Formant-frequency matching between sounds with different bandwidths and on different fundamental frequencies

Pascal Dissard and C. J. Darwin[a)]
Experimental Psychology, University of Sussex, Brighton BN1 9QG, United Kingdom

(Received 28 April 2000; accepted for publication 10 April 2001)

The two experiments described here use a formant-matching task to investigate what abstract representations of sound are available to listeners. The first experiment examines how veridically and reliably listeners can adjust the formant frequency of a single-formant sound to match the timbre of a target single-formant sound that has a different bandwidth and either the same or a different fundamental frequency ($F0$). Comparison with previous results [Dissard and Darwin, J. Acoust. Soc. Am. **106**, 960–969 (2000)] shows that (i) for sounds on the same $F0$, introducing a difference in bandwidth increases the variability of matches regardless of whether the harmonics close to the formant are resolved or unresolved; (ii) for sounds on different $F0$'s, introducing a difference in bandwidth only increases variability for sounds that have unresolved harmonics close to the formant. The second experiment shows that match variability for sounds differing in $F0$, but with the same bandwidth and with resolved harmonics near the formant peak, is not influenced by the harmonic spacing or by the alignment of harmonics with the formant peak. Overall, these results indicate that match variability increases when the match cannot be made on the basis of the excitation pattern, but match variability does not appear to depend on whether ideal matching performance requires simply interpolation of a spectral envelope or also the extraction of the envelope's peak frequency. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1379085]

PACS numbers: 43.66.Jh, 43.71.Es [RVS]

I. INTRODUCTION

In the experiments reported here, listeners match the timbre of two single-formant sounds by adjusting the formant frequency of one of the sounds. We ask whether, by varying the other dimensions along which the sounds differ, we can force listeners to make judgments based on different levels of abstraction of sound.

In an earlier paper (Expt. 2, Dissard and Darwin, 2000) we examined how reliably (as measured by the standard deviation of an individual listener's matches) listeners could match the timbre of a single-formant target sound by adjusting the formant frequency of a second (''match'') sound. We showed (unsurprisingly) that when the target and the adjustable sound had the same fundamental frequency ($F0$), matches were more reliable than when they differed in fundamental frequency. However, we also showed that this difference was substantially larger for sounds on high $F0$'s than it was for sounds on low $F0$'s.

When the $F0$ of both target and match sounds was the same, matches were both veridical (in that the match formant frequency was close to the target formant frequency) and reliable (across trials for a particular listener). Listeners here are performing a match that is based on making the two sounds identical, and so the match could be made on the basis of making either the total neural activity from the two sounds identical, or some subset of the activity such as that corresponding to the excitation pattern (Moore and Glasberg, 1983).

When the $F0$'s of the target and match sounds were different, however, the strategies open to listeners differed depending on whether the harmonics in the formant region were resolved by the auditory system or unresolved. With unresolved harmonics in the formant region, listeners can make a reliable and veridical match on the basis of identity of the auditory excitation pattern. With resolved harmonics, however, harmonic ripple in the excitation pattern disrupts formant matches, encouraging listeners to make the match on the basis of a more abstract property such as an interpolated envelope of the excitation pattern, which would smooth out harmonic ripple (for excitation patterns, see Fig. 1). We found that giving target and match sounds a different $F0$ changed the reliability of matches in a way that reflected these two different strategies. For sounds with unresolved harmonics in the formant region, matches were slightly less reliable than when $F0$ was the same on both sounds, but for sounds with resolved harmonics in the formant region, matches were much less reliable than when the $F0$ was the same on both sounds.

We argued against the possibility that this latter result was due to listeners trying to match excitation patterns from sounds with different $F0$'s. We modeled performance that minimized the rms error between excitation patterns and showed that this strategy predicted a much larger increase in variability than we actually found. We concluded that the increased variability of matches on different $F0$'s for sounds with resolved harmonics in the region of their single formant reflected the perceptual cost of interpolating an envelope across harmonic peaks in the excitation pattern. This claim—that matches made at more abstract levels are more variable—forms the basis for the present paper.

In the present paper we present new experiments in

[a)]Author to whom correspondence should be addressed.

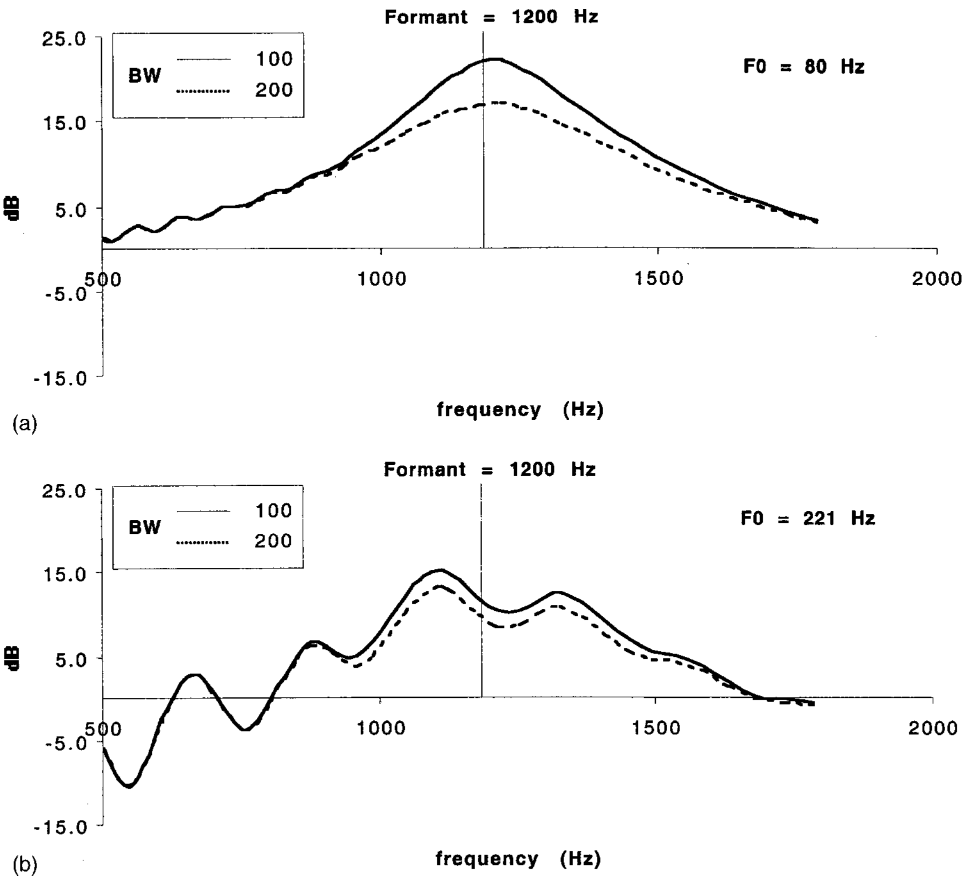

FIG. 1. Excitation patterns of single-formant (1200 Hz) sounds with different formant bandwidths. The upper panel shows sounds with different bandwidths on a fundamental frequency of 80 Hz. The lower panel shows similar sounds with a fundamental frequency of 212 Hz.

which we also manipulate formant bandwidth (BW). The point of this additional manipulation is that it could force listeners to use a more abstract representation than the interpolated spectral envelope—such as the formant frequency—even in the absence of a difference in $F0$.[1]

The rationale behind the manipulations made in this and our previous experiments can be understood by reference to Table I. Here, we show the stimulus conditions used in the present experiment 1 and in our previous paper (Expt. 2, Dissard and Darwin, 2000). The left-hand column provides acronyms for the particular stimulus conditions that we will refer to later (the letters refer to the contents of the next three columns, that is whether the target and match sounds are the same or different in harmonic resolution, $F0$ and BW, respectively). The second column indicates whether the harmonics near to the formant peak is resolved or unresolved by the normal ear. This property is illustrated in Fig. 1, which shows excitation patterns (Moore and Glasberg, 1983) for two single-formant sounds differing in bandwidth, with both sounds on either a low (80 Hz) or a high (212 Hz) $F0$. For sounds on a high $F0$, the excitation pattern resolves individual harmonic components, whereas for sounds on a low $F0$ this ripple is absent, indicating that the individual harmonics are not resolved. The third and fourth columns indicate whether the two sounds in a trial (the target and the match sound) share the same or have different $F0$'s and bandwidths. Finally, the fifth column describes the least abstract property that listeners could use to ensure a reliable and veridical (i.e., same formant frequency) match. The table aims to provide a hierarchy of abstraction for matches, without necessarily committing the reader to specific abstractions.[2]

When target and match sounds have the same $F0$ and bandwidth (USS, RSS), the formant-matching task can be

TABLE I. Conditions used in experiment 1 and in a previous experiment.

	Harmonics near formant	$F0$	BW	Match property	Old Expt. 2	Expt. 1
USS	Unresolved	Same	Same	Identity	x	
RSS	Resolved	Same	Same	Identity	x	
UDS	Unresolved	Different	Same	Excitation	x	
RDS	Resolved	Different	Same	Envelope	x	
USD	Unresolved	Same	Different	Excitation peak		x
RSD	Resolved	Same	Different	Envelope peak		x
UDD	Unresolved	Different	Different	Excitation peak		x
RDD	Resolved	Different	Different	Envelope peak		x

performed veridically by listeners making the sounds identical—the least abstract level of matching.

Introducing a difference in $F0$ requires a more abstract level of matching which differs depending on the resolution of the harmonics near to the formant peak. With unresolved harmonics (UDS), listeners can achieve a veridical match by making the excitation patterns identical (which will not show any harmonic ripple) in the vicinity of the formant peak. With resolved harmonics (RDS), this strategy will also fail in general, so that listeners will need to perform a more abstract match that generalizes across different frequencies of harmonic ripple; one such abstraction is the envelope of the excitation pattern.

If we now consider sounds that differ in bandwidth, the level of abstraction of all matches is increased to similarly abstract levels whether the sounds have the same or different $F0$'s. For sounds with unresolved harmonics near the formant frequency (USD, UDD), the peak in the excitation pattern provides the least abstract criterion for a veridical match. For resolved harmonics (RSD, RDD) the match must be made at a more abstract level, such as by first interpolating an envelope for the excitation pattern and then determining the peak of this envelope.

Although a task may theoretically require a particular level of abstraction in order to perform a veridical match, listeners may adopt a suboptimal strategy that uses a less abstract representation. In the previous paper we tested whether listeners were using a strategy of minimizing the rms error between excitation patterns as a matching strategy for sounds on different $F0$'s. A simulation of this strategy predicted a much higher variability of subjects' performance in matching sounds on different, high fundamentals than was actually found in the experiment. We will make similar arguments in this paper that the strategy of minimizing excitation-pattern rms error also does not explain subjects' performance in the present experiments.

While a difference in bandwidth can in principle substantially alter the criteria that listeners use in a matching experiment, differences in formant bandwidth have generally had rather little effect on listeners' phonetic judgments of vowel-like speech sounds. Carlson *et al.* (1979) found that manipulation of formant bandwidth had less of an influence on listeners' dissimilarity ratings than did changes to formant frequency. In addition, using similar sounds, Klatt (1979) showed that when making phonetic comparisons listeners pay far less attention to changes in formant bandwidths than they do when making psychophysical judgments. The exception to this pattern is that a reduction in the prominence of the first formant peak has been proposed as the primary perceptual correlate of vowel nasality (Hawkins and Stevens, 1985). However, in the experiments described here single-formant sounds are used where the formant frequency is well above the first-formant range, so a percept of changing nasality is unlikely to influence listeners' judgments.

The first experiment asks how reliably listeners can match single-formant sounds that differ in bandwidth (USD, RSD) or in both $F0$ and bandwidth (UDD, RDD). In particular, we are interested to discover whether the reliability of listeners' matches mirrors the hierarchy of levels of abstraction that we introduced in Table I, thereby producing perceptual evidence for different levels of abstraction for the processing of complex sounds.

II. EXPERIMENT 1

A. Stimuli and procedure

The general procedure for the experiment was similar to that used in our previous paper. Listeners had to adjust, by moving a roller-ball up or down, the formant frequency of a 500-ms periodically excited single-formant complex sound to match the timbre of a similar sound with a formant frequency of either 1100 or 1200 Hz. The pair of sounds could be repeated as often as necessary on each trial by pressing the roller-ball's button. The two sounds could have the same or different $F0$'s. Three factors were varied orthogonally across 8 blocks of 20 trials: whether both sounds in a trial were from the low (80 and 90.4 Hz) or the high (221.2 and 250 Hz) $F0$ range, whether the target sound had a narrow (100 Hz) or a wide (200 Hz) bandwidth, and whether target and match sounds had the same or different $F0$'s. These $F0$ manipulations, coupled with the difference in bandwidth, allow us to separate direct effects on matching reliability of a difference in $F0$ (such as distracting listeners from making the match) from the level of abstraction needed to make the match. In the previous paper (Dissard and Darwin, 2000; experiments 2, 3, and 4) matches which were nonidentical always had a difference in $F0$. For each pair of sounds, the direction of pitch was always upward (i.e., the target's $F0$ was lower than the adjustable sound's $F0$), with a frequency ratio of 1:1.13. Within a block the target formant frequency was randomly set to either 1100 or 1200 Hz. Each target sound was matched ten times in a quasirandom order. Each block took about 30 min to complete and the order of experimental blocks was randomized across subjects.

Sounds were synthesized in real time at 22.05 kHz using the parallel branch of SENSYN PPC™ (Sensimetrics, Cambridge, MA) incorporated into custom software. Voice source parameters were set to their default values, which are the same as described in Klatt (1979). Sounds were output through a Digidesign Protools board and presented through Sennheiser HD414 headphones in an IAC booth. An Apple Power Macintosh 7100 computer controlled the experiment; overall output level was around 60 dB SPL. At the beginning of each trial the formant frequency of the adjustable sound was chosen at random from the permitted range (850 to 1500 Hz). As subjects moved the roller-ball to adjust the formant frequency of the comparison sound a screen cursor also moved. The cursor was recentered after each button press so that subjects could not base their adjustment on the cursor's position. Moving the cursor by half a screen led to a maximum change of about 33 Hz (fine adjustment, minimum change: 0.1 Hz) or 100 Hz (coarse adjustment, minimum change: 0.3 Hz). Subjects could toggle between the coarse and fine adjustments. If the formant frequency was adjusted outside the permitted range, it was reset to a random value within the range and a warning sound played.

Nine subjects (including the first author) participated in the experiment. Subjects were university students or staff

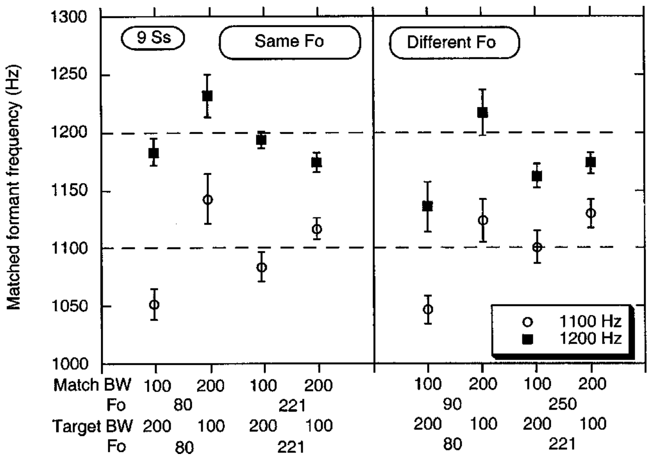

FIG. 2. Mean matched formant frequencies in experiment 1 for target formant frequencies of 1100 and 1200 Hz. Error bars are standard errors of the mean over nine subjects. The left panel shows matches made with stimuli on different $F0$; the right panel shows matches with stimuli on the same $F0$.

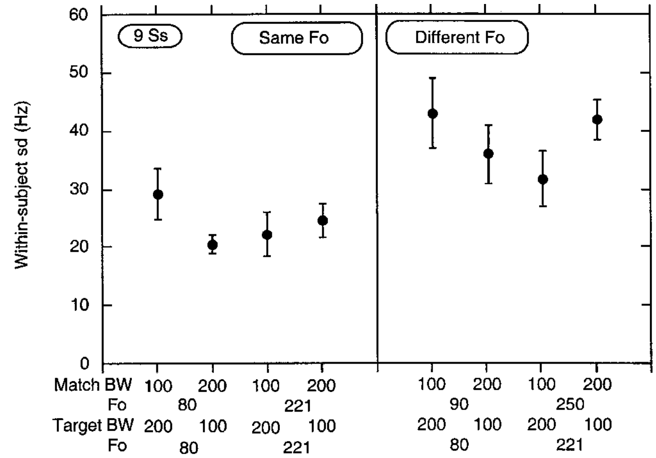

FIG. 3. Average within-subject standard deviations of matches in experiment 1 with their standard errors over nine subjects. The left panel shows matches made with stimuli on different $F0$; the right panel shows matches with stimuli on the same $F0$.

and were paid for their services. All had pure-tone thresholds within the normal range at octave frequencies between 250 Hz and 4 kHz. All the subjects had participated in previous matching experiments and were experienced in formant-matching tasks.

B. Results

1. Mean matches (veridicality)

Mean matches to the 1100- and 1200-Hz targets (Fig. 2) are well-separated and the standard errors across listeners of these mean matches are generally small regardless of whether the sounds differ in $F0$.

Nevertheless, there are some systematic deviations from veridicality. For unresolved harmonics, matches are higher in frequency when the target has a narrow bandwidth (and the match a wide one) than when it has a wide bandwidth (and the match a narrow one), regardless of whether the sounds are on the same $F0$ or not. The direction of this effect is equivalent to the narrow bandwidth sound being heard as having a higher formant frequency, and may be due to the duller sound of a wider-bandwidth adjustable sound being compensated for by it being adjusted to a higher formant frequency (and vice versa). The resolved-harmonic sounds show a similar though weaker tendency.

2. Match variability (reliability)

The mean within-subject standard deviations of matches across the different conditions are shown in Fig. 3.

The main results from this experiment, where subjects made matches across sounds on different bandwidths, are (i) that matches are more variable for sounds that have different $F0$'s than for those with the same $F0$ [$F(1,8)=9.95$, $p<0.02$], but (ii) that this difference does not depend on whether the sounds have resolved or unresolved harmonics near the formant peak [$F(1,8)=0.1$, $p>0.5$].

If we contrast these present results with those from experiment 2 in Dissard and Darwin (2000), which used similar conditions but with sounds that always had the same bandwidth, an interesting pattern then emerges which is illustrated in Fig. 4 for the eight listeners who took part in both experiments. For sounds with the same $F0$ (left-hand panel of Fig. 4), making the bandwidths different increased the match variability equally for sounds with unresolved and resolved harmonics [main effect of bandwidth: $F(1,7)=36.9$, $p=0.0005$; bandwidth$\times F0$ interaction: $F(1,7)<1$]. But, when the sounds had a different $F0$ (right-hand panel of Fig. 4), making the bandwidths different increased the match variability more for sounds with unresolved than with resolved harmonics [bandwidth$\times F0$ interaction: $F(1,7)=13.4$, $p=0.008$]. The three-way interaction reflecting the different patterns in the two panels of Fig. 4 is marginally significant [$F(1,7)=4.1$, $p=0.08$].

This pattern of results can be interpreted as follows. Matches show low variability if they can be made directly from the excitation pattern (i.e., same-bandwidth matches for unresolved harmonics on either the same or different $F0$'s, or same-bandwidth matches for resolved harmonics provided that they are on the same $F0$). If listeners are prevented from making a match directly from the excitation pattern either by putting resolved harmonics on a different $F0$, or by giving the sounds a different bandwidth, then variability increases;

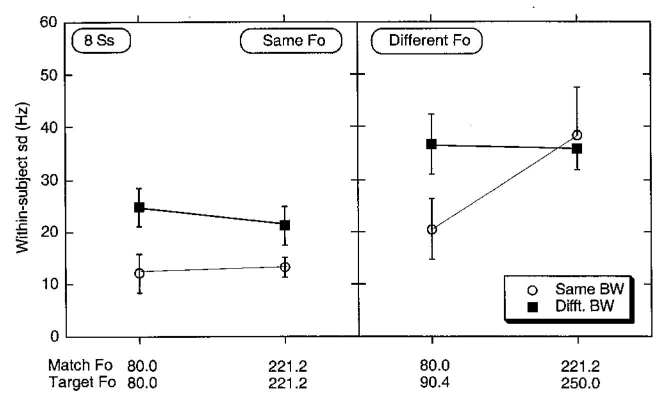

FIG. 4. Comparison of within-subject standard deviations between the present experiment 1, in which target and match sounds had the different bandwidths and experiment 2 of Dissard and Darwin (2000), in which they had the same bandwidth.

TABLE II. Comparison of experimental matching accuracy in experiment 1 with predictions of a model based on best-fitting excitation patterns.

Condition	Match BW/ target BW	Match criterion	Model formant range (Hz)	±1 within-subject s.d. (Hz)	Ratio col.4/col.5
USD	100/200	Excit peak	156	56	2.8
USD	200/100	Excit peak	163	40	4.1
RSD	100/200	Env peak	84	44	1.9
RSD	200/100	Env peak	146	52	2.8
	Average		137	48	2.9
UDD	100/200	Excit peak	145	86	1.7
UDD	200/100	Excit peak	172	72	2.4
RDD	100/200	Env peak	182	64	2.8
RDD	200/100	Env peak	313	84	3.7
	Average		203	76.5	2.7
	Grand Av		170	62	2.8

but having a difference both in $F0$ and in bandwidth gives no further increase in variability over having a difference in either one.

An additional point: sounds which have the same $F0$ but differ in bandwidth are matched with about the same reliability whether they have resolved or unresolved harmonics near the formant peak. This similarity is interesting since the simplest way that the matches could be made differs between the resolved and unresolved cases. With unresolved harmonics, the match can be made through the peak in the explicit excitation pattern. But, with resolved harmonics the match can only be made via the peak in the interpolated envelope.

Taken together, these observations suggest that the metric of match reliability that we have used is not sensitive to differences in the level of abstraction of a match beyond the simple distinction between making a match at the level of the excitation pattern, and making a match at a more abstract level.

Before adopting this somewhat negative conclusion, we should make sure that listeners are not just making the matches via the raw excitation pattern. In Dissard and Darwin (2000) we addressed this problem by simulating a strategy of minimizing the error between the target and the match excitation patterns. In the earlier paper we found that when listeners were matching identical stimuli (same $F0$, same bandwidth) the range of formant frequencies, for which the mean-square error difference between the excitation patterns was 1 dB above its minimum, was about 1.4 times the experimentally determined within-subject standard deviations. Following this approach we use the same model as in the earlier paper to compare the expected variability of their matches from the model with those from the experimental data. The fourth column of Table II shows the range of formant frequencies for which the mean-square error difference between the excitation patterns of target and matched sounds was 1 dB above its minimum. The fifth column of Table II shows ±1 within-subject standard deviation from our experimental data, and the sixth column shows the ratio of these two values. The average ratio of the two measures is 2.9, double the expectd ratio of 1.4. Listeners are thus making the matches twice as reliably as predicted from a simple excitation-pattern-based model. It is thus unlikely that listeners are simply finding the best match of the raw excitation patterns.

III. EXPERIMENT 2

In our previous paper (Dissard and Darwin, 2000) we showed that the perceptual cost (as measured by match variability) of interpolating a spectral envelope in one-formant complex sounds was higher for stimuli with resolved harmonics than for stimuli with unresolved harmonics.

The goal of experiment 2 was to assess whether this difference is in fact a continuous function which decreases smoothly as the number of harmonics that sample the formant envelope increases, or a discrete one that reflects the difference between resolved and unresolved harmonics. We test these two possibilities by increasing the frequency of $F0$ within the resolved-harmonic range. If match variability increases gradually, then the effect is due simply to the density of harmonic sampling. If there is no increase in match variability within the resolved-harmonic range, then the difference that we obtained previously is likely to be due to the discrete difference between resolved and unresolved harmonics.

A complicating factor in designing the experiment was that formant frequency difference limens are generally smaller when the formant peak is located symmetrically between two harmonics than when it lies on a harmonic frequency (Lyzenga and Horst, 1997). We control for any effect of the position of harmonics under the formant peak on the variability of matches by selecting $F0$ values so that both the target and the adjustable sound on a particular trial had the formant frequency either at a harmonic frequency, or midway between two harmonic frequencies.

A. Stimuli and procedure

As in the previous experiment, subjects matched single-formant (1100-Hz) sounds that differed in $F0$. Within a condition, the $F0$ of the target and the match sound were constructed so that both sounds either had a harmonic at the formant frequency, or had the formant frequency symmetrically between the harmonics. Half the stimuli had either the third, fourth, fifth, or sixth harmonic aligned to the formant

TABLE III. $F0$ values used in experiment 2.

Block /||\		Block /|\	
Harmonic number	$F0$ (Hz)	Harmonic number	$F0$ (Hz)
3.5	314	3	366.7
4.5	244	4	275
5.5	200	5	220
6.5	169.2	6	183.3

peak of the target sound (/|\); the other half had a formant located halfway between two harmonics (/||\). The exact $F0$ values are given in Table III.

Because of the impossibility of keeping an exactly constant $F0$ ratio between match and target, we made sure that the average $F0$ ratio within each of the two groups was very similar (see Table IV).

Nine subjects (including the first author) participated in the experiment. Subjects were university students or staff and were paid for their services. All had pure-tone thresholds within the normal range at octave frequencies between 250 Hz and 4 kHz and were already trained in formant-matching tasks.

B. Results

1. Mean matches (veridicality)

The mean matched formant frequencies for each condition across the six subjects are shown in Fig 5. All matches are relatively accurate; no main effect of the $F0$ factor is observed $[F(5,8)=1.15]$; in addition, accuracy of matches is the same for stimuli with a formant peak between two harmonics and for those with a harmonic at the formant frequency.

2. Match variability (reliability)

Figure 6 shows the mean within-subject standard deviation of matches within each block of trials, together with the standard error of these means across the nine subjects. Variability of matches does not differ significantly either for the six $F0$ conditions $[F(5,8)=1.84]$ or between the two formant-alignment conditions (/|\ vs /||\), $[F(1,8)=1.79]$.

C. Discussion

Experiment 2 has found no effect on the accuracy of formant matches of either the density or the alignment of harmonics under the formant envelope. Since there is no increase in match variability within the resolved-harmonic range, then the difference that we obtained previously between conditions with resolved and unresolved harmonics is likely to be due to the discrete difference between resolved and unresolved harmonics rather than simply to the density of harmonic sampling. Our previous conclusion, that matches have lower variability if they can be made directly from the excitation pattern rather than requiring the interpolation of a spectral envelope, is therefore justified by the results of experiment 2.

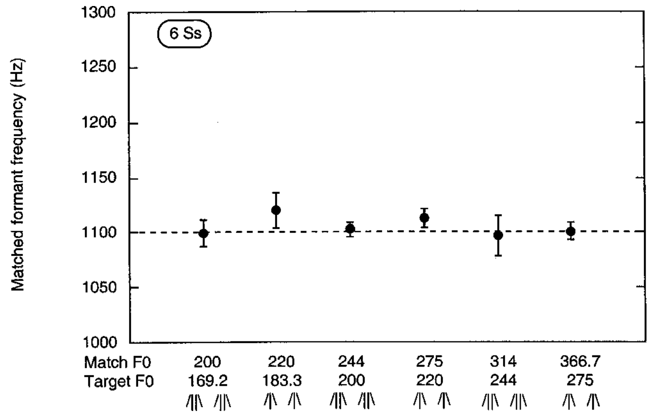

FIG. 5. Mean matched formant frequencies in experiment 2 for a target formant frequency of 1100 Hz. Error bars are standard errors of the mean over six subjects.

The experiment also found no difference in match variability between conditions where a harmonic coincided with the formant frequency (/|\) and those where the formant frequency lay midway between two harmonics (/||\). This result contrasts with the general finding of Lyzenga and Horst (1997) that formant frequency difference limens are higher in the former condition than in the latter. The discrepancy, however, is only apparent since in the specific conditions of their experiment that most closely match the stimuli that we have used (Expt. 1, Fig. 3, Klatt envelope, $F0=200$ Hz, F_1) Lyzenga and Horst also find no reliable difference between the two harmonic alignments.

IV. GENERAL DISCUSSION

The results of these experiments confirms the conclusions from our previous paper that formant matches that can be made on the basis of the matched sounds' excitation patterns are less variable than are matches that require a more abstract representation of the sound, such as its envelope. The present experiments have varied the bandwidth as well as $F0$ across the target and match sounds and have shown that a difference in bandwidth between the target and match sounds generally increases the variability of the matches. The exception to this general rule is when the sounds differ

TABLE IV. $F0$ values and ratios for target and match sounds in experiment 2.

Block /||\			Block /|\		
Match $F0$	Target $F0$	Ratio M/T	Match $F0$	Target $F0$	Ratio M/T
200	169.2	1.18	220	183.3	1.2
244	200	1.22	275	220	1.25
314	244	1.29	366.7	275	1.33
mean=1.23			mean=1.26		

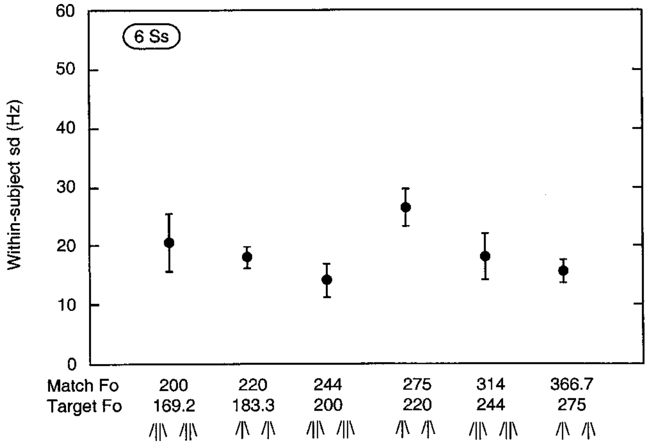

FIG. 6. Average within-subject standard deviations of matches in experiment 2 with their standard errors over six subjects.

in $F0$ and also have $F0$'s that give resolved harmonics around the formant frequency. Under these conditions, giving additionally a difference in bandwidth does not increase matching variability.

This pattern of results implies that although the matching task is sensitive to the difference between matches that can be made on the basis of the excitation pattern and those that need a more abstract representation, it is not sensitive to the difference between matches that can be made on the basis of an interpolated spectral envelope, and those that require the peak of the spectral envelope (or formant frequency) to be extracted.

This result is compatible with the possibility that listeners are basing their matches on the peak of the spectral envelope even when the spectral envelope itself can provide an adequate basis for a match. Although attractive, this conclusion needs more supporting evidence. Further work on this question could make use of the hierarchy of properties shown in Table I to explore the sensitivity of other tasks to different levels of abstractness needed to perform the task.

ACKNOWLEDGMENTS

The first author was supported by Grant No. GR/L03422 from the UK EPSRC. Paul Russell wrote the software for formant matching based on SENSYN for the PowerPC. We are grateful to Rob Hukin for technical assistance, and to Peter Assmann for helpful reviewer's comments.

[1]An alternative strategy for forcing listeners to use the formant peak in their matching is to make the matching sound a sine wave. This technique was used by Hermansky (1987) but suffers from the problem that it encourages listeners to hear out individual partials of a complex (particularly when they are close to the formant frequency) and to match the frequency of a particular partial. By keeping both the target and the match as complex sounds, we hope to dissuade listeners from hearing out the individual components of a complex.

[2]For a comparison of various other metrics for comparing speech spectra see Nocerino *et al.* (1985).

Carlson, R., Granström, B., and Klatt, D. H. (**1979**). ''Vowel perception: The relative perceptual salience of selected acoustic manipulations,'' STL-QPSR **3–4**, 73–83.

Dissard, P., and Darwin, C. J. (**2000**). ''Extracting spectral envelopes: Formant frequency matching between sounds on different and modulated fundamental frequencies,'' J. Acoust. Soc. Am. **107**, 960–969.

Hawkins, S., and Stevens, K. N. (**1985**). ''Acoustic and perceptual correlates of the non-nasal–nasal distinction for vowels,'' J. Acoust. Soc. Am. **77**, 1560–1575.

Hermansky, H. (**1987**). ''Why is the formant frequency difference limen asymmetric?,'' J. Acoust. Soc. Am. **81**, S18.

Klatt, D. H. (**1979**). ''Perceptual comparisons among a set of vowels similar to /ae/: Some differences between psychophysical distance and phonetic distance,'' J. Acoust. Soc. Am. **66**, S86.

Lyzenga, J., and Horst, J. W. (**1997**). ''Frequency discrimination of stylized synthetic vowels with a single formant,'' J. Acoust. Soc. Am. **102**, 1755–1767.

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.

Nocerino, N., Soong, F. K., Rabiner, L. R., and Klatt, D. H. (**1985**). ''Comparative study of several distortion measures for speech recognition'' Speech Commun. **4**, 317–331.