Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement

Alain de Cheveigné, Stephen McAdams, Jean Laroche, and Muriel Rosenberg

Citation: The Journal of the Acoustical Society of America **97**, 3736 (1995);
View online: https://doi.org/10.1121/1.412389
View Table of Contents: http://asa.scitation.org/toc/jas/97/6
Published by the Acoustical Society of America

Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement

Alain de Cheveigné
Laboratoire de Linguistique Formelle (CNRS, URA1028), Université Paris 7, 2 place Jussieu, case 7003, F-75251 Paris Cédex 05, France

Stephen McAdams
Laboratoire de Psychologie Expérimentale (CNRS, URA316), Université René Descartes, EPHE, 28 rue Serpente, F-75006 Paris, France, and IRCAM, 1 place Stravinsky, F-75004 Paris, France

Jean Laroche and Muriel Rosenberg
Département Signal, Télécom Paris (ENST/CNRS), 46 rue Barrault, F-75634 Paris Cédex 13, France

(Received 15 March 1994; revised 27 September 1994; accepted 20 December 1994)

The improvement of identification accuracy of concurrent vowels with differences in fundamental frequency (ΔF_0) is usually attributed to mechanisms that exploit harmonic structure. To decide whether identification is aided primarily by selecting the target vowel on the basis of its harmonic structure ("harmonic enhancement") or removing the interfering vowel on the basis of its harmonic structure ("harmonic cancellation"), pairs of synthetic vowels, each of which was either harmonic or inharmonic, were presented to listeners for identification. Responses for each vowel were scored according to the vowel's harmonicity and that of the vowel that accompanied it. For a given target, identification was better by about 3% for a harmonic ground unless the target was also harmonic with the same F_0. This supports the cancellation hypothesis. Identification was worse for harmonic than for inharmonic targets by 3%–8%. This does not support the enhancement hypothesis. When both vowels were harmonic, identification was better by about 6% when the F_0's differed by 1/2 semitone, consistent with previous experiments. Results are interpreted in terms of harmonic enhancement and harmonic cancellation, and alternative explanations such as waveform interaction are considered.

PACS numbers: 43.66.Ba, 43.66.Lj, 43.71.Es

INTRODUCTION

When two voices are present at the same time, differences in fundamental frequency (F_0) can help listeners attend to one or the other voice and understand what is being said. This has been verified for natural and synthetic speech (Brokx and Nooteboom, 1982) and for pairs of synthetic vowels (Scheffers, 1983; Culling and Darwin, 1993). One interpretation is that differences in F_0 allow the voices to be perceptually segregated from each other. Various models and methods have been proposed to explain or reproduce this process (see de Cheveigné, 1993a, for a review). Some make use of the harmonic structure of a voice to identify its components within the composite spectrum. The voice is then isolated by *enhancing* those components relative to the ground. Others make use of the harmonic structure of the interfering voice, which is then removed by *cancelling* its components. Either strategy (or both) can be used if both voices are harmonic, as long as they have different F_0's. Both strategies fail if the vowels have the same F_0, which explains why performance in double-vowel identification experiments is not as good in this case.

Each strategy has its advantages and disadvantages. Harmonic enhancement allows harmonic sounds such as voiced speech to emerge from any type of interference (except harmonic interference with the same F_0 as the target). Harmonic cancellation, on the other hand, allows any type of target to emerge from harmonic interference. Enhancement works best when the signal-to-noise ratio is high, because the F_0 of the target is then relatively easy to estimate. However, separation is probably most needed when the signal-to-noise ratio is low, in which case cancellation should be easier to implement. Cancellation removes all components that belong to the harmonic series of the interference, and may thus distort the spectrum of the target. Enhancement should cause no spectral distortion to the target, as long as it is perfectly harmonic. Cancellation of perfectly harmonic interference can be obtained using a filter with a short impulse response, whereas enhancement requires a filter with a long impulse response to be effective (de Cheveigné, 1993a). The dynamic nature of speech may limit the effectiveness of such a filter.

The aim of this paper is to study the degree to which each strategy is used by the auditory system in a double-vowel identification experiment. An answer to this question may allow us to better understand auditory processes of sound organization, and refine our models of harmonic sound separation. We first review the literature on mixed vowel identification experiments and present the rationale and predictions for our experiment.

 0001-4966/95/97(6)/3736/13/$6.00

A. Double-vowel identification experiments

A mixture of several voices poses to a listener what Cherry (1953) called the "cocktail party problem." Cherry showed that among the cues useful to the listener trying to track a source is its spatial position, which creates binaural information that the auditory system uses to segregate the source. Another important cue for the separation of natural speech is the fundamental frequency. Brokx and Nooteboom (1982) found that this cue helped listeners separate competing speech streams and better reproduce the message carried by one stream or another. The effects of fundamental frequency differences are reinforced by dichotic presentation (Summerfield and Assmann, 1991; Zwicker, 1984).

Another cue that might be expected to reinforce F_0 differences is frequency modulation (FM), particularly if competing streams are modulated incoherently. McAdams (1989) and Marin and McAdams (1991) showed that FM increased the perceptual prominence of a vowel presented concurrently with two other vowels at relatively large F_0 separations (five semitones, or 33%). However, they also found that this increase did not depend on whether the vowels were modulated coherently or separately. Subsequent studies showed that effects once attributed to FM incoherence can be accounted for by the instantaneous differences in F_0 that it causes (Demany and Semal, 1990; Carlyon, 1991; Summerfield, 1992; Summerfield and Culling, 1992a). These results suggest a crucial importance of harmonicity, exploited by the auditory system when there are differences in fundamental frequency (ΔF_0) between constituents of an acoustic mixture. The effects of ΔF_0 have been studied in detail by a number of authors (Assmann and Summerfield, 1989, 1990; Scheffers, 1983; Summerfield and Assmann, 1991; Zwicker, 1984; Chalikia and Bregman, 1989, 1993; Darwin and Culling, 1990; Culling and Darwin, 1993). In these studies, two synthetic vowels were presented simultaneously at various ΔF_0 values and subjects were requested to identify both vowels from a predetermined set of five to eight vowels. Identification scores reflecting the ability to identify both vowels (combinations-correct score) for several of these studies are plotted in Fig. 1.

There are large differences in overall identification rate between studies that may be attributed to differences in training of subjects, presence or absence of feedback, size of vowel set, inclusion of pairs of identical vowels, stimulus duration, level, etc. A common trend is a rapid increase in identification performance with ΔF_0 up to between 1/2 and 2 semitone separation (3%–12% difference in F_0), followed by an asymptote. This effect is usually explained by assuming that the mechanism that exploits the harmonic structure of the vowel spectrum is effective when the F_0's are different but fails when they are the same and the harmonic series of both vowels coincide. However, a question that none of these studies has addressed is whether it is primarily the harmonicity of the vowel being recognized that aids its segregation and subsequent identification, or that of the background vowel. This leaves unresolved many issues involved in the design of voice separation models. The primary aim of the present study is to directly test the effect of the harmonicity of both the target vowel and the background vowel on the target's identification.

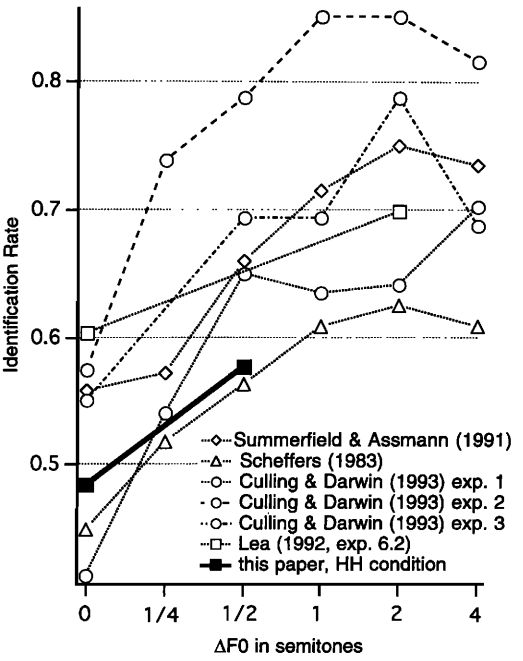

FIG. 1. Dotted lines: combination-correct identification rates as a function of ΔF_0 reported in previous studies. Continuous line: combination correct rates obtained in this study for pairs of harmonic vowels (*H*/*H* condition).

One study that approached this question was conducted by Lea (1992; Lea and Summerfield, 1992). He presented listeners with pairs of vowels of which each could be either voiced or whispered, and requested them to identify both vowels. He scored results according to the harmonicity of the vowel being named (the target) and that of the other vowel (the ground), and found that targets were better identified when the ground was voiced than when it was whispered. There was no significant advantage when the target itself was voiced rather than whispered. However, with a slightly different method, Lea and Tsuzaki (1993a,b) found that targets were better recognized when they were voiced.

A difficulty with this experiment is that it requires voiced and whispered vowels to be equivalent in both "phonetic quality" and "masking power" (except insofar as these depend on harmonicity). This is a difficult requirement because it is not evident how one should go about matching the continuous spectrum of a whispered vowel to the discrete spectrum of a voiced vowel. Lea (1992) used a model of basilar membrane excitation to match the vowels, but the possibility remains that some imbalance, for example of level, might have affected the results. In the experiment to be described in this paper, whispered vowels were replaced by inharmonic vowels with spectral structure and density closer to those of harmonic vowels.

Summerfield and Culling (1992b) measured relative intensity thresholds for identification of synthetic vowels in the presence of vowel-like maskers with spectra that were either harmonic or inharmonic (with partials displaced randomly). They found that thresholds were lower for harmonic than for

inharmonic maskers, by about 5–7 dB. This is strong evidence for cancellation. In contrast, the harmonic state of the target had little effect on the threshold, suggesting that enhancement plays a minor role, if any. However, we noted earlier that enhancement is difficult to perform at low signal-to-noise ratios, so their experiment is perhaps not the most sensitive test.

B. Experimental rationale and predictions

We wished to determine whether the auditory system uses the harmonicity of the target or that of the ground to segregate the target from the mixture. For that purpose we used stimuli consisting of pairs of vowels, each of which was either harmonic or inharmonic. Inharmonic vowels were obtained by perturbing the frequencies of the components of a harmonic vowel by small random amounts, as explained in Sec. I B and Appendix B. We define the "fundamental frequency" of an inharmonic vowel as the fundamental frequency of the harmonic series before perturbation. In addition to harmonicity states we introduced differences in fundamental frequency (ΔF_0) in order to compare their effects and study their interaction, and allow comparisons with previous studies. Pairs of vowels were presented together. Subjects were asked to identify both vowels and respond with an unordered pair of vowel names. For each vowel in the stimulus, the answer was deemed correct if the vowel's name appeared within the response pair. This answer was classified according to the harmonic state of that vowel (the target), the state of the other vowel (the ground), and the F_0 difference between them. This process was repeated for the second vowel in the pair, reversing the roles of target and ground.

In this paper, the notation H/I, for example, indicates a harmonic target with an inharmonic ground, and $R(H/I)$ indicates the identification rate for that target. Other combinations are denoted I/H, H/H, and I/I. Where necessary, the relation between the F_0's may also be specified: $H/I0$ signifies the same F_0 and H/Ix signifies a different F_0 (H/I signifies a regrouping of $H/I0$ and H/Ix together). For each hypothesis concerning the strategy that is used by the auditory system to separate harmonic sounds, specific predictions can be made concerning the outcome of this experiment.

1. Enhancement

According to this hypothesis, harmonicity of the target promotes segregation from the ground (unless the ground is also harmonic and has the same F_0). All else being equal, a target should be better identified if it is harmonic:

$$R(H/I0)>R(I/I0),$$

$$R(H/Ix)>R(I/Ix),$$

$$R(H/Hx)>R(I/Hx).$$

If the hypothesis is false, these differences should be insignificant.

2. Cancellation

According to this hypothesis, harmonicity of the ground allows the target to be segregated (unless it is also harmonic and has the same F_0). All else being equal, identification should be better when the ground is harmonic:

$$R(I/H0)>R(I/I0),$$

$$R(I/Hx)>R(I/Ix),$$

$$R(H/Hx)>R(H/Ix).$$

If the hypothesis is false, the differences should be insignificant. In addition to these two hypotheses that our experiment was specifically designed to test, there are others that are worth considering.

3. Symmetric mechanisms

According to Bregman (1990), a characteristic of primitive segregation is the symmetry of its effects: segregation causes both parts of a mixture to become equally accessible. Thus vowels in a pair should be equally affected by factors that promote segregation. In that case we expect

$$R(I/H0)=R(H/I0), \quad R(I/Hx)=R(H/Ix).$$

Several cues or mechanisms might show that behavior:

a. Component mismatch. According to this explanation, harmonicity *per se* is unimportant; segregation is limited by the proximity of components and thus increases when harmonic structures are different. In the $H/H0$ condition harmonic series coincide, whereas all other conditions introduce a mismatch between component frequencies that should ease identification of both constituents. Accordingly,

$$R(\text{all conditions other than } H/H0)>R(H/H0).$$

b. Beating between partials. Culling (1990), Culling and Darwin (1993, 1994), and Assmann and Summerfield (1994) suggested that beating between partials in the F_1 region might explain improvements in identification with ΔF_0. Beating occurs, for example, if two partials belonging to different vowels fall within the same auditory filter: the output fluctuates at a rate that depends on the difference in frequency between the partials. Fluctuations may allow the amplitudes of the two partials to be better estimated, as long as they are neither too slow to be appreciable within the duration of the stimulus, nor too fast to be resolved temporally by the auditory system. Beating is likely to affect identification in a complex fashion, but insofar as it depends only on the absolute frequency difference between partials of both vowels, both should be equally affected.

c. Quality differences (pitch, timbre). Vowels that share the same pitch and harmonic nature (such as constituents of the $H/H0$ and $I/I0$ conditions) may "sound alike" and thus be difficult to segregate when mixed. Differences in quality should promote segregation:

$$R(\text{conditions other than } H/H0, I/I0)>R(H/H0, I/I0).$$

This prediction differs from that of the component-mismatch hypothesis in that here the $I/I0$ condition does not promote segregation (if one assumes that all the inharmonic stimuli evoke a similar quality).

For all hypotheses, ΔF_0 effects are likely to be smaller when either vowel is inharmonic than when both are harmonic. For example, in the *I/H* condition the effectiveness of enhancement would be limited, whereas that of cancellation should change relatively little with ΔF_0 (much of the ΔF_0 effect in the *H/H* condition is due to the fact that when $\Delta F_0 = 0$ all target components fall precisely on the ground vowel's harmonic series, and are canceled together with those of the ground). Component mismatch or beating should also be less affected by ΔF_0 than in the *H/H* condition, leading to smaller effects when either vowel is inharmonic.

When both vowels are harmonic, all hypotheses predict alike:

$$R(H/Hx) > R(H/H0).$$

It is for this reason that classic double-vowel experiments do not allow us to choose between these various hypotheses.

I. STIMULI

A. Spectral envelopes

Vowels belonged to a set of French vowels—/a/, /e/, /i/, /o/, /u/—which have equivalents in many different languages. The spectral envelopes were derived from natural voiced speech by a screening procedure that produced a set of ten allophones for each vowel (see Appendix A). Envelopes for each experimental condition were drawn at random from the allophone set. By using allophones that were selected randomly for each trial, we hoped (a) to reduce the likelihood that a listener might learn the spectra of particular combinations of synthetic vowels and respond correctly without using separation mechanisms, (b) to make the task more difficult in conditions such as equal F_0 and thus obtain larger effects, and (c) to lower the overall recognition rate to avoid ceiling effects. We reasoned that intraclass variability would make the task more typical of situations in which human beings recognize speech.

B. Harmonic structure

Vowels were synthesized in one of two harmonicity states (harmonic and inharmonic) and at three nominal fundamental frequencies (125 Hz and ±1/4 semitone, i.e., ±1.45% of the F_0). Harmonic vowels had component frequencies equally spaced at multiples of the F_0. For inharmonic vowels, each component frequency was shifted from the harmonic series by an amount drawn at random from a uniform distribution bounded by ±3% of the harmonic frequency, or half the spacing between adjacent harmonics, whichever was smaller (see Appendix B for more details). The F_0 of an inharmonic vowel is by definition that of the harmonic series before modification. We chose to use a rather mild perturbation to ensure that the spectral density was similar to that of a harmonic vowel shaped by the same envelope. Different inharmonic component frequency patterns were used for different allophones, but for each allophone the same pattern was used at different F_0's. An example of the derivation of an inharmonic pattern is illustrated in Fig. 2.

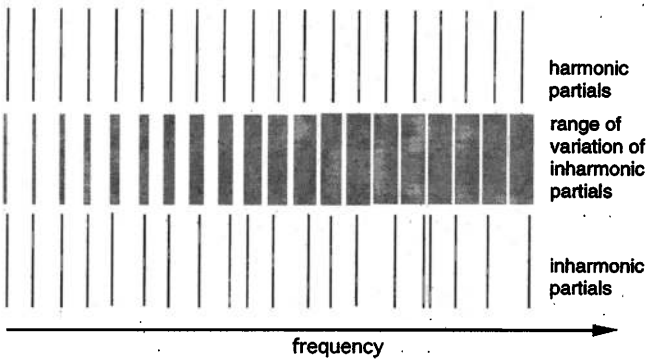

FIG. 2. Top: harmonic series; middle: range of frequencies from which inharmonic partials are drawn; bottom: a particular inharmonic series.

The F_0 values we chose allow ΔF_0's of 0% and 2.9% (1/2 semitone) to be investigated. Based on previous studies (Fig. 1), such values should ensure an effect large enough to be significant while leaving room for improvement with other factors. The maximum frequency shift of the partials of our inharmonic vowels is also about 1/2 semitone. This happens to be the mistuning up to which individual low-frequency partials still make a full contribution to virtual pitch, as estimated by Moore *et al.* (1985).

C. Synthesis

Individual vowels were generated by additive synthesis at a sampling rate of 16 kHz. Their spectra comprised 45 components with amplitudes determined by interpolated look-up in a spectral envelope table corresponding to a given allophone. There was an additional −5-dB/component deemphasis from the 30th to the 45th component. All components started in sine phase.

II. PRETEST: SINGLE VOWEL IDENTIFICATION

The purpose of the pretest was to verify that listeners could correctly identify all allophones of the synthesized vowels used in the experiment. We also wished to check for any systematic effects of harmonicity or F_0 on the identifiability of vowels, as such effects might interfere with the effects studied in the main experiments.

A. Subjects

Subjects were 21 male and 11 female caucasian *homo sapiens* volunteers recruited from the staff and students at IRCAM and ENST (including the four authors). Their ages ranged from 23 to 50 yr. None of the subjects reported having a hearing disorder. The subjects had French as either their mother tongue (23) or as a highly fluent second language which they practiced on a daily basis in their professional lives (9). Most had extensive experience producing and listening to synthesized sounds. Nineteen of the subjects had participated in a similar pilot experiment about two months prior to this one.

B. Stimuli

Ten allophones of the vowels /a/, /e/, /i/, /o/, and /u/ were each synthesized at the three F_0s to be used in the main experiment (123.208, 125.0, 126.818 Hz). Each was synthesized in both harmonic and inharmonic versions. All stimuli were equalized for rms level, up sampled to 44.1 kHz, sent through the NeXT Cube D–A converters and presented diotically over Sennheiser HD 520 II earphones. The sound system was calibrated using a flat-plate coupler connected to a Bruel & Kjær 2209 sound level meter to obtain a level of approximately 60 dBA. Stimuli were 200 ms in duration including 25-ms raised cosine onset and offset ramps.

C. Procedure

Subjects were seated in a Soluna SN-1 double-walled soundproof booth, in front of a computer terminal that was used for prompting and to collect responses. Subjects were informed that they would hear individual vowel sounds and were to identify them as one of /a/, /e/, /i/, /o/, or /u/ by typing the appropriate key on the computer keyboard (a, e, i, o, u, respectively). They were informed that they needed to attain a criterion performance level of 95% to continue on to the main experiment. Each combination of allophone, nominal F_0, and harmonicity was presented once for a total of 300 trials in random order.

D. Results

All but two of the subjects attained 95% criterion performance and continued on to participate in the main experiment. The identification rates for the two subjects rejected were 91% and 94%. Mean performance for all allophones but three was better than 95%. One allophone fell between 90% and 95% (an /u/) and two below 90% (an /e/ and an /o/).

A multivariate repeated measures analysis of variance on factors vowel class (5)× harmonicity (2)×F_0 (3) was performed with, as the dependent variable, proportion correct identifications across allophones by each subject within a given condition. Each data point was based on ten judgments per subject. The analysis revealed no significant effect of fundamental frequency nor any significant interactions involving this factor. There was also no main effect of harmonicity but the interaction between vowel and harmonicity was significant [$F(4,124)=6.4$, $p=0.0002$, $GG=0.88$],[1] indicating an effect of harmonicity on vowel identification that is limited to certain vowels. Contrasts for each vowel class showed that harmonic stimuli were better identified than inharmonic ones for /e/ by 2.8% [$F(1,124)=19.5$, $p<0.0001$, $GG=0.88$] and the reverse was true by 1.4% for /u/ [$F(1,124)=4.5$, $p=0.041$, $GG=0.88$]. Differences for other vowels were not significant.

III. MAIN EXPERIMENT: DOUBLE-VOWEL IDENTIFICATION

A. Subjects

Subjects were the 30 who attained criterion performance on the pretest.

B. Stimuli

The stimulus set consisted of pairs of synthesized vowel allophones belonging to the set /a/, /e/, /i/, /o/, /u/. Vowels within a pair were always different, yielding ten unordered combinations. Each vowel within a pair was either harmonic or inharmonic, yielding four combinations of harmonicity. Finally, there were two conditions of F_0 difference: 0 and 1/2 semitone (2.9%). All factors, vowel pair (ten), harmonicity (four), and ΔF_0 (two), were crossed, giving 80 different combinations.

In addition to the factors that interest us, the design contained others that might also influence the phonetic quality of the target or the masking power of the ground: absolute F_0, choice of inharmonic pattern, choice of allophone, or presentation order. To avoid any systematic bias due to these factors, the following precautions were taken: (a) Pairs were duplicated so that each vowel of each pair occurred once at the higher and once at the lower F_0 when $\Delta F_0 \neq 0$. Duplication of ΔF_0 conditions resulted in a 160-stimulus set. (b) For each inharmonic allophone, the same component pattern was used to synthesize different F_0 conditions. (c) Allophones were assigned in a balanced fashion across conditions. For example, the subset of allophones representing the eight repetitions of the vowel /a/ (2 positions ×4 other vowels) in the *H/H*0 condition within a run of the stimulus set also represented that vowel in all other main conditions (*H/Hx*, *H/I*0, etc.). Other subsets were chosen for other runs. (d) Stimuli were presented in random order, and this order was renewed for each run and each subject.

In the inharmonic state a different component pattern was used for each allophone. Since vowels within a pair were different, *component patterns within inharmonic–inharmonic pairs were always different.* As noted above, all conditions used the same set of allophones, but for practical reasons it was not possible to guarantee that the occurrence of allophone pairs was similarly balanced. Allophones were paired at random, and the pairing was renewed for each presentation and subject.

Preliminary experiments had shown that when vowels are mixed at equal rms signal levels, one vowel might dominate the pair due to unequal mutual interference, as noted by McKeown (1992). In that case, the identification probability of one vowel is likely to be at its "floor" and the other at its "ceiling," both being thereby insensitive to the conditions of interest. To avoid such a situation, we performed a preliminary experiment to determine levels of equal "mutual interference" (see Appendix C). From these results we derived a level correction factor for all pairs, such that identification rates for both vowels were the same. Vowel levels were adjusted according to this factor, the vowels were summed, and the rms signal level of the sum was set to a standard level for all pairs, corresponding to a stimulus presentation level of about 60 dBA.

C. Procedure

The experimental apparatus was the same as in the pretest. Subjects were informed that they would hear a complex sound composed of two different vowels from the set /a/, /e/,

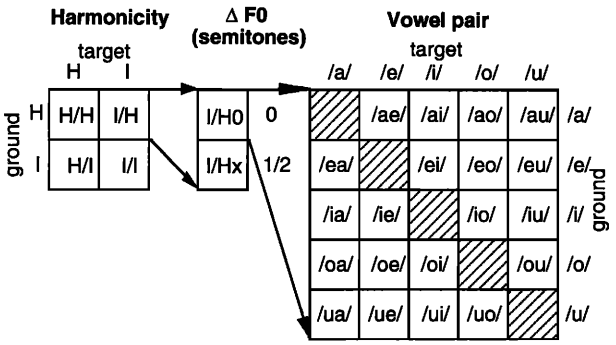

FIG. 3. Response conditions: target harmonicity×ground harmonicity $\times\Delta F_0\times$ vowel pairs.

/i/, /o/, /u/. Each vowel pair was presented once, followed by a visual prompt on the terminal screen. Subjects were required to hit two keys in succession, corresponding to the two vowels heard (two of a, e, i, o, u)—or else Q to quit temporarily. Any other response produced a message reminding the subject of the options, and requesting a new response. A response with two identical vowels produced a message reminding the subject that the vowels were different, and requesting a new response. Aside from information about response constraints, no feedback was given concerning the correct response. Subjects were presented with three consecutive runs of all combinations of vowel, harmonicity, and ΔF_0 in randomized order for a total of 480 stimuli.

The response to each stimulus was scored twice, once for each vowel present within the stimulus. A vowel was deemed correctly identified if its name appeared within the response pair. This partial response was classified according to the harmonic state of that vowel (the target), the state of the other vowel (the ground), the F_0 difference between the two, and the names of both vowels. This procedure was repeated for the other constituent vowel, reversing the roles of target and ground, leading to a total of 960 identifications for each subject. Figure 3 summarizes these conditions and their notation. This method of scoring is equivalent to that used by Lea (1992) to obtain "constituents-correct" scores.

D. Results

Within each harmonicity and ΔF_0 condition, proportion-correct identification measures for each target vowel were calculated for every subject across all vowel combinations, yielding eight data points per subject. Each data point was based on 120 judgements (20 vowel pairs ×2 vowel identifications ×3 repetitions). A multivariate repeated measures analysis of variance was performed on factors ΔF_0 (two), target harmonicity (two), and ground harmonicity (two). All interactions were statistically significant,[2] meaning that the effect of each factor differs according to the values of the other factors. There is little advantage in averaging such incongruous effects: we shall therefore ignore the main effects and consider only partial effects. Subsequent discussion will focus on tests of these partial effects in relation with the various hypotheses outlined in the Introduction.

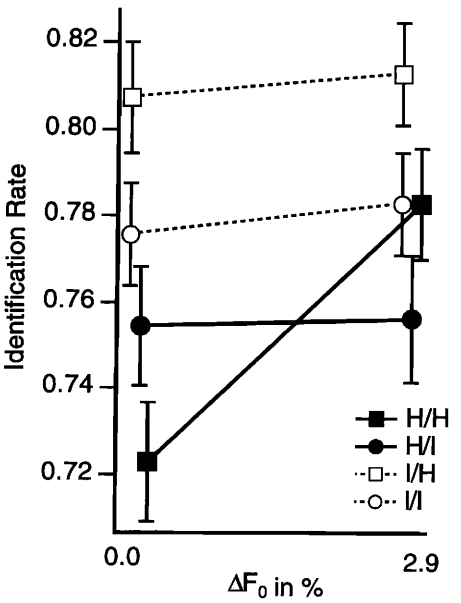

FIG. 4. Identification rate as a function of ΔF_0 for each of the harmonicity conditions. Error bars represent ±1 standard error of the mean. The standard deviations vary between 0.066 and 0.081. Data points for *H/H* and *H/I* are displaced horizontally for visibility.

1. Effect of ΔF_0

In Fig. 4 the means across subjects are plotted as a function of ΔF_0. Each line represents one of the four combinations of target and ground harmonicity. Filled symbols represent harmonic targets and open symbols inharmonic targets. Squares represent harmonic grounds and circles inharmonic grounds. When both vowels are harmonic, performance increases with ΔF_0, as predicted by all the hypotheses mentioned in the Introduction. Planned contrasts show that this effect, about 6%, is highly significant [$F(1,29)=50$, $p<0.0001$]. When at least one vowel is inharmonic, the effect is not significant [for H/I:$F(1,29)=0.1$; for I/H:$F(1,29)=0.4$; for I/I:$F(1,29)=0.4$]. We take advantage of this fact to group these conditions across ΔF_0 in subsequent contrasts.

2. Effect of harmonicity of ground

The data are replotted in Fig. 5 to emphasize the effects of ground and target harmonicity. Contrasts planned to test the cancellation hypothesis (Introduction, Sec. B 1) show that identification is significantly higher when the ground is harmonic, unless the target is also harmonic and $\Delta F_0=0$ [$R(I/H)$ vs $R(I/I)$:$F(1,29)=26$, $p<0.0001$; $R(H/Hx)$ vs $R(H/I)$:$F(1,29)=14$, $p=0.0008$]. The improvement in identification rate is about 3%. These results are compatible with the cancellation hypothesis. An additional contrast shows that when the target is harmonic and $\Delta F_0=0$, performance is significantly worse with a harmonic ground, also by about 3% [$R(H/H0)$ vs $R(H/I0)$:$F(1,29)=13$, $p=0.0009$].

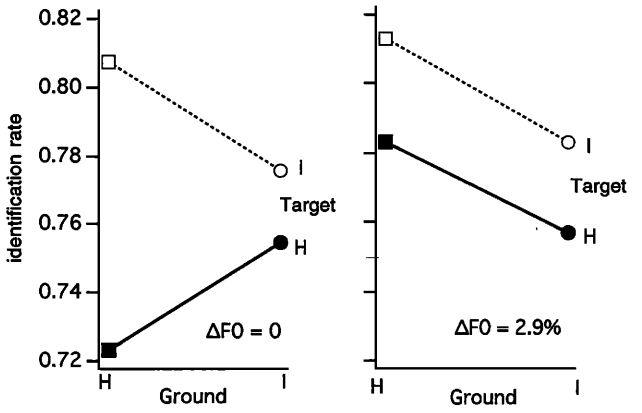

FIG. 5. Identification rate of target as a function of ground harmonicity, for harmonic and inharmonic targets and F_0 differences of 0 and 1/2 semitone.

3. Effect of harmonicity of target

Whatever the ΔF_0 and whatever the nature of the ground, identification is worse when the target is harmonic. Contrasts planned to test the enhancement hypothesis (Introduction, Sec. B 2) are highly significant [$R(H/I)$ vs $R(I/I)$:$F(1,29)=15$, $p=0.0004$; $R(H/Hx)$ vs $R(I/H)$: $F(1,29)=13$, $p=0.0008$], but the direction of the effects observed is *opposite* to that predicted by that hypothesis. The effect is similar in size, about 3%, to what was observed for ground harmonicity. An additional contrast shows that the larger effect (about 8%) obtained when the ground is harmonic and $\Delta F_0=0$ is also significant [$R(H/H0)$ vs $R(I/H0)$:$F(1,29)=99$, $p<0.0001$].

4. Evidence of symmetrical segregation

A contrast planned to test the hypothesis of symmetrical segregation (Introduction, Sec. B 3) shows that, contrary to a hypothesized lack of difference, performance is significantly better for I/H than for H/I conditions [$R(H/I)$ vs $R(H/I)$:$F(1,29)=96$, $p<0.0001$], by about 5% (Fig. 5). Symmetric segregation mechanisms cannot account for our results. They might, however, coexist with other asymmetric mechanisms, so it is of interest to consider contrasts specific to the various symmetric segregation hypotheses.

Performance for $H/H0$ is worse than for all other conditions [$R(H/I)$ vs $R(H/H0)$:$F(1,29)=19$, $p<0.0001$; $R(I/H)$ vs $R(H/H0)$:$F(1,29)=142$, $p<0.0001$; $R(I/I)$ vs $R(H/H0)$:$F(1,29)=59$, $p<0.0001$]. This would be consistent with the component-mismatch hypothesis, were it not for the asymmetry between $R(H/I)$ and $R(I/H)$ mentioned above.

Performance is better for I/H than for I/I [$F(1,29)=26$, $p<0.0001$] but worse for H/I than for I/I [$F(1,29)=15$, $p=0.0004$]. This is inconsistent with the quality differences hypothesis, already weakened by the asymmetry between $R(H/I)$ and $R(I/H)$.

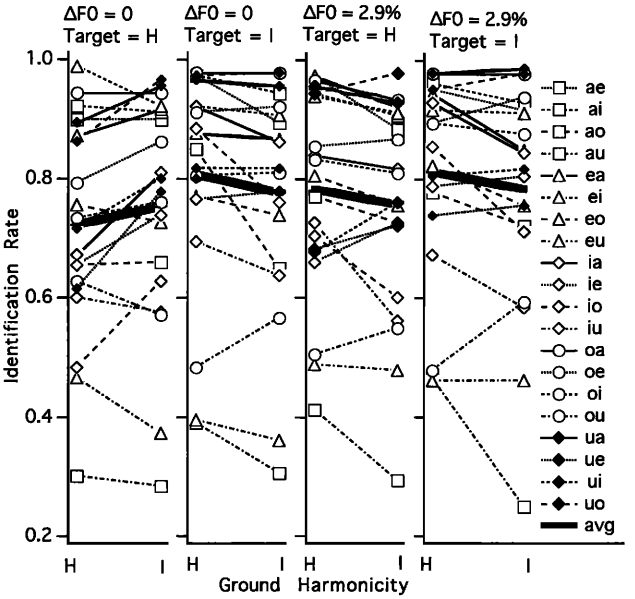

FIG. 6. Identification rate of target vowel as a function of ground harmonicity for each vowel pair and for all four conditions of ΔF_0 and target harmonicity. The thick lines without markers represent the effect averaged over vowel pairs, also plotted in Fig. 5.

5. Confusion matrix

Examination of the confusion matrix reveals a slight bias toward responses containing o (22.0%) and e (21.0%), rather than those containing i (19.1%), u (19.1%), or a (18.8%). The unordered response pair ou was recorded most often (14.2%), and au least often (7.1%). The vowel /u/ appears to be correctly identified most often (85%), followed by /o/ (80%), /e/ (76%), /a/ (73%), and /i/ (72%). Vowels paired with /a/ are identified correctly most often (91%), followed by those paired with /i/ (86%), /e/ (82%), /o/ (79%), and /u/ (49%). The poor rate for vowels paired with /u/ is almost certainly due to the excessive level emphasis given to /u/ relative to other vowels (see Appendix).

Results of the pretest suggested that harmonicity might have an effect on the identifiability of /e/ and /u/ targets, independent from any segregation effects. However, excluding either of these vowels from analysis does not affect the main pattern of results.

6. Dependency of effects on vowel pair

Our experiment was designed assuming that data would be averaged over vowel pairs (and thus over allophone pairs and component pattern pairs), because we had no theoretical reason to expect major differences in the way different vowel pairs, allophone pairs, or pattern pairs might affect the dependency of identification rate on our main conditions: ground harmonicity, target harmonicity, and ΔF_0. It is nevertheless of interest to note such effects. Figure 6 displays the identification rate as a function of ground harmonicity for each of the 20 vowel pairs, for both conditions of ΔF_0 and both conditions of target harmonicity. Vowel pairs differ considerably in overall identification rate, as well as in the size and direction of the effects of ground harmonicity. These

differences may reflect an actual dependency of these effects on vowel pair, or some effect of the level correction factors that we applied, or possibly differences between the component patterns used to synthesize each vowel pair (each allophone had its own inharmonic pattern when it was synthesized in an inharmonic state, so each vowel was represented by a different set of patterns). Our experimental design does not allow us to decide which of these factors are responsible for the differences. It is, however, of interest to keep them in mind when interpreting our main effects. For example, it may be that the population of "inharmonic" patterns that we treat as homogeneous is actually made up of members with widely differing properties.

IV. DISCUSSION

A. Effect of ΔF_0 in comparison with previous studies

Most previous studies report the proportion of responses for which both vowels in a pair were correctly identified (combinations-correct rates). To allow comparisons to be made, similar scores were calculated from our data for the *H/H* conditions and plotted in Fig. 1 together with data from those studies. The effect of ΔF_0 is quite similar. Although our task was relatively easy [chance level is 10%, as in Culling and Darwin (1993) and Lea (1992), compared to 3.8% for Scheffers (1983), or 6.7% for Summerfield and Assmann (1991)], our rates are relatively low. This probably reflects the greater variability of our stimulus material, and differences in training (we used a large number of relatively untrained subjects).

B. Evidence for cancellation

At a ΔF_0 of 1/2 semitone, whatever the target, and at $\Delta F_0 = 0$ when the target is inharmonic, identification is better when the ground is harmonic. This is consistent with the cancellation hypothesis. No advantage was to be expected for a harmonic ground in the *H/H0* condition, but identification was actually *worse* when the ground was harmonic than when it was inharmonic [$R(H/H0) < R(I/H0)$], an outcome that the cancellation hypothesis does not predict. One possible explanation is that our inharmonic stimuli were approximately harmonic with a "pseudoperiod" that differed from their nominal period (on informal listening they often appeared to have a pitch different from that of a harmonic vowel of same F_0). A harmonic sieve tuned to reject the peudoperiod might partially remove the inharmonic ground without completely removing the target, whereas that target would be eliminated if both vowels were harmonic and had the same F_0. Another possible explanation is that other mechanisms are at work together with cancellation.

Lea (1992) also found evidence for cancellation: when the target was a 112-Hz voiced vowel, identification rates were better by 3% for a 100-Hz voiced ground than for a whispered ground. When the target was a whispered vowel, the advantage was 8%. Subsequent experiments (Lea and Tsuzaki, 1993a,b) gave similar results. The smaller size of the effects we found (≈3%) may be due to the fact that our inharmonic vowels were more "harmonic" than the whispered vowels used by Lea.

C. Evidence for enhancement

Our results do not support enhancement. In fact, identification rates are *worse* when the target is harmonic, whereas in the absence of enhancement we predicted a null effect. This result is unexpected. It is worth considering in more detail at this point the assumptions upon which we based our predictions. We assumed that both vowels could be retrieved simultaneously via independent processing channels involving enhancement and/or cancellation, and thus that both hypotheses could be tested independently. If instead the auditory system must *choose* between strategies, factors that favor one may penalize the other. If, for example, cancellation is used systematically, it may tend to "lock" onto whatever happens to be harmonic within the stimulus, and thus impair the identification of harmonic targets. Inharmonic targets would be relatively immune. Thus the unexpected outcome of our experiment may be due to the mutual interference between segregation mechanisms. If so, we cannot rule out the eventuality that enhancement *is* used, but its effects are swamped by the side effects of cancellation. Enhancement would eventually show up in tasks in which cancellation is less likely to come into play. Our results contrast with those of Lea (1992), who found no significant difference between whispered and voiced targets, and Lea and Summerfield (1992), who found an advantage for targets that were voiced rather than whispered. Summerfield and Culling (1992b) also found no effect of target harmonicity on masking level thresholds.

An explanation for the apparent preference of the auditory system for cancellation over enhancement may be found in an experiment by McKeown (1992). He requested subjects to identify both vowels within a pair, and at the same time judge which vowel was "dominant," and which was "dominated." Improvements in identification with ΔF_0 only occurred for the dominated vowel. If we suppose that it is easier to estimate the F_0 of a dominant vowel than that of a dominated vowel, it should follow that cancellation is likely to have segregated the dominated vowel (de Cheveigné, 1993a). It is then reasonable that factors upon which cancellation depends should affect the scores. Another explanation may be found in an experiment of de Cheveigné (1993b) and de Cheveigné *et al.* (1994). Harmonic enhancement and cancellation were implemented in a speech recognition system to reduce the effects of cochannel speech interference. Cancellation was more effective, presumably because it was less affected than enhancement by the nonstationary nature of speech (as explained in the Introduction, effective enhancement requires a filter with a relatively long impulse response). The synthetic vowels used in our experiments were stationary, so this consideration should not apply here. However, the auditory system may have evolved to use only strategies that are robust for natural stimuli.

D. Compatibility with F_0-guided models of concurrent vowel perception

A variety of models make use of explicit F_0 information. Some clearly take sides for either enhancement (Frazier *et al.*, 1976) or cancellation (Childers and Lee, 1987; Hanson

and Wong, 1984; Naylor and Boll, 1987), but most other models are capable of both. Models come in three sorts: spectral, spectro-temporal, and temporal.

The harmonic sieve employed by spectral models based on Parson's harmonic selection method (Assmann and Summerfield, 1990; Denbigh and Zhao, 1992; Parsons, 1976; Scheffers, 1983; Stubbs and Summerfield, 1988, 1990, 1991) can be used in either of two modes: to retain components that fall close to a harmonic series, or else to remove them. These modes correspond to enhancement and cancellation, respectively. However, the sieve may be applied in turn to each harmonic series to select correlates of one voice among those rejected from the other. In that case each voice retrieved is actually a product of *both* strategies. Similar remarks can be made concerning models derived from Weintraub's spectro-temporal model (Assmann and Summerfield, 1990; Lea, 1992; Meddis and Hewitt, 1992; Weintraub, 1985): channels dominated by the period of a voice can be retained (enhancement) or else removed (cancellation). If both operations are applied in turn, each voice retrieved is really the product of both strategies. In the model of Meddis and Hewitt (1992), only one F_0 was used, so one voice (the dominant one) was purely the product of enhancement, whereas the other voice was purely the product of cancellation. However, this model is easily modified to use both strategies to segregate both voices. Finally, de Cheveigné (1993a) proposed a time-domain comb-filtering model implemented by neutral circuits involving inhibition. That model was also capable of either enhancement or cancellation.

Since most models allow both strategies, our results do not allow us to choose among them, but they do allow us to better understand how each model functions.

E. Compatibility with other models of concurrent vowel perception

A number of models that do not require explicit extraction of F_0 have been proposed to explain improvement of identification with ΔF_0. Summerfield and Assmann (1991) suggested that such an improvement might be explained by misalignment between partials of constituent vowels. At unison the partials of both vowels coincide, and their relative contributions to the combined spectrum are obscured by phase-dependent vector summation. Misaligned partials, on the other hand, may show up as independent peaks within a high-resolution spectrum and thus template-matching strategies might be more successful. Summerfield and Assmann (1991) found some evidence for an effect of component misalignment for vowels with widely spaced components (200 Hz), but none for monaurally presented vowels at 100 Hz. On the other hand, in a masking experiment in which thresholds were determined for synthetic vowels masked by vowel-like maskers, Summerfield (1992) attributed up to 9 dB of a 17-dB release from masking to component misalignment. The remaining 8 dB were attributed to F_0-guided mechanisms. Our results certainly cannot be explained solely in terms of component misalignment. *H/I* and *I/H* conditions involve the same intercomponent intervals, yet they produce identification rates that are very different. However, if harmonic misalignment were involved together with a mechanism such as cancellation, it might help explain, for example, why the *H/H*0 condition was significantly worse than the *H/I*0 condition. Our experiments used *I/I* pairs in which the inharmonic patterns of the vowels were different, and thus partials did not coincide at $\Delta F_0 = 0$. It would be worth investigating a similar condition in which both vowels have the *same* inharmonic pattern. Comparisons between the two would allow us to factor out possible effects of component misalignment.

If a vowel's period is long relative to time constants of integration within the auditory system, the vowel's auditory representation may fluctuate during the period. Mutual interference between concurrent vowels may be more or less severe according to whether the fluctuations of their respective representations line up in time. A small F_0 difference is equivalent to a gradually increasing delay of one vowel relative to the other, and this might allow the auditory system to select some favorable interval on which to base identification. Differences in F_0 might thus enhance identification. Summerfield and Assmann (1991) investigated the effects of pitch period asynchrony on identification rate using vowels with same F_0 but varying degrees of phase shift. They found a significant effect at 50 Hz, but none at 100 Hz, presumably because the integrating properties of the auditory representation smooth out fluctuations at this rate. Our vowels had even higher F_0's, so this explanation is unlikely to account for our data.

Slower fluctuations may occur in the *compound* representation of the vowel pair. Two partials falling within the same peripheral channel produce beats with a depth that depends on their relative amplitudes, and a rate equal to their difference frequency. Three or more partials produce yet more complex interactions. These fluctuations may cause the auditory representation to take on a shape that momentarily allows one vowel or the other, or both together, to be better identified. Culling and Darwin (1993, 1994) suggested that such beats might explain increases of identification rate with differences in F_0. Assmann and Summerfield (1994) found that successive 50-ms segments excised from a 200-ms stimulus composed of two vowels with different F_0's were not equally identifiable. For small ΔF_0's, identification of the whole stimulus could be accounted for assuming it was based on the "best" of the segments that composed it. This result is compatible with the notion that F_0 differences cause the auditory representation to fluctuate (as indeed the short-term spectrum itself fluctuates), and provide the auditory system with various intervals upon which to base identification, one of which may be particularly favorable to either vowel or both.

Inharmonicity or F_0 differences between vowels can be interpreted as slowly varying phase relationships between partials of harmonic vowels with the same F_0. The "best interval" provided by beating can be interpreted simply as a phase relationship that is particularly favorable for identification. The harmonic vowels used in our experiments were all synthesized in sine phase, whereas the partials of inharmonic vowels can be interpreted as progressively moving out of this phase relationship. *If* the masking power of vowels in sine phase were relatively small, and the resistance to mask-

ing of vowels in sine phase relatively poor, then harmonic vowels would appear to be both less well recognized and less effective as maskers, as indeed we found. Such phase effects, if they exist, thus constitute a possible alternative explanation of our results.

F. Harmonicity and the cohesion of sound

The lack of a positive effect of harmonicity on target vowel identification is the most surprising result of this study. It has been suggested that harmonicity labels parts of a sound as belonging together in several ways: continuity of F_0 indicates that successive parts of speech belong to the same voice; the same F_0 indicates that different formants belong to the same vowel; a common F_0 signals that partials within a formant belong together (Bregman, 1990; Broadbent and Ladefoged, 1957; Cutting, 1976; Darwin, 1981). Without this "harmonic glue" components would fall apart, and the sound might lose its intelligibility or be more easily masked. Nevertheless, Darwin (1981) found in several cases that speech sounds synthesized with different formants on different F_0's retained their phonetic quality. Culling and Darwin (1993) synthesized vowels with a difference in F_0 between their first and higher formants, and paired them so that the components making up the first formant of one vowel belonged to the same harmonic series as the higher formants of the other. In other words, the F_0's were swapped between vowels at the transition between the F_1 and higher formant regions. Identification was as good as for vowels with unswapped F_0's for all but the largest ΔF_0's, from which Culling and Darwin concluded that small differences in F_0 between *formants* do not affect how they are grouped together. Our results go a step further. They suggest that small differences in F_0 between *partials* have no negative effect (and apparently even a positive effect) on the identification of the sound that they form. This result is counterintuitive, and is contradicted by some other studies. For example, Darwin and Gardner (1986) found that mistuning a single partial within a formant affected the phonetic quality of a vowel. However, at small ΔF_0's the effect of mistuning (which was phase dependent) did not always go in the direction expected on the basis of harmonic grouping.

A common F_0 does have one important effect: it produces the impression of a single source. The presence of multiple F_0's within a sound, what Marin (1991) calls "polyperiodicity," produces the impression of multiple sources, and thus tells the auditory system that segregation is called for. In the absence of such a cue, the auditory system may fail to "notice" that there are several sounds, and no segregation will occur. If so, the cue is important for segregation in everyday situations. In our experiment the task was such that subjects were forced to consider each stimulus as containing exactly two sounds. The presence or absence of "multiple sound" cues related to harmonicity or delta F_0 would have made no difference to the listener's responses.

V. SUMMARY AND CONCLUSION

(1) Vowels within pairs synthesized in sine phase were identified better by about 3% when they were inharmonic than when they were harmonic, except when the ground was harmonic and $\Delta F_0 = 0$, in which case the advantage was 8%. These results are contrary to what one would expect if a strategy of harmonic enhancement was used to segregate the vowels.

(2) Vowels within pairs synthesized in sine phase were identified better by about 3% when the vowels accompanying them were harmonic than when they were inharmonic, except when the target vowel was also harmonic and $\Delta F_0 = 0$, in which case they were *less* well identified by about 3%. These results are consistent with the hypothesis of harmonic cancellation.

(3) When both vowels within a pair were harmonic, they were better identified by about 6% when there was a difference in F_0 of 1/2 semitone. This result is similar to those of previous studies. When either vowel was inharmonic, a difference in F_0 did not affect identification.

(4) When one vowel within a pair was harmonic and the other inharmonic, the inharmonic component was identified significantly better than the harmonic component. Effects did not follow the symmetric pattern that is sometimes assumed to be characteristic of primitive segregation.

(5) Our experiments employed a particular starting phase pattern (sine) to synthesize all vowels. In the light of recent results that demonstrate the role of beats in the identification of concurrent vowels (Assmann and Summerfield, 1994; Culling and Darwin, 1994), we cannot rule out the possibility that our results are partly specific to this phase pattern.

Fundamental frequency had two putative roles for Darwin (1981): to "group consecutive sounds together into the continuing speech of a single talker" and to "group together the harmonics from different formants of one talker, to the exclusion of harmonics from other sources" (p. 186). Our results suggest a third role: to group together components that belong to an interfering source to better eliminate it. The lack of benefit of target harmonicity for identification is surprising, as it can in principle be exploited by a majority of harmonic sound separation models. The question merits further examination, perhaps using tasks in which the effects are less likely to be dominated by cancellation.

ACKNOWLEDGMENTS

Thanks to Gérard Bertrand for technical assistance, to Laurent Ghys for guiding some of us through the mysteries of the NeXT Machine, and to Nina Fales for assistance during the experiments. Thanks to John Culling and Quentin Summerfield for providing data on which Fig. 1 was based, and to Andrew Lea for useful discussions. Thanks to three reviewers, Chris Darwin, John Culling, and Brian Moore, for detailed comments on prior drafts. This research was supported by a grant from the "Cognitive Sciences" program of the French Ministry of Research and Space.

APPENDIX A: PREPARATION OF SPECTRAL ENVELOPES

We wished to use stimuli with high intraclass variability in order to make the identification task more difficult and

more typical of real speech communication. We reasoned that the best place to look for such variability is in natural, continuous speech. We systematically extracted voiced (quasiperiodic) tokens from a multispeaker speech database to obtain samples of a wide range of spectra. We then screened them in several stages to obtain a set of spectral envelopes that were consistently identifiable as given vowels after resynthesis. The thresholds of acceptance in these screening tests were chosen to strike an (arbitrary) balance between the goals of variability and consistent identifiability.

The database consisted initially of 50 phonetically balanced French sentences pronounced by 11 adult speakers (5 male, 6 female), belonging to the CD6_GRECO1 disk of the GRECO1 database (GRECO, 1987). To this initial database we later added 16 sentences containing mainly /u/ vowels and a set of CVCV (V=/u/) words from the same database. Data were sampled at 16 kHz with 16-bit resolution. The database was processed by an F_0 estimation algorithm based on the average magnitude difference function algorithm (described in Appendix B-2 of de Cheveigné, 1993a), that produces as a by-product a measure of periodicity. The F_0 and periodicity measure were used to label portions of voiced speech as follows: wherever the periodicity measure was above an arbitrary threshold (2.0) for more than 50 ms, and the F_0 was within the range 111–141 Hz, an index was set every 50 ms. A total of 1788 indices were thus set, of which 572 were retained after a first informal listening test. For each index, a 512-point 0- to 8-kHz spectral envelope was calculated. The envelopes served to synthesize periodic synthetic vowels that were further screened to obtain 75 /a/, 49 /e/, 35 /i/, 13 /o/, and 13 /u/ allophones. A clustering algorithm was used to choose from each set ten allophones with spectral envelopes as different from each other as possible.

A pilot version of our experiments, conducted with 20 subjects, served as a final screening test. Analysis of the results revealed an abnormally high error rate for four /u/ allophones that tended to be systematically identified as /o/, even by subjects that had consistently classified them as /u/ in previous screening tests (a result that no doubt illustrates effects of stimulus set on vowel identification). We eliminated these allophones, duplicated four of the remaining allophones, renamed them, and proceeded as if /u/ had the same number (ten) of allophones as the other phonemes.

We repeatedly met difficulties with /u/. For some reason, very few portions of speech isolated from our database sounded like /u/ after resynthesis, even those taken from sentences labeled as containing mainly /u/ phonemes. A tentative explanation is that in French /u/ is articulated with a protrusion of the lips. The target position may require some time to be attained, and the resulting spectral transition may in fact be necessary for identification. Evidently no such transition is present in the resynthesized vowel. This does not explain, however, why a few tokens *do* sound reasonably /u/-like after synthesis. Overall, surprisingly few of the original voiced speech tokens were identified consistently as vowels after resynthesis: less than 10% of the original tokens survived the final screening. In real speech, vowel identity is probably largely determined by contextual or dynamic features that are absent from the resynthesized vowels (Hillenbrand and Gayvert, 1993).

APPENDIX B: SYNTHESIS OF INHARMONIC COMPONENT PATTERNS

We wished to obtain vowels that were inharmonic, but with a spectral density close to that of a harmonic vowel. The frequency of each component of a harmonic series was shifted by a random amount drawn from a uniform distribution bounded by ±3% of the harmonic frequency, or half the spacing between adjacent harmonics, whichever was smaller. We synthesized twice the required number of component patterns (50), then screened out the "least inharmonic" half by choosing those with the greatest values of the following measure of inharmonicity:

$$\sum_{\mathbf{n}=1}^{44} [\mathbf{n}/(\mathbf{n}+1)-\mathbf{f}_n/\mathbf{f}_{n+1}]^2,$$

where f_n is the frequency of the nth component.

APPENDIX C: LEVEL CORRECTION FACTORS

When vowels are mixed at equal rms signal levels, one vowel may dominate the pair due to unequal mutual interference. We wished to avoid this situation. Informal listening showed that an equal rms level results in approximately equal loudness; we concluded that matching for equal loudness was unlikely to fulfill our goal. Instead, we decided to experimentally determine a correction factor to balance mutual interference.

We first informally determined, for each of the ten vowel pairs, the rms level differences for which either vowel appeared to be absent. We then centered a scale with 4-dB steps and ten levels on the mean of these two differences, and synthesized pairs of unison harmonic vowels according to this scale. There were ten such scales, one for each vowel pair. The stimuli were presented five times each in random order to four subjects (the four authors). At each presentation the stimulus was repeated twice; after each repetition the subject had to identify one constituent. A response could be any of the five vowels, or "x" if no vowel could be heard, but the two responses had to be different. Psychometric functions were plotted for each component of a pair, and their intercept was taken as the correction factor. The correction factors for all pairs are shown in Table CI.

These results are roughly compatible with those reported by McKeown (1992) for three of his four subjects: /a/ tends to dominate all other phonemes while /u/ tends to be domi-

TABLE CI. Level correction factors for vowel pairs, in dB. The level after correction of one vowel relative to another is shown at the intersection of the row and column that they label, respectively. For example, to synthesize /æ/ the rms level of /a/ should be set to be 5.0 dB less than that for /e/.

	/e/	/i/	/o/	/u/
/a/	−5.0	−7.5	−17.5	−31.0
/e/		1.0	−11.5	−17.0
/i/			−2.0	−16.0
/o/				−16.5

nated by all others. Other phonemes are intermediate: /o/, /i/, /e/ in order of increasing dominance. However, our factors were determined before the final screening that eliminated four allophones of /u/. Levels are therefore certainly biased too far in favor of /u/ to compensate for the poor quality of those allophones. This is evident in the identification rates as a function of ground vowel which were particularly low when /u/ was ground (Sec. III D 6), but it should not have affected our main conclusions concerning the effects of harmonicity or ΔF_0: they remain quite similar when pairs containing /u/ are removed from the analysis. We do not recommend that these particular level correction factors be used in other studies.

[1]In all reports of F statistics in this article the probabilities reflect, where necessary, an adjustment of the degrees of freedom by the Greenhouse–Geisser factor to correct for the inherent correlation of repeated measurements (Geisser and Greenhouse, 1958). *GG* indicates the epsilon factor by which degrees of freedom were multiplied to determine the probability level. This is a conservative correction factor.

[2]The data were also reanalyzed after transformation by an arcsine function to obtain distributions that better satisfy the assumptions of ANOVA, with similar results.

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).

Broadbent, D. E., and Ladefoged, P. (**1957**). "On the fusion of sounds reaching different sense organs," J. Acoust. Soc. Am. **29**, 708–710.

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. (**1991**). "Discriminating between coherent and incoherent frequency modulation of complex tones," J. Acoust. Soc. Am. **89**, 329–340.

Chalikia, M. H., and Bregman, A. S. (**1989**). "The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation," Percept. Psychophys. **46**, 487–496.

Chalikia, M. H., and Bregman, A. S. (**1993**). "The perceptual segregation of simultaneous vowels with harmonic-shifted or random components," Percept. Psychophys. **53**, 125–133.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech with one, and with two ears," J. Acoust. Soc. Am. **25**, 975–979.

Childers, D. G., and Lee, C. K. (**1987**). "Co-channel speech separation," Proc. IEEE ICASSP, 181–184.

Culling, J. (**1990**). "Exploring the conditions for the perceptual segregation of concurrent voices using F_0 differences," Proc. Inst. Acoust. **12**, 559–566.

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.

Cutting, J. E. (**1976**). "Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening," Psychol. Rev. **83**, 114–140.

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

Darwin, C. J., and Culling, J. F. (**1990**). "Speech perception seen through the ear," Speech Commun. **9**, 469–475.

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. (**1993a**). "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. (**1993b**). "Time-domain comb filtering for speech separation," ATR Human Information Processing Laboratories, Tech. Rep. TR-H-016.

de Cheveigné, A., Kawahara, H., Aikawa, K., and Lea, A. (**1994**). "Speech separation for speech recognition," J. Phys. Paris **IV**, C5-545–548.

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

Denbigh, P. N., and Zhao, J. (**1992**). "Pitch extraction and separation of overlapping speech," Speech Commun. **11**, 119–125.

Frazier, R. H., Samsam, S., Braida, L. D., and Oppenheim, A. V. (**1976**). "Enhancement of speech by adaptive filtering," Proc. IEEE ICASSP, 251–253.

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

GRECO. (**1987**). "BDSONS, base de donnees des sons du francais, GRECO 1," edited by Jean-Francois Serignat and Ofelia Cervantes, ICP, Grenoble, France.

Hanson, B. A., and Wong, D. Y. (**1984**). "The harmonic magnitude suppression (HMS) technique for intelligibility enhancement in the presence of interfering noise," Proc. IEEE ICASSP **2**, 18A.5.1–4.

Hillenbrand, J., and Gayvert, R. T. (**1993**). "Identification of steady-state vowels synthesized from the Peterson and Barney measurements," J. Acoust. Soc. Am. **94**, 668–674.

Lea, A. (**1992**). "Auditory models of vowel perception," Doctoral dissertation, University of Nottingham, UK.

Lea, A., and Tsuzaki, M. (**1993a**). "Segregation of competing voices: Perceptual experiments," Proc. Acoust. Soc. Jpn., Spring session, 361–362.

Lea, A. P., and Tsuzaki, M. (**1993b**). "Segregation of voiced and whispered concurrent vowels in English and Japanese," J. Acoust. Soc. Am. **93**, 2403 (A).

Lea, A. P., and Summerfield, Q (**1992**). "Monaural segregation of competing voices," Proc. Acoust. Soc. Japan Committee on Hearing H-92-31, 1–7.

Marin, C. (**1991**). "Processus de séparation perceptive des sources sonores simultanées," Doctoral dissertation, Université de Paris III, France.

Marin, C., and McAdams, S. (**1991**). "Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width," J. Acoust. Soc. Am. **89**, 341–351.

McAdams, S. (**1989**). "Segregation of concurrent sounds. I: Effects of frequency modulation coherence," J. Acoust. Soc. Am. **86**, 2148–2159.

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.

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

Naylor, J. A., and Boll, S. F. (**1987**). "Techniques for suppression of an interfering talker in co-channel speech," Proc. IEEE ICASSP, 205–208.

Parsons, T. W. (**1976**). "Separation of speech from interfering speech by means of harmonic selection," J. Acoust. Soc. Am. **60**, 911–918.

Scheffers, M. T. M. (**1983**). "Sifting vowels," Doctoral dissertation, University of Gröningen, The Netherlands.

Stubbs, R. J., and Summerfield, Q. (**1988**). "Evaluation of two voice-separation algorithms using normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **84**, 1236–1249.

Stubbs, R. J., and Summerfield, Q. (**1990**). "Algorithms for separating the speech of interfering talkers: Evaluations with voiced sentences, and normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **87**, 359–372.

Stubbs, R. J., and Summerfield, Q. (**1991**). "Effects of signal-to-noise ratio, signal periodicity, and degree of hearing impairment on the performance of voice-separation algorithms," J. Acoust. Soc. Am. **89**, 1383–1393.

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 deGruyter, 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. (**1992a**). "Auditory segregation of competing voices: Absence of effects of FM or AM coherence," Philos. Trans. R. Soc. London, Ser. B **336**, 357–366.

Summerfield, Q., and Culling, J. F. (**1992b**). "Periodicity of maskers not targets determines ease of perceptual segregation using differences in fundamental frequency," 124th meeting of the ASA [J. Acoust. Soc. Am. **92**, 2317 (A)].

Weintraub, M. (**1985**). "A theory and computational model of auditory monaural sound separation," Doctoral dissertation, Stanford University, Stanford, CA.

Zwicker, U. T. (**1984**). "Auditory recognition of diotic and dichotic vowel pairs," Speech Commun. **3**, 256–277.