Vowel quality changes produced by surrounding tone sequences

C. J. DARWIN, HELEN PATTISON, and ROY B. GARDNER
University of Sussex, Brighton, England

In three experiments, we examined whether energy at the same frequency as one of a vowel's harmonics in the F1 region can be captured by a preceding or following sequence of tones. The position of the /I/-/e/ phoneme boundary along an F1 continuum was used to assess the extent of capture. The first two experiments showed that a sequence of tones at 500 Hz (56-msec duration at 10/sec) can perceptually remove added energy at 500 Hz from a steady vowel ($F_0 = 125$ Hz) that forms part of the sequence. The effect is detectable with one preceding tone, asymptotes with four, and is greater when two tones follow the vowel than when none do. Rising and falling sequences of tones (at 62.5-Hz intervals or at whole-tone intervals) differ in their effect. Falling sequences behave much like constant tones at 500 Hz but with less effect, whereas rising sequences show no evidence of removing the added tone. The second experiment replicated the first and also showed that when the vowel is embedded in a rising or a falling sequence of tones that continue after it, the following tones have no effect. The third experiment suggested that the different effects found with rising versus falling sequences are qualitatively predictable on the basis of the additive effects of their constituent tones rather than by virtue of their contour. The experiments indicated that sequences of repeating tones are much more effective at capturing a harmonic from a vowel than are sequences that follow a simple pattern. This result may reflect the operation of a principle of least commitment in auditory grouping.

In the experiments described here, we addressed the general question of how one distributes among potential sound sources the energy of the various frequency components arriving at the ears. Apart from the unusual circumstance of the soundproof room, it is rare for the sound arriving at the ears to have come from a single sound source. When more than one sound source is present, the energy at any particular frequency could be due to either source or to both.

The problem of recognizing particular sounds when they are heard against a background of other sounds would be simplified if the attribution of energy to one sound source or another could be made on the basis of simple organizational principles. Such principles could then provide possible groupings of different frequency components that were more likely than the original to correspond to separate sources.

Evidence for the operation of such organizational principles has been gathered by Bregman and his colleagues (see review in McAdams & Bregman, 1979) for tonal sequences and tonal complexes. Their results have been extended by us (e.g., Darwin, 1984b) to speech sounds.

In our previous experiments, we have looked at simple properties of sound such as onset time (Darwin, 1984a, 1984b; Darwin & Sutherland, 1984) and whether simultaneous frequency components are all harmonics of a common fundamental (Darwin, 1981; Darwin & Gardner, 1986). We have shown that these variables can influence vowel quality by perceptually removing some or all of the energy of a particular frequency component or group of components from the vowel percept to give an additional percept of an extra sound source. For example, if energy is added to an /I/ vowel at a harmonic frequency just above its first formant frequency, then the vowel sounds more /e/-like. But if that added energy starts or stops a few tens of milliseconds before or after the rest of the vowel, the added energy is heard as a separate tone, and the vowel's phonetic quality reverts back to the original /I/. Here the perceptual system is exploiting the fact that sounds from a common source tend to start at the same time, in order to segregate perceptually the added tone from the original vowel. Similar segregation can be achieved by mistuning one of the harmonics of a vowel; it is then heard as a different sound source, and thus makes a reduced contribution to the phonetic quality of the vowel.

Our experiments so far have looked only at simple variables that are local in frequency and time. Such experiments have probably exposed grouping mechanisms that are sensitive to very general properties of sound (such as the simultaneous onset of the components produced by a single simple source). We now turn our attention to more arbitrary, configurational properties. If a tone contained within a vowel also continues or completes a simple tem-

Pattison's and Gardner's salaries were paid by grants from the SERC. The first experiment was summarized in Darwin and Gardner (1987) and the first and third were presented to the 112th Meeting of the Acoustical Society of America (Pattison, Gardner, & Darwin, 1986). C. J. Darwin and Roy B. Gardner are at the Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, England; Helen Pattison's present address is: Department of Human Sciences, University of Technology, Loughborough LE11 3TU, England.

poral pattern, will perception remove it from the vowel? There has been no work on this question with respect to speech sounds, although Bregman and his colleagues have looked at the effect of various types of pattern on auditory grouping with nonspeech tonal sounds.

Steiger and Bregman (1981) have summarized the results of their own and earlier experiments with nonspeech sounds as providing "no evidence that the auditory system extrapolates trajectories" (p. 434). In their experiments, they looked at the effectiveness of different preceding sounds in capturing one of a pair of simultaneous, brief, harmonically related tone glides. They found that the most effective captor was a tone glide that was identical to the captured glide; in particular, such a glide was more effective than one that could be extrapolated into the captured glide.

Auditory grouping mechanisms may be more likely to perform interpolation rather than extrapolation. The clearest evidence comparing interpolation and extrapolation comes from Dannenbring's (1976) experiments on auditory continuity. A rising tone that is briefly masked near its center can be heard as continuing to rise behind the masker, while a tone that alternately rises and falls, and is masked at its points of inflection, is heard as an interpolation between the most extreme unmasked endpoints, rather than as an extrapolation to more extreme frequencies.

In the present experiments, we used sequences of discrete tones to ask whether their pattern can be extrapolated or interpolated to capture added tonal energy from a vowel.

EXPERIMENT 1

The basic paradigm was the same as that used previously by us (e.g., Darwin, 1984b). Energy is added at a harmonic frequency near to the first formant (F1) in a continuum of vowels that differ in F1. The original continuum gives a percept from /I/ (as in "bit") to /ε/ (as in "bet"), with a clear phoneme boundary at an F1 value a little below 500 Hz. When energy is added to this continuum at 500 Hz, the perceived first formant frequency changes, moving closer to 500 Hz. For sounds with an F1 of less than 500 Hz, the perceived F1 increases. For the /I/-/ε/ continuum used here, the phoneme boundary lies at around 480 Hz, so the effect of adding energy at 500 Hz is to shift the phoneme boundary to a lower nominal F1 value. Experimental manipulations can then be applied to try to regain the original position of the phoneme boundary, by encouraging the added energy to be perceived as an additional sound rather than as part of the vowel.

In this first experiment, we compared the effectiveness of various lengths and types (rising, falling, steady) of tone sequence at capturing this added energy. With rising and falling sequences, two tones were always included after the vowel so that the frequency of the tone during the vowel could be estimated by interpolation rather than by having to resort to extrapolation.

Method

Stimuli. The *original continuum* had 9 sounds, each 56 msec long with 16-msec rise/fall times, synthesized on a fundamental of 125 Hz. The first formant varied from 375 Hz to 543 Hz in 21-Hz steps giving /I/-like sounds at low F1s and /ε/-like sounds at high ones. The second through fifth formants had frequencies of 2300, 2900, 3800, and 4600 Hz, respectively. The synthesizer program calculated the transfer function of a cascade of five formant poles (after Klatt, 1980) and added together sine waves with the appropriate amplitudes and phases to give the original continuum. The amplitude of the 500-Hz component was increased by 6 dB to give the *+6-dB continuum*. Increasing the 500-Hz harmonic's energy by 6 dB is equivalent to adding in phase an extra tone with the same energy as the original. This added energy changed with position along the continuum, because of the changing F1 values.

The +6-dB continuum was then embedded in a variety of different tone sequences. The pure tones that made up the sequences had the same duration as the vowel (56 msec) and had the same energy as the added 500-Hz tone for a sound close to the phoneme boundary. They were played at a rate of 10/sec—that is, with 44 msec of silence separating them.

The 500-Hz sequences had 1, 2, 4, or 8 500-Hz tones before the vowel, and either 0 or 2 after it.

The *ascending sequences* had 1, 2, 4, or 8 tones before the vowel and 2 after. The tones rose either in equal-tempered whole tone intervals (frequency ratio of 1.12), or in half-harmonic intervals (frequency difference of 62.5 Hz). All the sequences converged on 500 Hz when they reached the vowel. The ascending half-harmonic sequence with 8 preceding tones was not used, since its initial tones would have been too low in frequency to be clearly audible.

The *descending sequences* were similar to the ascending sequences, with the necessary changes (see Figure 1).

Procedure. Seven of the nine members of each continuum were chosen on the basis of pilot data to estimate the phoneme boundaries, the same seven being used for each subject. Each member of each continuum was played 10 times in each of the 25 conditions, giving a total of 1,750 trials. The order was completely randomized to prevent range effects from influencing the boundaries for the different conditions.

The 25 conditions were as follows: original continuum and +6-dB continuum with no surrounding tones; 1, 2, 4, or 8 500-Hz tones before the +6-dB continuum, with either no or 2 tones after; ascending or descending sequences of 1, 2, 4, or 8 at half-harmonic or semitone intervals (with 2 tones after), but excluding the ascending 8 half-harmonic condition.

Twelve subjects heard these trials in a random order in two blocks, each with five replications of each stimulus. The subjects were graduate students, faculty, and technical staff of the School of Biology, aged between 23 and 42. Most had little or no musical training, all had experience of listening to synthetic vowels of the type used in this experiment, and none reported any hearing defects. They were tested individually in a sound-treated booth, and they listened over Sennheiser 414 headphones to the sounds directly from a VAX-11/780 computer (12-bit DACs, 10-kHz sampling rate, low-pass filtered at 4.5 kHz at a level of around 80 dB SPL for the original continuum). They responded with the "I" and "ε" buttons of a normal keyboard; each trial was presented 1 sec after the keypress for the previous trial. A practice session preceded the main experiment. Each subject's phoneme boundary was estimated in each condition by probit analysis (50% point), and the probit-estimated boundaries were checked graphically against the original data. The boundaries are expressed in terms of the F1 value used to synthesize the original continuum.

Results

We first consider the two continua that had no surrounding tones. The original continuum gives a phoneme bound-

Figure 1. Example of an ascending and a descending tone-sequence intersecting at the 500-Hz harmonic of a vowel that has had 6 dB of energy added to it. The sequences have four preceding and two following tones.

ary at around 480 Hz (the upper dotted line in Figure 2). Adding 6 dB to the original continuum gives the expected downward shift in the phoneme boundary to a value of around 460 Hz (the lower dotted line). These data points are replotted in Figure 3. We now look at the changes the various experimental conditions produced in this shift. Unless otherwise stated, the statistics are the results of analyses of variance on the individual phoneme boundaries. Analysis of variance generally were performed on the change in phoneme boundary from the +6-dB condi-

tion induced by 1-, 2-, 4-, and 8-tone sequences of two different types (e.g., with/without following tones). The factors were thus subjects (12) $\times$ number of preceding tones (4) $\times$ factor of interest.

500-Hz sequences. The number of preceding tones [$F(3,33) = 7.9, p < .0004$] and whether there were two following tones or not [$F(1,11) = 5.5, p < .04$] significantly influenced the phoneme boundary. The upward shift in the boundary from the +6-dB value (indicating perceptual segregation of the added tone from the vowel)

Figure 2. Phoneme boundaries from Experiment 1 for the /I/-/ε/ distinction for the original continuum differing in F1, one with +6 dB of energy at 500 Hz, and a variety of other continua where the +6-dB vowels have been embedded in a sequence of 500-Hz tones. The solid line shows the effect of the number of preceding tones when there are also two following tones. The dashed line shows the effect when there are no following tones.

Figure 3. Phoneme boundaries as in Figure 2, but for ascending or descending tone sequences (with two following tones) at whole-tone intervals.

was significant with one preceding tone [$t(11) = 4.33$, $p < .01$], and asymptoted by four. The following tones also served to increase the segregation of the added tone.

Descending sequences. The descending sequences behaved very similarly, whether at half-harmonic or whole-tone intervals, so only the whole-tone data are shown in Figure 3. The overall effect of the descending sequences was in the same direction as that of the 500-Hz sequences, showing that they also tended to capture the additional energy. Longer sequences produced a greater shift [$F(3,33) = 8.3$, $p = .003$], and there was no difference between the harmonic and the tonal sequences ($F < 1$). An asymptotic shift was again reached by four preceding tones.

Ascending sequences. The ascending sequences again showed no difference between the harmonic and the tonal conditions, but the overall pattern of their influence was very different from the 500-Hz and the ascending conditions. The 2, 4, and 8 preceding-tone conditions showed no significant shift from the +6-dB baseline, with only the single preceding tone showing an effect, which was broadly similar to that shown by the other single preceding-tone conditions. The ascending and descending sequences differed significantly as a function of their length [$F(5,55) = 7.5$, $p < .0001$].

Discussion

The most obvious pattern of results comes from the 500-Hz sequences. The two longest sequences (4 and 8 tones before, with two tones following) show boundaries indistinguishable from that of the original continuum. In terms of perceptual segregation, they have succeeded in capturing all of the tone added to the vowel. But an alter-

native explanation could be constructed around the concept of adaptation. Suppose that the effect of a preceding tone were to weaken the auditory system's response to subsequent tones at, or near its frequency. The response to the 500-Hz component in the vowel would then be reduced by adaptation to a previous tone.

A variety of auditory effects that demonstrate adaptation-like properties have been discovered (see Summerfield & Assmann, 1987), some of which are formally similar to what we found in the present experiment. The extent to which adaptation contributes in the present paradigm can be estimated by comparison with a previous experiment (Darwin, 1984a, pp. 204-205). There, the 500-Hz component of a 320-msec /I/-/ε/ continuum was increased by 8 dB. The additional tone necessary to do this either started and stopped at the same time as the vowel or was extended forward to start 1 sec before it. The 1-sec precursor abolished the phoneme boundary shift produced by the simultaneous +8-dB boost. However, if a short silent period of 30 msec was introduced between the 1-sec precursor tone and the vowel, the change in phoneme boundary induced by the 1-sec precursor was substantially reduced. A 30-msec silence is thus sufficient to reduce substantially adaptation effects that could be contributing to the effect in this paradigm. Since the silent intervals between each tone and between them and the vowel are 44 msec, it is unlikely that adaptation is making a substantial contribution to the effect of a single-tone precursor.

Further evidence that the effects of the tones in this experiment are not substantially due to a peripheral mechanism such as adaptation is demonstrated by the difference found between two and four preceding tones and by the

decreased effect when there are no following tones. So a sequence of tones *at the same frequency* can capture energy that has been added to a vowel at that frequency.

The ascending and descending series present a more complex picture. Although the descending series behave similarly to the corresponding 500-Hz ones, if less effectively, the ascending sequences have almost no effect. There are two types of explanation for this difference. First, it could be due to some difference in the way that the system treats ascending and descending sequences. For example, the system might find ascending sequences harder to extrapolate. But such an explanation is clearly ad hoc. Second, it could be that the difference arises because of the individual effects of the component tones in the sequences. The ascending sequence approaches the vowel from lower frequencies, and so the individual tones could be exerting some effect on the lower harmonics of the vowel. By contrast, the descending sequence approaches the vowel from higher frequencies. If perceptual grouping or adaptation were occurring on the basis of the individual tones' weakening the representation of their appropriate harmonics in the vowel, then we might obtain effects similar to those that we have found here. This line of thought was pursued after the second experiment had replicated and extended the results of the first.

EXPERIMENT 2

The first experiment showed that ascending and descending tone sequences differed radically in their apparent ability to capture a tone that had been added to a vowel. It also showed that a sequence of 500-Hz tones was more

effective at capturing a 500-Hz tone if the vowel was followed by two tones at the same frequency than if it was not.

In the second experiment, we looked at the effect of different numbers of 500-Hz, ascending or descending tones with and without following tones. Since in the first experiment we failed to find any difference between half-harmonic and whole-tone sequences, in the second experiment we used only half-harmonic sequences. Also, since the first experiment showed that the effects had asymptoted by four preceding tones, the eight preceding-tone sequence was not used.

Method

Twenty different conditions were used in this experiment: the original and +6-dB continua from the first experiment with no surrounding tones; the 500-Hz, ascending harmonic and descending half-harmonic sequences with 1, 2, or 4 preceding tones and either 0 or 2 following tones.

The procedure was similar to that of Experiment 1, with 12 subjects drawn from the same population as the first experiment, some of whom participated in both.

Results

The main features of Experiment 1 were replicated. In particular, the 500-Hz sequences with following tones gave the largest effect, with the original boundary being reached with four preceding tones (Figure 4). The following-tone sequences gave a significantly larger shift than the 500-Hz sequences without them [$F(1,11) = 7.1$, $p = .022$].

The striking difference between the ascending and descending sequences was also replicated (Figures 5 and

Figure 4. Phoneme boundaries as in Figure 2 but from Experiment 3.

Figure 5. Phoneme boundaries from Experiment 3 for descending 62.5-Hz intervals. The solid line shows the effect of the number of preceding tones when there are also two following tones. The dashed line shows the effect when there are no following tones.

6), both for the sequences with following tones [$F(2,22) = 6.37, p < .01$] and for those without them [$F(2,22) = 7.5, p < .005$]. The presence of following tones did not increase the effect of the ascending or descending tone sequences ($F < 1$), which asymptoted by two preceding tones. If anything, the effect of the following tones was to reduce the apparent capturing effect of the descending sequence.

Discussion

The main features of the first experiment were replicated. The 500-Hz sequences gave easily interpretable results. In both experiments the longer the sequence of tones preceding the vowel (up to four) the more likely it was to capture the tone added to the vowel. In addition, if two tones also followed the vowel, capture was even more likely.

Figure 6. Phoneme boundaries from Experiment 3 for ascending 62.5-Hz intervals.

The striking difference between ascending and descending sequences with two following tones was also replicated, with the descending series behaving as if it were capturing the tone (rather more weakly than the 500-Hz sequences), and the ascending series only showing weak evidence of capture with one preceding tone.

The new contribution of this experiment was to show that the two following tones at the end of the ascending and descending series did not enhance the capture, although as we have just seen, they did enhance it for the 500-Hz tone series.

In trying to explain the effects produced by the ascending and the descending tone series, let us propose that any perceptual capture that they show is not the result of their configurational properties, but rather simply the sum of the effects of their individual tones' acting by relatively peripheral processes (such as adaptation or perceptual contrast), which were confined to preceding tones.

If this is the case, then the reason for the difference between the ascending and the descending sequences is that they presented different frequency tones just before the vowel, which altered its internal representation. The descending sequences presented tones above the formant frequency, whereas the ascending sequences presented tones below it.

We could test this attempted explanation by comparing the results that we have obtained with different length sequences with the effect of preceding the vowels with single tones at different frequencies from the series at different times before the vowel. If our explanation was correct, then the shift for a particular sequence would be predicted by the sum of the shifts produced by its constituent tones. In Experiment 3, we made this test.

EXPERIMENT 3

In this experiment, we looked specifically at whether the previous results that we had obtained with two-tone ascending and descending sequences could be predicted from the effects of the individual tones making up those sequences. We again look at shifts in the phoneme boundary of the same /I/-/ε/ continuum with 6 dB of energy added at 500 Hz. This time we preceded the vowel with a single tone at various frequencies, starting 100, 200, or 400 msec before the vowel. A subset of these conditions could be used to assess the role of the individual tones that formed part of the two-tone ascending and descending series, and the whole set of results should give some insight into the possible mechanisms responsible for the effect of single preceding tones.

Method

As in Experiment 2, only half-harmonic (not whole-tone) intervals were used. The original and +6-dB continua were used again, as were the two-tone ascending and descending sequences without following tones. In addition, single-tone sequences were used, which presented 375-, 437.5-, 500-, 562.5-, and 625-Hz tones 100, 200, and 400 msec before the vowel. These frequencies and times were appropriate for the tones of the 1-, 2-, and 4-tone sequences.

The procedure was similar to that of the previous two experiments, with 12 subjects again drawn with partial overlap from the same pool as the previous two.

Results and Discussion

Overall effects of single preceding tones. The effects of preceding tones at different frequencies and at different times on the phoneme boundary are shown in Figures 7, 8, and 9. The effects produced by the individ-

Figure 7. Phoneme boundaries for Experiment 3, comparing the effect of preceding the +6-dB vowels with single tones at different frequencies starting 100 msec before the vowel.

Figure 8. Phoneme boundaries for Experiment 3, comparing the effect of preceding the +6-dB vowels with single tones at different frequencies starting 200 msec before the vowel.

ual tones that start 200 and 400 msec before the vowel vary with frequency [$F(4,44) = 6.5$ and 5.8 ; $p < .001$] in a predictable way. Tones higher than the phoneme boundary formant value tend to raise the boundary; those lower tend to lower it. These effects could be due to adaptation or grouping reducing the apparent energy in the vowel at the preceding tone's frequency. The effects produced by the tones starting 400 msec before the vowel

are a little weaker than those shown by the tones starting 200 msec before. The joker among the results, though a consistent one across all three experiments, is that the effect of tones starting 100 msec before the vowel (45-msec silent interval before the vowel) is *independent* of their frequency [$F = 1.5$, $p > .2$]. They all show a shift to higher formant frequencies in the phoneme boundary, indicating a lower perceived formant. If adaptation mecha-

Figure 9. Phoneme boundaries for Experiment 3, comparing the effect of preceding the +6-dB vowels with single tones at different frequencies starting 400 msec before the vowel.

nisms alone (cf. Summerfield & Assmann, 1987; Summerfield, Haggard, Foster, & Gray, 1984) had been operating, we might expect the shortest time-interval to show the largest, most frequency-dependent shifts. The effects at the shortest time-interval are large, but they do not vary with frequency. It is possible that adaptation effects are less frequency-selective than grouping effects and that they dominate the response to the short time-interval condition, but we know of no independent evidence that this should be the case.

Additive effect of individual tones. In addition to the conditions whose results are shown in the figures, Experiment 3 also included an ascending and a descending two-tone sequence before the vowel. These conditions can be thought of as made up of one tone with the shortest gap and one with an intermediate gap. The boundaries for the ascending and descending sequences were at 453 and 472 Hz, respectively.

If our previous explanation was correct, and the ascending and descending sequences were behaving just like a collection of isolated tones, then we should have found a particular pattern of results.

First, the shift produced by the two-tone ascending sequence should have been similar to the sum of the shifts produced by its two component tones presented in isolation. Second, the shift produced by the two-tone descending sequence should have been similar to the shifts produced by its two component tones.

The data for the descending sequence bear out this prediction well. The shift produced by the two-tone descending sequence (18.5 Hz) is close to the sum of the shifts (17.2 Hz) produced by the 625-Hz tone 200 msec before the vowel (10.1 Hz) and the 562.5-Hz tone presented 100 msec before it (7.1 Hz).

The ascending sequence gives qualitative support to the hypothesis. The positive shift produced by the 437.5-Hz tone 100 msec before (11.6 Hz) being reduced by a negative shift produced by the 375-Hz tone 200 msec before (−5.7 Hz), although not sufficiently to match the almost zero shift produced by the ascending two-tone sequence.

The differences observed in all three experiments between ascending and descending sequences can be explained at least qualitatively by the additive effect of the individual tones making up the sequences rather than by any gestalt effect of a patterned sequence.

In summary, the different effects of ascending and descending sequences of tones on the quality of a vowel containing an extra tone embedded in those sequences can plausibly be explained by the additive effects of the tones that make up each sequence. There is no evidence that the overall contour of the tone sequence leads to perceptual grouping.

Individual tones at 500 Hz gave large positive shifts only at the two shorter time-intervals; at 400 msec, the effect of an individual tone at 500 Hz was negligible. The increasing effect seen in the first two experiments for more

distant 500-Hz tones as part of a sequence is likely, then, to be due to higher processes than those responsible for the effect of the individual tones.

SUMMARY AND GENERAL DISCUSSION

On the basis of the three experiments, we can make the following points:

1. A sequence of tones at 500 Hz (each 56 msec long at 10/sec) can perceptually remove a tone at the same frequency added to a vowel that is embedded in the tone sequence.

2. Longer sequences of 500-Hz tones before the vowel remove more of the energy than do shorter sequences, asymptoting at four preceding tones.

3. If two 500-Hz tones also follow the vowel, more energy is removed than if none do.

4. Ascending and descending tone sequences have different effects, which asymptote by two preceding tones. These differences may be due to the additive effect of their individual tones.

5. Two tones following the vowel as part of a rising or falling sequence do not have any effect.

6. Individual tones played a short time before the vowel have, paradoxically, an effect on vowel quality that does not depend on their frequency, whereas those played a longer time before the vowel do have a frequency-specific effect.

There is clear evidence from these results that a repeating sequence of 500-Hz tones was capturing energy added to a vowel at 500 Hz, both more efficiently and by different mechanisms than those responsible for the effects that we have found with rising or falling sequences of tones.

The 500-Hz sequences were progressively more effective as they were extended before the vowel and after it, whereas the rising or falling sequences showed a more complex pattern of results. Their results can perhaps be explained by the additive effects of their individual tones, rather than by any configurational properties that they had. By contrast, the 500-Hz tone sequences showed increased effects with the addition of distant tones at time intervals beyond those at which individual tones have a detectable effect.

A possible conclusion from these experiments is that perceptual organization is more sensitive to straight repetitions of sounds than to the somewhat arbitrary rising and falling tone sequences that we used. Although an almost identically repeating sound is not uncommon in nature (for instance, as some resonator is repeatedly struck), sounds that rise in the intervals that we used only occur as a result of cultural convention or experimental ingenuity! If the auditory system follows the principle of least commitment—only making decisions that do not prejudice higher levels of processing—then we might expect the results that we have obtained. It is an open and interest-

ing question whether higher-order structures such as well-known tunes would allow more cognitive mechanisms to permit the capture of tones from embedded vowels.

REFERENCES

- DANNENBRING, G. L. (1976). Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology*, **30**, 99-114.
- DARWIN, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset-time. *Quarterly Journal of Experimental Psychology*, **33A**, 185-208.
- 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., & 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., & GARDNER, R. B. (1987). Perceptual separation of vowels from concurrent sounds. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 112-124). Dordrecht, The Netherlands: Martinus Nijhoff.
- 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.
- KLATT, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.
- MCADAMS, S., & BREGMAN, A. S. (1979). Hearing musical streams. *Computer Music Journal*, **3**, 26-43, 60. [Reprinted in C. Roads & J. Strawn (Eds.), *Foundations of computer music*. Cambridge, MA: MIT Press, 1985]
- PATTISON, H., GARDNER, R. B., & DARWIN, C. J. (1986). Effects of acoustical context on perceived vowel quality. *Journal of the Acoustical Society of America*, **80**, (Suppl. 1), S110-111.
- STEIGER, H., & BREGMAN, A. S. (1981). Capturing frequency components of glided tones: Frequency separation, orientation, and alignment. *Perception & Psychophysics*, **30**, 425-435.
- SUMMERFIELD, A., & ASSMANN, P. (1987). Auditory enhancement in speech perception. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 140-150). Dordrecht, The Netherlands: Martinus Nijhoff.
- SUMMERFIELD, Q., HAGGARD, M. P., FOSTER, J., & GRAY, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory effect. *Perception & Psychophysics*, **35**, 203-213.

(Manuscript received July 1, 1988;
revision accepted for publication October 7, 1988.)