# Concurrent vowel identification. II. Effects of phase, harmonicity, and task[a)]

Alain de Cheveigné
*Centre National de la Recherche Scientifique/Université Paris 7, 2 place Jussieu, case 7003, F-75251 Paris Cédex 05, France and ATR Human Information Processing Research Laboratories, 2-2 Hikaridai, Seika-cho Soraku-gun, Kyoto 619-02, Japan*

Stephen McAdams
*Laboratoire de Psychologie Expérimentale (Centre National de la Recherche Scientifique, URA316), Université René Descartes, EPHE, 28 rue Serpente, F-75006 Paris, France and Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), 1 place Stravinsky, F-75004 Paris, France*

Cécile M. H. Marin
*IRCAM, 1 place Stravinsky, F-75004 Paris, France*

(Received 19 December 1995; revised 12 August 1996; accepted 22 November 1996)

Subjects identified concurrent synthetic vowel pairs in four experiments. The first experiment found that improvements in vowel identification with a difference in fundamental frequency ($\Delta F_0$) do not depend on component phase. The second investigated more precisely whether phase patterns resulting from ongoing phase shifts in inharmonic stimuli can by themselves produce effects similar to those attributed to differences in harmonic state of component vowels. No such effects were found. The third experiment found that identification was better for harmonic than for inharmonic backgrounds, and that it did not depend on target harmonicity. The first three experiments employed a task in which subjects were free to report one or two vowels for each stimulus. The fourth experiment reproduced several conditions with a more classic task in which subjects had to report two vowels. Compared to the classic task, the new task gave larger effects and provided an additional measure of segregation: the number of vowels reported per stimulus. Overall, results were consistent with the hypothesis that the auditory system segregates targets by a mechanism of harmonic cancellation of competing vowels. They did not support the hypothesis of harmonic enhancement of targets. The lack of a phase effect places strong constraints on models that exploit pitch period asynchrony (PPA) or beats. © *1997 Acoustical Society of America.*
[S0001-4966(97)04004-6]

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

## INTRODUCTION

Speech is easier to understand when there is a difference in fundamental frequency ($\Delta F_0$) between a target voice and an interfering voice (Brokx and Nooteboom, 1982). When two steady-state synthetic vowels are presented simultaneously, identification is better when their fundamental frequencies ($F_0$) are different than when they are the same (Scheffers, 1983; Darwin, 1981; Zwicker, 1984; Assmann and Summerfield, 1990; McKeown, 1992; Culling and Darwin, 1993, 1994). A variety of models and methods of ''$F_0$-guided segregation'' have been proposed to explain or emulate this effect [see de Cheveigné (1993) for a review]. They may be classified according to whether they exploit target harmonicity (harmonic *enhancement*) or background harmonicity (harmonic *cancellation*). Some evidence has been found in favor of harmonic cancellation (Lea, 1992; Summerfield and Culling, 1992; de Cheveigné, 1994; de Cheveigné *et al.*, 1995a), but so far there is little to support the harmonic enhancement hypothesis. Recently, other mechanisms have been proposed that do not depend directly on harmonicity or $F_0$, but rather on waveform interactions that co-occur with $F_0$ differences.

### A. Pitch period asynchrony (PPA)

An $F_0$ difference is equivalent to a gradually increasing time shift of one waveform relative to another. A natural vowel's short-term energy is pulsatile, so the masking it causes or receives may vary with time alignment relative to the other vowel. A $\Delta F_0$ might in this way cause either vowel or both to be better perceived. This is known as the pitch period asynchrony (PPA) mechanism (Assmann and Summerfield, 1988, 1994; Summerfield and Assmann, 1991; Carlyon and Shackleton, 1994). Summerfield and Assmann (1991) investigated whether a time shift *per se* is sufficient to produce segregation in the absence of mistuning. They presented subjects with synthetic vowels at the same $F_0$ (50 or 100 Hz), with and without a time shift of half a period (both vowels were ramped on and off simultaneously, so the shift did not affect onset times). The time shift produced a significant improvement at 50 Hz, but not at 100 Hz. Although in a later experiment Assmann and Summerfield (1994) did find a significant effect of intervowel alignment at 100 Hz, as well

[a)] Portions of this work were presented at a meeting of the Acoustical Society of Japan and as an ATR Human Information Processing Research Laboratories technical report (de Cheveigné, 1995; de Cheveigné *et al.*, 1995b).

as other indirect evidence that PPA contributes to segregation, they failed to replicate this effect in a further experiment with inexperienced subjects.

Estimates of the equivalent rectangular duration (ERD) of the auditory temporal window (Plack and Moore, 1990) are of the same order (6–13 ms) as the fundamental periods used in double-vowel experiments. One might therefore expect features of a 10-ms period to be smoothed out too much for PPA to be effective. However, Kohlrausch and Sander (1995) found that masking of a short pure-tone target varied by as much as 17 dB within the period of a 100-Hz masker. The variation was smaller but still appreciable (about 6 dB) at a fundamental of 220 Hz. The effect was dependent on the component phase of the masker and largest for a phase relationship designed to produce highly modulated patterns of activity within auditory channels.

Several experiments suggest that vowel identification might depend on uneven masking within a masker's fundamental period. Moore and Alcántara (1995, 1996) synthesized harmonic ''vowels'' with a fundamental of 100 Hz and a spectral envelope that was flat on average. ''Formants'' were defined by amplitude modulation of groups of two consecutive harmonics at a rate of 10 Hz. For cosine phase, the stimuli could be identified as vowels, despite their flat average spectrum. For random phase, identification was at chance level. Stimuli with cosine phase and a flat spectrum have a peaked waveform that produces strongly modulated activity within peripheral channels, provided the $F_0$ is low enough and the channel CF high enough (Horst *et al.*, 1986). Within the dips of this modulation, masking may be relatively weak. Raising or lowering the level of a group of components is equivalent to adding those components (in same or opposing phase) so that they stand out during the modulation dips.

Palmer *et al.* (1987) observed a change with phase of the position of the $F1$ phoneme boundary along a /e/–/I/ continuum (Darwin and Gardner, 1986). The manipulated partial was the fourth harmonic (500 Hz) of a 125-Hz fundamental. The boundary moved down from 450 to 430 Hz when the phase shifted by 90° relative to the phase produced by a Klatt synthesizer. In other words, this shift is equivalent to a 20-Hz rise in the perceived $F1$ of the stimuli. The authors also performed a physiological experiment in which similar stimuli (with a fundamental of 100 Hz) were presented to guinea pigs and the response was recorded from a population of auditory-nerve fibers. Without the phase shift fibers below 1 kHz were equally dominated by frequencies of 400 or 500 Hz. With the 90° phase shift they were dominated mainly by the higher component, which is congruent with a rise in perceived $F1$ in the human subjects. Stimuli in Klatt phase (phase produced by the Klatt synthesizer) produce highly modulated patterns of activity in auditory channels. Shifting the phase of a component is equivalent to adding a component with the same frequency and suitable phase and amplitude. The added component may be perceptible within the valleys of modulation.

Traunmüller (1987) modulated the amplitude spectrum of a glottal source with the phase spectrum of a glottal tract (simulated as a cascade synthesizer) to produce nine Swedish ''vowels.'' There were no spectral amplitude peaks present to signal the formants, but several subjects could label the stimuli consistently if the $F_0$ was low enough (71 or 100 Hz). Labeling was less consistent at higher frequencies (141 and 200 Hz) and at 283 Hz it fell to chance level. The ''phase vowels'' were intelligible via earphones, but not when presented through a loudspeaker in an ordinary room. ''Flat-spectrum'' diphthongs produced by Schroeder and Strube (1986) were also unintelligible if presented via loudspeakers in a reverberant room, because of phase randomization.

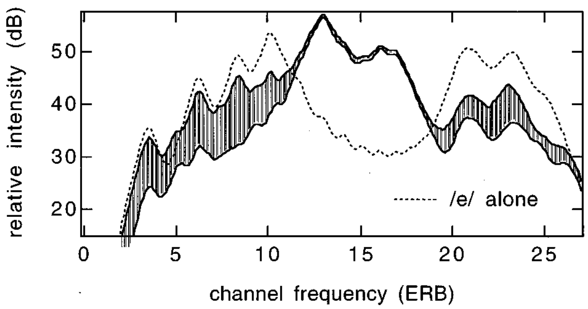

FIG. 1. Striped zone: Range of variation of the excitation pattern due to beats within a vowel pair /a/+/e/ in which /e/ is 12 dB weaker than /a/. Thin dotted line: Excitation pattern produced by /e/ alone. Excitation patterns were calculated by taking the FFT of a 16-ms Hanning-shaped window and applying spectral smoothing according to formulas of Moore and Glasberg (1983). The origin of the ordinate (dB scale) is arbitrary.

These experiments all show that phase may in some circumstances affect vowel identification. The PPA hypothesis depends on the particular phase patterns that produce peaked waveforms. One might therefore suspect that the $\Delta F_0$ effects observed in ''double vowel'' experiments are specific to the particular phase employed. If so, they should be reduced for random phase stimuli that lack the temporal cues upon which PPA depends.

## B. Waveform interaction (beats)

The PPA explanation involves intravowel phase patterns that produce peaked waveforms, together with the particular intervowel phase relationship that is equivalent to a time shift. It also supposes that the temporal resolution of the auditory system is fine enough to resolve patterns on the time scale of a period. However, waveform interaction may also produce patterns that are static (at $\Delta F_0=0$) or that vary on a slower time scale. When a $\Delta F_0$ is introduced between vowels, beats occur between corresponding partials at a rate equal to their frequency difference, and with a depth that depends on their relative amplitudes. The short-term spectrum thus varies with time, and it may assume a shape that favors the identification of one vowel or the other, or possibly both together. Alternatively, the pulsation itself might reveal spectral cues too weak to stand out in the average spectrum. Figure 1 illustrates this idea. The vowel /e/ at 132 Hz was added to the vowel /a/ at 124 Hz, with a 12-dB mismatch in favor of /a/. The two vowels have the same spectrum level at the formants $F1$ and $F2$ of /e/, causing the spectrum of their sum to undergo relatively deep beats near those formants. The excitation pattern for the sum varies over the range shown in Fig. 1. The pulsation might reveal the formants, despite the fact that the average spectrum does not show clear evidence of their presence at any instant.

Culling and Darwin (1994) suggested that beats in the low-frequency ($F1$) region might explain improvements in identification performance with $\Delta F_0$'s smaller than 1 semitone. In agreement with this hypothesis, Assmann and Summerfield (1994) found that successive 50-ms intervals excised from a 200-ms double vowel were not equally identifiable. Identification rates for the best interval were consistent with the idea that the auditory system takes advantage of beats to choose, within the 200-ms stimulus, a favorable interval on which to base identification.

If the $\Delta F_0$ is small or zero, the overall spectrum of a double-vowel stimulus depends on the pattern of intervowel starting phase. In particular, the spectrum of the $\Delta F_0=0$ condition of double-vowel experiments is phase-dependent. It is conceivable that the commonly used Klatt or sine phase patterns might produce at $\Delta F_0=0$ a spectrum that is particularly unfavorable for identification, contributing artificially to the size of the $\Delta F_0$ effects observed. Like PPA, the beats hypothesis leads us to suspect that the classic $\Delta F_0$ effect might be phase dependent.

In a previous experiment (de Cheveigné *et al.*, 1995a) we presented subjects with double-vowel stimuli in which each vowel was either harmonic or inharmonic (with partials randomly mistuned). Our aim was to determine whether $F_0$-guided segregation mechanisms used the harmonic structure of the vowel being identified, that of the competing vowel, or both. We found that harmonicity of the competing vowel improved identification of the target, but that harmonicity of the target itself did not. However, all our stimuli were synthesized with a sine starting phase. Each partial of an inharmonic vowel can be interpreted as gradually shifting in phase, due to its mistuning. Consequently, inharmonic vowels shifted towards a random phase pattern, whereas harmonic vowels kept their original phase throughout the stimulus. If phase affected identification, it might conceivably have been responsible for the pattern of results that we attributed to harmonicity.

### C. The present investigation

The experiments described in this paper were designed to reveal phase effects and to test the generality of our results on harmonicity. Experiment 1 examined whether the classic $\Delta F_0$ effect depends on intra- or intervowel phase relationships. Experiment 2 investigated more particularly whether phase effects could have constituted an artifact in our previous harmonicity experiment. Experiment 3 reproduced three crucial conditions of that experiment with stimuli designed to minimize the usefulness of beat or PPA cues. Experiment 4 compared the particular task we used (one-or-two response task) to the task classically used in double-vowel experiments (two vowel forced-choice task).

## I. GENERAL METHODS

### A. Stimuli

The subjects (six), vowel set (five), synthesis method, and presentation conditions were described in a companion paper (de Cheveigné *et al.*, 1997). The present experiments used a different choice of $F0$, phase, duration, intervowel amplitude, harmonic state, and task, as described here. Single vowels were synthesized at $F_0$'s of 124 and 132 Hz, with a duration of 270 ms including 20-ms raised-cosine onset and offset ramps. Both $F_0$'s were chosen to be multiples of 4 Hz, the reciprocal of the effective stimulus duration (250 ms between $-6$-dB points), so that all beat patterns between partials would have an integer number of periods, and the overall spectrum would be the same whatever the starting phases of beating partials. The spectrum did, however, depend on the relative phase of partials that had the *same* frequency. This was the case of all partials at $\Delta F_0=0$, but of only one partial (4092 Hz) when the $F_0$'s differed (this frequency is beyond the range that largely determines vowel identification). Partials started either in sine phase or in one of two ''random'' phase patterns $(R,R')$. The amount of modulation produced by different phase patterns within output channels of a peripheral filter model (Holdsworth *et al.*, 1988) is illustrated in Fig. 2 for a fundamental of 100 Hz. Modulation was estimated by taking the largest ratio between rms output calculated over two consecutive frames, each one-half period in length (5 ms). A large value of this measure indicates that the energy is localized within the period. Modulation is small up to about 14 equivalent rectangular bandwidth (ERB) (845 Hz) for all phases. It increases rapidly for Klatt, sine and cosine phase patterns, but remains small for both ''random'' phase patterns. Single vowels were added to obtain double vowels, with an amplitude mismatch of 15 dB to reduce ceiling effects for the weaker vowel (de Cheveigné *et al.*, 1997).

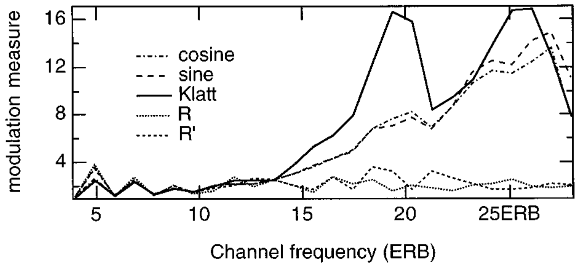

FIG. 2. Modulation of output channels of an auditory model as a function of CF for several phase patterns, averaged over vowels. $F_0$ was 100 Hz.

Stimuli were harmonic in experiments 1 and 2, and either harmonic or inharmonic in experiments 3 and 4. Inharmonic stimuli were produced by shifting partial frequencies of a harmonic vowel by random amounts less than 6.45%, or less than half the spacing between partials, whichever was smaller. The shifts obeyed further constraints that were designed to reduce the usefulness of phase or beat cues: (1) All partial frequencies had to be multiples of 4 Hz to ensure that the effective length of the stimulus (250 ms) was a superperiod of all beat patterns; (2) any given partial deviated by at most $F_0/2$, or $8*n$ Hz (where $n$ was the partial's rank) whichever was smaller, from the harmonic series, to ensure that the spectral density was not too different from that of a harmonic stimulus; (3) Each partial was at least 16 Hz from any other component in the stimulus to ensure that all beats between adjacent partials were faster than 16 Hz; and (4) within these constraints, the partial was chosen at random. Care was taken to ensure that the constraints did not introduce a systematic shift towards either higher or lower frequencies.

In order to satisfy constraint 3, different patterns had to be used at each of the nominal frequencies employed. Constraint 3 was relaxed for the second harmonic because it was incompatible with constraint 2. Since a random choice of frequencies may produce, by chance, patterns that are locally harmonic, a measure of inharmonicity was used to screen out such patterns. The measure was defined as the sum of absolute differences between consecutive partial frequencies divided by their rank. It is sensitive to local rather than cross-spectrum harmonicity patterns, and puts relatively less weight on higher partials. In this respect it differs from the measure used by de Cheveigné *et al.* (1995a).

The ‘‘$F_0$’’ of an inharmonic vowel is defined as the $F_0$ of the harmonic vowel from which it is derived. Inharmonic vowels sounded odd but were unambiguously identifiable as vowels. They had a relatively clear pitch that depended on the particular vowel.

### B. Task and experiment design

Experiments 1–3 had several stimulus conditions in common. In the interest of economy, their stimuli were pooled and presented together (in other words, they formed a single experiment that we describe as three in the interest of clarity). The stimulus set consisted of 200 single and 400 double vowels in random order. The subject's task was to report one or two vowels for each stimulus, as in de Cheveigné *et al.* (1997).

The stimuli of experiment 4 were pooled with stimuli of another experiment not reported here. The stimulus set consisted of 400 double vowels in random order. Each stimulus was presented once and the subjects had to report *two* vowels. Subjects were warned that identification of both vowels might sometimes be impossible and were asked to make their ‘‘best guess’’ in that case. Again there was no feedback. All six subjects participated in experiments 1–3. Of the six, five also participated in experiment 4. Each subject performed five sessions on different days.

### C. Scoring methods

Scoring methods are the same as used by de Cheveigné *et al.* (1997). Each double-vowel stimulus was scored twice, once for each vowel. A stimulus vowel was scored as correctly identified if it was matched by the response vowel (or either response vowel if two were given). Each single-vowel stimulus was scored once, in a similar fashion. Responses were classified according to the vowel's nature (phoneme, $F_0$, phase, harmonicity), the nature of the competing vowel, and their relationship ($\Delta F_0$, relative amplitude) to obtain *target-correct* identification rates for each of these conditions. Results for the more intense vowel (15 dB) were essentially perfect and are not reported. We report only rates for the weaker ($-15$ dB) vowel. From previous results (de Cheveigné *et al.*, 1997) no effect of absolute $F_0$ (low/low versus high/high or low/high versus high/low) was expected, so scores were averaged over that factor. For all stimuli, the number of vowels reported was noted.

## II. EXPERIMENT 1: PHASE DEPENDENCY OF THE $\Delta F_0$ EFFECT

Experiment 1 was designed to check whether the classic $\Delta F_0$ effect depends on the component phases of constituent vowels. All vowels were harmonic. There were two $\Delta F_0$ conditions: 0% and 6.45%. There were also three phase conditions: S/S (sine/sine), R/R (random/random, same pattern) and R/R′ (random/random, different patterns). In the S/S condition both vowels have peaked waveforms that could support a PPA mechanism. The R/R condition lacks peaked waveforms, but might conceivably support a weak form of the PPA hypothesis, based on the alignment of whatever temporal features are present. These features would be aligned at $\Delta F_0=0\%$ and misaligned otherwise. The R/R′ condition should defeat PPA altogether: the waveforms lack large peaks and have no features in common. As far as waveform interaction is concerned, S/S and R/R are equivalent: Both have the same intervowel phase pattern (0) and produce the same particular spectrum at $\Delta F_0=0$ (sum of the spectra of the constituents). In the R/R′ condition, intervowel phase is random and remains so with ongoing phase shifts due to $\Delta F_0$. The spectrum produced at $\Delta F_0=0$ is the result of random vector summation.

There were $(2\Delta F_0\text{'s})\times(3\text{ phase conditions})\times(10\text{ unordered vowel pairs})\times(2\ F_0\text{'s})\times(2\text{ amplitudes})=240$ different stimuli, repeated within each of five sessions. The scoring process described in Sec. I C retained responses for only one amplitude ($-15$ dB), but distinguished 20 ordered vowel pairs. Identification rates, averaged over $F_0$'s (2) and sessions (5), were subjected to a repeated-measures analysis of variance (ANOVA) with factors $\Delta F_0$ (2), PHASE (3), and ordered vowel PAIR (20). Probabilities reflect, where necessary, an adjustment of the degrees of freedom by a factor that corrects for the inherent correlation of repeated measurements (Geisser and Greenhouse, 1958). The main effects of $\Delta F_0$ [$F(1,5)=28.07$, $p=0.003$] and PAIR [$F(19,95)=5.03$, $p=0.006$, GG$=0.2$] were significant, indicating that same-$F_0$ pairs were identified with more difficulty than different-$F_0$ pairs (15% vs 66% overall) and that identification rate varied across vowel pairs (from 27% to 54% overall). Their interaction was not significant, nor were the main effect of PHASE and its interactions with the other factors. Identification rates averaged over pairs and subjects are plotted in Fig. 3(a). The $\Delta F_0$ effect is large, and phase effects are negligible. These data do not support the hypothesis that the $\Delta F_0$ effect observed in classic ‘‘double-vowel’’ experiments is specific to the phase patterns (Klatt or sine) that were employed. The average number of vowels reported per stimulus [Fig. 3(b)] is also strongly dependent on $\Delta F_0$ but not on phase.

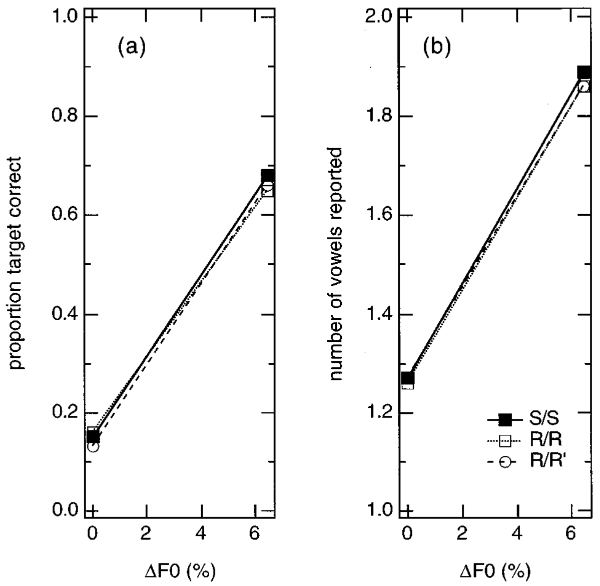

FIG. 3. (a) Target vowel identification rate as a function of $\Delta F_0$, for several phase patterns. (b) Number of vowels reported.

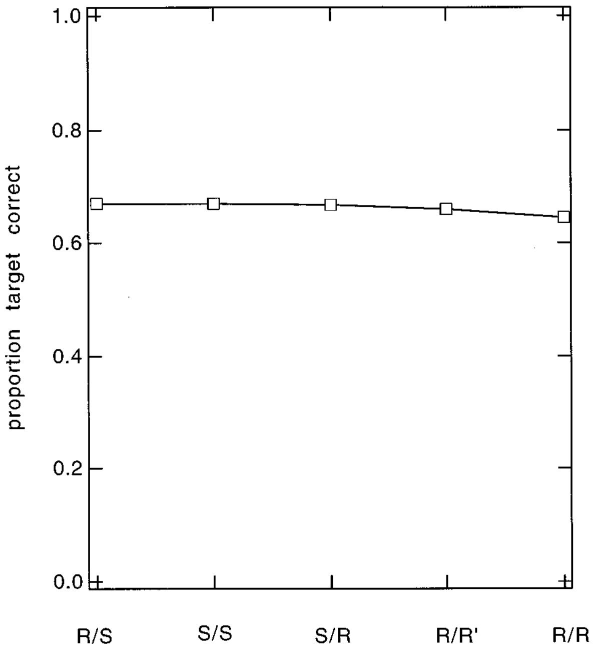

FIG. 4. Identification rate as a function of target and background vowel phase, at $\Delta F_0$=6.45%.

## III. EXPERIMENT 2: TESTING FOR A PHASE ARTIFACT

In a previous experiment (de Cheveigné *et al.*, 1995a), we presented subjects with vowel pairs in which each vowel was either harmonic or inharmonic. When the $F_0$'s differed by 2.9%, we found that target identification was better when targets were inharmonic rather than harmonic. It was also better when the competing vowel was harmonic rather than inharmonic. We attributed that pattern of results to a particular segregation strategy (harmonic cancellation) that is sensitive to harmonicity. However, as pointed out in the Introduction, harmonic vowels used in that experiment were in sine phase, whereas inharmonic vowels shifted to random phase. If phase *per se* were sufficient to explain the results, then we should expect similar results for harmonic vowels with the same phase patterns. If such an outcome were observed, it would cast doubt on the generality of the conclusions of the harmonicity experiment.

Experiment 2 tested four phase relations that arose in the harmonicity experiment (the notation $x/y$ means target phase $x$ with background phase $y$): S/S (sine/sine), S/R (sine/random), R/S (random/sine), and R/R′ (random/random, different random pattern), plus a fifth one: R/R (random/random, same random pattern), at a $\Delta F_0$ of 6.45%. Identification rates were averaged over $F_0$'s (2) and sessions (5), and were subjected to a repeated-measures ANOVA with factors PHASE (5) and ordered vowel PAIR (20). The main effect of PAIR barely missed the 5% significance level [$F(19.95)=2.89$, $p=0.06$, GG=0.18]. Neither PHASE nor its interaction with PAIR were significant. Identification rates averaged over subjects, pairs, and sessions are plotted in Fig. 4. Phase does not appear to affect identification at this $\Delta F_0$, and there is no evidence of the hypothesized artifact.

## IV. EXPERIMENT 3: HARMONICITY

Experiment 2 argued against the role of a phase artifact in the experiment reported by de Cheveigné *et al.* (1995a). Experiment 3 confirmed this conclusion by reproducing three crucial conditions of that experiment, using stimuli designed to reduce the usefulness of PPA or beat cues as follows. (1) Intravowel starting phase was ''random,'' to reduce the salience of temporal cues, and each vowel had a different random phase, so residual temporal features, if any, were not common to both vowels. (2) Intervowel phase was also ''random'' and remained random with ongoing phase shifts due to $\Delta F_0$ or inharmonicity. Beats were not eliminated, but as they occurred with random phases within different channels, there is no reason why the pattern arising in any particular condition should favor that condition over others. (3) Pairs containing inharmonic vowels had no partials closer than 16 Hz. To use spectral cues caused by beats, the auditory system would therefore have had to sample the beat pattern with a resolution better than about 30 ms. This cannot be excluded, but we expect it to be more difficult than with slower beats (note that making all beats faster than 16 Hz would have required a larger minimum spacing between partials, which is hard to reconcile with other constraints described in Sec. I A). (4) As explained in Sec. I A, all components of all vowels were multiples of 4 Hz, so all beats admitted the effective length of the stimulus (250 ms) as a period or subperiod. The long-term spectrum of the stimulus was therefore independent of starting phase.

The stimulus conditions were I/H, H/H, and H/I, with a $\Delta F_0$ of 6.45% and an R/R′ phase pattern. Following the scoring process described in Sec. I C, identification rates for the weaker vowel were averaged over $F_0$'s (2) and sessions (5), and subjected to a repeated-measures ANOVA with factors HARMONICITY (3) and ordered PAIR (20). The main factors of HARMONICITY [$F(2,10)=47.89$, $p=0.0004$, GG =0.46] and PAIR [$F(19,95)=3.10$, $p0.04$, GG=0.21] were significant, as was their interaction [$F(31,190)=2.89$,

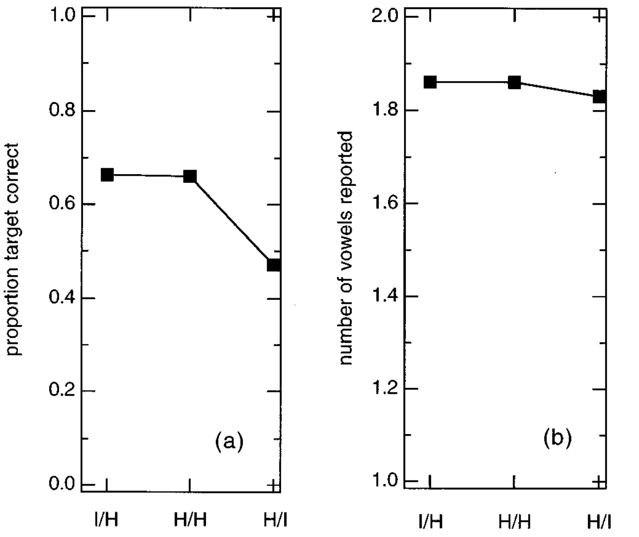

FIG. 5. (a) Identification rate as a function of target/ground harmonic state. (b) Number of vowels reported.

$p$=0.04, GG=0.12]. Identification rates averaged over pairs, subjects, and sessions are plotted in Fig. 5(a). Identification was better by about 21% when the background was harmonic rather than inharmonic. This effect is in the same direction as found by de Cheveigné *et al.* (1995a), but seven times larger. However, whereas they found that identification was better for inharmonic targets, here we observed no effect of target harmonicity.

In all conditions the compound stimulus was inharmonic. Inharmonicity of the *stimulus* seems to function as a strong multiplicity cue: The proportion of two-vowel responses was greater (86%) than when the stimulus was harmonic (27% in experiments 1 and 2 at $\Delta F_0$=0). This is evident also in the tendency of inharmonic single vowels to evoke more two-vowel responses than harmonic vowels (next paragraph). On the other hand for the inharmonic stimuli of experiment 3 it made no difference whether the *component* vowels were harmonic or not: the number of vowels reported did not differ significantly between conditions [Fig. 5(c)].

## V. SINGLE VOWELS

The stimulus set used in experiments 1–3 comprised 200 single vowels in addition to 400 double vowels. All single vowels were identified correctly more than 99% of the time; there is nothing to suggest that the phonetic quality of the constituents of the double vowels used in experiments 1–3 was affected by their component phases or harmonicity (despite the fact that inharmonic vowels sounded unnatural). On the other hand, fewer than 9% of harmonic vowels but more than 63% of inharmonic single vowels evoked two-vowel responses.

## VI. EXPERIMENT 4: TASK

The one-or-two response task we used differs from the classic two-response task used in double-vowel experiments. It is sensitive to ''multiplicity'' cues that influence the number of vowels reported, and also the identification rate. To see how the task affected identification rate, and permit comparison with previous reports, we reproduced several conditions of experiments 1–3 with the same subjects, but using the classic two-response task.

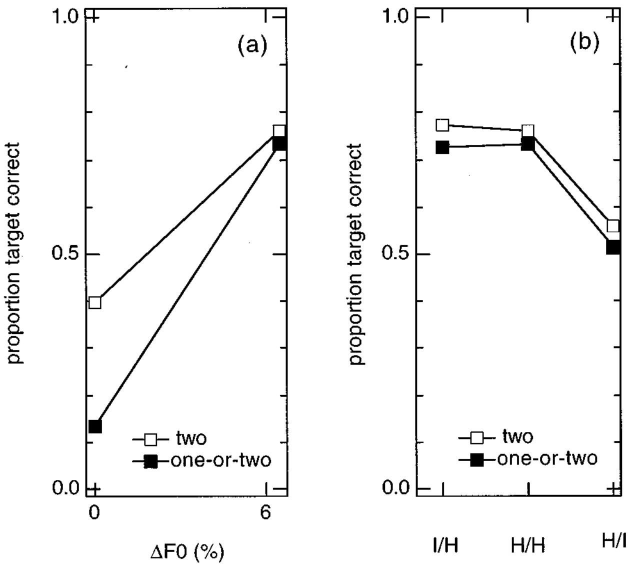

FIG. 6. (a) Identification rate as a function of $\Delta F_0$ in the $H/H$ condition for the two response task (open symbols) and the one-or-two response task (filled symbols). (b) Identification rate as a function of target/ground harmonic state at $\Delta F_0$=6%, for both tasks.

Subjects were five of the six that participated in experiments 1–3. Conditions were H/H at $\Delta F_0$=0, and I/H, H/H, and H/I at $\Delta F_0$=6.45%. Phase was R/R′. Within the stimulus set there were (10 unordered pairs)×(4 conditions)×(2$F_0$ orders)×(2 amplitude orders). After scoring as explained in Sec. I C, identification rates were paired with those obtained in experiments 1–3 by the same subjects for the same conditions, and submitted to a repeated-measures ANOVA with factors TASK (2), CONDITION (4), and ordered PAIR (20). The main effect of CONDITION [$F$(3,12)=56.47, $p$=0.0008, GG=0.39] was significant as was the TASK by CONDITION interaction [$F$(3,12)=21.37, $p$=0.008, GG =0.36]. Results are plotted in Fig. 6. Identification rates at $\Delta F_0$=0 in the H/H condition were higher when the subjects were forced to report two vowels [$F$(1,12)=115.20, $p$=0.0007, GG=0.36]. The result is a smaller $\Delta F_0$ effect size for the classic task [Fig. 6(a)]. Task had no significant influence on the pattern of results for harmonicity [Fig. 6(b)].

## VII. DISCUSSION

### A. $\Delta F_0$

The $\Delta F_0$ effect plotted in Fig. 3(a) is large compared to $\Delta F_0$ effects usually observed. This results from the combined benefit of the 15-dB amplitude mismatch (de Cheveigné *et al.*, 1997) and the one-or-two-vowel task (Sec. I E). The effect was not reduced with intravowel phase patterns that eliminated waveform cues required by the PPA hypothesis, nor was it affected by the intervowel phase pattern that determines both the relative phase of beats within different channels, and the spectrum of the double-vowel stimulus at $\Delta F_0 = 0\%$.

### B. Phase

We found no measurable effect of phase at either $\Delta F_0$. Effects for factors other than phase were large, and our experiments did not lack statistical power. This result is surprising and hard to reconcile with the PPA hypothesis that presumably requires peaked waveforms. A possible explanation is that the 15-dB mismatch was too great for a PPA-type unmasking effect to occur, even with peaked waveforms. Another explanation is that harmonic cancellation was highly effective because the background $F_0$ was easy to estimate, and all other effects were dwarfed. If so, the amplitude mismatch that we introduced to enhance sensitivity actually had the opposite result.

The beat hypothesis was introduced to explain effects of $\Delta F_0$'s smaller than one semitone (Culling and Darwin, 1994), but that mechanism might be expected to still have some effect at one semitone as in our experiment. In its simplest form, the beat hypothesis supposes that ongoing waveform interaction due to $F_0$ differences produces spectra that temporarily favor the identification of one vowel or the other (or perhaps the two together). If such were the case, one might expect identification to be affected by static, phase-dependent differences in waveform interactions at $\Delta F_0 = 0\%$, at least for individual vowel pairs. Instead, we found neither a main effect of phase, nor an interaction between phase and vowel pair. Once again, a possible explanation is that the 15-dB amplitude mismatch reduced spectral differences between phase conditions. An alternative form of the beat hypothesis is that identification depends on *dynamic* features of the beat pattern not present in our fixed-phase stimuli. Dynamic features also imply frequency cues exploitable by $F_0$-guided mechanisms, so it is difficult to design an experiment that triggers one type of mechanism and not the other. Further evidence against the beat hypothesis may be found in the results of experiment 3. The I/H and H/I conditions are symmetrical and should produce similar beats, so a segregation mechanism based on beats cannot explain the asymmetry observed between these two conditions.

### C. Background harmonicity

We found a strong effect of background harmonicity for both tasks [Figs. 5(a) and 6(b)]. The effect is the same as found previously (de Cheveigné *et al.*, 1995a), but about seven times larger. Several factors may explain the difference in effect size: (1) the larger $\Delta F_0$ (6.45% rather than 2.9%); (2) the different inharmonic patterns, with larger mistunings; (3) the 15-dB amplitude mismatch that may have made harmonic cancellation more effective. There were also differences in vowel set, stimulus generation, and subjects. As previously, the results support the hypothesis that the auditory system uses a strategy of *harmonic cancellation* to separate vowels. Vowels that we called ''inharmonic'' were only mildly so (they retained a relatively clear pitch), which may explain why the background harmonicity effect [Fig. 6(b), open symbols] was only about half the size of the $\Delta F_0$ effect [Fig. 6(a), open symbols].

### D. Target harmonicity

We found no effect of target harmonicity. This result contradicts our previous finding that a target was easier to identify when it was inharmonic rather than harmonic (de Cheveigné *et al.*, 1995a). That effect was paradoxical in that it was the opposite of the effect predicted by the hypothesis of harmonic enhancement. A tentative explanation that we offered was that cancellation is employed indiscriminately by the auditory system whenever segregation is called for. Harmonic targets are more likely to be victims of cancellation than inharmonic targets, so they are less well identified, hence the paradoxical effect. In the present experiment, targets were weak so the cancellation system would have found it more difficult to lock onto their harmonic structure, which may account for the lack of effect. In any case, neither experiment supported the hypothesis of *harmonic enhancement.*

### E. Task

Subjects found the one-or-two response task in experiments 1–3 natural and easy to perform, and complained when they were forced to report two vowels in experiment 4. The one-or-two response task is sensitive to segregation cues that signal the *multiplicity* of sources. The classic task ignores these cues, since the subject must report two vowels whether they are heard or not. The one-or-two task produced larger $\Delta F_0$ effects [Fig. 6(a)], mainly because identification was less good at $\Delta F_0 = 0$ where subjects tended to report only one vowel. Conditions that elicited double responses were less affected by the change of task [Fig. 6(b)].

One can object to the one-or-two task on the grounds that it taps into two different processes that both affect identification (one which senses the ''multiplicity'' of sources, the other which performs ''unmasking''). Different subjects may give different weights to each, so one is not sure exactly what is being measured. Indeed, de Cheveigné *et al.* (1997) found that the pattern of identification conditional on two-vowel responses varied between subjects, suggesting differences in strategy. The classic two-response task is easier to interpret because subjects are encouraged to ignore ''multiplicity'' cues, so identification rates depend only on the ''unmasking'' process. Similar remarks might be made for identification thresholds measured by an adaptive technique used by Summerfield and Assmann (1991), Summerfield (1992), Summerfield and Culling (1992), Culling and Darwin (1994), and Culling and Summerfield (1995). In those experiments, subjects had to decide which interval contained the target and to identify the target. The background was a random vowel-like sound, different for each trial and each interval. It is possible that identification of the correct interval was aided by the presence of a ''multiplicity'' cue similar to those discussed here. However, according to J. Culling (personal communication), listeners rarely made errors with regard to target interval in this paradigm, even at identification threshold.

The two measures (identification rate and number of vowels reported) are neither independent nor equivalent. In some cases they covaried [Fig. 3(a) and (b)]. In others, the response count was constant while identification rate varied [Fig. 5(a) and (b)]. In others the opposite was true: harmonic and inharmonic single vowels were recognized with the same accuracy, but the latter evoked double responses more often than the former (63% vs 9%). The number of vowels reported may be a useful measure in future studies of segregation.

## VIII. CONCLUSIONS

(1) The $\Delta F_0$ effect measured in a double-vowel experiment was not affected by the particular phase patterns chosen. This suggests that segregation was not the result of mechanisms sensitive to phase-dependent waveform interactions due to pitch-period asynchrony or beats.

(2) Phase effects did not constitute an artifact in a previous experiment on harmonicity. We reproduced our previous finding that identification is better when background vowels are harmonic. The result is consistent with the hypothesis of *harmonic cancellation*.

(3) We failed to reproduce our previous paradoxical finding that identification was better when targets were inharmonic rather than harmonic. Here we found no effect of target harmonicity. In either case the conclusion is the same; we found no evidence of *harmonic enhancement*.

(4) A task in which subjects may report one or two vowels is easier to perform and tends to produce larger $\Delta F_0$ effects than the two-response task used in classic experiments. The *number of vowels reported* is an interesting measure of segregation, sensitive to cues that signal the multiplicity of sources.

## ACKNOWLEDGMENTS

The experiments were carried out at ATR Human Information Processing Laboratories, within a research collaboration agreement with the Centre National de la Recherche Scientifique. The first author thanks ATR for its kind hospitality and the CNRS for leave of absence. Hideki Kawahara, Minory Tsuzaki, Kiyoaki Aikawa, Hiroaki Kato, and Ikuyo Masuda contributed ideas and advice and Rieko Kubo supervised the experiments. Thanks to John Culling of the MRC Institute of Hearing Research for providing the software for stimulus synthesis, and to him and Peter Assmann for detailed comments on previous versions of the manuscript.

Assmann, P. F., and Summerfield, Q. (**1988**). ''Pitch-pulse asynchrony and the perceptual segregation of competing voices,'' Speech 88 Conference (7th FASE), Institute of Acoustics, Edinburgh, pp. 531–538.

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.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). ''Intonation and the perceptual separation of simultaneous voices,'' J. Phon. **10**, 23–36.

Carlyon, R. P., and Shackleton, T. M. (**1994**). ''Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms?,'' J. Acoust. Soc. Am. **95,** 3541–3554.

Culling, J. F., and Darwin, C. J. (**1993**). ''Perceptual separation of simultaneous vowels: Within and across-formant grouping by $F_0$,'' 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. (**1996**). Personal communication.

Culling, J. F., and Summerfield, Q. (**1995**). ''The role of frequency and modulation in the perceptual segregation of concurrent vowels,'' J. Acoust. Soc. Am. **98**, 837–846.

Darwin, C. J. (**1981**). ''Perceptual grouping of speech components differing in fundamental frequency and onset-time,'' Q. J. Exp. Psychol. A **33**, 185–207.

Darwin, C. J., and Gardner, R. B. (**1986**). ''Mistuning of a harmonic of a vowel: Grouping and phase effects on vowel quality,'' J. Acoust. Soc. Am. **79,** 838–845.

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. (**1994**). ''Strategies for voice separation based on harmonicity,'' Proceedings of the International Conference Speech and Language Processing, Yokohama (Acoustical Society of Japan), pp. 1071–1074.

de Cheveigné, A. (**1995**). ''Experiments in vowel segregation'' ATR Human Information Processing Research Labs Tech. Report TR-H-154 (unpublished).

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

de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (**1995b**). ''Sensitive experimental techniques for the study of sound segregation,'' ASJ Autumn meeting (Acoustical Society of Japan), pp. 373–374.

de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (**1997**). ''Concurrent vowel segregation I. Effects of relative amplitude and $F_0$ difference,'' J. Acoust. Soc. Am. **101,** 2839–2847.

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.

Holdsworth, J., Nimmo-Smith, I., Patterson, R. D., and Rice, P. (**1988**). ''Implementing a GammaTone filter bank (SVOS final report, annex C),'' MRC Applied Psychology Unit Tech. Report (unpublished).

Horst, J. W., Javel, E., and Farley, G. R. (**1986**). ''Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms,'' J. Acoust. Soc. Am. **79,** 398–416.

Kohlrausch, A., and Sander, A. (**1995**). ''Phase effects in masking related to dispersion in the inner ear. II Masking period patterns of short targets,'' J. Acoust. Soc. Am. **97,** 1817–1829.

Lea, A. (**1992**). ''Auditory models of vowel perception,'' Ph.D. thesis, Nottingham (unpublished).

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

Moore, B. C. J., and Alcántara, J. I. (**1995**). ''Identification of flat-spectrum vowels on the basis of amplitude modulation,'' J. Acoust. Soc. Am. **97,** 3274.

Moore, B. C. J., and Alcántara, J. I. (**1996**). ''Vowel identification based on amplitude modulation,'' J. Acoust. Soc. Am. **99,** 2332–2343.

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.

Palmer, A. R., Winter, I. M., Gardner, R. B., and Darwin, C. J. (**1987**). ''Changes in the phonemic quality and neural representation of a vowel by alteration of the relative phase of harmonics near $F1$,'' in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Martinus Nijhoff, Dordrecht), pp. 371–376.

Plack, C. J., and Moore, B. C. J. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87,** 2178–2187.

Scheffers, M. T. M. (**1983**). ''Sifting vowels,'' Ph.D. thesis, Gröningen (unpublished).

Schroeder, M. R., and Strube, H. W. (**1986**). ‘‘Flat-spectrum speech,’’ J. Acoust. Soc. Am. **79,** 1580–1583.

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 Assmann, P. F. (**1991**). ‘‘Perception of concurrent vowels: Effects of harmonic misalignment and pitch-period asynchrony,’’ J. Acoust. Soc. Am. **89,** 1364–1377.

Summerfield, Q., and Culling, J. F. (**1992**). ‘‘Periodicity of maskers not targets determines ease of perceptual segregation using differences in fundamental frequency,’’ J. Acoust. Soc. Am. **92,** 2317(A).

Traunmüller, H. (**1987**). ‘‘Phase vowels,’’ in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Martinus Nijhoff, Dordrecht), pp. 377–384.

Zwicker, U. T. (**1984**). ‘‘Auditory recognition of diotic and dichotic vowel pairs,’’ Speech Commun. **3**, 256–277.