Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification

R. W. HUKIN and C. J. DARWIN
University of Sussex, Brighton, England

Previous experiments have shown that when a slightly mistuned harmonic of a complex tone starts more than about 80 msec before the remaining components, it makes a reduced contribution to the pitch of the complex. This contribution decreases to zero by about 300-msec onset asynchrony. In vowel perception, however, analogous experiments have shown that a much shorter asynchrony (around 40 msec) is enough to ensure that a component does not influence a vowel's phonemic category. The three experiments reported here demonstrate that this difference in the utility of onset time as a grouping cue does not arise because of differences in stimulus structure, but rather is due to the perceptual task. They show that the onset asynchrony needed in a pitch-matching experiment to remove the contribution that a mistuned component makes to the pitch of a vowel is the same as that needed to remove the contribution to the pitch of a flat-spectrum complex tone. They further show that a much smaller onset asynchrony is needed to perceptually remove the same harmonic from a vowel for the calculation of vowel quality. The implication of this result for models of auditory grouping is discussed.

A major computational problem facing the auditory system is to interpret the waveform reaching the ears as sound-producing events. Each event must be assigned the appropriate instantaneous properties such as location, timbre, and pitch, and their variation over time tracked to obtain such properties as melodic line, speech articulation, or spatial trajectory.

The ability to assign appropriate properties to each of potentially many simultaneous sound sources may be helped by the use of low-level primitive grouping mechanisms (Bregman, 1990) that heuristically exploit properties such as harmonicity and common onset time. For example, mistuning one harmonic of a complex tone by about 8% is sufficient to remove it from the perceptual calculation of the pitch of the complex (Moore, Glasberg, & Peters, 1985) and also to prevent it from contributing fully to the phonetic quality of a vowel (Darwin & Gardner, 1986). Such quantitative agreement argues for primitive grouping constituting an independent early stage in perception.

Onset asynchrony between one frequency component of a complex and the remainder acts as a powerful heuristic for perceptual grouping. Components that start or stop at roughly the same time are more likely to

be perceived to have originated from the same sound source than are components that start or stop at different times. However, the amount of onset asynchrony that is needed to perceptually segregate one component from a complex appears to vary with the experimental procedure.

Bregman and colleagues (Dannenbring & Bregman, 1978; Bregman & Pinker, 1978) played subjects an alternating sequence of two sounds—a pure tone and a two-tone complex. When the upper tone led the lower tone by 29 or 58 msec, the subjects were more likely to hear the upper note join the pure tone to form a separate auditory stream.

An onset asynchrony of about 30 msec is also sufficient to remove either an integral (Darwin, 1984b; Darwin & Sutherland, 1984) or an additional (Roberts & Moore, 1991) frequency component from the first formant region of a vowel, as measured by a change in the phonetic quality of the vowel along an /I/-/ε/ continuum. This result is compatible with a finding by Green and Dai (1992) in a profile analysis task. They showed that listeners are substantially worse at detecting an intensity increment in one component of an inharmonic complex when that component leads the rest of the complex by about 50 msec than when all the components are simultaneous. The profile analysis task ensures that the increment can only be detected by comparing levels across different frequencies.

Thus, when a subject has to judge which auditory stream a component belongs to, or has to compare the levels of components in order to judge timbre or vowel quality, an onset asynchrony of 30–50 msec is sufficient

Quentin Summerfield and Brian Moore made useful comments on an earlier version of the paper. We are grateful to Quentin Summerfield for drawing our attention to the relevance of Ken Robinson's work. Hukin's salary was paid by MRC Grant G9116710N. Equipment and software support were provided by grants from the SERC. The authors are at the Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, England.

to substantially influence their performance. However, a substantially greater onset asynchrony is needed to remove a leading frequency component when the subjects' task is to judge the pitch of a complex tone.

If one of the low-numbered harmonics of an otherwise harmonic complex is mistuned by a few percent of its frequency, the pitch of the complex will change slightly in the direction of the mistuning. However, with larger mistunings, the shift in pitch of the complex reaches a maximum (at around 3% mistuning) and then returns (by about 8% mistuning) to its original value (Moore et al., 1985). This result has been interpreted as showing that the auditory system can discount, for the purposes of calculating the pitch of a complex tone, components whose frequencies are implausibly distant from a harmonic value. The data can be modeled well by assuming that the contribution that a progressively mistuned frequency component makes to the pitch of a complex declines according to a Gaussian function (Darwin, 1992; Darwin, Ciocca, & Sandell, 1994). The size of the pitch shift at $\pm 3\%$ mistuning can be used to evaluate whether other grouping cues contribute to pitch perception.

If a 3% mistuned component starts before the remaining harmonic complex, its contribution to the pitch of that complex decreases (Ciocca & Darwin, 1993; Darwin & Ciocca, 1992). However, the striking difference with the previous results on onset asynchrony is that an onset asynchrony of more than between 40 and 80 msec is needed before *any* effect on pitch is shown; the complete removal of a mistuned component from the calculation of pitch is not complete until about 300-msec onset asynchrony.

In summary, a substantially greater onset asynchrony appears to be needed to remove a leading component from a harmonic complex for the purposes of calculating the pitch of the complex than is required to remove a leading component from a harmonic complex for the purposes of calculating its phonetic quality. Such lack of quantitative agreement is important, since it indicates at the very least that the information passed from a hypothetical primitive grouping mechanism to higher perceptual modules is graded rather than all-or-none.

The purpose of the three experiments reported here is to establish that the apparent difference between the two types of paradigm does not reflect differences in the individual complex sounds used. In particular, this paper seeks to establish whether the onset asynchronies needed to perceptually remove a mistuned component from a vowel and from a spectrally flat sound for the purpose of calculating pitch are the same.

In the first experiment, we establish that similar pitch shifts are obtained when the fourth harmonic of either a flat-spectrum complex tone or a vowel is mistuned. We then use this result to show in the second experiment that similar amounts of onset asynchrony are needed to remove this mistuned component from either the flat-spectrum sound or the vowel. In the third experiment, we show that less asynchrony is required to remove the contribution of this harmonic in a vowel identification task,

using a stimulus similar to those used in the first two experiments.

EXPERIMENT 1

In this experiment we compared the pitch shifts produced by mistuning the 4th harmonic of either a 12-harmonic flat-spectrum sound or an / ϵ / vowel. The aim was to confirm that the maximum pitch shift occurs at the same degree of mistuning for both sounds.

Method

Procedure. The two experiments follow the general paradigm introduced by Moore (Moore et al., 1985). On each trial, the subject heard two complex sounds separated by 500 msec of silence and adjusted the pitch of the second to match the first. The first complex (target) could contain a partial that was shifted in frequency from its harmonic value. The fundamental of the second sound was adjusted by moving a roller ball up or down. The pair of sounds could be repeated as often as required by the subject. A more detailed description of this procedure can be found in Darwin and Ciocca (1992).

The adjustable sound in all cases contained the first 12 harmonics of a 125 ± 8 Hz fundamental, with a spectral envelope that rolled off at -6 dB/octave. The target sound was either a synthesized vowel or a complex containing 12 equal-amplitude harmonics. In each case, the 4th harmonic could be mistuned. The vowel used in the experiment was synthesized with a fundamental frequency of 125 Hz and formant frequencies for F_1 , F_2 , and F_3 of 500 Hz, 2100 Hz, and 2900 Hz, respectively; it gave the percept of the short vowel / ϵ / and was similar to the vowels used previously by Darwin and Gardner (1986) in experiments on onset-time effects on vowel identification. The level of the target sounds was such that the amplitude of the 500-Hz component in each, presented separately, corresponded to 60 dB SPL.

Stimuli were presented monaurally through Sennheiser HD414 headphones in a double-walled IAC booth. Sounds were synthesized in real time at 44.1 kHz using custom software (Russell & Darwin, 1991) written for the 56001 processor on the Digidesign Audiomedia board, and they were played through the board's 16-bit DACs and antialiasing filters. The board was attached to an Apple MacIIx computer, which also controlled the experiment.

The frequency of the mistuned component of the complex tone or vowel was 460, 475, 485, 495, 500 (harmonic frequency), 505, 515, 525, or 540 Hz, corresponding to mistunings of -8% , -5% , -3% , -1% , 0% , 1% , 3% , 5% , and 8% of the harmonic frequency. To avoid bias effects due to subjects' remembering a particular pitch from previous trials, the fundamental frequency of the stimuli was varied randomly $\pm 4\%$ across trials. All stimuli were 90 msec in duration (including 5-msec raised-cosine attack and decay times). Five repetitions of each match were blocked by target type (complex tone or vowel), and the order of the blocks was randomized across subjects.

Subjects. Eight subjects took part in the experiment—7 university students, who were paid for their services, plus the first author. All had normal pure-tone thresholds (within 15 dB HL) around the stimulus frequencies used in the experiments, and all were experienced at pitch matching.

Results

Mean pitch shifts (calculated as half the difference between pitch shifts to positive and negative mistunings) and their standard errors across subjects are shown in Figure 1 as a function of the absolute percentage of mistuning of the fourth harmonic. The effect of mistuning

Figure 1. Mean pitch shifts (with standard errors over 8 subjects) produced in Experiment 1 by a mistuned fourth harmonic of a 125-Hz fundamental, 90-msec target, as a function of target type. The target was either a 12-harmonic flat-spectrum complex tone or the synthesized vowel / ϵ /.

was highly significant, as shown by a two-way analysis of variance (ANOVA) with repeated measures [$F(4,28) = 9.0, p < .001$]. There was no significant difference between the vowel and the tone in their pitch shifts as a function of mistuning [$F(4,28) = 0.08, p > .5$]. The maximum pitch shift occurs for a $\pm 3\%$ mistuning. Overall, the standard errors both across subjects and across trials (not shown) are slightly larger than those found in previous experiments. This is probably attributable to the differences in timbre between the target and matching stimuli (Moore & Glasberg, 1990).

EXPERIMENT 2

The second experiment addresses the main issue of the paper. Will the onset asynchrony needed to perceptually remove a mistuned harmonic from the calculation of pitch for the / ϵ / vowel be the same as that for the flat-spectrum sound? To answer this question, sounds simi-

Figure 2. Mean pitch shifts (with standard errors over 8 subjects) produced in Experiment 2 by $\pm 3\%$ mistuning of the fourth harmonic of a 125-Hz fundamental, 90-msec target, as a function of onset asynchrony and target type. The target was either a 12-harmonic complex tone or the synthesized vowel / ϵ /.

lar to those used in the first experiment were used, but with the fourth harmonic extended forward by various amounts to give different onset asynchronies between that harmonic and the rest of the sound.

Method

The fourth harmonic in the target tone or vowel was mistuned -3% , 0% , or 3% of the harmonic frequency. A mistuning of $\pm 3\%$ was chosen, since it was found in Experiment 1 (and previously) to give the maximum pitch shift. The mistuned component started 0, 40, 80, 160, or 320 msec before the remaining harmonics, but always stopped with them. The main part of the target sound lasted 90 msec. A total of 150 trials were used in the experiment (5 repetitions of 2 target types $\times$ 3 mistunings $\times$ 5 onset asynchronies). Eight subjects took part in the experiment, all of whom had participated in Experiment 1.

Results

Mean pitch shifts for $\pm 3\%$ mistunings are plotted as a function of onset asynchrony in Figure 2, together with their standard errors across the 8 subjects. For both the tone and vowel targets, the pitch shift decreases as onset asynchrony is increased beyond 80 msec. The main effect of asynchrony was highly significant, as shown by a two-way ANOVA with repeated measures [$F(4,28) = 9.6, p < .0001$]; the interaction between onset and target type was not significant [$F(4,28) = 1.2, p = .33$].

The results for the flat-spectrum complex tone replicate those of Darwin and Ciocca (Ciocca & Darwin, 1993; Darwin & Ciocca, 1992) in showing that upwards of 80 msec of onset asynchrony are required to obtain significant removal of a mistuned harmonic from the calculation of the pitch of a complex. The experiment extends this result to an / ϵ / vowel.

EXPERIMENT 3

Previous investigations into the effects of onset asynchrony in vowel perception have produced results indicating that, for some conditions, contributions by the asynchronous component may still be significant at asynchronies as great as 80 msec. When the vowel is of relatively short duration (56 msec), an onset asynchrony of 32 msec is enough to completely remove a component (Darwin, 1984b; Darwin & Sutherland, 1984), whereas for a long (320 msec) vowel, the asynchronous component still makes some contribution when it starts 320 msec before the remaining harmonics (Darwin, 1984a). Discrepancies between the contribution that an asynchronous component of a vowel makes in an identification rather than a pitch matching task may be to some extent attributable to differences in stimulus duration rather than to differences in task. Experiment 3 re-examines the effect of onset asynchrony in a vowel identification task, using the same stimulus duration as that for Experiments 1 and 2 (90 msec).

Method

The experimental stimuli consisted of seven members of a vowel continuum that varied in first formant (F_1) frequency to give a percept that changed from / I / at lower F_1 values to / ϵ / at

higher values. The first formant frequencies ranged from 396 Hz to 521 Hz in steps of about 21 Hz. The second and third formants were fixed at 2100 and 2900 Hz, respectively. The bandwidths of the first three formants were 90, 110, and 170 Hz, and the fundamental frequency was 125 Hz. The 500-Hz component of each vowel started 0, 40, 80, 160, or 320 msec before the remaining harmonics, but always stopped with them. The main part of the target sound lasted 90 msec. A further condition, in which the 500-Hz component was completely removed in each of the seven vowels, was also included.

The subjects were told that they would hear a vowel that could either be /I/ as in *pit*, or /ɛ/ as in *pet*; they might also hear a tone, which they were to ignore. Responses were entered using the "i" and "e" keys on a keyboard. Each sound followed 500 msec after the response to the previous stimulus.

A total of 420 trials were used in the experiment (10 repetitions of the 7 vowels in the original continuum $\times$ 5 onset asynchronies, plus 10 repetitions of the 7 vowels with the 500-Hz component removed). Eight subjects took part in the experiment, 7 of whom had participated in Experiment 1, plus the second author.

Results

Phoneme boundaries were estimated from the number of /I/ responses given by each subject to the 10 replications of each condition. Boundaries were calculated from a rescaled tanh function¹ fitted to the data. Figure 3 shows the average boundaries for the asynchronies used in the experiment together with their standard errors across the 8 subjects. As onset asynchrony is increased, the vowel boundary moves toward the "no 500" condition (corresponding to complete removal of the 500-Hz component). The main effect of asynchrony was highly significant, as shown by a two-way ANOVA with repeated measures [$F(4,28) = 7.5, p < .001$]. The shift in vowel boundary away from the 0-msec condition becomes significant at 80-msec asynchrony ($p < .01$; planned comparison), at which point the boundary is not significantly different from the "no 500" condition ($p > .5$; planned comparison).

Figure 3. Mean /I/-/ɛ/ phoneme boundaries along an F1 continuum (with standard error over 8 subjects) from Experiment 3. The single filled square is the phoneme boundary when the fourth harmonic (500 Hz) has been physically removed from the vowel. The open circles show the phoneme boundary for vowels in which the fourth harmonic leads the remaining harmonics by between 0 and 320 msec.

DISCUSSION

Experiment 2 showed that a flat-spectrum complex tone and a vowel need similar amounts of onset asynchrony to remove a mistuned fourth harmonic in a pitch-matching task. In neither case was there any reduction in the contribution that the mistuned component made to the pitch of the complex until the mistuned component started more than 80 msec before the rest of the sound. Onset asynchrony between 160 and 320 msec was needed before the contribution of the mistuned component to the pitch of the complex was reduced to zero.

These results contrast with those of Experiment 3, in which 80-msec onset asynchrony was sufficient to completely remove the fourth harmonic of a vowel from the calculation of vowel quality. Since the stimuli in Experiment 3 were, *mutatis mutandis*, identical to those of Experiment 2, the large quantitative difference between the effect of onset asynchrony in the pitch-matching task of Experiment 2 and the vowel identification task of Experiment 3 cannot be attributed to stimulus differences between the two experiments. Rather, the difference must lie in the nature of the task.²

The amount of onset asynchrony needed to perceptually remove one component from a vowel thus depends on the task. Less asynchrony is needed if the subject's task is to identify the vowel than if it is to match the vowel's pitch. Any grouping produced by primitive grouping mechanisms that precede the identification of pitch or the assignment of vowel identity cannot then be all-or-none. Rather, grouping processes have to provide graded information about the strength of a particular grouping to the mechanisms responsible for pitch or vowel identification.

The strength of this grouping may, however, be tied to a particular grouping parameter (such as onset asynchrony or harmonicity). The reasoning behind this possible restriction becomes clear if we compare the amount of mistuning that is needed to remove a component from the calculation of pitch or of vowel quality with the amount of onset asynchrony that is needed. Around 8% mistuning is needed for a component to be substantially removed from the calculation (1) of the pitch of either a flat-spectrum complex tone (Moore et al., 1985) or of a vowel (the present Experiment 1), or (2) of vowel quality (Darwin & Gardner, 1986). If anything, a mistuned harmonic is *more* likely to be incorporated into judgments of vowel quality than into judgments of pitch. By contrast, as we have confirmed in the present paper, an asynchronous harmonic is much *less* likely to be incorporated into judgments of vowel quality than into judgments of pitch. This dissociation argues against primitive grouping mechanisms merely indicating a graded overall strength of grouping between components without also indicating what cue is responsible for that grouping.

One possible explanation³ for why the pitch-matching task should be less sensitive to onset asynchrony than the vowel identification task, whereas the opposite may be true for sensitivity to mistuning, concerns the time course of decisions and the different durations of stimu-

lus needed by the listener to perform different identification tasks.

Robinson (Robinson, 1993; Robinson & Patterson, 1994) has shown that vowel quality can be identified at a much shorter stimulus duration (a single period of a voiced vowel) than is necessary for pitch chroma to be identified (asymptotic performance to distinguish the notes C, D, E, and F is eight periods). Our own experiments indirectly confirm this observation. We have found that a stimulus duration of about 90 msec is the minimum at which subjects can perform the pitch-matching task accurately, whereas a vowel stimulus duration of 56 msec still allows subjects to show sharp vowel identification boundaries.

The argument makes three assumptions: first, that the strength of the evidence that a *leading* harmonic does not belong in a complex gets *weaker* the longer the complex (since the harmonic has coexisted with the other harmonics for longer); second, that the strength of the evidence that a *mistuned* harmonic does not belong in a complex gets *stronger* the longer the complex (since the longer the complex the more accurately the mistuned pitch can be estimated). According to these two assumptions, at a short time into the complex, evidence that a leading component does not belong in the complex will be relatively strong, whereas evidence concerning a mistuned component will be relatively weak. Conversely, at a longer time into the complex, evidence that a leading component does not belong in the complex will be weaker, whereas evidence concerning a mistuned component will be stronger. If the decision about vowel category is made after only a short duration of a vowel has been heard, then asynchrony will have a strong effect and mistuning will have a weak one. If the decision about pitch is made after a longer duration of a vowel has been heard, then asynchrony will have a weaker effect and mistuning will have a stronger one. Our data indicate this pattern of results.

The third assumption concerns the time at which decisions are made. The argument requires that decisions are made early in the stimulus at different times. If they were based on the actual length of the utterance, then the argument would make no differential predictions on the basis of task. However, there is some evidence—which provided the motivation for Experiment 3—that a longer lead time is necessary for a component to be removed from a long vowel than from a short vowel. This result suggests that decisions about vowel identity are *not* made later for longer stimuli. This result, if true, does not necessarily imply that decisions about different aspects of sound are not made at different times, but it does show that they are affected by a physical rather than a task parameter.

The evidence that the effect of onset asynchrony in vowel identification is weaker for longer stimuli comes from comparing the results of Experiment 3 with those of previous experiments. With vowels that are 56 msec long, the original 500-Hz component of a vowel is effectively removed at an onset asynchrony of 32 msec (Darwin, 1984b). Vowels of 90 msec require about 80 msec

for complete removal (the present Experiment 3), but a longer onset asynchrony is needed for 320-msec vowels to remove additional energy at 500 Hz (Darwin, 1984a). Because of methodological differences between the last and the first two of these experiments, the comparison is not a water tight one, but the evidence is at least suggestive. It should also be noted that Roberts and Moore (1991) found that 40-msec onset asynchrony was sufficient for the asymptotic removal of two extraneous tones from a 150-msec vowel. More work is needed to resolve the time course of decisions.

If such an argument based on the time course of decisions fails to account for the data, then it might prove necessary instead to invoke arguments based on the nature of the categories being perceived. One might propose that the reason that only a short onset asynchrony was necessary to segregate a single harmonic from a vowel in a vowel identification task was that speech sounds almost never occur naturally preceded by an isolated fourth harmonic (although the onset of periodic voicing after aspirated stops may occur in the lowest four or five harmonics several centiseconds before it spreads to higher frequencies; Fischer-Jørgensen & Hutter, 1981). Constraints on periodic sounds in general are much less severe, with, for example, different musical instruments showing complex differences in the onset times of different harmonics (Risset & Wessel, 1982).

Whichever approach turns out to be the closer to the truth, the results reported in this paper are incompatible with the view that auditory grouping is carried out in an all-or-none fashion prior to and independently of the subsequent perceptual classification. The results are compatible with the view that auditory grouping mechanisms provide information on the relative strength of grouping, probably tied to the particular parameters (such as onset asynchrony or mistuning) responsible for the potential segregation.

REFERENCES

- BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organisation of sound*. Cambridge, MA: MIT Press, Bradford Books.
- BREGMAN, A. S., & PINKER, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, **32**, 19-31.
- CIOCCA, V., & DARWIN, C. J. (1993). Effects of onset asynchrony on pitch perception: Adaptation or grouping? *Journal of the Acoustical Society of America*, **93**, 2870-2878.
- DANNENBRING, G. L., & BREGMAN, A. S. (1978). Streaming vs. fusion of sinusoidal components of complex tones. *Perception & Psychophysics*, **24**, 369-376.
- DARWIN, C. J. (1984a). Auditory processing and speech perception. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 197-210). Hillsdale, NJ: Erlbaum.
- DARWIN, C. J. (1984b). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Society of America*, **76**, 1636-1647.
- DARWIN, C. J. (1992). Listening to two things at once. In M. E. H. Schouten (Ed.), *The auditory processing of speech: From sounds to words* (pp. 133-147). Berlin: Mouton de Gruyter.
- DARWIN, C. J., & CIOCCA, V. (1992). Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *Journal of the Acoustical Society of America*, **91**, 3381-3390.

- DARWIN, C. J., CIOCCA, V., & SANDELL, G. R. (1994). Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic. *Journal of the Acoustical Society of America*, **95**, 2631-2636.
- DARWIN, C. J., & GARDNER, R. B. (1986). Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. *Journal of the Acoustical Society of America*, **79**, 838-845.
- DARWIN, C. J., & SUTHERLAND, N. S. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic? *Quarterly Journal of Experimental Psychology*, **36A**, 193-208.
- FISCHER-JØRGENSEN, E., & HUTTERS, B. (1981). Aspirated stop consonants before low vowels—a problem of delimitation, its causes and consequences. *Annual Report of the Institute of Phonetics, University of Copenhagen*, **15**, 77-102.
- GREEN, D. M., & DAI, H. (1992). Temporal relations in profile comparisons. In Y. Cazals, L. Demany, & K. Horner (Eds.), *Auditory physiology and perception* (pp. 471-478). Oxford: Pergamon.
- MOORE, B. C. J., & GLASBERG, B. R. (1990). Frequency discrimination of complex tones with overlapping harmonics. *Journal of the Acoustical Society of America*, **87**, 2163-2177.
- MOORE, B. C. J., GLASBERG, B. R., & PETERS, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America*, **77**, 1853-1860.
- RISSET, J.-C., & WESSEL, D. L. (1982). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *The psychology of music* (pp. 25-58). New York: Academic Press.
- ROBERTS, B., & MOORE, B. C. J. (1991). The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels under conditions of asynchrony. *Journal of the Acoustical Society of America*, **89**, 2922-2932.
- ROBINSON, K. (1993). *Studies in timbre and pitch*. Unpublished doctoral dissertation, Cambridge University.
- ROBINSON, K., & PATTERSON, R. D. (1994). *Is pitch required in the extraction of timbre?* Manuscript submitted for publication.
- RUSSELL, P., & DARWIN, C. J. (1991). Real-time synthesis of complex sounds on a Mac II with 56001 DSP chip. *British Journal of Audiology*, **25**, 59-60.

NOTES

1. $y = 1/(1+e^{-\sigma(a-x)})$, where a is the boundary and σ is the slope parameter.
2. For the sake of completeness, it should be pointed out that the tasks also differ in that: (1) the spectral envelope was fixed in the pitch-matching task but was necessarily variable in the vowel identification task, and (2) the fundamental frequency was constant in the vowel identification task but varied slightly in the pitch-matching task.
3. We are grateful to Quentin Summerfield for this suggestion.

(Manuscript received April 27, 1994;
revision accepted for publication August 16, 1994.)