Journal of Phonetics (1996) **24,** 187–208

On explaining certain male-female differences in the phonetic realization of vowel categories

Randy L. Diehl,* Björn Lindblom,† Kathryn A. Hoemeke and Richard P. Fahey

University of Texas, Austin, TX 78712, U.S.A.

Received 11th October 1994, and in revised form 10th July 1995

The scaling between female and male formant frequencies tends to be highly nonuniform across vowel categories with the result that female vowels exhibit greater between-category dispersion in the $F_1 \times F_2$ plane than male vowels. Vocal tract modeling studies strongly suggest that this greater dispersion of female vowels is partly behavioral, rather than purely anatomical, in origin. The present study tested one explanation for this behavioral difference between females and males, *viz.,* that without the compensatory effect of greater dispersion, the typically higher fundamental frequency (f_0) of female talkers would yield reduced identifiability of vowels because of sparser harmonic sampling of spectral envelopes. The specific question addressed was whether, all else being equal, a higher f_0 has the assumed deleterious effect on vowel identifiability. In two experiments, the overall effect of increasing f_0 beyond 150 Hz was to reduce vowel labeling accuracy. Across individual vowel categories, the effect of raising f_0 varied. Auditory modeling suggests that this category variation is partly attributable to differing degrees to which a high f_0 obscured the distinctive auditory properties of each vowel category. Consistent with the spectral undersampling account, the performance decline at high f_0s was reduced or eliminated when f_0 was time-varying rather than constant.

1. Introduction

Adult female vocal tracts tend to be shorter than those of adult males, and accordingly female formants tend to be higher in frequency. One might naively assume that if a given female's vocal tract were, say, 20% shorter than a given male's, then her formant frequencies would be uniformly scaled upward by about 20% relative to his. However, as Fant (1966, 1975) and others have shown, the scale factor relating female and male formant values is decidedly *non*uniform across

* Please address all correspondence to Randy L. Diehl, Department of Psychology, 330 Mezes, University of Texas at Austin, Austin, TX 78712, U.S.A.

† Current address: Department of Linguistics, University of Stockholm, S-106 91 Stockholm, Sweden.

0095-4470/96/020187 + 22 $18.00/0

different vowel categories and across formants. Moreover, this nonuniform scaling appears to be fairly consistent across languages (Fant, 1975).

Fig. 1 illustrates the nonuniformity with Korean data from 10 males and 10 females collected by Yang (1990, 1992) at the University of Texas Phonetics Laboratory. A simple description of these data is that as the frequency of a given formant increases, so does the scale factor relating female to male values. The dotted lines show the predicted effect of a uniform scaling of F_1 and F_2 based on acoustically derived estimates of vocal-tract length;[1] the solid regression lines were derived from the actual data. Notice that the slopes of these regression lines are greater than the slopes of the uniform scaling functions, and the regression intercepts are both negative. The effect of this nonuniform scaling is shown in Fig. 2. Whereas the female /i/ and /a/ differ considerably from the male versions in F_2 and F_1, respectively, the female /u/ is quite similar to the male /u/. Thus, relative to a uniform female-to-male scaling across vowels, the nonuniform scaling results in greater between-category dispersion of the female vowels. This paper addresses the question: what is the basis of this nonuniform scaling?

Fant (1966, 1975) has suggested that part of the answer to this question lies in a characteristic anatomical difference between the sexes: in adult males the pharynx takes up a greater proportion of overall vocal tract length than in adult females. Apparently, however, anatomical differences of this sort cannot fully account for the pattern of observed formant differences. A vocal tract modeling study by Nordström (1977) showed that even when sex differences in relative pharynx length are taken into account, female values for F_1 and F_2 are not well predicted on the basis of the corresponding male values. Using a more realistic vocal tract model, Goldstein (1980) similarly concluded that anatomical differences account for only part of the vowel formant differences between males and females. Further simulations by Traunmüller (1984) yielded a somewhat closer agreement with observed male and female differences for F_2 and F_3, but not for F_1. Fant, Nordström & Goldstein all attribute the anatomically unexplained formant differences (or, in the case of /u/, formant similarities) to sex differences in articulatory behavior. For example, in producing the vowel /u/ female talkers may offset the acoustic effects of a shorter vocal tract by creating smaller constriction areas at the lips and tongue hump (Edholm X-ray data cited by Fant, 1975), yielding a formant pattern similar to that of male talkers.[2] Other evidence that sex differences in vowel formant frequencies are partly behavioral rather than purely anatomical in origin has been reported by Mattingly (1966), Sachs, Lieberman & Erickson (1973), and Henton (1992a,b).[3]

We are left, then, with the problem of accounting for these sex differences in

[1] Following Nordström & Lindblom (1975), whose procedure was adapted with modification from Fant (1973: p. 52), the uniform scale factor was derived by calculating the ratio of the average female F_3 for low vowels to the average male F_3 for the same low vowels. For our purposes, ''low vowel'' was defined as any category for which the average adult male F_1 was greater than 550 Hz. The Nordström & Lindblom procedure assumes that F_3 of low vowels provides a reasonable acoustic estimate of vocal tract length.

[2] Helmholtz (1885/1954: p. 105) was perhaps the first to observe that females may use more extreme vocal tract constrictions than males such that formant differences between males and females expected on purely anatomical grounds are reduced for certain vowels.

[3] Bennett (1981; Bennett & Weinberg, 1979a,b) reported that preadolescent males and females show formant frequency and f_0 differences in the same direction as—but considerably smaller than—those observed for adult males and females. These differences were assumed to be partly behavioral in origin.

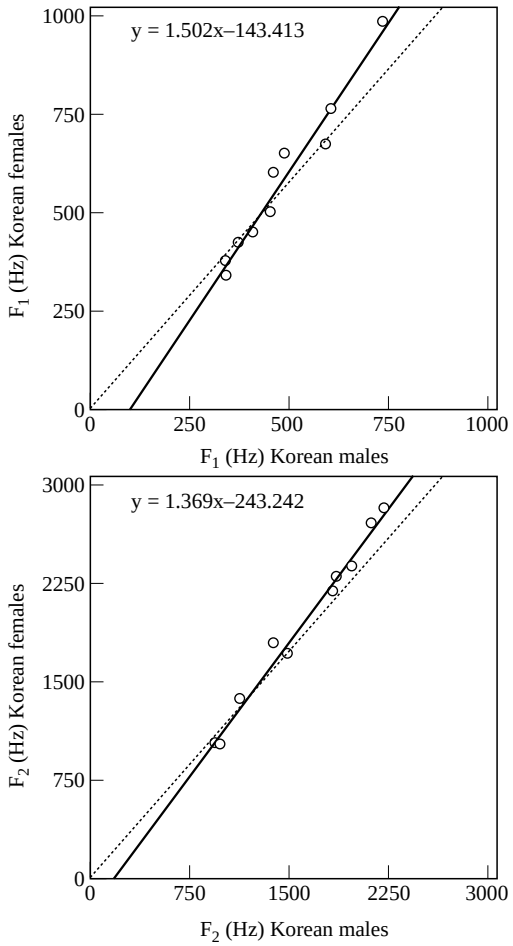

Figure 1. Female formant frequencies plotted as a function of male formant frequencies for Korean vowels spoken by 10 adult talkers of each sex (adapted from Yang, 1990). The solid regression lines are derived from the data; the dotted lines correspond to the predicted effect of uniform scaling of formant values based on acoustically derived estimates of vocal-tract length (see footnote 1).

articulatory behavior. One hypothesis, offered by Sachs *et al.* (1973), following Mattingly (1966), is that the differences reflect a strategy of talkers to supplement or exaggerate the acoustic effects of anatomical differences so as to achieve a more masculine-sounding vocal quality in males and a more feminine-sounding vocal quality in females. Consistent with Darwin's (1871) theory of sexual selection, tendencies to exaggerate sexual dimorphism both anatomically and behaviorally are widely observed among animal species, and this is particularly true of dimorphism related to vocalization (Ohala, 1984). (In fact, Ohala, 1984, suggested that the relatively elongated pharynx of adult human males evolved to enhance the lower formant frequencies associated with the larger male vocal tract.)

A main weakness of the sexual dimorphism hypothesis is that it does not explain the nonuniformity of the scaling between male and female vowels. If the behavioral

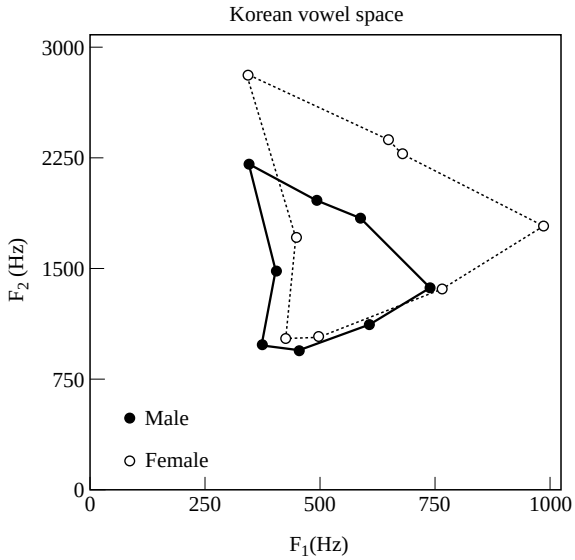

Figure 2. $F_2 \times F_1$ plot of Korean vowels produced by 10 male and 10 female talkers (adapted from Yang, 1990). Vowels in the interior of each vowel space were omitted in this figure.

aim were to exaggerate those acoustic differences that are owing to sex-related differences in vocal tract size, it is unclear why the scale factor for F_1 and F_2 would be greatest for vowels with a high F_1 and F_2, respectively. As noted, for example, the vowel /u/ is produced by females in a way that actually reduces the male-female formant differences expected on anatomical grounds. This is, of course, opposite to what is predicted by the hypothesis that male and female speech behavior serves to exaggerate the acoustic effect of sex-related anatomical differences. In the case of /i/ the male-female scale factor is much larger for F_2 than for F_1, whereas just the reverse is true in the case of /a/ (Fant, 1975; Yang, 1990, 1992). These facts suggest that the behavioral aim of female talkers is not to sound as if they have smaller vocal tracts than they actually do; rather, the aim appears to be to produce vowels that are relatively more dispersed in the $F_1 \times F_2$ space than those of males.

But this again raises the question of why males and females differ in their articulatory behavior. As a partial answer, Fant (1975) raised the possibility ''of universal 'feministic' preference in vowel qualities'' (p. 18). Goldstein (1980) also suggested that women may ''prefer a wider vowel triangle because of a tendency to speak more clearly'' (p. 235), but noted that the tendency appears to be cultural, since female speakers of at least one language—Arabic—do not reliably produce clearer speech than their male counterparts. Although it may well be true that clear speech, particularly in the form of more dispersed vowels, is considered more ''feminine'' in most cultures, this may be more of an effect than a cause. What needs to be explained is why females tend to prefer clear speech in the first place.

One possible account of this preference is based on the requirement of maintaining sufficient auditory contrast among vowel categories. The fundamental frequencies (f_0s) of adult females average about 75%–90% higher than those of adult males (Hollien & Paul, 1969; Hollien & Shipp, 1972), and this means that the

spectral envelopes of female vowels tend to be more sparsely sampled harmonically. Ryalls & Lieberman (1982) hypothesized that the greater between-category dispersion of female vowels in $F_1 \times F_2$ space relative to those of males may be a way for female talkers to compensate for the otherwise poorer resolution of their spectral peaks. Although the point was not raised by Ryalls & Lieberman (1982), their hypothesis implies that the tendency toward greater vowel dispersion may be associated with higher f_0s per se and not simply with female speech. Evidence consistent with their hypothesis (and not predicted by hypotheses stressing anatomical and behavioral tendencies toward sexual dimorphism) was obtained by Cleveland (1977): F_1 and F_2 of vowels sung by tenors were related to those of vowels produced by bass singers in very much the same (nonuniform) way that average male formant values are related to average female values.[4]

The hypothesis of Ryalls & Lieberman (1982) (hereafter the *sufficient contrast* hypothesis) presupposes that the acoustic and perceptual distance among vowels of different categories will tend to decrease at higher f_0 values and that, all else being equal, vowel intelligibility will consequently be reduced. (Were this not the case, there would be no need for talkers with high f_0s to produce more dispersed vowel categories.)

In a preliminary investigation, we applied a Euclidean acoustic distance metric to 10 synthetic English monophthongs at f_0s ranging from 90 to 390 Hz. For every pair of vowels at a given f_0, the metric computes the square root of the sum of squared amplitude differences at each harmonic over a frequency range of 0 to 5 kHz. The female formant frequencies were equal to mean adult female values reported by Peterson & Barney (1952). For the male vowel set, the Peterson & Barney adult male formant values were uniformly scaled upward by a factor of 1.156 to neutralize the effects of estimated vocal tract length differences between the sexes (see footnote 1). Thus, whatever male-female differences are present in the distance calculations reflect only the effects of nonuniform scaling. Fig. 3 plots the mean acoustic distance for both a male and female vowel set. Consistent with the sufficient contrast hypothesis, there was a marked reduction in the acoustic distance score at higher f_0 values.[5] Also as predicted, the female vowel set yielded a greater mean distance than the male vowel set at every f_0.

A number of studies have examined vowel identifiability as a function of f_0, and most of the evidence suggests that performance declines at higher f_0 values. This trend was observed with both synthesized vowels (Ryalls & Lieberman, 1982; but see Lehiste & Meltzer, 1973, for a less consistent f_0 effect) and sung vowels (Stumpf, 1926; Howie & Delattre, 1962; Morozov, 1965; Nelson & Tiffany, 1968; Sundberg, 1977a,b; Smith & Scott, 1980; Gottfried & Chew, 1986).

Two other characteristics of identification performance were commonly observed in these studies. First, identification accuracy was not strictly a monotonically

[4] Cleveland's (1977) finding casts doubt not only on the sexual dimorphism hypothesis but also on any other account of nonuniform formant frequency scaling that is based on sex or gender differences per se. Thus, it appears unlikely that the greater dispersion typical of female vowels is simply a reflection of the greater tendency of females to be aware of and to produce standard forms of the language (Labov, 1990; Chambers, 1992).

[5] It might be argued that a more appropriate distance metric would compute the *average* squared difference across harmonics rather than using the sum of the squared differences at each harmonic. However, an average measure fails to capture the intuition that a greater number of harmonics within a fixed frequency range carries more information about the shape of the spectral envelope.

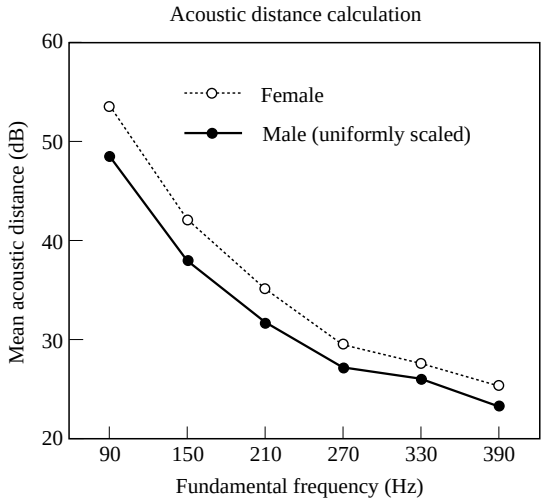

Figure 3. Mean intervowel acoustic distance for a male and female set of 10 synthetic English monophthongs as a function of fundamental frequency. The formant frequencies of the female set were equal to the mean adult female values reported by Peterson & Barney (1952). For the male vowel set, the Peterson & Barney adult male values were uniformly scaled upward by a factor of 1.156 (see footnote 1) to neutralize the effects of estimated vocal-tract length differences between the sexes.

decreasing function of f_0; rather, peak performance tended to occur slightly above the lowest f_0 value of the range used (see, especially, Morozov, 1965, and Ryalls & Lieberman, 1982). Second, peak performance occurred at a higher f_0 value for female than for male vowels (Morozov, 1965; Lehiste & Meltzer, 1973; Ryalls & Lieberman, 1982). Both of these effects suggest that identification accuracy is somewhat impaired by combinations of f_0 and formant frequencies that are unlikely to occur in natural speech. Although Ryalls & Lieberman (1982) describe the effects of mismatching formant frequencies and f_0 as ''secondary'', nothing in their reported results (or in the results of other studies cited) excludes the possibility that such mismatching may explain the *entire* f_0 effect on vowel identification accuracy. Thus, the decline in accuracy of vowel identification at higher f_0s could, in principle, derive from increased unnaturalness or atypicality of the formant-f_0 pairing rather than from undersampling of the spectral envelope per se (as assumed by the sufficient contrast hypothesis).

Another potential difficulty in evaluating the results of the above investigations is that, apart from any effect of undersampling, a higher f_0 can alter identification performance by shifting perceived vowel height upward (Potter & Steinberg, 1950; Slawson, 1968; Fujisaki & Kawashima, 1968; Hoemeke & Diehl, 1994). This problem is most critical when the stimulus set includes vowel categories of different heights and correspondingly different intrinsic f_0 values. For example, a token synthesized as an /ɛ/ at a lower f_0 may sound like an /ɪ/ when the f_0 is substantially raised, and if it is so identified, the response will spuriously be counted as an error. It is worth noting that in vowel perception studies where f_0 is systematically varied (e.g., Gottfried & Chew, 1986), height confusions tend to predominate among incorrect identification responses.

The present study was designed to test further the assumption of the sufficient contrast hypothesis that, all else being equal, higher f_0s yield reduced vowel identifiability because of sparser harmonic sampling of spectral envelopes. To mitigate the potential problem of height-related biasing effects of f_0, the stimulus set was limited to synthesized male and female formant patterns corresponding to /ɪ/ and /ʊ/, two vowels that have approximately the same height and intrinsic f_0 (Peterson & Barney, 1952; Lehiste & Peterson, 1961). In Experiment 1, f_0 was varied across vowel tokens but remained constant within a token, and all stimuli were presented in quiet. Although the results were generally consistent with the sufficient contrast hypothesis, several alternative explanations, including one based on the atypicality of certain f_0 and formant frequency pairings, could not be ruled out. In Experiment 2, both constant and time-varying f_0 trajectories were used, and stimuli were presented in quiet as well as in background noise (the latter to eliminate possible ceiling effects in identification performance). If an observed decrement in identification accuracy at higher f_0s is produced simply by the atypicality of the formant-f_0 pairing, then there is no reason to predict that within-token variation in f_0 per se would significantly improve performance. However, if the decrement in performance is attributable to harmonic undersampling of the spectral envelope, as claimed by the sufficient contrast hypothesis, then a time-varying f_0 should help to alleviate the problem. This is because frequency-modulated harmonics sweep continuously through portions of the spectral envelope, defining its shape more completely than in the constant f_0 condition. Experiment 2 was designed to distinguish empirically between these (and one other) alternative theoretical accounts of reduced vowel identfication accuracy at high f_0s.

2. Experiment 1

2.1. *Method*

2.1.1. *Stimuli*

The cascade-formant branch of the Klatt (1980) synthesizer, implemented on a DEC VAXstation 3500, was used to prepare male-like and female-like formant patterns corresponding to the English vowels /ɪ/ and /ʊ/ at 6 steady-state values of f_0 ranging from 90 Hz to 390 Hz in steps of 60 Hz. The first three formant frequencies, which equalled the average adult male and adult female values reported by Peterson & Barney (1952) for these vowel categories, were (in Hz): 390, 1990, and 2550 (male-like /ɪ/); 430, 2480, and 3070 (female-like /ɪ/); 440, 1020, and 2240 (male-like /ʊ/); and 470, 1160, and 2610 (female-like /ʊ/). F_4 and F_5 were fixed at 3300 Hz and 3850 Hz for all stimuli. These formant values yielded stimuli that were judged by the experimenters to be quite natural sounding tokens of their respective categories.

The stimuli were 100 ms in duration, with the initial and final 15 ms multiplied by cosine-squared onset and offset functions. Each stimulus was normalized to a constant RMS amplitude as measured during the interval separating the rise and decay segments. Formant frequencies and f_0s were verified using a spectrum analyzer.

2.1.2. *Procedure and subjects*

Eleven subjects identified 10 randomized blocks of the 24 stimuli (2 vowel categories $\times$ 2 sexes $\times$ 6 f_0s) by pressing either of two response keys corresponding to /ɪ/ or /ʊ/. A sample word containing the appropriate vowel was written next to each response key. Subjects were given 2 s to respond after which another 1 s elapsed before the next stimulus token was presented.

The stimuli, stored on an IBM AT computer, were output at a 10 kHz sampling rate via a 16-bit D/A converter, low-pass filtered at a 4.9 kHz cut-off frequency, and presented to subjects binaurally through Beyer DT-100 earphones at 72 dB SPL. Up to four subjects, seated at separate response stations in a double-walled sound-attenuated chamber (Industrial Acoustics Corporation), participated in each experimental session.

Subjects were introductory psychology students at the University of Texas at Austin who participated in the experiment to satisfy a course requirement. All were native speakers of English and reported having normal hearing.

2.2. *Results and discussion*

Fig. 4 plots the average correct identification for all vowel stimuli as a function of f_0. Beginning at an f_0 value of 150 Hz, there was an overall monotonic decline in labeling accuracy. Fig. 5 displays the results separately for the male-like and female-like formant patterns of each vowel category. A repeated measures analysis of variance on the percent correct labeling responses showed significant main effects of f_0 [$F(5, 50) = 8.19$, $p < 0.001$], and vowel category [$F(1, 10) = 6.13$, $p < 0.05$], as well as significant interactions between f_0 and vowel category [$F(5, 50) = 16.62$, $p < 0.001$] and between f_0, vowel category, and male-like *vs.* female-like formant pattern (hereafter referred to as *sex*) [$F(5, 50) = 4.25$, $p < 0.01$]. Analysis of simple effects showed that the effect of f_0 was significant ($p < 0.05$) for the male- and

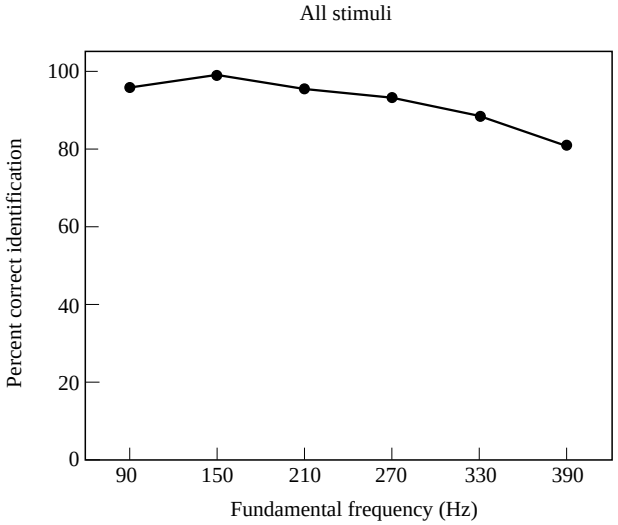

Figure 4. Mean correct identification for all vowel stimuli as a function of f_0 (Experiment 1).

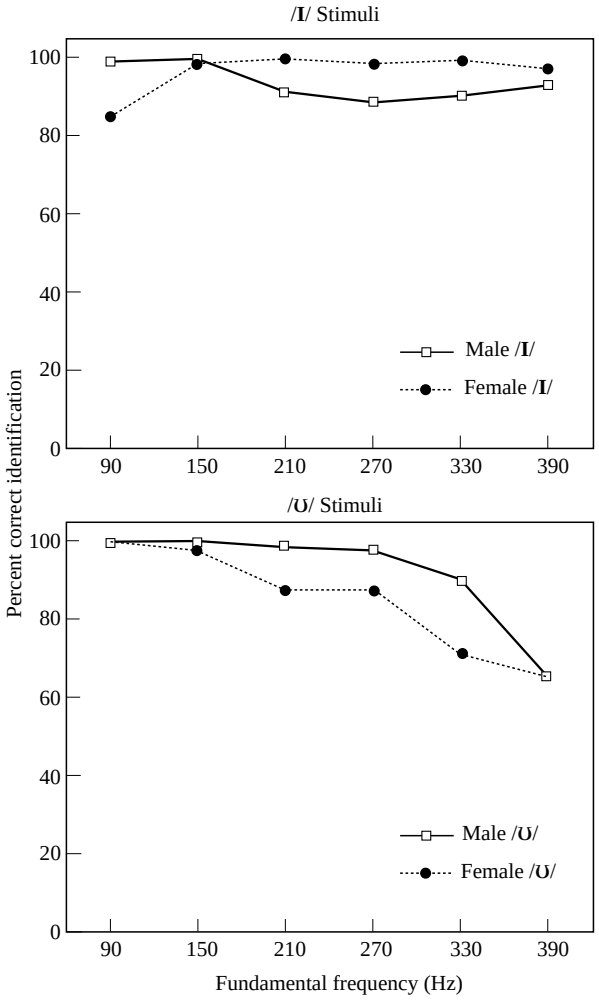

Figure 5. Mean correct identification for each vowel category and sex (Experiment 1).

female-like /ʊ/ formant patterns and for male-like /ɪ/. The only formant pattern that did not show a significant effect of f_0 was female-like /ɪ/. It should be noted that the decline in average performance for this vowel at the 90 Hz f_0 value was largely attributable to a single subject who labeled this item consistently as /ʊ/. Listeners performed nominally better on the female-like tokens in the case of /ɪ/ and nominally better on the male tokens in the case of /ʊ/, but the interaction between vowel category and sex was not statistically significant [$F(1, 10) = 1.67$, $p > 0.2$].

Consistent with a basic assumption of the sufficient contrast hypothesis, the overall effect of increasing f_0 was to reduce identification accuracy. However, the magnitude and reliability of this effect were considerably greater for the /ʊ/ than for the /ɪ/ category. For the female-like /ɪ/ the effect was nominally (but not significantly) in the opposite direction, while for male-like /ɪ/, the downward trend

in identification accuracy bottomed out at 270 Hz. The significant main effect of vowel category is consistent with earlier findings that /ɪ/ tends to be identified more accurately than /ʊ/ (Lehiste & Meltzer, 1973; Gottfried & Chew, 1986), while the significant interaction between vowel category and f_0 bears some resemblance to Ryalls & Lieberman's (1982) finding that /ɪ/ was identified more accurately than /ʊ/ only at a high f_0. However, these parallels must be interpreted cautiously since far more vowel categories were used in the earlier studies than in the present one.

Recall that a main reason for restricting the stimulus set in this study to vowel categories of similar height was to reduce the likelihood that identification accuracy might reflect height-related biasing effects of f_0. The interaction between vowel category and f_0 raises the possibility that a different kind of biasing effect of f_0 may nevertheless have occurred. In cases where vowel identity is relatively uncertain, listeners may tend to adopt a labeling criterion based on an abstract analogy between fundamental frequency and the frequency of the upper formants. In particular, an otherwise ambiguous stimulus with a higher f_0 might be interpreted as more /ɪ/-like by analogy to the high-frequency upper formants of the /ɪ/ category.[6] (The analogy is very abstract indeed inasmuch as the frequency range of the higher f_0s in this study lies far below the upper formant range of /ɪ/ and even below the upper formant range of /ʊ/.) Such a biasing effect might account for the decline in identification accuracy for the /ʊ/ category at higher f_0s. An /ɪ/-biasing effect of higher f_0s would also be expected to elevate the hit rate for the /ɪ/ category. Unfortunately, the results of Experiment 1 are difficult to interpret in this respect because of a ceiling effect in the case of female-like /ɪ/ stimuli. The slight upturn in labeling accuracy for the male-like /ɪ/ stimuli at f_0 values higher than 270 Hz does offer at least modest support for the biasing account.

Drawing together the above remarks and earlier discussion, we are left with three possible (and not necessarily incompatible) accounts for the decline in identification accuracy, especially for /ʊ/, at higher f_0s:

(a) The spectral envelopes are harmonically undersampled, yielding poor definition of peaks and valleys.

(b) The pairings of formant values and higher f_0s are atypical of normal male and female vowels and therefore unnatural sounding.

(c) By abstract analogy to relatively high-frequency upper formants, higher f_0s bias vowel labeling toward high, front, unrounded categories such as /ɪ/.

One potential way to distinguish empirically between these accounts—which we refer to as the *undersampling, formant-f_0 mismatching,* and *biasing* accounts, respectively—is to include time-varying as well as constant f_0 contours among the stimulus set. As noted earlier, if the undersampling account is correct, then a time-varying f_0 should improve labeling performance since the changing frequencies of the harmonics will sweep through portions of the spectral envelope, defining its shape more completely. In contrast, neither the formant-f_0 mismatching nor biasing accounts predicts greater identification accuracy owing to variation in f_0 per se. Provided that the mean frequency values for the time-varying f_0 stimuli are roughly

[6] Kuhl, Williams & Meltzoff (1991) reported that when adult listeners were instructed to match pure tones to vowel sounds, they selected higher frequency tones for /i/ than for /a/. This confirms the possibility that listeners can form an analogy between higher formant frequencies and fundamental frequencies.

equivalent to the values for the corresponding constant f_0 stimuli, there is no obvious reason to expect reduced effects of formant-f_0 mismatching or f_0 biasing.

These predictions were tested in the next experiment. The design was similar to that of Experiment 1 except that the stimulus set varied in f_0 contour (constant *vs.* time-varying), and stimuli were presented in two conditions of background noise as well as in quiet. The noise conditions were included to reduce the likelihood of ceiling effects such as were observed in Experiment 1.

3. Experiment 2

3.1. *Method*

3.1.1. *Stimuli*

Two sets of male-like and female-like /ɪ/ and /ʊ/ formant patterns were prepared using the Klatt (1980) synthesizer. One set had constant f_0 contours identical to those used in Experiment 1, while the other set had linear falling f_0 contours. For each stimulus in the latter set, the f_0 contour was centered about the frequency of the corresponding constant f_0 stimulus, beginning 10% above and ending 10% below that frequency. Instead of the Peterson & Barney (1952) formant frequency values used for the stimuli in Experiment 1, F_1, F_2, and F_3 values were modeled after adult male and adult female averages reported by Yang (1990) for American English productions of /ɪ/ and /ʊ/, which were considered to be more typical of the dialects spoken by the subject population used for the present experiment. These values were (in Hz): 398, 2035, and 2690 (male-like /ɪ/); 448, 2394, and 3029 (female-like /ɪ/); 438, 1321, and 2366 (male-like /ʊ/); and 479, 1489, and 2828 (female-like /ʊ/). F_4 and F_5 were fixed at 3300 Hz and 3850 Hz. The stimuli had the same durations and amplitude envelopes as those used in Experiment 1.

3.1.1. *Procedure and subjects*

The procedure, facilities, and equipment were the same as in Experiment 1, except that the stimuli were presented in two levels of background noise as well as in quiet. The noise band ranged from 0–4.9 kHz and had a flat spectrum up to 600 Hz beyond which the power density fell by 6 dB/octave. The noise and vowel segments had simultaneous onsets and offsets. In the quiet condition, the vowel stimuli were presented binaurally at 72 dB SPL. In the two noise conditions, the combined output signal (vowel plus noise) was also 72 dB SPL, and the signal-to-noise (S/N) ratios were 6 dB and 3 dB. (For convenience, *S/N ratio* will be used to refer to the variable encompassing the quiet as well as the two noise conditions.) The design of Experiment 2 thus included 144 stimulus/presentation conditions (2 vowel categories $\times$ 2 sexes $\times$ 6 f_0s $\times$ 2 f_0 contours $\times$ 3 S/N ratios). S/N ratio was a between-subjects variable; all others were within-subjects variables. As in Experiment 1, subjects identified 10 randomized blocks of the stimulus set.

Subjects were drawn from the same pool as that used for Experiment 1. All were native speakers of English and reported having normal hearing. Sixteen subjects served in the quiet condition, 14 in the 6 dB S/N condition, and 18 in the 3 dB S/N condition, and no subject served in more than one of these conditions.

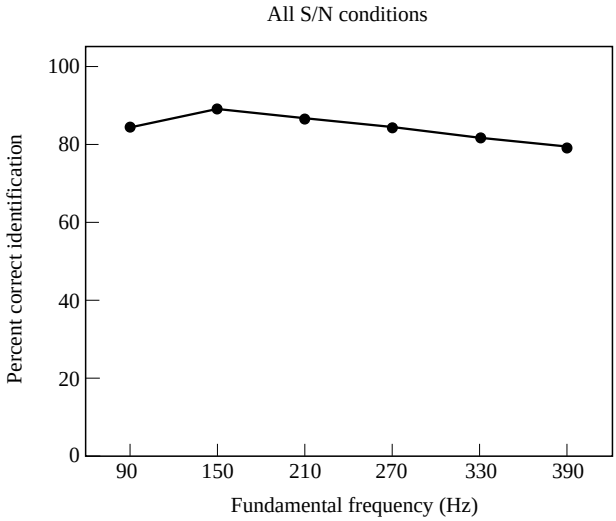

Figure 6. Mean correct identification for all vowel stimuli and S/N conditions as a function of f_0 (Experiment 2).

3.2. *Results and discussion*

Fig. 6 shows the average correct labeling performance for all vowel stimuli and S/N conditions as a function of f_0. As in Experiment 1, there was an overall monotonic decline in performance starting at an f_0 value of 150 Hz. Figs 7, 8, and 9 display the results separately for the male-like and female-like formant patterns of each vowel category, f_0 contour, and S/N condition. A repeated-measures analysis of variance on the percentage correct labeling responses yielded significant main effects of f_0 [$F(5, 225) = 24.65$, $p < 0.0001$], vowel category (/ɪ/ > /ʊ/) [$F(1, 45) = 25.00$, $p < 0.0001$], sex (male-like > female-like) [$F(1, 45) = 20.16$, $p < 0.0001$], f_0 contour (time-varying > constant) [$F(1, 45) = 20.56$, $p < 0.0001$], and S/N ratio (quiet > 6 dB S/N > 3 dB S/N) [$F(2, 45) = 16.82$, $p < 0.0001$]. (The ">" sign here means that labeling performance was better in the left-indicated condition.)

There were also significant interactions between: f_0 and vowel category [$F(5, 225) = 48.82$, $p < 0.0001$], f_0 and sex [$F(5, 225) = 8.84$, $p < 0.0001$], f_0 and f_0 contour [$F(5, 225) = 17.07$, $p < 0.0001$], f_0 and S/N ratio [$F(10, 225) = 4.58$, $p < 0.0001$], vowel category and sex [$F(1, 45) = 44.62$, $p < 0.0001$], f_0, vowel category, and sex [$F(5, 225) = 23.60$, $p < 0.0001$], f_0, vowel category, and f_0 contour [$F(5, 225) = 8.62$, $p < 0.0001$], f_0, sex, and f_0 contour [$F(5, 225) = 3.74$, $p < 0.01$], f_0, vowel category, and S/N ratio [$F(10, 225) = 9.73$, $p < 0.0001$], f_0, sex, and S/N ratio [$F(10, 225) = 4.34$, $p < 0.0001$], vowel category, sex, and f_0 contour [$F(1, 45) = 15.44$, $p < 0.001$]. To summarize the most important of these interactions, labeling accuracy tended to: (a) decrease as a function of f_0 for /ʊ/ but to increase for /ɪ/, (b) decrease more as a function of f_0 for female-like formant patterns than for male-like ones, (c) be higher for the male-like than for the female-like /ʊ/, but higher for the female-like than for the male-like /ɪ/, (d) decrease more as a function of f_0 at the more favorable S/N ratios, (e) decrease more as a function of f_0 for the constant than for the time-varying f_0 contours, (f) be enhanced by the time-varying f_0 contours only in those conditions where performance declined at higher f_0 values. It

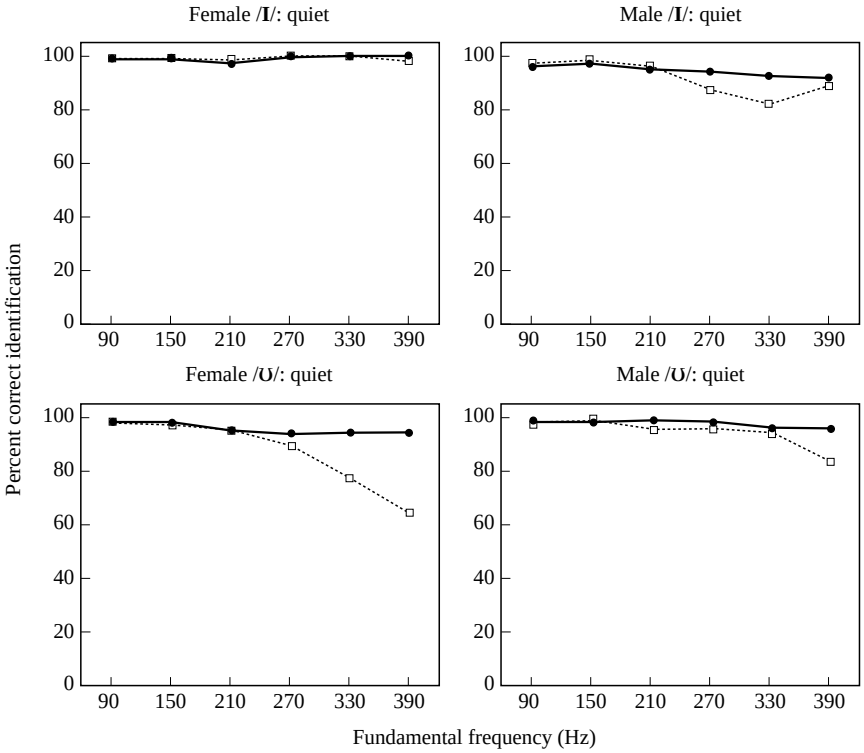

Figure 7. Mean correct identification for each vowel category and f_0 contour, in quiet (Experiment 2). —●—, Falling f_0; ---□---, level f_0.

is also worth noting that the perceptual advantage conferred by time-varying f_0 contours did not differ significantly across S/N ratios (i.e., there was no interaction between f_0, f_0 contour, and S/N ratio [$F(10, 225) = 0.59$, $p > 0.8$]).

How do these results bear on the three accounts—formant-f_0 mismatching, biasing, and undersampling—of the effect of f_0 on vowel identifiability? In respect to the formant-f_0 mismatching account, the present results were generally unfavorable. First, in conditions where identification accuracy declined with higher average f_0s, the perceptual advantage for stimuli with time-varying f_0s was not predicted by this account. Second, in most of the remaining conditions, identification accuracy was greatest at the highest average f_0 (i.e., 390 Hz) for both male-like and female-like vowels, which is directly contrary to the predictions of the formant-f_0 mismatching account.

Certain patterns in the data appear to favor the biasing over the undersampling account.[7] In particular, for the /ɪ/ formant patterns, identification accuracy tended to increase as a function of f_0, not decrease as predicted by the undersampling hypothesis. Two other characteristics of the /ɪ/ results also appear to support the biasing account. First, the increase in labeling accuracy as a function of f_0 was most

[7] Given the nature of the biasing account, analyses based on signal detection theory, which permits independent assessment of the effects of sensitivity and response bias, would appear to be applicable here. However, such analyses are problematic when individual subject performance is at 100% accuracy, as was the case in many of the conditions in Experiments 1 and 2.

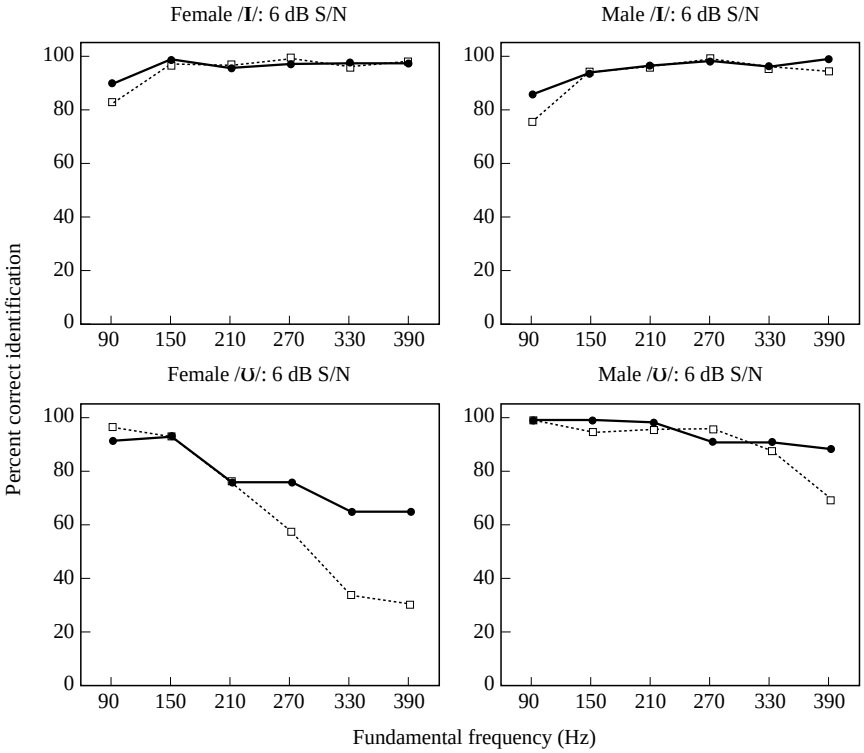

Figure 8. Mean correct identification for each vowel category and f_0 contour, at 6 dB S/N (Experiment 2). —●—, Falling f_0; ---□---, level f_0.

marked at the less favorable S/N ratios. Although this was no doubt partly attributable to the elimination of ceiling effects which would tend to obscure such differences (see results of Experiment 1), the results for the male-like /ɪ/ showed a decrease in labeling accuracy as a function of f_0 in quiet but an increase as a function of f_0 in both noise conditions. This suggests that as the /ɪ/ tokens become more ambiguous at less favorable S/N ratios, listeners are more likely to show a biasing effect of high f_0s toward more /ɪ/ responses. Second, the undersampling account predicts that stimuli with constant f_0s will be misidentified more often than those with time-varying f_0s (at least at higher average f_0s). This was invariably true for the /ʊ/ formant patterns. However, for the /ɪ/ formant patterns, this prediction was clearly confirmed only for the male-like stimuli presented in quiet (which, as noted, was also the only /ɪ/ condition in which labeling accuracy declined with f_0).

In view of the results for the /ɪ/ formant patterns, is it likely that the decline in labeling accuracy at higher f_0s that was consistently observed for the /ʊ/ formant patterns merely reflects a biasing effect toward /ɪ/ responses when a stimulus is relatively ambiguous? The answer appears to be "no", even if the biasing account may explain some portion of this decline. The reason is that, whenever a decline in labeling accuracy with higher f_0s was observed, the effect was significantly reduced (i.e., performance remained more accurate) when the f_0 contour was time-varying rather than constant. As discussed earlier, such an advantage for the time-varying stimuli is consistent with the undersampling account but is not predicted by either

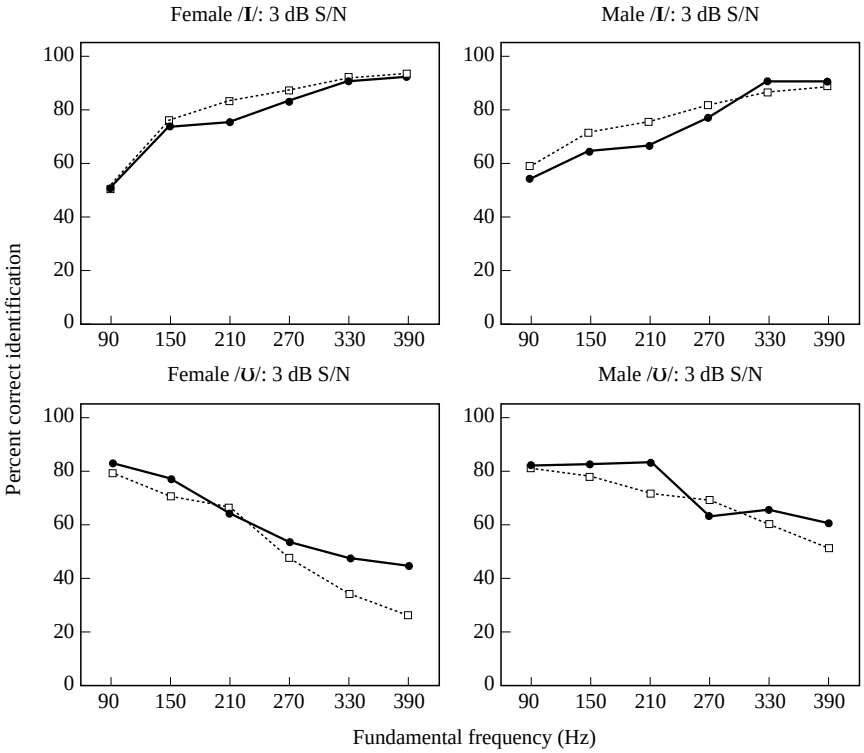

Figure 9. Mean correct identification for each vowel category and f_0 contour, at 3 dB S/N (Experiment 2). —●—, Falling f_0; ---□---, level f_0.

the biasing or formant-f_0 mismatching account. Recall also that the size of the perceptual advantage for time-varying f_0 contours was not significantly affected by the S/N ratio. This suggests that the advantage for time-varying stimuli is not itself the result of an f_0 biasing effect, since such an effect would be expected to be greater at low S/N ratios where the stimuli are more ambiguous. Thus, although certain patterns in the identification data favored the biasing account of the effect of f_0, there was also clear evidence of a detrimental effect of undersampling. This supports the main hypothesis under test in the present study.

One other aspect of the labeling results deserves comment. As in Experiment 1, female-like formant patterns were not consistently identified more accurately than male-like ones; rather female-like patterns were more accurately labeled in the case of /ɪ/ but less accurately labeled in the case of /ʊ/. In Experiment 2 (though not in Experiment 1) this cross-over interaction between vowel category and sex was statistically significant. Such a result might appear to be incompatible with the assumption of the sufficient contrast hypothesis that female formant patterns are more dispersed in the vowel space than their male counterparts and are therefore less confusable, all else being equal. However, the direction of interaction between vowel category and sex was, in fact, consistent with the claim that greater dispersion yields more accurate identification. In both Experiments 1 and 2, the male-like /ʊ/ and the female-like /ɪ/ were more distant in the $F_1 \times F_2$ plane from the centroid of the four formant patterns than were the female-like /ʊ/ and the male-like /ɪ/,

respectively. Thus, the restriction to just two vowel categories in the present study did not permit a fair test of the claim that female formant patterns are *in general* more identifiable than male ones.

3.3. *Effects of f_0 on auditory representations of the vowel stimuli*

According to the undersampling account of the effects of f_0 on vowel identifiability, the sparser distribution of harmonics at high f_0s yields poorer definition of the peaks and valleys in the spectral envelope, creating a more ambiguous stimulus. Given the significant interaction between vowel category and f_0 observed in Experiments 1 and 2, it appears that certain vowel categories are more susceptible than others to adverse effects of undersampling.

In order to investigate the likely auditory effects of increasing f_0 for the vowel categories used in this study, tokens of the two categories were input to LUTEar (O'Mard, Hewitt & Meddis, 1994), a computational model of peripheral auditory processing. Following pre-emphasis intended to replicate the transfer function of the outer and middle ear, the model uses a bank of gammatone filters to simulate spectral analysis performed by the basilar membrane. The gammatone filter has amplitude characteristics similar to the Roex filter commonly used to model the human peripheral auditory filter shape (Schofield, 1985), but has the added advantage of closely matching the impulse response of mammalian primary auditory neurons (de Boer & de Jongh, 1978), and it can be economically implemented. Once the channel density and the frequency span are chosen (our implementation used 89 channels between 100 Hz and 4800 Hz), filters are spaced in equal equivalent-rectangular-bandwidth (ERB) steps between the frequency extremes. Following basilar membrane simulation, LUTEar incorporates the Meddis hair cell model (Meddis, Hewitt & Shackleton, 1990), which simulates the neurotransmitter production process thought to occur in the hair cells. In our implementation, the hair cell model corresponded to the medium spontaneous rate fibers in Meddis *et al.* (1990). The output for each channel is a string of neural spike probabilities over time that closely matches the output of real hair cells (Meddis, 1986, 1988). The final stage of LUTEar averages these spike probabilities within each channel over the duration of the stimulus and converts to a measure of spikes/second.

Fig. 10 displays the LUTEar output representations for the female-like /ʊ/ and /ɪ/ stimuli used in Experiment 2 at constant f_0 values of 150 Hz, 270 Hz, and 390 Hz. Spikes/second are plotted as a function of channel frequency (scaled with equal ERB steps). The output representations clearly reflect the influence of both the spectral envelopes (e.g., locations of formant peaks) and the harmonic fine structure of the stimuli. Notice that the /ɪ/ and /ʊ/ representations are fairly similar in the region of F_1 and also in the region above about 2500 Hz; they differ mainly in the intermediate region owing to the fact that /ʊ/, but not /ɪ/, has a formant (F_2) near 1500 Hz. Thus, to reliably identify an /ʊ/ stimulus in the present study, the listener presumably must detect a peak in the spectral envelope in that intermediate region. It may be seen that as f_0 is increased, the shape of the spectral envelope becomes less well defined in the auditory representation while the harmonic fine structure becomes more salient. In the case of /ʊ/, the effect of raising f_0 is to make the F_2 peak near 1500 Hz considerably less distinctive, since the level of auditory response to the harmonic closest to that peak becomes more

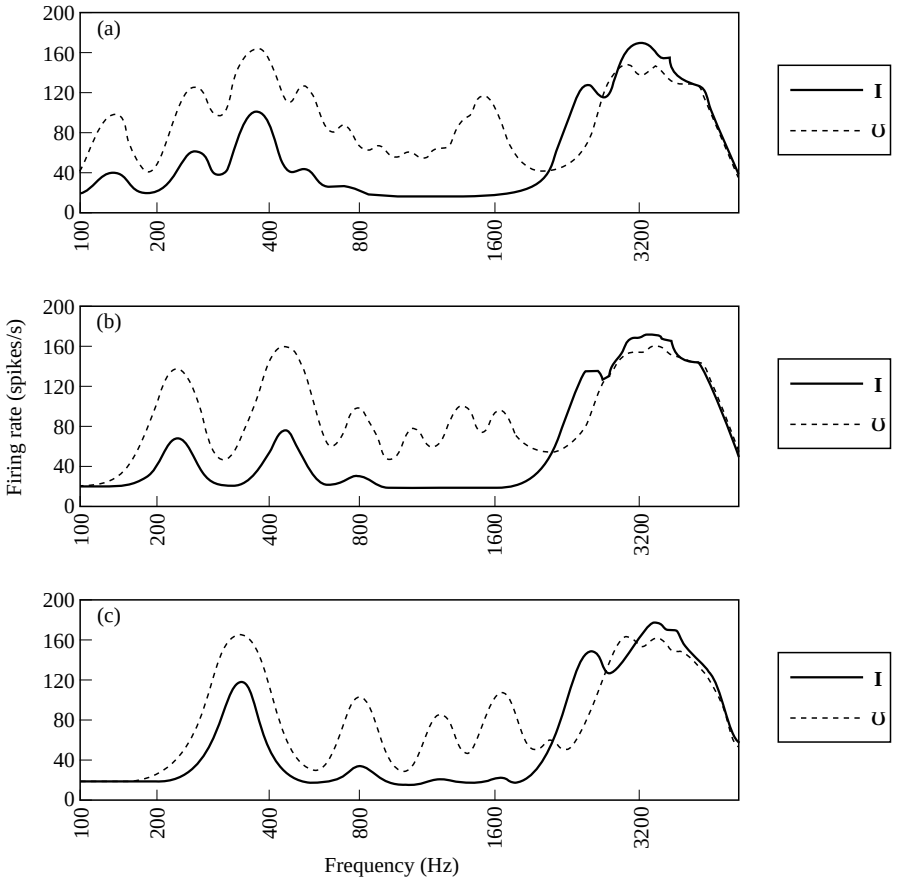

Figure 10. LUTEar output representations for three constant-f_0 female-like /ʊ/ (----) and /ɪ/ (—) stimuli used in Experiment 2. (a) $f_0 = 150$ Hz; (b) $f_0 = 270$ Hz; (c) $f_0 = 390$ Hz.

similar to that of harmonics outside the bandwidth of F_2. In the case of /ɪ/, the effect of raising f_0 appears to be less deleterious because the distinctive acoustic property of this vowel *vis-à-vis* /ʊ/ is the *absence* of a formant peak near 1500 Hz. A sparser harmonic sampling of the spectral envelope would not be expected to yield misleading cues that such a peak is actually present in an /ɪ/ stimulus. This asymmetry between /ɪ/ and /ʊ/ in the acoustic/auditory effects of raising f_0 offers one plausible explanation (in addition to the biasing account) for the significant interaction in Experiment 2 between vowel category and f_0.

4. Summary and concluding remarks

The present study was designed to test one explanation for the cross-language tendency of female talkers to exhibit greater between-category vowel dispersion than male talkers. That explanation—the sufficient contrast hypothesis—assumes that without the compensatory effect of greater dispersion, the higher f_0s of female talkers would yield reduced identifiability of vowels because of sparser harmonic

sampling of spectral envelopes. The specific question addressed here was whether, all else being equal, higher f_0s do indeed have the assumed adverse effect on vowel identifiability.

In both Experiments 1 and 2, the overall effect of increasing f_0, at least in the region beyond 150 Hz, was to reduce vowel labeling accuracy. The fact that peak performance did not occur at the very lowest f_0 value (i.e., 90 Hz) is consistent with earlier findings (Morozov, 1965; Ryalls & Lieberman, 1982). When labeling results were broken down by vowel category, it became clear that most of the decline in performance at higher f_0s occurred for /ʊ/ stimuli; for /ɪ/ stimuli, performance was generally either flat or improved with increasing f_0.

In addition to the spectral undersampling account of the effect of f_0 on vowel identifiability, two other accounts were considered. According to the formant-f_0 mismatching account, vowel identifiability declines as the combination of f_0 and formant frequencies departs from typical values for male and female talkers. The biasing account claims that when vowel identity is ambiguous (because of, e.g., poor resolution of spectral peaks and valleys or presence of background noise), listeners may judge an item with a higher f_0 as /ɪ/ on the basis of an abstract analogy between the relatively high f_0 and the relatively high upper formant frequencies of /ɪ/. To evaluate the three accounts of the effect of f_0, Experiment 2 included both constant and time-varying f_0 conditions. Because the mean f_0 values of the time-varying stimuli equalled those of the corresponding constant f_0 stimuli, neither the formant-f_0 mismatching account nor the biasing account predicts that performance in the time-varying f_0 conditions will significantly exceed that in the constant f_0 conditions. In contrast, the spectral undersampling account predicts that any performance decline at higher constant f_0s will be reduced or eliminated with time-varying f_0s because the changing harmonic values will sweep through portions of the spectral envelope providing greater definition of peaks and valleys. The results of Experiment 2 were consistent with the prediction of the spectral undersampling account: in every instance where performance declined significantly at higher constant f_0s (i.e., in all of the /ʊ/ conditions and in the condition where the male-like /ɪ/ was presented in quiet), the decline was significantly reduced in the corresponding time-varying conditions. Thus, although the biasing account may be necessary to explain the tendency for /ɪ/ labeling performance to improve at higher f_0s, it offers at best only a partial explanation for the decline in /ʊ/ labeling performance at higher f_0s.

If the statistical interaction between vowel category and f_0 is partly explained by the biasing account, another contributing factor appears to be a difference in the way higher f_0s affect the auditory representations of /ʊ/ and /ɪ/. A main distinguishing property of the former vowel *vis-à-vis* the latter is the presence of a formant peak (f_2) near 1500 Hz. Fig. 10 showed that this peak becomes auditorily less salient as f_0 increases. Because /ɪ/ is characterized by an absence of a spectral peak in the 1500 Hz region, a negative property more or less unaffected by increasing f_0, an interaction between vowel category and f_0 is the expected outcome. In any case, it may be concluded on the basis of the present study that the effects of raising f_0 on vowel identifiability are not uniformly distributed across vowel categories.

The results of the present study support the assumption of the sufficient contrast hypothesis that higher f_0s may interfere with vowel identification because of poorer

definition of spectral envelopes. Thus, the findings suggest that the greater between-category dispersion of female vowels may be plausibly explained as a means of offsetting the deleterious effects on vowel identifiability of (typically) higher f_0s. Before accepting such a conclusion, we must briefly address two possible objections to the sufficient contrast hypothesis. One of these, raised by Goldstein (1980), citing the work of Chen (1980), is that talkers tend to raise f_0 when trying to increase their speech intelligibility. A related observation is that f_0 is often relatively high on sentence constituents that carry stress or pitch accent. If higher f_0s yield reduced identifiability, such observations would appear to be troublesome for the sufficient contrast hypothesis, since focused constituents usually carry a high information load. A closer examination of f_0 contours associated with focused constituents shows, however, that the higher f_0s are almost always part of relatively large f_0 excursions (see, e.g., f_0 contours in Pierrehumbert, 1980). Our findings demonstrate that large variations in f_0 within, say, a syllable, greatly reduce the negative effect of high average f_0s. Thus, the large f_0 excursions associated with focused constituents apparently serve not only to highlight these constituents but also to counteract any adverse perceptual effects of raising f_0.

A second possible objection to the sufficient contrast hypothesis is that if dynamic f_0 contours mitigate the spectral undersampling effect of high f_0s, then there would appear to be little motivation for female talkers to produce a more dispersed vowel set than males. This objection assumes that even relatively small variation in f_0 will suffice to improve vowel identification. However, Sundberg (1977b) failed to find a beneficial effect on vowel identifiability when f_0 was varied by $\pm 2.93\%$ to simulate vibrato. It is fair to say that only a subset of vowels produced in natural speech exhibit f_0 changes of the magnitude used in the present study (an 18% drop from the onset value). Therefore, f_0 changes per se provide only a partial solution to the problem of potential vowel confusability at high f_0 values.

Some limitations of the present study must be acknowledged. The restriction of the stimulus set to only two vowel categories was intended to eliminate height-related biasing effects of f_0, which might contribute spuriously to identification errors. Although there are clear advantages to this design, there are also disadvantages. First, the restricted set of vowel categories makes it difficult to generalize the findings to the rest of the vowel inventory. Second, it might be argued that a two-category identification task reduces to a kind of simple psychophysical discrimination task that is a poor model of natural phonetic categorization. Third, as noted earlier, the present design was ill-suited to testing one central claim of the sufficient contrast hypothesis, namely, that the greater between-category dispersion of female vowel formant patterns yields greater identification accuracy, all else being equal.

In our view, these limitations do not undermine the basic conclusions drawn from the study. Concerning the issue of generalizability, the deleterious effect of higher f_0s observed in Experiments 1 and 2 (particularly at the more favorable S/N ratios where putative biasing effects were minimal) essentially replicates results of studies in which multiple vowel categories were used (e.g., Morozov, 1965; Ryalls & Lieberman, 1982; Gottfried & Chew, 1986). Although some portion of the identification errors reported in these earlier studies may be attributable to height-related biasing effects of f_0, it is unlikely that the entire influence of f_0 can be thus explained. Moreover, it is not a requirement of the sufficient contrast

hypothesis that all vowel categories are perceptually degraded at high f_0s, only that a significant subset of them are.

There is no doubt that a two-vowel category labeling task differs in significant ways from vowel identification under less restricted conditions. However, in the present study each vowel category included twelve different stimulus tokens (2 formant patterns $\times$ 6 f_0s), and so it is very unlikely that listeners were performing some kind of simple discrimination rather than a genuine phonetic categorization task. Additional evidence on this point is the fact that a changing f_0 contour reduced the deleterious effect of higher f_0s on labeling accuracy. While this result makes sense on the assumption that listeners were identifying vowel categories on the basis of estimates of formant location or overall spectral shape, it is unclear how the result might arise from processes of simple psychophysical discrimination.

The sufficient contrast hypothesis involves two claims: (1) that the higher f_0s typical of female talkers result in vowel spectral envelopes that are harmonically undersampled and thus harder to identify, all else being equal, and (2) that the greater between-category dispersion of female vowels helps to offset the otherwise deleterious perceptual effects of the spectral undersampling. Experiments 1 and 2 were designed to test only the first claim. The limited sample of formant patterns used in this study did not permit a fair test of the second claim; such a test would require a more complete set of peripheral vowel categories so that the greater dispersion of female vowels can be acoustically realized. Despite the absence of direct evidence from this study, we think that the second claim—that greater between-category dispersion reduces confusability—has a high degree of face validity.

Additional studies using a larger sample of vowel categories are now being planned in order to test both claims comprising the sufficient contrast hypothesis.

This work was supported by a grant from the Advanced Research Program of the Texas Board of Coordination to the first two authors, a grant from the National Institutes of Health (No. DC00427) to the first author, and grants from the Council for Research in the Humanities and the Social Sciences, Sweden (No. F 149/91) and the National Science Foundation (BNS-9011894) to the second author.

References

Bennett, S. (1981) Vowel formant frequency characteristics of preadolescent males and females, *Journal of the Acoustical Society of America,* **69**, 231–238.

Bennett, S. & Weinberg, B. (1979a) Sexual characteristics of preadolescent children's voices, *Journal of the Acoustical Society of America,* **65**, 179–189.

Bennett, S. & Weinberg, B. (1979b) Acoustic correlates of perceived sexual identity in preadolescent children's voices, *Journal of the Acoustical Society of America,* **66**, 989–1000.

Chambers, J. K. (1992) Linguistic correlates of gender and sex, *English World-Wide,* **13**, 173–218.

Chen, F. R. (1980) *Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level.* S.M. Thesis, M.I.T.

Cleveland, T. F. (1977) Acoustic properties of voice timbre types and their influence on voice classification, *Journal of the Acoustical Society of America,* **61**, 1622–1629.

Darwin, C. R. (1871) *The descent of man, and selection in relation to sex.* London: Murray.

de Boer, E. & de Jongh, H. R. (1978) On cochlear encoding: potentialities and limitations of the reverse-correlation technique, *Journal of the Acoustical Society of America,* **63**, 115–135.

Fant, G. (1966) A note on vocal tract size factors and non-uniform F-pattern scalings, *STL-QPSR* **4**/1966, 22–30.

Fant, G. (1973) *Speech sounds and features.* Cambridge, MA: M.I.T. Press.

Fant, G. (1975) Non-uniform vowel normalization, *STL-QPSR* **2-3**/1975, 1–19.
Fujisaki, H. & Kawashima, T. (1968) The roles of pitch and higher formants in the perception of vowels, *IEEE Transactions on Audio and Electroacoustics,* **AU-16**, 73–77.
Goldstein, U. (1980) *An articulatory model for the vocal tracts of growing children.* Doctoral dissertation, M.I.T.
Gottfried, T. L. & Chew, S. L. (1986) Intelligibility of vowels sung by a countertenor, *Journal of the Acoustical Society of America,* **79**, 124–130.
Helmholtz, H. L. F. (1885/1954) *On the sensations of tone as a physiological basis for the theory of music.* (Translated and with an additional Appendix by A. J. Ellis. Reprinted in 1954). New York: Dover.
Henton, C. (1992a) The abnormality of male speech. In *New departures in linguistics* (G. Wolf, editor), pp. 27–55. New York: Garland Publishing, Inc.
Henton, C. (1992b) Acoustic variability in the vowels of female and male speakers. Paper presented at the 123rd meeting of the Acoustical Society of America, Salt Lake City, May 13.
Hoemeke, K. A. & Diehl, R. L. (1994) Perception of vowel height: the role of F_1–F_0 distance, *Journal of the Acoustical Society of America,* **96**, 661–674.
Hollien, H. & Paul, P. (1969) A second evaluation of the speaking fundamental frequency characteristics of post-adolescent girls, *Language and Speech,* **12**, 119–124.
Hollien, H. & Shipp, T. (1972) Speaking fundamental frequency and chronological age in males, *Journal of Speech and Hearing Research,* **15**, 155–159.
Howie, J. & Delattre, P. (1962) An experimental study of the effect of pitch on the intelligibility of vowels, *The NATS Bulletin,* **4**, 6–9.
Klatt, D. H. (1980) Software for a cascade/parallel formant synthesizer, *Journal of the Acoustical Society of America,* **67**, 971–995.
Kuhl, P. K., Williams, K. A. & Meltzoff, A. N. (1991) Cross-modal speech perception in adults and infants using nonspeech auditory stimuli, *Journal of Experimental Psychology*: *Human Perception and Performance,* **17**, 829–840.
Labov, W. (1990) The intersection of sex and social class in the course of linguistic change, *Language Variation and Change,* **2**, 205–254.
Lehiste, I. & Meltzer, D. (1973) Vowel and speaker identification in natural and synthetic speech, *Language and Speech,* **16**, 356–364.
Lehiste, I. & Peterson, G. E. (1961) Some basic considerations in the analysis of intonation, *Journal of the Acoustical Society of America,* **33**, 419–425.
Mattingly, I. G. (1966) Speaker variation and vocal-tract size. Paper presented at the 71st meeting of the Acoustical Society of America, Boston, June 1, 1966.
Meddis, R. (1986) Simulation of mechanical to neural transduction in the auditory receptor, *Journal of the Acoustical Society of America,* **79**, 702–711.
Meddis, R. (1988) Simulation of auditory-neural transduction: further studies, *Journal of the Acoustical Society of America,* **83**, 1056–1063.
Meddis, R., Hewitt, M. J. & Shackleton, T. M. (1990) Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse, *Journal of the Acoustical Society of America,* **87**, 1813–1816.
Morozov, V. P. (1965) Intelligibility in singing as a function of fundamental voice pitch, *Soviet Physics-Acoustics,* **10**, 279–283.
Nelson, H. D. & Tiffany, W. R. (1968) The intelligibility of song: research results with a new intelligibility test, *The NATS Bulletin,* **25**, 22–33.
Nordström, P.-E. (1977) Female and infant vocal tracts simulated from male area functions, *Journal of Phonetics,* **5**, 81–92.
Nordström, P.-E. & Lindblom, B. (1975) A normalization procedure for vowel formant data. Paper 212 presented at the International Congress of Phonetic Sciences, Leeds, August 1975.
Ohala, J. J. (1984) An ethological perspective on common cross-language utilization of f_0 of voice, *Phonetica,* **41**, 1–16.
O'Mard, L. P., Hewitt, M. J. & Meddis, R. (1994) LUTEar core routines library: a flexible auditory simulation development computing system. Loughborough, U.K.
Peterson, G. E. & Barney, H. L. (1952) Control methods used in the study of the vowels, *Journal of the Acoustical Society of America,* **24**, 175–184.
Pierrehumbert, J. B. (1980) *The phonology and phonetics of English intonation.* Doctoral disseration, M.I.T.
Potter, R. K. & Steinberg, J. C. (1950) Toward the specification of speech, *Journal of the Acoustical Society of America,* **22**, 807–820.
Ryalls, J. H. & Lieberman, P. (1982) Fundamental frequency and vowel perception, *Journal of the Acoustical Society of America,* **72**, 1631–1634.
Sachs, J., Lieberman, P. & Erickson, D. (1973) Anatomical and cultural determinants of male and female

speech. In *Language attitudes*: *current trends and prospects* (R. W. Shuy & R. W. Fasold, editors), pp. 74–84. Washington, D.C.: Georgetown University Press.
Schofield, D. (1985) *Visualizations of speech based on a model of the peripheral auditory system.* NPL Report DITC 62/85.
Slawson, A. W. (1968) Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency, *Journal of the Acoustical Society of America,* **43**, 87-1-1.
Smith, L. A. & Scott, B. L. (1980) Increasing the intelligibility of sung vowels, *Journal of the Acoustical Society of America,* **67**, 1795–1797.
Stumpf, C. (1926) *Die Sprachlaute.* Berlin: Springer.
Sundberg, J. (1977a) Studies of the soprano voice, *Journal of Research on Singing,* **1**, 25–35.
Sundberg, J. (1977b) Vibrato and vowel identification, *Archives of Acoustics,* **2**, 257–266.
Traunmüller, H. (1984) Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels, *Speech Communication,* **3**, 49–61.
Yang, B. (1990) A comparative study of normalized English and Korean vowels. Unpublished doctoral dissertation. University of Texas at Austin.
Yang, B. (1992) An acoustical study of Korean monophthongs produced by male and female speakers, *Journal of the Acoustical Society of America,* **91**, 2280–2283.