were no significant terms involving the ipsilateral/contralateral factor, indicating that side of presentation of the mistuned component did not influence the size of the pitch shifts.

Figure 4 shows the mean change in matched fundamental expressed as half the difference between the matched fundamentals for corresponding positive and negative mistunings. The data have also been averaged across the ipsilateral/contralateral factor (which again was insignificant) and are plotted against percentage absolute mistuning. On these mean data, the main effect of the isolated/sequence factor is significant ($F_{1,6}=25.6$, $p<0.003$) as is its interaction with mistuning ($F_{2,12}=12.4$, $p<0.002$), showing again that pitch shifts are substantially reduced when the target tone is preceded by the tone sequence.

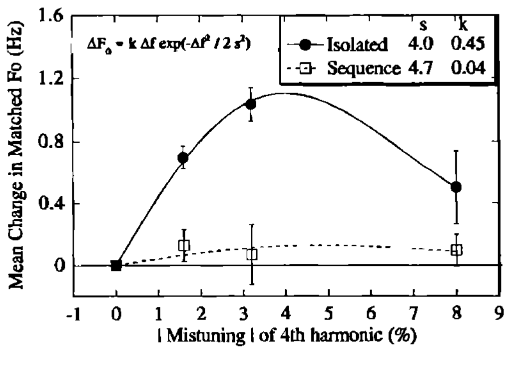

FIG. 4. Mean pitch shifts (with standard errors) produced by different mistunings of the 4th harmonic of a 155-Hz fundamental 12-harmonic complex tone. The complex tone was played on its own (isolated) or was preceded by a captor tone sequence (sequence).

Following previous papers (Darwin, 1992; Darwin *et al.*, 1994), the mean data of Fig. 4 were fitted with a Gaussian derivative function:

$$\Delta F_0 = k \Delta f \exp(-\Delta f^2/2s^2),$$

where Δf is the amount of mistuning, s is the standard deviation of the underlying Gaussian (giving the width of the harmonic sieve) and k is a scaling factor proportional to the contribution that a harmonic makes to the pitch of the complex. The standard deviation s remains roughly constant at between 4% and 5% for the isolated and sequence conditions, but the scaling factor k is reduced from 0.45 to 0.04 by the presence of the preceding tone sequence. The width of the harmonic sieve is thus unaffected by the presence of the tone sequence, but the contribution that the mistuned harmonic makes to the pitch of the target complex is substantially reduced.

C. Discussion

These results extend those of Darwin *et al.* (1989) and of Hukin and Darwin (in press) to pitch perception. They show that the sequential context can influence the extent to which a particular component contributes to the pitch of a complex tone.

However, an alternative explanation to auditory grouping that could be responsible for the reduced contribution that the mistuned component makes to pitch in the sequence conditions is that its neural representation is reduced during the complex by peripheral adaptation. This possibility has been discussed in a number of previous papers on grouping effects both in pitch perception (Darwin and Ciocca, 1992; Ciocca and Darwin, 1993) and in vowel identification (Darwin, 1984; Darwin and Sutherland, 1984; Darwin *et al.*, 1989), and the general conclusion has been that adaptation effects do not provide a complete explanation either for the simultaneous groupings effects of onset time, or for the sequential grouping effects of a preceding tone sequence.

The next experiment exploits the fact that pitch shifts are very little affected by presentation of the mistuned component to the contralateral ear to provide some further evidence against adaptation effects being responsible for the reduction in pitch shifts found with a preceding tone sequence. We ask whether an *actual* reduction in the level of an ipsilateral or contralateral mistuned component can reduce the contribution that it makes to the pitch of a complex sound. The effects of a tone sequence (experiment 1) and of onset asynchrony (Darwin and Ciocca, 1992; Ciocca and Darwin, 1993) are very similar with contralateral presentation, so an explanation in terms of adaptation would be rendered less likely if an actual reduction in level of a mistuned component were to have substantially less effect with contralateral than with ipsilateral presentation.

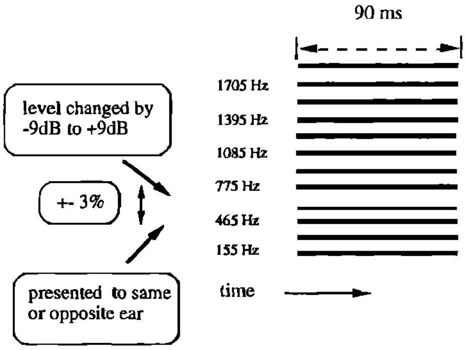

FIG. 5. Stimulus configuration in experiment 2 for the target complex tone with the mistuned component presented either ipsilaterally or contralaterally and at different levels.

II. EXPERIMENT 2

The effect on the contribution to pitch of varying the level of a mistuned second harmonic of a 12-harmonic 200-Hz fundamental complex was investigated by Moore *et al.* (1985a). They measured the contribution of the second harmonic as a function of its level over a −9 dB to +9 dB range relative to the (equal) levels of the other components. In two of their subjects, the contribution increased almost linearly as a function of level, whereas in the third subject (SK), it increased for the lowest levels but asymptoted by 0 dB. The following experiment replicates their experiment on a larger number of subjects but using stimuli similar to those used in experiment 1.

A. Method

The target stimuli in this experiment had the same fundamental frequency (155 Hz) and duration (90 ms) and level as the target stimuli used in experiment 1. The frequency of the mistuned fourth harmonic was either 600, 620 (harmonic), or 640 Hz, corresponding to mistunings of −3.2%, 0%, and +3.2%, respectively. A mistuning of about 3% was chosen to give the maximum perceived pitch shift for the complex. The mistuned component was presented either contralaterally or ipsilaterally at an amplitude of −9, −6. −3, 0, +3, +6, and +9 dB relative to the other (equal-amplitude) harmonics. A total of 42 conditions (illustrated in Fig. 5) were used in the experiment: 21 conditions with the ipsilateral mistuned component (7 levels by 3 mistunings) and 21 conditions with the contralateral mistuned component.

Five replications of each match were blocked by ear of presentation of the mistuned component, the order of the blocks being randomized across subjects. Each block was immediately preceded by a short "familiarization" session, in which the subjects were given one example of each condition used in the block that followed. Eight subjects participated in the experiment: six university students, who were paid for their services and the first two authors. All subjects were musically trained and had normal pure-tone thresholds (within 15 dB HL) in each ear around the stimulus frequencies used in the experiment.

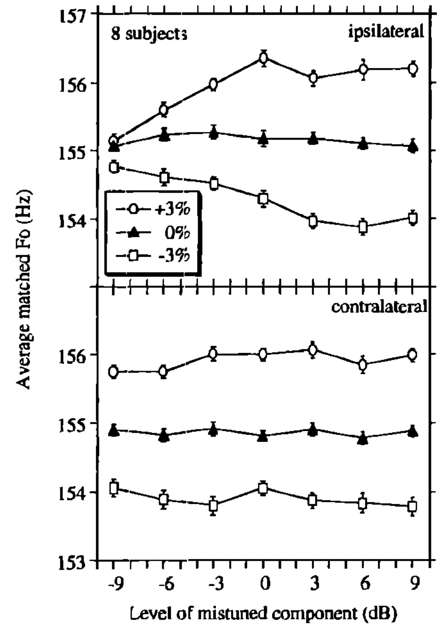

FIG. 6. Average matched fundamental frequencies to a 12-harmonic complex with a fundamental frequency of 155 Hz, when the fourth harmonic was mistuned by −3%, 0%, or +3% and presented at various levels relative to the other components. The fourth harmonic was presented either in the same ear (upper panel) or the opposite ear (lower panel) to the other 11 harmonics.

B. Results

Matched fundamental frequencies and their standard errors across subjects are shown in Fig. 6 as a function of the level of the mistuned component. The pitch matches to positive and negative mistunings are roughly symmetrical about the in-tune baseline for both contralateral and ipsilateral presentation. For the ipsilateral condition, pitch matches to the mistuned conditions converge on the baseline as the level of the mistuned component is reduced ($F_{12,84}=10.5$, $p<0.0001$), whereas this effect is not present with contralateral presentation ($F_{12,84}=1.3$, $p>0.2$). This difference between ipsilateral and contralateral presentation is highly significant ($F_{12,84}=4.8$, $p<0.0001$).

Figure 7 replots the results as mean shifts (half the dif-

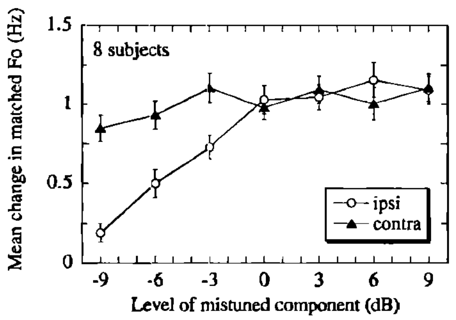

FIG. 7. Mean change in matched fundamental frequency (with standard errors) produced by mistuning the fourth harmonic of a 155 Hz, 12 harmonic complex by ±3% presented either ipsilateral or contralateral to the remaining harmonics. The abscissa shows the level of the mistuned component (relative to the other harmonics).

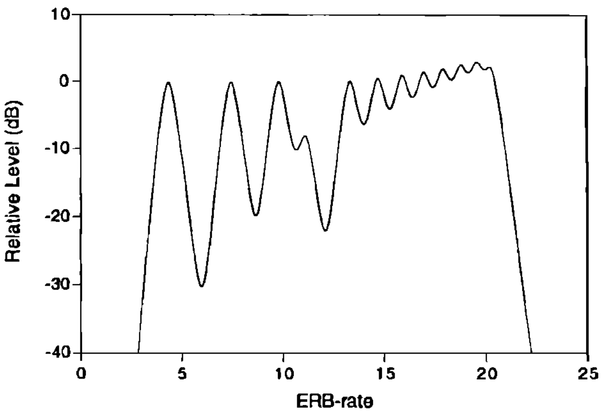

FIG. 8. Excitation pattern for the −3%, −9-dB ipsilateral condition in experiment 2.

ference between matches to corresponding positive and negative mistunings). As the level of the mistuned harmonic is reduced below the level of the others the mean shift decreases for the ipsilateral condition ($F_{6,42}=12.5$, $p<0.0001$), but not for the contralateral ($F_{6,42}=2.1$, $p>0.05$). Again this difference is highly significant ($F_{6,42}=6.0$, $p<0.0001$).

C. Discussion

The effect of changing the level of the mistuned component on its contribution to the pitch of the complex varies markedly with whether the mistuned component is presented ipsilaterally or contralaterally to the remainder of the complex. With ipsilateral presentation, reducing its level below that of the remaining components (0 dB) substantially reduces its contribution to the pitch of the complex; increasing its level above 0 dB does not further increase its contribution. However, with contralateral presentation, neither an increase nor a decrease in level of up to 9 dB has any significant effect on the pitch of the complex.

The substantial effect of a decrease in level in the ipsilateral but not contralateral conditions can be understood in terms of ipsilateral masking by neighboring components on the mistuned one. Figure 8 shows an excitation pattern (Moore and Glasberg, 1983) for the −3.2% mistuning, −9-dB ipsilateral condition. The mistuned fourth harmonic is still visible in the excitation pattern but is substantially less prominent than the neighboring peaks. With contralateral presentation the excitation level of the peak would be broadly similar, but the peak itself would be more prominent. A similar argument could be made in terms of the extent to which auditory-nerve fibers tuned to frequencies around that of the mistuned harmonic phase-locked to its frequency, rather than to those of neighbors. Our ipsilateral results confirm Moore *et al.*'s results, and with them argue against theories of pitch perception (Goldstein, 1973) which weight resolved frequency components independently of the degree to which they are resolved. Our contralateral results indicate that the level of presentation *per se* of a component is only important when it allows that component to be (partially) masked. The difference between our ipsilateral and contralateral results explains an apparent contradiction in earlier data brought to our attention by Adrian Houtsma. Houtsma (1981) found that substantial changes in the relative amplitude of the two (consecutive) harmonics of a two-tone complex had little effect on the accuracy with which subjects identified the pitch of the complex. By contrast, one study has found direct evidence, and another, indirect evidence that the relative amplitude of harmonics is important in their contribution to pitch.

First, (as described above) Moore *et al.* (1985b) found that changes in the amplitude of a single mistuned component in an otherwise harmonic 12-component complex influenced the size of the pitch shift caused by the mistuning. Second, Eggen and Houtsma (1986), compared their data on the pitch of the strike-note of a bell with that from a previous study by Greenhough (1976). Eggen and Houtsma examined the pitch changes to the strike-tone pitch caused by mistuning each of the first nine components of a natural bell sound. They found quite small pitch shifts when the double-octave component was mistuned. By contrast, Greenhough (1976) had found much larger pitch shifts when mistuning the double octave partial of a bell-like sound that was synthesized with equal-amplitude harmonics. Eggen and Houtsma concluded that their smaller pitch shifts were probably due to the low level of the double octave partial relative to the other components.

These apparently conflicting claims—that amplitude levels either are or are not important in determining the contribution that a particular harmonic makes to the pitch of a complex—are compatible with the findings of the present experiment 2. They can be reconciled by noting that Houtsma's 1981 experiment used dichotic presentation, whereas the others used either diotic or monotic.

The results of experiment 2 also address the issue of the contribution that adaptation might have made to the results of experiment 1. To the extent that adaptation effects can be equated to a decrease in physical level, the absence of any effect on pitch of reducing the level of a contralateral mistuned component argues against the effects that we found in experiment 1 being due to adaptation. Since the effects of onset asynchrony on pitch perception are also found with contralateral presentation of the asynchronous, mistuned component (Darwin and Ciocca, 1992), the present results add to the arguments (Ciocca and Darwin, 1993) that effects of onset asynchrony on pitch are not due to adaptation.

III. SUMMARY AND CONCLUSIONS

The two experiments presented here have demonstrated that:

(i) The pitch shifts induced in a complex tone when its fourth harmonic is mistuned are substantially reduced in size when the complex is preceded by a sequence of tones at the frequency of the mistuned component.

(ii) This reduction occurs whether the mistuned component and the preceding sequence are both presented to the same ear as the rest of the complex or to the opposite ear.

(iii) When the level of a mistuned component of a complex is changed by up to ±9 dB, the pitch of the complex only changes when the mistuned component is presented ipsilaterally and is reduced in level.

These results indicate that pitch perception can be influenced by sequential grouping constraints, and present additional evidence that adaptation plays little role in these and other grouping phenomena in pitch perception. The results of experiment 2 argue against models of pitch perception which weight resolved frequency components independently of the degree to which a harmonic is resolved.

ACKNOWLEDGMENTS

This work was supported by MRC Grant G9116710N to the first author. The second experiment was abstracted in Hukin and Darwin (1993). We are grateful to Adrian Houtsma for a number of helpful comments, and in particular for pointing our the relevance of experiment 2 to his earlier experiments alone and with Eggen.

Assmann, P. F., and Paschall, D. D. (1944). "Pitches of concurrent vowels," J. Acoust. Soc. Am. **95**, 2965.

Beerends, J. G. (1989). "The influence of duration on the perception of pitch in single and simultaneous complex tones," J. Acoust. Soc. Am. **86**, 1835–1844.

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.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organisation of Sound* (MIT, Cambridge, MA).

Carlyon, R. P., and Shackleton, T. M. (1994). "Some evidence that two separate mechanisms encode the FO's of resolved and unresolved harmonics," J. Acoust. Soc. Am. **95**, 2966.

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. (1992). "Listening to two things at once," in *The Auditory Processing of Speech: From Sounds to Words*, edited by M. E. H. Schouten (Mouton de Gruyter, Berlin), pp. 133–147.

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.

Darwin, C. J., Ciocca, V., and Sandell, G. R. (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., Pattison, H., and Gardner, R. B. (1989). "Vowel quality changes produced by surrounding tone sequences," Percept. Psychophys. **45**, 333–342.

Darwin, C. J., and Sutherland, N. S. (1984). "Grouping frequency components of vowels: when is a harmonic not a harmonic?," Q. J. Exp. Psychol. **36** A, 193–208.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (1982). "Measurement of pitch in speech: an implementation of Goldstein's theory of pitch perception," J. Acoust. Soc. Am. **71**, 1568–1580.

Eggen, J. H., and Houtsma, A. J. M. (1986). "The pitch perception of bell sounds," IPO Ann. Prog. Rep. **21**, 15–23.

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

Greenhough, M. (1976). "The analysis and synthesis of bell sounds," Proc. Acoust. **3**, 4:1–4:4.

Houtsma, A. J. M. (1981). "Pitch of unequal-amplitude dichotic two-tone complexes," J. Acoust. Soc. Am. **69**, 1778–1785.

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

Hukin, R. W., and Darwin, C. J. (1993). "Grouping in pitch perception: effects of level and ear of presentation of a mistuned component," Br. J. Audio. **27**, 349–350.

Hukin, R. W., and Darwin, C. J. (in press). "Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel," J. Acoust. Soc. Am.

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

Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

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

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1985b). "Thresholds for the detection of inharmonicity in complex tones," J. Acoust. Soc. Am. **77**, 1861–1868.