Vowel-specific effects in concurrent vowel identification[a)]

Alain de Cheveigné[b)]
Laboratoire de Linguistique Formelle, CNRS/Université Paris 7, 2 place Jussieu, case 7003, 75251, Paris, France and ATR Human Information Processing Research Laboratories, 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan

(Received 5 May 1998; accepted for publication 15 March 1999)

An experiment investigated the effects of amplitude ratio (−35 to 35 dB in 10-dB steps) and fundamental frequency difference (0%, 3%, 6%, and 12%) on the identification of pairs of concurrent synthetic vowels. Vowels as weak as −25 dB relative to their competitor were easier to identify in the presence of a fundamental frequency difference (ΔF_0). Vowels as weak as −35 dB were not. Identification was generally the same at $\Delta F_0 = 3\%$, 6%, and 12% for all amplitude ratios: unfavorable amplitude ratios could not be compensated by larger ΔF_0's. Data for each vowel pair and each amplitude ratio, at $\Delta F_0 = 0\%$, were compared to the spectral envelope of the stimulus at the same ratio, in order to determine which spectral cues determined identification. This information was then used to interpret the pattern of improvement with ΔF_0 for each vowel pair, to better understand mechanisms of F_0-guided segregation. Identification of a vowel was possible in the presence of strong cues belonging to its competitor, as long as cues to its own formants $F1$ and $F2$ were prominent. ΔF_0 enhanced the prominence of a target vowel's cues, even when the spectrum of the target was up to 10 dB below that of its competitor at all frequencies. The results are incompatible with models of segregation based on harmonic enhancement, beats, or channel selection. [S0001-4966(99)01307-7]

PACS numbers: 43.66.Ba, 43.71.An, 43.71.Es [WS]

INTRODUCTION

Our everyday environment contains multiple acoustic sources, some of which we must attend to, others which we must ignore. Their amplitudes are rarely the same. With luck, the target might have a greater amplitude than its competitors, and be easy to hear. Or, at least such might be the case during some *time interval* that the auditory system can attend to, or within some *spectral region* that it can isolate by simple filtering. However, it often occurs that a target is weaker than the interference most of the time, and in most frequency regions, in which case the auditory system must employ a more sophisticated process to segregate one from the other.

Experiments that investigate the usefulness of harmonicity or ΔF_0 cues are usually performed with pairs of vowels of approximately equal amplitude (see de Cheveigné *et al.*, 1995, 1997b for reviews). Exceptions are an experiment by McKeown (1992), who introduced systematic intervowel amplitude differences in addition to the ΔF_0 parameter, similar experiments by Meyer and Berthommier (1995), and a series of experiments that used an adaptive paradigm to determine identification thresholds (Demany and Semal, 1990; Assmann and Summerfield, 1990; Summerfield, 1992; Summerfield and Culling, 1992; Culling, Summerfield, and Marshall, 1994; Culling and Summerfield, 1995). Thresholds with that paradigm ranged typically between −5 and −25 dB, depending upon the condition, demonstrating that relatively accurate identification (70%) of weak targets is possible, especially when segregation cues are present.

In a recent study (de Cheveigné *et al.*, 1997a), we performed a constant-stimuli experiment similar to that of McKeown, with ΔF_0's of 0% and 6% and target/competitor amplitude ratios of −20 to 20 dB in 10-dB steps. The effects of ΔF_0 turned out to be particularly strong for weak targets. The experiment to be reported here is an extension of that experiment to a wider range of amplitude ratios (−35 to 35 dB, in 10-dB steps), and a larger set of ΔF_0's (0%, 3%, 6%, 12%). In the previous experiment, vowels were synthesized in ''Klatt phase,'' which simulates the pattern of starting phases of natural vowels (Klatt, 1980). Corresponding partials had different starting phases, resulting in a complex pattern of vector summation at $\Delta F_0 = 0\%$. In the current experiment, corresponding partials of both vowels are given identical starting phases, with the result that the spectral envelopes of double-vowel stimuli are easier to predict. Beyond a general curiosity about performance over this larger parameter space, the experiment was motivated by several specific questions. (a) Is there a minimum target/competitor ratio below which a ΔF_0 is ineffective? (b) Is there a tradeoff between amplitude ratio and ΔF_0? In other words, might a large ΔF_0 be effective at target levels so low that a small ΔF_0 is ineffective? (c) What spectral features determine vowel identification in this situation? When the relative amplitude is varied, the spectrum of the stimulus varies between that of one vowel and that of the other. Features that allow identification of the first vowel (whatever they may be: formants, center of gravity, etc.) fade away, while those that allow identification of the other are strengthened. By correlating these changes with response rates, it is possible to know which cues are necessary for vowel identification. The pattern of errors itself may be instructive. The set of detailed responses for individual vowel pairs is thus a good test bed

[a)]Part of this work was described in an ATR technical report (de Cheveigné, 1997a).
[b)]Electronic mail: cheveign@ircam.fr

TABLE I. Formant frequencies (Hirahara and Kato, 1992) and bandwidths (BW).

	/a/	/e/	/i/	/o/	/u/	BW
$F1$	750	469	281	468	312	90
$F2$	1187	2031	2281	781	1219	110
$F3$	2595	2687	3187	2656	2469	170
$F4$	3781	3375	3781	3281	3406	250
$F5$	4200	4200	4200	4200	4200	300

for models of vowel perception, in the footsteps of Assmann and Summerfield (1989).

This paper is based on one experiment, the results of which are described in three parts. First, responses are averaged over vowel pairs and analyzed as a function of ΔF_0 and target/competitor amplitude ratio. Next, response rates for individual vowel pairs at $\Delta F_0 = 0\%$ are compared to their spectral envelopes, to determine what spectral features are important for perception of vowels in competition. Finally, the *increase with* ΔF_0 of the percentage of correct vowel responses is analyzed for each vowel pair. The aim is to make sense of ΔF_0 effects based on what the previous analysis told us about cues to identification. Such detailed analysis should provide insight as to the nature of F_0-guided segregation.

I. METHODS

Methods were similar to those of de Cheveigné *et al.* (1997a, b). Stimuli were based on synthetic tokens of Japanese vowels /a/, /e/, /i/, /o/, /u/, with frequencies and bandwidths listed in Table I [Single vowels are denoted between slashes (/a/), and so are vowel pairs (/ae/) when the target/competitor relationship is unimportant. When necessary to distinguish a target and a competitor, the vowels are separated by a slash (a/e) (target/competitor).] Vowels were synthesized at a sampling rate of 20 kHz. Their duration was 270 ms, including 20-ms raised-cosine onsets and offsets (250-ms ''effective duration'' between-6-dB points). Fundamental frequencies (F_0's) ranged from 124 to 140 Hz. The starting phase of each harmonic was given a ''random'' value that was the same for all vowels in all conditions. Such a random phase pattern produces vowel waveforms that are less peaky than sine, cosine, or Klatt phase (produced by the formulas of Klatt, 1980). All vowel tokens were scaled to the same standard rms amplitude after synthesis.

Vowels were paired, one vowel was scaled with respect to the other by a factor chosen between -35 and 35 dB in 10-dB steps, the two were added, and their sum was scaled to the standard rms amplitude. Double-vowel stimuli are thus weighted sums of their constituents. At $\Delta F_0 = 0$, the starting phase of each harmonic is the same in both vowels, and therefore its amplitude in the sum can be calculated without considering complex phase-dependent vector summation. The double-vowel spectral envelope is therefore the weighted sum of the individual vowel spectra, with the same weighting factors as the vowels themselves. This is illustrated for the vowel pair /oe/ in Fig. 1(a). Thick lines represent spectra of individual vowels, and each of the eight thin lines represents the spectrum of a weighted sum. Some are

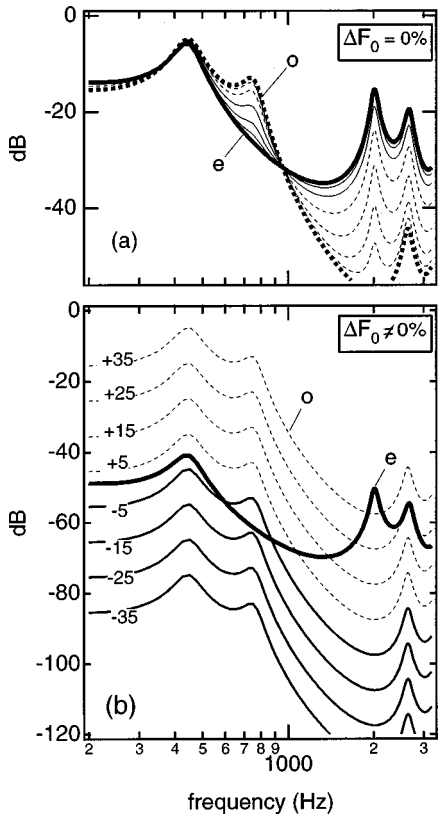

FIG. 1. (a) Spectral envelopes of single vowels /o/ and /e/ (thick lines) and weighted sums with amplitude ratios ranging from -35 to 35 dB in 10-dB steps (thin lines). Dotted lines correspond to stimuli dominated by /o/, full lines to stimuli dominated by /u/. In some places the thin lines are indistinct because they merged with the thick lines. Single vowels were not used as stimuli. The plot includes the first three formants ($F1$, $F2$, and $F3$), but excludes $F4$ and $F5$. (b) Spectral envelopes determining the amplitudes of partials of an /oe/ double vowel when $\Delta F_0 \neq 0\%$. Partials belonging to the harmonic series of /e/ are determined by the thick line, and those belonging to /o/ by the thin lines for each value of the o/e amplitude ratio. Dotted lines correspond to stimuli dominated by /o/, full lines to stimuli dominated by /e/. The origin of the ordinate scale is arbitrary: what counts is the relative levels between envelopes.

indistinct in places because they merged with one or the other thick lines. As the amplitude ratio varies from -35 to 35 dB, the spectral envelope of the pair gradually metamorphoses from that of one vowel to that of the other. Another example is the vowel pair /ou/ in Fig. 2(a). Single vowels (represented as thick lines) were not used as stimuli.

At $\Delta F_0 \neq 0$, the spectrum of the vowel pair is made up of two interleaved harmonic series, each a scaled version of the spectrum of a constituent vowel. This is illustrated for vowel pair /oe/ in Fig. 1(b). The thick line determines the amplitude of partials of /e/, and the thin lines determine that of partials of /o/ for each of the values of the o/e amplitude ratio. Dotted lines are for positive values of the o/e ratio, and full lines for negative values. The origin of the ordinate in this plot is arbitrary: of interest is the relative amplitude of one vowel versus the other. Notice for example that when the o/e amplitude ratio is -35, -25, or -15 dB, the spectrum is dominated everywhere by /e/. For other values, it is dominated by the first vowel in some places, and by the second in others. It is never dominated by /o/ everywhere. Another example is the vowel pair /ou/ in Fig. 2(b). These two vowel

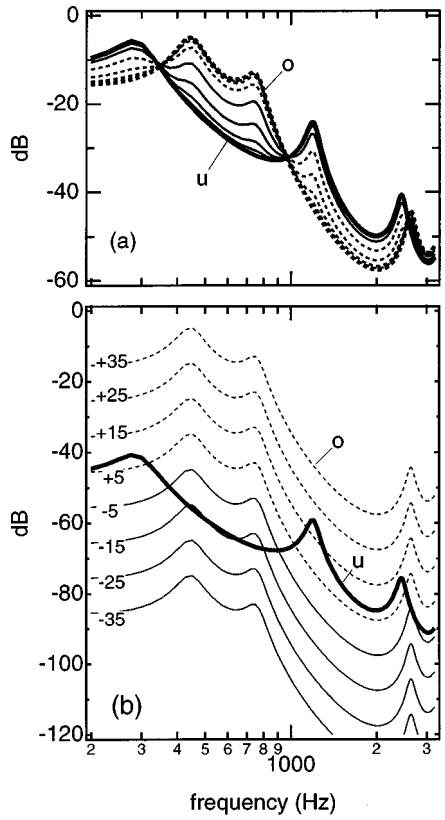

FIG. 2. Same as Fig. 1, for /ou/.

pairs are used to discuss pair-wise results in Sec. II. The other eight vowel pairs are described in the Appendix.

Fundamental frequencies were chosen by pairs arithmetically centered on 132 Hz, to obtain ΔF_0's of approximately 12% (124, 140 Hz), 6% (128, 136 Hz), 3% (130, 134 Hz), and 0% (132 Hz). F_0's were placed symmetrically about 132 Hz to avoid any influence of the mean F_0, and all conditions were repeated with both F_0 orders (low/high and high/low).

Vowels within a pair were always different. There were a total of (vowel pair=/ae/, /ai/, /ao/, /au/, /ei/, /eo/, /eu/, /io/, /iu/, /ou/)×(amplitude ratio=−35, −25, −15, −5, 5, 15, 25, 35 dB) ×(2F_0 orders)×(ΔF_0=0%, 3%, 6%, 12%)=640 double-vowel stimuli within a stimulus set. Ideally, the stimulus set should have contained a proportion of single vowels to make it consistent with the description made to the subjects (see below). However, the 25- and 35-dB amplitude ratio conditions are very similar to single vowels, and *stricto sensu* single vowels were omitted to reduce the stimulus set size.

Subjects were 15 Japanese students (seven male and eight female, aged 18 to 22 years) recruited for a series of ten experiments on concurrent vowel identification and paid for their services. The experiment described in this paper was the third of that series. The subjects were told that a stimulus could be either a single vowel, or two simultaneous vowels that were not the same, and they were instructed to report freely one or two vowels according to what they heard. There was no feedback. Stimuli were presented diotically via earphones, at a level of 63 to 70 dBA SPL (depending on the pair).

The response for each double-vowel stimulus was scored twice: each vowel in turn was nominated the ''target'' and the other the ''competitor,'' and then the roles were reversed. A target was deemed identified if its name was among the one or two vowels reported by the subject. The proportion of vowels correctly identified (constituent-correct, or target-correct identification rate) was calculated for each condition. The average number of vowels reported per stimulus was also recorded. For pair-wise analysis, response rates were recorded for all five vowels: the two constituent vowels and the three not present in the stimulus but sometimes reported nevertheless.

II. RESULTS

A. Results averaged over pairs

Identification rates were submitted to a repeated-measures analysis of variance (ANOVA) with factors amplitude ratio and ΔF_0. Probabilities reflect, where necessary, a correction factor applied to the degrees of freedom to compensate for the correlation of repeated measures (Geisser and Greenhouse, 1958). Amplitude ratio was highly significant [$F(7,98)=420.12$, $p<0.0001$, $GG=0.34$], as was ΔF_0 [$F(3,42)=96.32$, $p<0.0001$, $GG=0.45$]. Their interaction was also highly significant [$F(21,294)=27.60$, $p<0.0001$, $GG=0.20$] (a trivial consequence of the fact that the scores had a limited range). Identification rates are plotted in Fig. 3(a), together with rates measured in a previous experiment with different subjects (de Cheveigné *et al.*, 1997a). Results were similar between experiments, except that weak targets at $\Delta F_0=0\%$ were better identified in the present experiment. Identification rates increased as a function of target/competitor amplitude ratio. ΔF_0 had no measurable effect at −35, 15, 25, and 35 dB, but at all other amplitude ratios identification increased with ΔF_0. The greatest increase was between 0% and 3%. The step from 3% to 6% was significant at −15 dB [$F(1,294)=19.81$, $p=0.007$, $GG=0.20$], but not at other ratios. The difference between 6% and 12% was not significant at any ratio.

With respect to the first two questions raised in the Introduction, it can be said: (a) ΔF_0 effects exist down to a target/competitor ratio of −25 dB, but at −35 dB they are too small to be measurable. (b) At all amplitude ratios, large ΔF_0's (6% or 12%) are no more effective than small ΔF_0's (3%). The plateau beyond 3% or 6% in the improvement of identification with ΔF_0, previously observed for equal amplitude vowels, persists at target/competitor ratios down to −25 dB. There is no tradeoff between ΔF_0 and amplitude ratio.

The average number of vowels reported per stimulus was also submitted to a repeated-measures ANOVA with factors amplitude ratio and ΔF_0. Amplitude ratio was highly significant [$F(3,42)=83.32$, $p<0.0001$, $GG=0.40$], as was ΔF_0 [$F(3,42)=42.57$, $p<0.001$, $GG=0.42$] and their interaction [$F(9,126)=16.11$, $p<0.0001$, $GG=0.31$]. The number of vowels reported is plotted in Fig. 3(b), together with data from the previous study (de Cheveigné *et al.*, 1997a). The number of vowels reported was greater at 3% than 0% at all amplitude ratios except 35 dB. The increment between

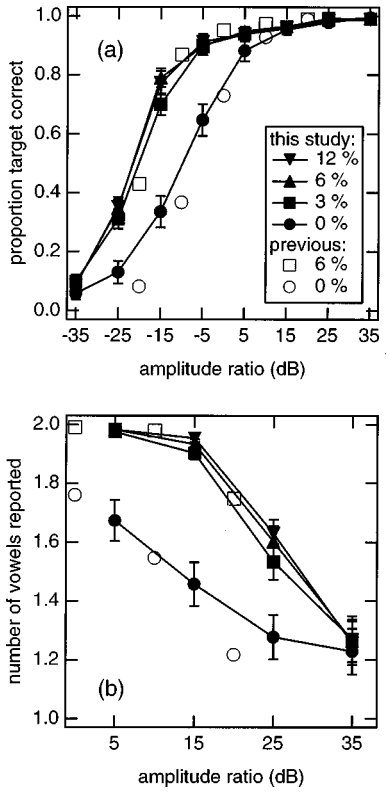

FIG. 3. (a) Target identification rate as a function of target/competitor amplitude ratio, for ΔF_0's of 0%, 3%, 6%, and 12%. Error bars represent ±one standard error of the mean. Open symbols are from a previous study (de Cheveigné, 1997a). (b) Average number of vowels reported per stimulus as a function of the amplitude ratio between vowels.

3% and (6%, 12%) was not significant, except at 25 dB [$F(1,126)=8.47$, $p=0.027$, $GG=0.31$]. The increment between 6% and 12% was not significant at any ratio.

B. Pair-wise results

Originally, the experiment was designed to exploit data averaged over vowel pairs (as in most studies of this type). However the detailed vowel-pair-specific responses can be used to find what features underlie the perception of steady-state vowels (at least when they compete with other vowels). Response rates at each amplitude ratio can be compared to the shape of the spectrum at the same ratio, and responses for each vowel can be related to the presence and strength of particular spectral features. This proposal has at least three weaknesses: (a) The amount of information is considerable and expensive to describe. (b) Pair-specific data are less reliable than average data (30 responses per data point vs 600), and reliability itself is difficult to evaluate. (c) The spectral envelope, with respect to which the data are analyzed, may not be representative of spectral features available to the auditory system. These weaknesses are counterbalanced by the insight that vowel-pair-specific patterns provide concerning vowel identification.

To keep the description within reasonable limits, two pairs (/oe/ and /ou/) are selected for detailed presentation. The latter (/ou/) is typical of the majority of vowel pairs as far as ΔF_0 effects are concerned, whereas the former combines two exceptional patterns: o/e (/o/ target and /e/ background) showed the largest increase in correctly identified target vowel as a function of ΔF_0, whereas e/o showed the smallest. The pair /ou/ illustrated a previous modeling study (de Cheveigné, 1997b), and both pairs were described in detail in a previous experimental study (de Cheveigné *et al.*, 1997a). Other pairs are described in the Appendix.

Patterns of excitation along the basilar membrane (Moore and Glasberg, 1983) are an attractive alternative to spectral envelopes. They show *less* detail than spectral envelopes at high frequencies because of limited spectral resolution and/or masking, and *more* detail at low frequencies where individual harmonics are resolved by the ear. Spectral envelopes were nevertheless, preferred for the following reasons. (a) Formant positions, that are known to correlate well with vowel quality, are evident in spectral envelopes. (b) Excitation patterns display an F_0-dependent structure at low frequencies that makes their plots less legible. Vowel identification is known to be relatively insensitive to F_0, suggesting that this sort of detail is not helpful in the present discussion. (c) Excitation patterns embody one particular form of auditory processing, but there may be others that they do not reflect as well (for example, vowel quality might be derived from time-domain cues, as suggested by Meddis and Hewitt, 1992). A description closer to acoustics implies less commitment in this respect.

1. ANOVAs for the 35-dB subset

Prior to the description of pair-wise results, a series of ANOVAs was performed on a subset of conditions to derive a rough estimate of the variability of responses. Analysis of variance of the full data set was impractical because of its size and complexity, and possibly unjustified because of the severe inhomogeneity of variances (many data points were at ceiling or floor). In vowel pairs with a 35-dB amplitude ratio, the weaker vowel affects perception in a negligible way: responses do not vary according to its identity, and rates are hardly affected by ΔF_0 [Fig. 3(a)]. It is reasonable to assume that these stimuli are essentially identical to the stronger vowel by itself. Each vowel appeared as the 35-dB-stronger vowel four times in the stimulus set, and one can take advantage of this repetition to get an estimate of the variability of responses in these conditions.

Five independent ANOVAs were performed, one for each response category (vowel). For each, the dependent variable was the proportion of trials for which (a) that vowel was reported, and (b) the stronger vowel was some other vowel. The nature of the stronger vowel was the independent factor (four levels), while the weaker vowel served as a random factor (four levels). Results are given in Table II. The lower two lines of the table show that, with the exception of /i/, the number of times a spurious vowel was reported depended significantly ($p=0.05$) on the nature of the stronger vowel. The pattern of means is discussed later. The important thing to note for now is that the *standard deviation* of responses ranged between 3% and 5.5%. These values hold for the 35-dB ratio, but they will nevertheless be considered representative of variability of responses rates of weak target vowels and spurious vowels at other ratios. Based on this estimate, one can be sure that the effects described in this

TABLE II. Response rates for each of the five vowels, at a 35-dB amplitude ratio and $\Delta F_0 = 0\%$. The stronger vowel was reported on all trials: rates shown are for each of the four other vowels that could eventually be reported together with it. They are given as a function of the nature of the stronger vowel. The significance of that factor is tested by a set of ANOVAs, one per column, using the ''weaker vowel'' factor as a random factor (lower two lines). Standard deviations are calculated from the part of variance that is not accounted for by the ''stronger vowel'' factor.

Response vowel	/a/	/e/	/i/	/o/	/u/
stronger=/a/	···	7.1%	3.6%	0.9%	6.3%
stronger=/e/	8.0%	···	8.9%	1.8%	18.8%
stronger=/i/	4.5%	1.8%	···	0.0%	18.8%
stronger=/o/	0.0%	0.1%	2.7%	···	14.3%
stronger=/u/	13.4%	0.1%	3.6%	8.9%	···
standard deviation	3.62%	3.02%	3.52%	3.04%	5.48%
$F(3,3)$	9.83	3.95	2.6	7.19	4.61
p	0.0015	0.036	0.1	0.005	0.023

paper (always greater than 10%, and often consistent across several similar conditions) are reliable.

2. Pair-wise results at $\Delta F_0 = 0\%$

Results are presented in this section for two vowel pairs, /oe/ and /ou/. The other eight are described in the Appendix. Responses to the two constituent vowels and to the three vowels not present in the stimulus (spurious responses) are considered for each pair. For the vowel pair /oe/, response rates for each of the five vowels are plotted in Fig. 4(a). Thick lines represent the proportion of responses for which the constituent vowels were correctly identified. Thin lines represent response rates for the three other vowels.

Let us first consider responses for the first constituent, /o/. Identification was poor up to an o/e amplitude ratio of −15 dB, and good at 5 dB and above. By comparing responses for each value of the abscissa in Fig. 4(a) with the corresponding envelope in Fig. 1(a), changes in the spectrum can be related to changes in identification. $F1$ is the same for both vowels, and the shape and absolute amplitude of that part of the spectrum hardly change across conditions. Changes in identification can therefore be attributed to other features, presumably in the $F2$ region of /o/.

The spectral envelope for an o/e ratio of 5 dB (for which identification was good) is the lowest dotted line near formant $F2$ of /o/ in Fig. 1(a). There is quite a clear peak at $F2$ of /o/, and below 1 kHz the spectrum hardly differs from that of /o/. However, at 2 kHz there is a clear peak at $F2$ of /e/ that is not usually found in the spectrum of /o/. The presence of this spurious cue did not prevent /o/ from being identified (the cue did not lack perceptual prominence, judging from the high identification rates for /e/ at the same ratio). This illustrates a first principle: identification of a vowel is not prevented by the mere presence of cues belonging to another vowel.

Both vowels of the /oe/ pair were simultaneously identified over a range of relative amplitudes (5 and 15 dB), despite the fact that both vowels had the same $F1$. This illustrates a second principle: a formant peak may be shared between vowels. The rule of exclusive allocation (Bregman, 1990) does not apply. Bregman noted many other exceptions to this rule.

The spectral envelope for an o/e ratio of −5 dB (for which identification fell to 50%) is the uppermost full line near $F2$ of /o/. The envelope for −15 dB (for which identification fell to 20%) is the next-to-uppermost. The drop in identification rate was presumably caused by the fading away of the spectral peak at $F2$ of /o/. This illustrates a third principle: identification of a vowel requires the presence of cues to both its formants $F1$ and $F2$. The spectrum near $F2$ of /e/ hardly changed between these conditions: the drop in identification of /o/ had nothing to do with competition from cues to /e/.

Let us now consider responses for the second constituent vowel, /e/. Identification was relatively poor at o/e ratios of 25 and 35 dB, and relatively good at 15 dB and below. Referring to Fig. 1, an o/e ratio of 15 dB corresponds to the next-to-uppermost dotted line near $F2$ of /e/. Ratios of 25 and 35 dB correspond to the lowest dotted lines. The spectrum hardly differs between conditions in the common $F1$ region, or near formant $F2$ of the competing vowel /o/.

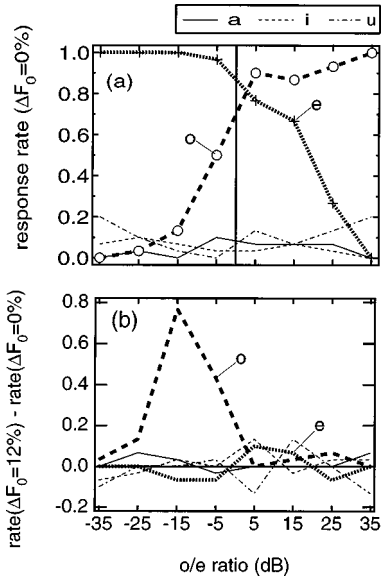

FIG. 4. (a) Response rates for vowel pair /oe/ as a function of the o/e amplitude ratio, for the constituent vowels /o/ and /e/ (thick lines) and the three vowels not in the stimulus (/a/, /i/, /u/). (b) Difference in response rates between $\Delta F_0 = 12\%$ and $\Delta F_0 = 0\%$.

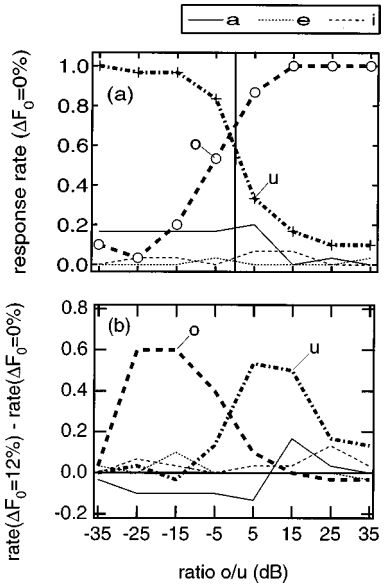

FIG. 5. Same as Fig. 4, for vowel pair /ou/.

Rather, identification seems to have varied with the spectral changes in the region of $F2$ of /o/. It is not certain what caused the loss of perceptual prominence of the $F2$ cue: upward masking from the lower part of the spectrum, masking by ambient noise, or some other cause.

Similar comparisons can be made for the pair /ou/, between responses [Fig. 5(a)] and spectral envelopes [Fig. 2(a)]. This pair is somewhat atypical in that the range over which both vowels are identified is rather restricted. There is a tradeoff between responses as if, contrary to other vowel pairs, simultaneous perception of both vowels were difficult.

It is interesting to note the large proportion of spurious /a/ responses for o/u amplitudes ranging from -35 to 5 dB. Referring to Fig. 2(a), it appears that the $F2$ formants of /o/ and /u/ were being interpreted as formants $F1$ and $F2$ of /a/ (see Fig. A1 for a picture of /a/, or Table I for its formant frequencies). This interpretation coexisted with /o/ responses (at low u/o ratios) and /u/ responses (at higher u/o ratios), implying that the same formant was being perceptually shared between /a/ and one of the constituent vowels (another counterexample for the rule of exclusive allocation). According to which vowel was reported in addition to /a/, this interpretation left the $F1$ of the other vowel ''orphaned;'' that is, not interpreted as belonging to a vowel. It is conceivable that this is what limited the number of times this otherwise plausible interpretation was made.

A final detail to notice is the high proportion of /u/ responses at high o/u ratios. This would be ascribed to chance if it weren't that /u/ responses were common for all pairs dominated by /o/ (Table II). Such /u/ responses were also common for pairs dominated by /e/ or /i/. The effect is easier to explain for /i/ (/u/ and /i/ have similar $F1$'s) than for /o/ or /e/: nothing in their spectrum seems to support perception of /u/.

As detailed in the Appendix, the three ''principles'' hold for most other vowel pairs. Identification is not disrupted by the mere presence of spurious cues, the same formant cue can be shared by both vowels, and identification of a vowel requires that cues for both its $F1$ and $F2$ be prominent. However, it is not clear what determines the prominence of a formant cue. Referring back to o/e when the amplitude ratio was reduced from 5 to -15 dB, the peak at $F2$ of /o/ underwent at least three changes: its absolute amplitude was reduced, its amplitude relative to neighboring peaks was reduced, and its local prominence (''sharpness,'' and height relative to the neighboring spectrum) was reduced (Fig. 1). The first factor is unlikely to be decisive (see the Appendix), but it is harder to decide between the other two factors.

3. Pair-wise patterns of ΔF_0 effects

The previous description of vowel identification at $\Delta F_0 = 0\%$ constitutes a good backdrop against which to analyze vowel-specific patterns of ΔF_0 effects. Results are again presented in detail for vowel pairs /oe/ and /ou/. For the vowel pair /oe/, the *difference* in response rates between $\Delta F_0 = 12\%$ and $\Delta F_0 = 0\%$ is plotted in Fig. 4(b). The difference is plotted as a function of the o/e amplitude ratio for both component vowels (/o/ and /e/, thick lines) as well as for the other three vowels (/a/, /i/, /u/, thin lines). A positive value implies that the response rate for that vowel increased when a ΔF_0 was introduced. These plots may be compared to the ''baseline'' rates measured at $\Delta F_0 = 0\%$ for the same pair [Fig. 4(a)].

Let us consider first the vowel /o/. At an o/e amplitude ratio of -15 dB, the proportion of stimuli for which /o/ was reported increased by almost 0.8, the largest effect of all vowel pairs. As /o/ has the same $F1$ frequency as its competitor, it was argued previously that identification was limited by the (lack of) prominence of cues to its $F2$. The spectral envelope of the vowel pair at $\Delta F_0 = 0\%$ is shown in Fig. 1(a). At an o/e ratio of -15 dB, the envelope shows a modest shoulder at $F2$ of /o/ (second-to-topmost full line). Apparently the ΔF_0 increased the prominence of this cue, allowing /o/ to be identified on almost all trials. ΔF_0 was also effective at -5 dB, but the effect is smaller because of a ceiling effect. It was also ineffective at -25 dB, presumably because the cue was too weak to be saved by the ΔF_0.

Figure 1(b) shows the spectral envelopes that determine the relative amplitudes of partials belonging to the two harmonic series that make up the double vowel at $\Delta F_0 \neq 0\%$. At an o/e amplitude ratio of -15 dB, the spectral envelope of /o/ at $F2$ is slightly below that of its competitor (-1.7 dB), and at -5 dB it peaks 8.3 dB above its competitor's envelope. The ΔF_0 was effective in both situations, but not at -25 dB when $F2$ of /o/ peaked at -11.7 dB below the envelope of its competitor. This gives a first hint of the conditions under which a ΔF_0 may be effective. A more complete discussion follows later.

Let us now consider the vowel /e/. The e/o target/competitor pair was exceptional in showing *no* increase of response rate with ΔF_0 [thick dotted line in Fig. 4(b)]. The lack of improvement for /e/ may be explained in at least two ways. First, the $F2$ of /e/ is prominent whatever the amplitude ratio (Fig. A1), and has little to gain from ΔF_0. Indeed, the /e/ response rate at $\Delta F_0 = 0\%$ was relatively high at all amplitude ratios. Second, its frequency is relatively high

(2031 Hz). Hypothetical within-channel segregation mechanisms (de Cheveigné, 1997b) might be less effective at higher frequencies.

For the vowel pair /ou/, the difference in response rates between $\Delta F_0 = 12\%$ and $\Delta F_0 = 0\%$ is plotted in Fig. 5. Let us consider first the vowel /o/. The ΔF_0 effect was negligible at -35 dB. It was strong at o/u ratios of -25 and -15 dB, and weaker for higher ratios because of ceiling effects. The spectral envelope of the vowel pair at $\Delta F_0 = 0\%$ is shown in Fig. 2(a). $F1$ and $F2$ of /o/ show up as shoulders that are modest at -15 dB (second-to-topmost full line), and even more modest at -25 dB. Apparently the ΔF_0 increased their prominence, allowing /o/ to be identified on 60% of all trials at -25 dB, and 80% at -15 dB.

Figure 2(b) shows the spectral envelopes that determine the harmonic series of the constituent vowels at $\Delta F_0 \neq 0\%$. At an o/u amplitude ratio of -15 dB the $F1$ and $F2$ of /o/ dominate the interformant valley of the envelope of /u/ by 1 and 3.7 dB, respectively. At -25 dB, they peaked at 9 and 6.3 dB, respectively, *below* that envelope, but ΔF_0 was nevertheless effective. It was, however, ineffective at -35 dB, where $F1$ and $F2$ of /o/ peaked, respectively, at 19 and 16.3 dB below their competitor's envelope.

Let us consider next the vowel /u/. The ΔF_0 effect was strong at o/u ratios of 5 and 15 dB, and weaker at lower o/u ratios because of the ceiling effect. For higher ratios (weaker /u/), the effect was also small. The spectral envelope of the vowel pair at $\Delta F_0 = 0\%$ is shown in Fig. 2(a). $F1$ and $F2$ of /u/ show up as small peaks at an o/u ratio of 5 dB, and as modest shoulders at 15 dB (topmost and second-to-topmost dotted lines near these formants). These cues are even more modest at 25 and 35 dB. Evidently the ΔF_0 increased their perceptual prominence. Referring to Fig. 2(b), at an o/u amplitude ratio of 5 dB, the $F1$ and $F2$ of /u/ dominate the envelope of /o/ by 3.4 and 14.7 dB, respectively. At 15 dB, $F1$ peaked at 6.6 dB below the envelope of /o/, and $F2$ at 4.7 dB above it.

It is interesting to note that ΔF_0 reduced the number of spurious /a/ responses at o/u ratios of -25 to 5 dB. This might be explained in two ways. According to one, when formants $F2$ of /o/ and /u/ are excited with different F_0's, they are harder to group together as formants $F1$ and $F2$ of /a/. According to the other, ΔF_0 increased the prominence of the $F1$ of /u/ that would be orphaned at -25, -15, and -5 dB, or that of $F1$ of /o/ that would be orphaned at 5 dB. The latter explanation is favored later on in the paper. At an o/u ratio of 15 dB, the proportion of /a/ responses *increased* with ΔF_0. This is probably a consequence of the increased prominence of $F2$ of /u/, interpreted as $F2$ of /a/.

Patterns of ΔF_0 effects for other pairs are detailed in the Appendix.

III. DISCUSSION

Results averaged over pairs were similar to those found in a previous experiment (de Cheveigné, 1997a). ΔF_0 increased both the average number of vowels reported and the proportion of vowels correctly identified, particularly for the weaker vowel of each pair. New is the observation that ΔF_0 effects were measurable down to -25 dB (-20 dB in the previous study), but not -35 dB. New also is the observation that response rates were roughly the same at ΔF_0's of 3%, 6%, and 12%, whatever the amplitude ratio. There was no tradeoff between ΔF_0 and amplitude ratio: a less favorable ratio could not be compensated by a larger ΔF_0. The ΔF_0 effect was greatest at -15 dB, both on average and for most vowel pairs, which justifies the choice of this ratio to reduce ceiling effects in this kind of experiment.

With some exceptions, pair-wise patterns of ΔF_0 effects were similar across vowel pairs. Identification usually improved over a range of target/competitor ratios (for example, -25 to -5 dB for o/u), which argues for the usefulness of this cue in everyday life. The pair-wise results may be interpreted with respect to various hypotheses and theories of sound organization and segregation.

The asymmetry of effects of ΔF_0 on both vowels of /oe/ is one striking exception to the auditory scene analysis rule, according to which effects of primitive segregation should be symmetric (Bregman, 1990). A similar exception was found in the identification of pairs of harmonic and inharmonic vowels (de Cheveigné *et al.*, 1995). Another principle, that of ''exclusive allocation,'' was also violated in the case of vowels that shared a formant but were nevertheless both well identified. Bregman (1990) noted many other exceptions to this rule.

Computer models have been proposed to ''segregate'' sound objects based on a two-dimensional time-frequency map. A first dimension (tonotopic) is an index into a filter bank representing cochlear filtering, and the second dimension is time. Spectro-temporal regions are assigned to different sources based on periodicity information (often estimated from the autocorrelation function). In response to a mixture of periodic sounds, the map may show regions that are dominated by one period or another. The map is partitioned on the basis of this information and the parts attributed to the various sounds (Weintraub, 1985; Cooke, 1991; Brown, 1992; Meddis and Hewitt, 1992; Ellis, 1996). Many conditions were found where the spectrum was *entirely* dominated by one vowel, and a partition was therefore not possible. Such models fail in those conditions and cannot explain the strong ΔF_0 effects that were observed.

Beats between partials of both vowels have been proposed to explain the effects of small ΔF_0's (Assmann and Summerfield, 1994; Culling and Darwin, 1994). In previous studies, we found evidence that weakens this hypothesis (de Cheveigné *et al.*, 1995, 1997a, b). The present study adds the observation that, for many vowel pairs, ΔF_0 effects are not greatest at amplitude ratios where beats have maximal amplitude near target formants. Effects are instead large at ratios for which the strongest beats should occur in regions unrelated to formant peaks.

In a previous study, it was speculated that a ''multiplicity'' cue might encourage subjects to report two vowels, and thus indirectly increase correct response rates in difficult conditions (such as low target/competitor ratio) for which cues to the *identity* of the weaker vowel are hard to obtain (de Cheveigné *et al.*, 1997a,b). The lack of any example of a uniform increase in response rate of all four vowels (distinct from the stronger one) argues against this hypothesis.

Across-spectrum grouping by F_0 is an attractive hypothesis. Formants belonging to the same vowel might be grouped by common F_0, and thus segregated from those of a competitor with a different F_0. The data do not support this hypothesis. Consider for example a target vowel that shares a formant with a stronger competing vowel. The F_0 of the common formant is that of the stronger vowel, and it therefore differs from the F_0 of the weak vowel's other formant when $\Delta F_0 \neq 0\%$. This should prevent correct grouping and produce a drop in identification, whereas an improvement was observed instead. Another example was found in the Appendix: for a/i, the ΔF_0 increased the number of incorrect /u/ responses more than that of correct /i/ responses, whereas grouping by F_0 should instead have favored /i/ and excluded /u/. Culling and Darwin (1993) also found evidence against this hypothesis: a vowel pair in which F_0's were swapped between first and second formants (the $F1$ of one vowel was excited with the same F_0 as the $F2$ of the other) benefited from a ΔF_0, despite the fact that different F_0's excited the $F1$ and $F2$ of each vowel.

Harmonic *enhancement* designates the general hypothesis according to which the F_0 of a target is exploited to segregate it from interference. Across-formant grouping is a variant of this hypothesis, but there are others. Previous work argued against this hypothesis, in particular because of the difficulty of estimating the target F_0 for weak targets (de Cheveigné *et al.*, 1997a). This argument is reinforced here: there is no way the auditory system could *directly* estimate the F_0 of a target vowel that is at least 10 dB beneath its competitor at all frequencies.

On the other hand, the results are consistent with the hypothesis that the *competitor's* F_0 is used to suppress it (harmonic cancellation). That F_0 is easy to estimate when the competitor is strong. One plausible model of harmonic cancellation is the channel selection model of Meddis and Hewitt (1992). However, it requires that at least *some* channels be dominated by the weaker vowel, and thus it cannot explain the entire data set. Another cancellation model is based on the time-domain ''cancellation filter'' of de Cheveigné (1993, 1997b). That model can operate at any amplitude ratio in principle, although in practice its implementation would face limits due to noise and imperfect linearity of transduction and neural processing. Within each channel, a neural cancellation filter suppresses the components of the competitor (based on its period), and the remainder of this operation constitutes evidence of the weaker vowel. The filter's dynamic range needs to be sufficient for this remainder to be distinguishable from noise. The limited amplitude range over which ΔF_0 was effective might reflect the limited dynamic range of such a mechanism.

Vowel identification was not impaired by the mere presence of cues to another vowel. A nonzero ΔF_0 rendered cues to both vowels more prominent, but there was no evidence that the increased prominence of cues to one vowel was detrimental to the other. There was no evidence of a cognitive competition between cues to each vowel. For pattern-matching, an extra formant peak poses a problem for a template-matching model using a Euclidean distance. Spectral differences are squared before summation, and locally large differences due to an extra formant have a disproportionate weight that may mask similarity over the rest of the spectrum. Pattern matching is likely to be unreliable.

The problem might be alleviated by replacing the square by a compressive nonlinearity, so as to emphasize small distributed differences over large localized ones. Another solution might be to synthesize *hybrid* templates (similar to the spectral envelopes in Fig. A1), and perform matching to them. A related scheme was proposed by Kopec and Bush (1989) for recognition of mixed speech. A third solution would be to somehow mark the spurious channels as unreliable, and exclude them from the distance calculation. A fourth would be to include only *positive* template-minus-token differences in the distance calculation. They alone constitute strong evidence of a mismatch: negative differences might be simply due to the dominance of part of the spectrum by the competitor. The last two schemes are among the ''missing data'' techniques proposed for speech recognition by Cooke, Morris, and Green (1996, 1997).

Our discussion of cues to vowel identification has limitations that were pointed out earlier. The spectral envelope differs from the representation available to the auditory system. Only formants $F1$ and $F2$ were considered, and higher formants were ignored. The description in terms of formants itself may be questioned (Rosner and Pickering, 1994). No account was taken of the fact that subjects usually had to choose the weaker vowel from a limited set of candidates (four), and that choosing one was equivalent to excluding the three others. Prominence of cues was evaluated qualitatively based on visual criteria, quantitatively based on amplitude differences between target and competitor at formant peaks of the target, and indirectly based on response rates. The relevance of those criteria to vowel perception is, however, unsure. The next step should be to construct quantitative models and test them on the response data. That is beyond the scope of the present paper, but raw data of this and other similar experiments are freely available to the interested reader.

IV. CONCLUSIONS

(1) ΔF_0 improved the identification of vowels as weak as -25 dB relative to their competitor, but at -35 dB the effects were no longer measurable.

(2) Improvement for weak targets was generally no greater at $\Delta F_0 = 6\%$ or 12 % than at $\Delta F_0 = 3\%$, whatever the target/competitor ratio. There was no tradeoff between factors: an unfavorable amplitude ratio could not be compensated by a larger ΔF_0.

(3) Identification of a vowel was generally not affected by the presence and prominence of cues belonging to the competing vowel. It depended only on the prominence of the target vowel's cues.

(4) Patterns of identification could generally be understood by assuming that the cues to identification of a vowel were its $F1$ and $F2$. However, the generality or uniqueness of this interpretation were not tested. In general, identification of a vowel required that both its formants be individually prominent, or else shared with the competitor, or at least potentially masked by the competitor.

(5) Sharing of a formant between vowels was not detrimental to the identification of either vowel. The principle of ''exclusive allocation'' did not apply.
(6) ΔF_0 improved target identification for certain vowel pairs over a wide range of amplitude ratios (at least 20 dB). This confirms the ecological value of F_0-guided segregation. On average, and for most vowel pairs, ΔF_0 effects were largest at -15 dB.
(7) ΔF_0 effects were observed when the target's spectral envelope was up to 10 dB below that of the competitor. Below that level, ΔF_0 effects were rare and small.
(8) Results were consistent with the hypothesis that segregation is based on a mechanism of harmonic cancellation. The limited dynamic range of ΔF_0 effects may be interpreted as reflecting the limited accuracy of that mechanism.

ACKNOWLEDGMENTS

The experiment was carried out at ATR Human Information Processing Research Laboratories, under a research agreement between ATR, the Center National de la Recherche Scientifique, and Paris 7 University. The author thanks ATR for its kind hospitality, and the CNRS for leave of absence. Hideki Kawahara participated in the preparation, and Rieko Kubo supervised the experiments. John Culling kindly provided the software for stimulus synthesis, and Willy Serniclaes kindly offered advice on statistical analysis. Thanks to two anonymous reviewers and the editor for very helpful advice on a previous version of this manuscript.

APPENDIX: DETAILS OF PAIR-WISE RESULTS

A. Spectral envelopes at $\Delta F_0 = 0\%$

Spectral envelopes of all pairs (other than /oe/ and /ou/) at $\Delta F_0 = 0\%$ are plotted in Fig. A1. Envelopes for /oe/ and /ou/ were plotted in Figs. 1 and 2. In each graph, isolated vowels (which were not presented as stimuli) are represented by thick lines, and mixtures (which were presented as stimuli) by the eight thin lines. Some of these are indistinct in places because they merged with one or the other thick lines.

B. Pair-wise results at $\Delta F_0 = 0\%$

Each vowel (target) is considered in competition with each of the other vowels (competitor). Response rates for all 20 target/competitor pairs at $\Delta F_0 = 0\%$ are plotted in Fig. A2 as a function of the target/competitor amplitude ratio. Each row of graphs corresponds to the same target vowel, and each column to the same competitor vowel. The diagonal is empty because vowels were not paired with themselves. Graphs that fall symmetrically relative to the diagonal correspond to both orderings of the same stimulus pair (for example, a/e and e/a both correspond to /ae/), with oppositely oriented abscissas. They share two abscissa values: 5 dB on one corresponds to -5 dB on the other.

Open circles are response rates for the target vowel (generally weaker), and crosses are rates for the competitor (generally stronger). Smooth lines represent a fit to these data by the following function:

$$R(x) = 1/(1 + e^{-(a/2)(x-b)}), \tag{A1}$$

where R is the response rate, x is the amplitude ratio in dB, b is the $R = 0.5$ intercept in dB, and a is the slope at intercept in dB^{-1}. For each vowel pair, the function was fit to eight response-rate data points (-35 to 35 dB) using the nonlinear fit procedure of the JMP statistics package (SAS Institute). Jagged lines near the abscissa are response rates for the three vowels not present in the stimulus. Whatever its role, each vowel (/a/, /e/, etc.) is coded in a distinctive line style (full, dotted, etc.) that is the same for all graphs.

Parameters of the fit for each target/competitor pair are given in Table AI. The smallest intercept (b) was -24 dB, for a/i (Fig. A2), and the greatest was -2 dB, for u/a. The smallest slope at intercept (a) was 0.21 for o/i and the greatest was 0.51 for a/o. Some general trends are evident in the table. Intercepts for /a/ were relatively low, indicating that it was robust with respect to interference. Intercepts for /u/ were relatively high, suggesting that it was vulnerable. The vowel /a/ was a relatively effective masker, while /o/ was a relatively weak masker (except when /u/ was the target).

1. Principles of vowel identification at $\Delta F_0 = 0\%$

In these data there are many examples of the ''first principle,'' according to which identification of a vowel is not prevented by cues belonging to a competing vowel. The target is usually well identified at ratios for which cues to the competing vowel are conspicuous in the spectral envelope (Fig. A1). When the target/competitor ratio is increased, these cues change while target identification does not. Conversely, when the ratio is decreased, they change very little while target identification changes greatly. One can wonder whether these cues, that are *visually* prominent (and also numerically on a log scale), are also *perceptually* prominent. An answer can be found in the fact that, in most panels of Fig. A2, the fit curves intersect at a high value. This indicates that over a certain range of amplitude ratios, cues to both vowels are perceptually effective. One might object that disruptive effects of cues to a competing vowel exist, but are masked by ceiling effects. This objection cannot be waived completely. One can, however, note that subjects were under no pressure to report two vowels, and that when they did so must indicate the presence of convincing cues.

The second principle is that both vowels can share a common formant, and nevertheless both be identified simultaneously. The principle of exclusive allocation described by Bregman (1990) does not hold. Four pairs out of ten share a formant: /ao/, /au/, /oe/, and /iu/ (Fig. A1).

The third principle is that identification of a vowel requires that both $F1$ and $F2$ be prominent. The necessity of the presence of cues to *both* formants is easiest to see in vowel pairs that share a formant. The spectral region of the common formant generally changes little across conditions, and identification appears to vary with the prominence of the other formant. Target/competitor pairs for which this is true

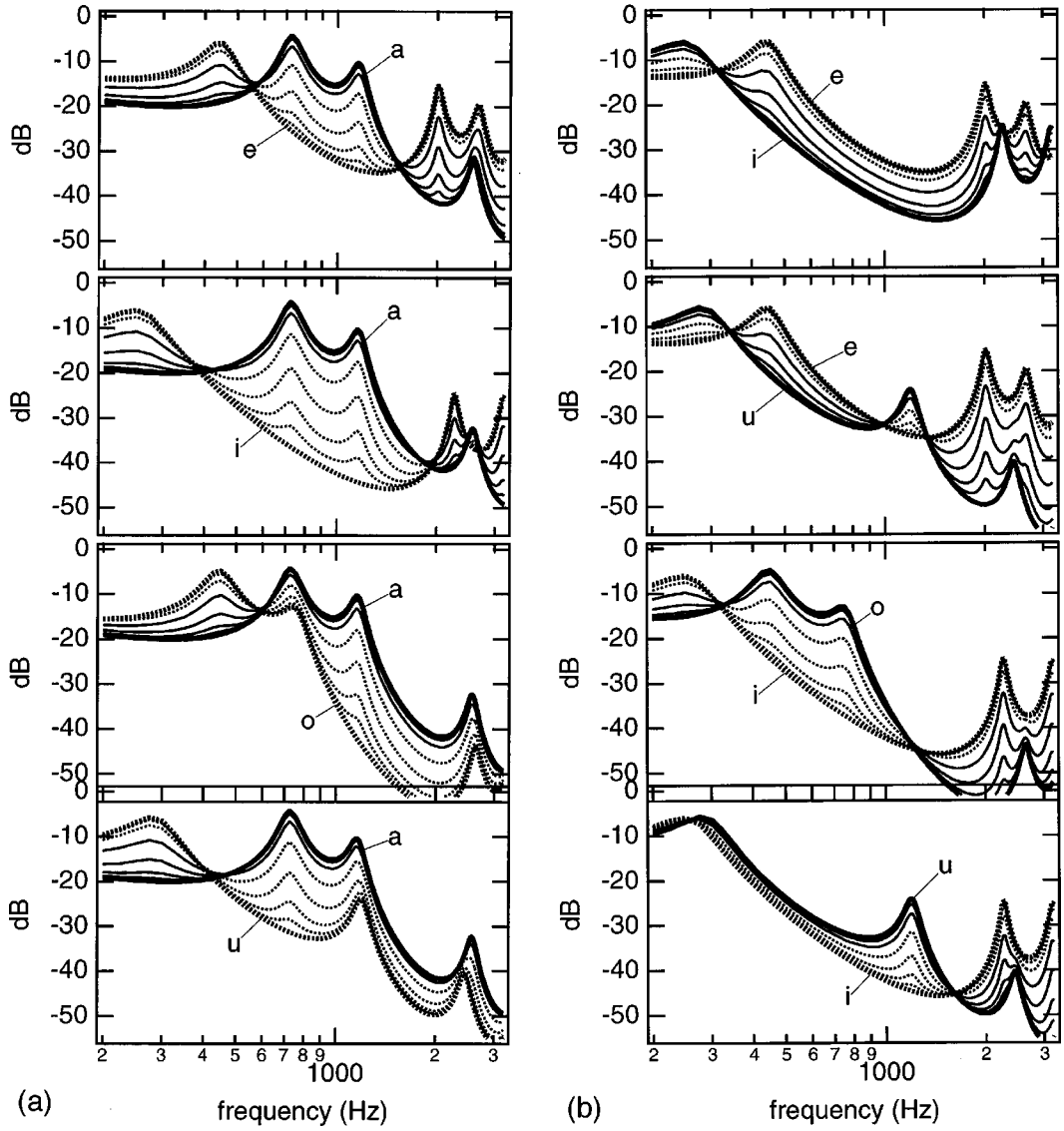

FIG. A1. Spectral envelopes of single vowels (thick lines) and weighted sums of pairs of vowels (thin lines), with amplitude ratios ranging from −35 to 35 dB in 10-dB steps. The plots include the first three formants ($F1$, $F2$, and $F3$), but exclude $F4$ and $F5$. Each graph corresponds to a different vowel pair. Together with Figs. 1(a) and 2(a), all ten vowel pairs are covered.

are a/o, o/a, a/u, u/a, o/e, e/o, i/u, and u/i. For a given vowel, the ''critical formant'' can be either $F1$ or $F2$. For example, in a/u identification varied with the prominence of $F1$ of /a/, whereas in a/o it varied with that of its $F2$.

Response rates decrease when cues to $F1$ and/or $F2$ become less prominent. However, prominence is not easy to define objectively. As noted in Sec. II B 2, at least three factors might underlie the prominence of formant peaks: their *absolute* amplitude, their amplitude *relative* to the competitor's formants, and their *local prominence* (amplitude relative to the neighboring spectral envelope, and ''sharpness''). Absolute amplitude is probably not a determining factor, as presentation levels were reasonably high, and thresholds for a given target vowel varied among competitors (Table AI). Support for local prominence comes from the fact that the lowest intercepts in Table AI are for a/i (−24 dB), a/u (−23 dB), and e/o (−18 dB), for which the deep interformant valleys of the competing vowels seem to favor the local prominence of target formants [Figs. A1 and 1(a)]. Evidence against it comes from e/o: the ''local prominence'' of the peak at $F2$ of /e/ hardly changed over the range over which identification of /e/ varied [Fig. 1(a)].

It is worth examining the spectral shapes for which formant cues were perceptually *ineffective*. Setting an arbitrary threshold at 10% and scanning through Fig. A2, one notes a certain number of data points that fall below this threshold. The corresponding spectral envelopes show cues to the target formants that are very modest indeed. The spectral envelope of the stimulus often differs by no more than a few dB from that of the competitor alone. An example is the −35-dB value of the a/e ratio (lowest dotted line near formants of /a/ in the appropriate plot of Fig. A1). Slightly more conspicuous ''noneffective'' cues are the shoulder at $F2$ of /a/ at −25 dB for a/o (second-to-lowest dotted line in Fig. A1), or the peak at $F2$ of /e/ at −25 dB for e/o (second-to-lowest dotted line in Fig. 1). Perceptually ineffective cues are objectively

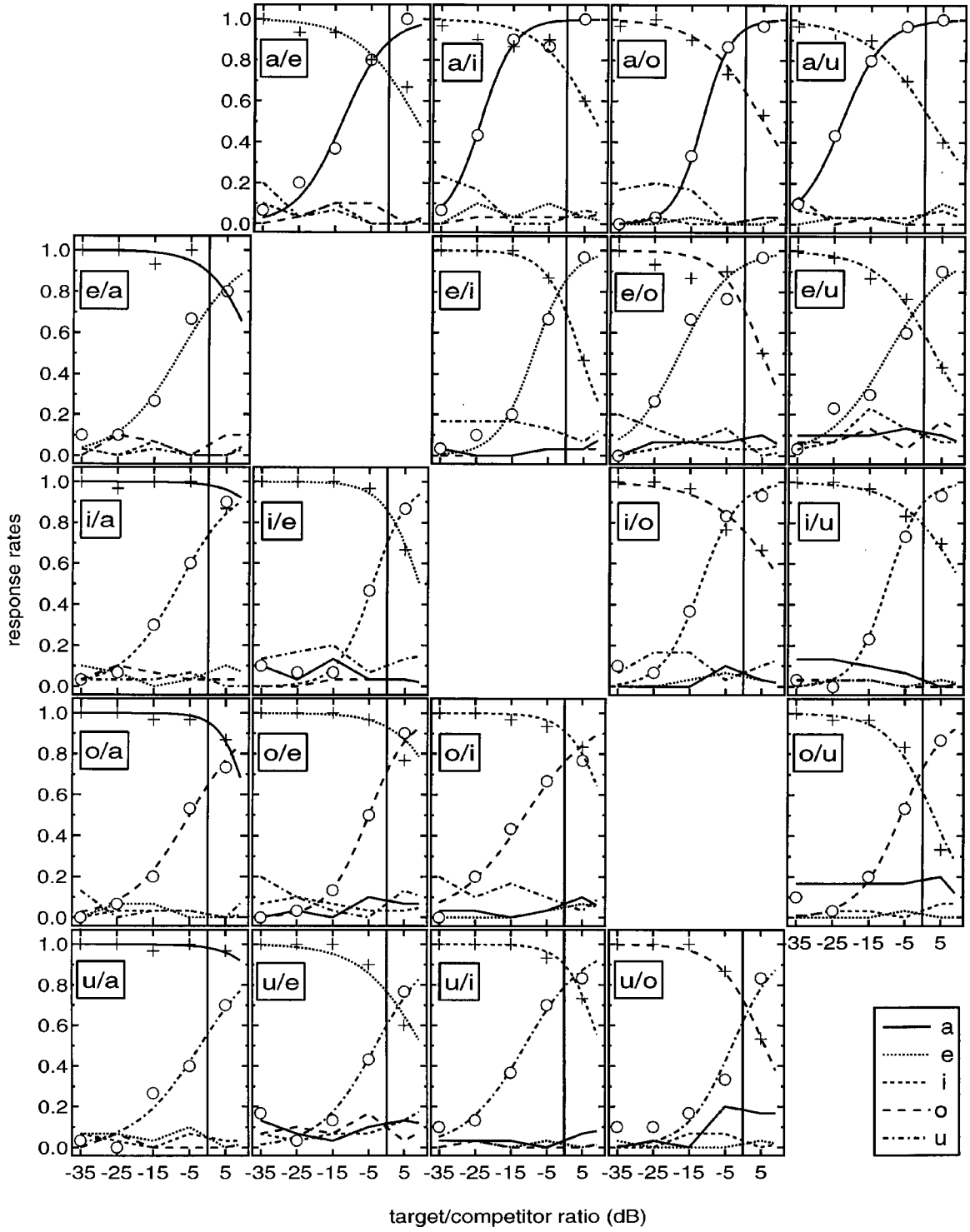

FIG. A2. Pair-wise responses at $\Delta F_0 = 0\%$. Each graph represents the response rates for all five vowels as a functions of the amplitude ratio. Graphs are given for each of the 20 target/competitor vowel pairs. Graphs on a row share the same target (generally weaker); graphs in a column share the same competitor (generally stronger). Open circles are data points for the target; crosses are data points for the competitor. Smooth lines are a function fit to these data (see the text). The jagged lines near the abscissa are response rates for the three vowels other than target and competitor (''spurious'' vowel responses).

very weak. By contrast, only slightly less modest spectral cues suffice to affect identification on a measurable number of trials.

2. Spurious responses

One might have imagined that the presence of a very weak vowel would have a rather indistinct effect, causing subjects to report at random a second vowel in addition to the stronger vowel. While there is a sort of ''noise floor'' evident in the spurious vowel responses in Fig. A2, it is not random: responses were usually dominated by one vowel, sometimes two, rarely three or four (the only example of four responses greater than 10% is the −15-dB point of e/u in Fig. A2). The dominant response (other than the strong vowel) was usually the weaker stimulus vowel, but at low amplitude ratios it was sometimes superseded by another

TABLE AI. First column: names of target/competitor vowel pairs. Second and third column: ratio between spectral envelopes of target and competitor at $F1$ and $F2$ of target (when both vowels have the same rms amplitude). Fourth and fifth column: intercept (abscissa at 50%) and slope at intercept of the fit to the target identification psychometric function. Last two columns: ratio between target and competitor spectral envelopes at $F1$ and $F2$ of target at criterion (see the text). Values are given only for target/competitor pairs for which the ΔF_0 effect size reached the criterion value (0.3), and only for formants that controlled target identification (shared formants were excluded).

	Ratio (dB)		Psychometric function		Ratio at criterion (dB)	
Pair	at $F1$	at $F2$	intercept (dB)	slope (dB^{-1})	at $F1$	at $F2$
a/e	21.3	23.8	−13	0.32	6.3	8.8
a/i	32.4	34.1	−24	0.44	7.4	9.1
a/o	8.9	32.2	−12	0.51	···	17.2
a/u	27.3	14.3	−23	0.35	···	···
e/a	13.3	26.3	−8	0.24	−1.7	11.3
e/i	18.6	23.5	−9	0.40	−6.4	···
e/o	−0.9	42	−18	0.29	···	···
e/u	15.0	34.6	−10	0.24	−10	9.6
i/a	13.6	16.1	−8	0.26	···	···
i/e	7.5	1.0	−4	0.42	−7.5	···
i/o	8.8	31.4	−12	0.39	−6.2	16.4
i/u	0.5	22.1	−10	0.44	···	7.1
o/a	14.2	−7.7	−5	0.24	−10.8	···
o/e	0.9	13.3	−5	0.36	···	−1.7
o/i	19.6	24.2	−11	0.21	4.6	9.2
o/u	16.0	18.7	−6	0.33	−9	−6.3
u/a	14.1	−11.7	−2	0.22	−0.9	···
u/e	7.3	10.3	−3	0.26	−7.7	−4.7
u/i	2.9	20.7	−11	0.24	···	5.7
o/u	8.4	19.7	−3	0.32	−6.6	4.7

vowel not present in the stimulus. Response rates for individual spurious vowels were as large as 20%.

The vowel /a/ was often reported for stimuli dominated by /u/ (last column in Fig. A2). The two vowels share similar $F2$'s, but it is hard to explain why /a/ was reported in the absence of evidence of its $F1$, except in the case of /ou/ where the $F2$ of /o/ can be interpreted as the $F1$ of /a/ (Sec. II B 2). The vowel /e/ was only rarely reported as a spurious response. The vowel /i/ was often reported for the pair e/u at −15 and −5 dB. Apparently the $F1$ of /u/ and $F2$ of /e/ were interpreted as belonging to /i/ (Fig. A1), an interpretation that leaves orphaned the $F1$ of /e/. The vowel /o/ was reported at certain levels for pairs e/u and e/a. This is easy to explain in the case of e/a: $F1$'s of /e/ and /a/ can be interpreted as $F1$ and $F2$ of /o/, and it is perhaps surprising that this interpretation was not made more often. Finally, the vowel /u/ was often reported for stimuli strongly dominated by /i/, /e/, or /o/ (Table II). The vowel /u/ has a similar $F1$ to /i/, but it shares no formants with /e/ or /o/. On the other hand, it has an $F2$ similar to that of /a/, but /u/ responses were less common with that vowel than with the other three. It is hard to account for the large number of /u/ responses for stimuli that lack either its $F1$ (a/o) or its $F2$ (/ei/, /io/), or both (/eo/). The vowel /u/ may possibly serve as a sort of ''default'' response. It is perhaps worth noting that the vowel /u/ is often devocalized in Japanese.

To summarize, spurious vowel responses were generally not random. They sometimes occurred despite the lack of specific cues to the vowel reported (typically /u/). More often, they were the result of cues present in the stimulus being reassembled in an incorrect but plausible way. In view of this fact, it is perhaps surprising that they did not occur more often.

C. Pair-wise patterns of ΔF_0 effects

1. ΔF_0 effects for constituent vowels

Differences in response rates between $\Delta F_0 = 12\%$ and $\Delta F_0 = 0\%$ are plotted in Fig. A3 for eight of the ten vowel pairs. The other two (/oe/ and /ou/) were plotted in Figs. 4(b) and 5(b). For most target/competitor pairs, ΔF_0 was effective over a range of amplitude ratios. For some (a/e, e/a, o/a, u/a, i/e, e/u, u/e, o/i, u/i, o/u, u/o) the range was at least 20 dB, for others it was more limited (a/o, i/o). The *upper* end of the range is the result of a ceiling effect: identification that is good at $\Delta F_0 = 0\%$ cannot be improved by a ΔF_0. The *lower* limit reflects the breakdown of segregation mechanisms when the target was too weak. This limit is worth examining in detail, as it may give cues to the nature of the segregation mechanisms. Let us quantify the ''lower limit'' as the lowest amplitude ratio for which the increase with ΔF_0 exceeded a criterion value (0.3). This criterion is purely arbitrary and unrelated to any judgment of significance. The criterion was not met for three target/competitor pairs out of 20 (i/a, a/u, and e/o). For the others it was met at either −25 dB (a/i, o/a, e/i, e/u, o/u) or at −15 dB (remaining pairs).

The amplitude ratios between spectral envelopes of target and competitor at formants $F1$ and $F2$ of the target, at

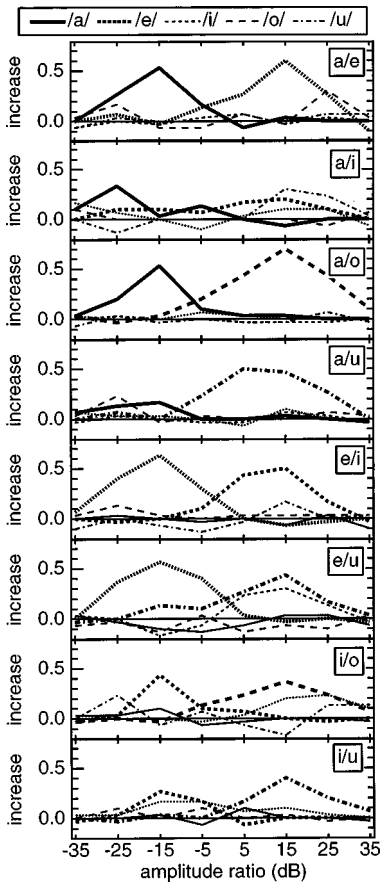

FIG. A3. Increase in response rates between $\Delta F_0 = 0\%$ and 12% as a function of intervowel amplitude ratio, for eight vowel pairs [the other two are plotted in Figs. 4(b) and 5(b)]. Thick lines: responses for vowels present in the stimulus. Thin lines: spurious vowel responses.

criterion, are given in the last two columns of Table AI. Excluded from this table are data for the three pairs that missed the criterion, and for formants that were shared between target and competitor and were therefore prominent in all conditions. For some pairs, the ratios at both target formants (or at the single formant that controlled identification) were positive (a/e, a/i, a/o, i/u, o/i, u/i). For others they were negative (e/i, i/e, o/a, o/e, o/u, u/a, u/e). For a few pairs, one ratio was negative, the other positive. The range of values was quite wide: -10.8 to 17.3 dB.

This pattern can be confronted with three possible models of concurrent vowel segregation: channel selection (Meddis and Hewitt, 1992), beats (Assmann and Summerfield, 1994; Culling and Darwin, 1994), and harmonic cancellation (de Cheveigné, 1997b). Channel selection requires that the target vowel dominate at least some parts of the spectrum: it cannot explain ΔF_0 effects for a target everywhere weaker than its competitor. It thus cannot account for the full data set.

Beats should be most effective when they are strong at formants of the target, that is, when the amplitude ratio at those formants is close to 0 dB. The -10.8-dB amplitude ratio at criterion for o/a is consistent with that hypothesis: the resulting modulation depth of about 30% is possibly effective, and it decreases with a further reduction in o/a ratio, consistent with the decrease in ΔF_0 effect size. The hypothesis fares less well for pairs that have a *positive* ratio at criterion. Starting from the ratio at criterion, a reduction should *increase* the size of ΔF_0 effects, contrary to what is observed. Conversely, an increase in ratio should diminish the amplitude of beats at formant peaks, thus diminishing ΔF_0 effect size, again contrary to what is observed. When a formant peak emerges above its competitor's spectrum, the locus of maximum beats splits and shifts along each flank of the formant peak. This might lead to confusion if beats were used to signal formant positions. ΔF_0 effects were nevertheless strong in such conditions. The beat hypothesis is thus hard to reconcile with these data.

The ideal, linear version of the harmonic cancellation model can exploit the presence of a mistuned target, however weak (de Cheveigné, 1997b). In practice, the dynamic range of the mechanism would be limited by noise and the imperfect linearity of a neural cancellation filter. The smallest ratio of envelopes at criterion (-10.8) might be interpreted as reflecting the limit of the dynamic range of such neural processing. However, in that case it is hard to explain why corresponding figures are not similar for other vowels.

Three vowel pairs gave ΔF_0 effects too small to meet the 0.3 criterion (i/a, a/u, and e/o). The case of e/o was discussed in Sec. II B 3. For a/u, identification of /a/ depended on the prominence of its $F1$, that emerged relatively well between formants of /u/ at all levels (Fig. A1), leading to good identification at $\Delta F_0 = 0\%$ and thus small ΔF_0 effects. For i/a the small effect size is partly due to a tendency to group the $F1$ of /i/ and the $F2$ of /a/ together and interpret them as /u/. This occurred despite the fact that these two formants were excited by different F_0's, an argument against the hypothesis of across-frequency grouping of formants by F_0.

A few data points show a *decrease* in identification rate with ΔF_0, but the effects were small and probably due to experimental noise. It seems safe to say that ΔF_0 was either beneficial or indifferent to identification according to the condition, but never detrimental.

2. ΔF_0 effects for spurious vowel responses

In many cases, a ΔF_0 increased the response rates for spurious vowels as well as (or instead of) correct vowels. In a few cases they were *reduced* instead. This section reviews the conditions in which a ΔF_0 affected spurious responses for each vowel pair.

For /ou/, the decrease in /a/ responses with ΔF_0 for o/u ratios of -25 to 5 dB (and its increase at 15 dB) were noted in Sec. II B 3. For /ae/ (Fig. A3), the number of /o/ responses increased at an a/e ratio of 25 dB. A possible explanation is that ΔF_0 increased the prominence of the $F1$ of /e/ that was interpreted as belonging to /o/. The amplitude of the $F1$ peak of /e/ relative to /a/ in that condition was -11.7 dB. For /ai/, at a/i ratios of 15 and 25 dB, the number of incorrect /u/ responses increased more than that of correct /i/ responses. Apparently, ΔF_0 enhanced the prominence of $F1$ of /i/, required for both interpretations. It is not clear why the incorrect interpretation was preferred. If anything, formant grouping by F_0 should have favored the correct interpretation

(another argument against this hypothesis). The ratio of envelopes at $F1$ of /i/ in these conditions was -1.4 and -11.4 dB, respectively.

For /eu/ at e/u ratios of 5, 15, and 25 dB, the number of incorrect /i/ responses increased almost as much as that of correct /u/ responses. Apparently, ΔF_0 enhanced the prominence of the $F1$ of /u/ required for both interpretations. The ratio of envelopes at $F1$ of /u/ in these conditions was 2.7, -7.7, and -17.7 dB, respectively. At some e/u ratios, response rates for spurious vowels decreased, possibly a simple effect of the increased prominence of cues to the ''correct'' vowels.

For /io/ at i/o ratios of 15 and 25 dB, the number of incorrect /e/ responses increased together with correct /i/ responses. Apparently, ΔF_0 enhanced the prominence of the $F1$ of /o/ required for both interpretations. For /iu/ at i/u ratios of -15 and -5 dB, the number of incorrect /e/ responses increased slightly together with correct /i/ responses. ΔF_0 enhanced the prominence of the $F2$ of /i/ required by both interpretations, but the increase in /e/ responses is hard to explain in the absence of any cue to $F1$ of /e/.

Overall, neither the increases in incorrect vowel response rates, nor their baseline rates themselves, were very large. A *uniform* increase over all four vowels (other than the stronger) was never observed. Such an increase was to be expected if a hypothetical ''multiplicity cue'' signaled the presence of an extra vowel without providing information about its identity. Rate increases with ΔF_0 usually concerned one vowel (correct or incorrect), sometimes two, and in rare cases three. In other cases, a decrease was observed instead of an increase. The hypothesis of a multiplicity cue is therefore improbable.

Assmann, P. F., and Summerfield, Q. (**1989**). ''Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency,'' J. Acoust. Soc. Am. **85**, 327–338.

Assmann, P. F., and Summerfield, Q. (**1990**). ''Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P. F., and Summerfield, Q. (**1994**). ''The contribution of waveform interactions to the perception of concurrent vowels,'' J. Acoust. Soc. Am. **95**, 471–484.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Brown, G. J. (**1992**). ''Computational auditory scene analysis: a representational approach,'' Sheffield, Department of Computer Science, unpublished doctoral dissertation.

Cooke, M., Morris, A., and Green, P. (**1996**). ''Recognizing occluded speech,'' Proceedings of the ESCA Workshop on the Auditory Basis of Speech Perception, Keele, edited by W. Ainsworth and S. Greenberg, pp. 297–300.

Cooke, M., Morris, A., and Green, P. (**1997**). ''Missing data techniques for robust speech recognition,'' Proceedings ICASSP, pp. 863–866.

Cooke, M. P. (**1991**). ''Modelling auditory processing and organization,'' Sheffield, Department of Computer Science, unpublished doctoral dissertation.

Culling, J. F. (**1996**). ''Signal processing software for teaching and research in psycholinguistics under UNIX and X-windows,'' Behav. Res. Methods Instrum. Comput. **28**, 376–382.

Culling, J. F., and Darwin, C. J. (**1993**). ''Perceptual separation of simultaneous vowels: Within and across-formant grouping by $F0$,'' J. Acoust. Soc. Am. **93**, 3454–3467.

Culling, J. F., and Darwin, C. J. (**1994**). ''Perceptual and computational separation of simultaneous vowels: Cues arising from low frequency beating,'' J. Acoust. Soc. Am. **95**, 1559–1569.

Culling, J. F., and Summerfield, Q. (**1995**). ''Perceptual segregation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay,'' J. Acoust. Soc. Am. **98**, 785–797.

Culling, J. F., Summerfield, Q., and Marshall, D. H. (**1994**). ''Effects of simulated reverberation on the use of binaural cues and fundamental frequency differences for separating concurrent vowels,'' Speech Commun. **14**, 71–95.

de Cheveigné, A. (**1993**). ''Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing,'' J. Acoust. Soc. Am. **93**, 3271–3290.

de Cheveigné, A. (**1997a**). ''Ten experiments in concurrent vowel segregation,'' ATR Human Information Processing Research Labs technical report, TR-H-217.

de Cheveigné, A. (**1997b**). ''Concurrent vowel segregation III: A neural model of harmonic interference cancellation,'' J. Acoust. Soc. Am. **101**, 2857–2865.

de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (**1997a**). ''Concurrent vowel identification I: Effects of relative level and $F0$ difference,'' J. Acoust. Soc. Am. **101**, 2839–2847.

de Cheveigné, A., McAdams, S., Laroche, J., and Rosenberg, M. (**1995**). ''Identification of concurrent harmonic and inharmonic vowels: A set of the theory of harmonic cancellation and enhancement,'' J. Acoust. Soc. Am. **97**, 3736–3748.

de Cheveigné, A., McAdams, S., and Marin, C. (**1997b**). ''Concurrent vowel identification II: Effects of phase, harmonicity and task,'' J. Acoust. Soc. Am. **101**, 2848–2856 (in preparation).

Demany, L., and Semal, C. (**1990**). ''The effect of vibrato on the recognition of masked vowels,'' Percept. Psychophys. **48**, 436–444.

Elllis D. (**1996**). Prediction-driven computational auditory scene analysis,'' MIT unpublished doctoral dissertation.

Geisser, S., and Greenhouse, S. W. (**1958**). ''An extension of Box's results on the use of the F distribution in multivariate analysis,'' Ann. Math. Stat. **29**, 885–889.

Hirahara, T., and Kato, H. (**1992**). ''The effect of $F0$ on vowel identification,'' in *Speech Perception, Production and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Ohmsha, Tokyo), pp. 89–1120.

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 838–844.

Kopec, G. E., and Bush, M. A. (**1989**). ''An LPC-based spectral similarity measure for speech recognition in the presence of co-channel speech interference,'' Proceedings IEEE ICASSP, pp. 270–273.

McKeown, J. D. (**1992**). ''Perception of concurrent vowels: The effect of varying their relative level,'' Speech Commun. **11**, 1–13.

Meddis, R., and Hewitt, M. J. (**1992**). ''Modeling the identification of concurrent vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **91**, 233–245.

Meyer, G., and Berthommier, F. (**1995**). ''Vowel segregation with amplitude modulation maps: a re-evaluation of place and place-time models,'' Proceedings ESCA Workshop on the Auditory Basis of Speech Perception, Keele, edited by W. Ainsworth and S. Greenberg, pp. 212–215.

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.

Rosner, B. S., and Pickering, J. B. (**1994**). *Vowel Perception and Production* (Oxford University Press, Oxford).

Summerfield, Q. (**1992**). ''Roles of harmonicity and coherent frequency modulation in auditory grouping,'' in *The Auditory Processing of Speech: From Sounds to Words*, edited by M. E. H. Schouten (Mouton de Gruyter, Berlin), pp. 157–166.

Summerfield, Q., and Culling, J. F. (**1992**). ''Auditory segregation of competing voices: absence of effects of FM or AM coherence,'' Philos. Trans. R. Soc. London, Ser. B **336**, 357–366.

Weintraub, M. (**1985**). ''A theory and computational model of auditory monaural sound separation,'' University of Stanford, unpublished doctoral dissertation.