Bandwidth of spectral resolution for two-formant synthetic vowels and two-tone complex signals[a)]

Qiang Xu,[b)] Ewa Jacewicz, and Lawrence L. Feth[c)]
Department of Speech and Hearing Science, Ohio State University, Columbus, Ohio 43210

Ashok K. Krishnamurthy
Department of Electrical Engineering, Ohio State University, Columbus, Ohio 43210

(Received 20 November 2002; accepted for publication 22 August 2003)

Spectral integration refers to the summation of activity beyond the bandwidth of the peripheral auditory filter. Several experimental lines have sought to determine the bandwidth of this ‘‘supracritical’’ band phenomenon. This paper reports on two experiments which tested the limit on spectral integration in the same listeners. Experiment 1 verified the critical separation of 3.5 bark in two-formant synthetic vowels as advocated by the center-of-gravity (COG) hypothesis. According to the COG effect, two formants are integrated into a single perceived peak if their separation does not exceed approximately 3.5 bark. With several modifications to the methods of a classic COG matching task, the present listeners responded to changes in pitch in two-formant synthetic vowels, not estimating their phonetic quality. By changing the amplitude ratio of the formants, the frequency of the perceived peak was closer to that of the stronger formant. This COG effect disappeared with larger formant separation. In a second experiment, auditory spectral resolution bandwidths were measured for the same listeners using common-envelope, two-tone complex signals. Results showed that the limits of spectral averaging in two-formant vowels and two-tone spectral resolution bandwidth were related for two of the three listeners. The third failed to perform the discrimination task. For the two subjects who completed both tasks, the results suggest that the critical region in vowel task and the complex-tone discriminability estimates are linked to a common mechanism, i.e., to an auditory spectral resolving power. A signal-processing model is proposed to predict the COG effect in two-formant synthetic vowels. The model introduces two modifications to Hermansky’s [J. Acoust. Soc. Am. **87**, 1738–1752 (1990)] perceptual linear predictive (PLP) model. The model predictions are generally compatible with the present experimental results and with the predictions of several earlier models accounting for the COG effect.
[DOI: 10.1121/1.1624066]

PACS numbers: 43.66.Ba, 43.71.Es, 43.66.Lj [DOS] Pages: 1653–1664

I. INTRODUCTION

The spectral envelopes of spoken vowels normally contain multiple peaks called formants. The interest in approximating phonetic quality by reducing the number of formants in synthetic vowels dates from the early 1950s. Delattre *et al.* (1952) experimented with back vowels whose first two formants are close in frequency and observed that the quality of these vowels is still preserved when a listener is presented with a single intermediate formant. This formant, located somewhere between the two, comprised an overall quality of the vowel, preserving its unique ‘‘color.’’ When formant separation was large, such as in front vowels, no single formant could be found that successfully approximated their quality. Delattre *et al.* (1952) concluded that an auditory mechanism must effectively average two formants which are relatively close in frequency. This phenomenon, known today as formant averaging or spectral integration, suggests that the auditory system performs an additional filtering of vowels beyond the level of the cochlea.

A further exploration of the ‘‘center of prominence’’ (or ‘‘center of gravity’’) of formant cluster in approximating phonetic quality of vowels led to the concept of the perceptually grounded ‘‘effective second formant’’ ($F2'$). $F2'$ was to substitute for both $F2$ and higher formants of a natural vowel in their two-formant approximation represented by $F1$ and $F2'$. Empirical formulas for $F2'$ computations from formant frequency values were proposed (Fant, 1959; Carlson *et al.*, 1970, 1975; Bladon and Fant, 1978). Center of gravity (COG) was understood as a possible correlate of $F2'$ (Carlson *et al.*, 1970). Further work on COG showed that when two formant peaks are separated by less than 3.5 bark, they are integrated into a single perceived peak, called the ‘‘perceptual formant’’ (Chistovich *et al.*, 1979). The next step in the conceptual development was to link the 3.5-bark ‘‘critical distance’’ with predicting $F2'$ values (Bladon, 1983), which resulted in further computational models of $F2'$ (Escudier *et al.*, 1985; Mantakas *et al.*, 1988). The model by Mantakas *et al.* was later applied to predicting the

[a)] Portions of this work have appeared in the first author’s Master’s thesis and in poster presentations given at the 3rd joint meeting of the Acoustical Society of America and Acoustical Society of Japan, 1996, Honolulu, Hawaii, and at the 133rd meeting of the Acoustical Society of America, 1997, State College, Pennsylvania.

[b)] Current address: Nokia Product Creation Center, NMP Dallas, Irving, TX 75039.

[c)] Author to whom correspondence should be addressed. Electronic mail: feth.1@osu.edu organizational trends in vowel systems of human languages (Schwartz *et al.*, 1997).

This line of research has established the validity of $F2'$ as a perceptual parameter in approximating vowel quality in two-formant models. It also brought to light the hypothesis that the auditory system performs a ''large-scale spectral integration'' over a limited frequency range of 3.5 bark (e.g., Schwartz and Escudier, 1989). This work focused necessarily on formant frequency values, assuming only implicit knowledge of the relation between the relative amplitudes of the formants. The role of formant amplitudes in spectral integration was explicitly addressed in a series of other studies (e.g., Bedrov *et al.*, 1978; Chistovich and Lublinskaja, 1979; Chistovich *et al.*, 1979; Chistovich, 1985). This line of research investigated the COG *effect*, i.e., the process restricted to a limited frequency range of 3.5 bark. Accordingly, when two existing vowel formants are perceptually integrated, the frequency of the perceived peak (''perceptual formant'') is closer to that of the stronger formant. The amplitudes of the two formants play an important role in that a change in their ratio is equivalent to a *frequency* change of a single-formant vowel which approximates their quality (Chistovich and Lublinskaja, 1979). The 3.5-bark critical distance indicates a possible limit on spectral integration in that the COG effect disappears with larger formant separation.

The present study is a continuation of the latter line of research. It first re-examines the COG effect in the traditional format of two-formant vowels to verify the role of the critical separation of 3.5 bark in spectral integration (experiment 1). It then presents experimental evidence from psychoacoustics, showing that spectral integration in two-tone complex signals occurs within the limit of 3.5-bark resolution bandwidth (experiment 2). It advocates the view that spectral integration is a fundamental property of the auditory system. Further, the 3.5-bark critical distance is used by the auditory system in general signal processing, and is not restricted to speech perception. The focus of this work is not on approximating the phonetic vowel quality but on the auditory processing of complex signals before higher-level decision processes apply, such as making phonetic decisions. The following sections sketch the background of the study.

A. The COG effect in two-formant vowels and the 3.5-bark critical distance

Chistovich and Lublinskaja (1979) reported that their two listeners matched a single-formant vowel of variable frequency (F_v) to a frequency between $F1$ and $F2$ of a two-formant reference signal, when the distance between the two formants was less than about 3.5 bark. The locus of F_v was called the ''perceptual formant.'' Depending on $A2/A1$ ratio (the ratio of the formant amplitudes), the matches fell midway between the frequencies of the two formants ($A2/A1 = 0$ dB) or were directed towards either one of the stronger formants ($A2/A1 = -20$ or 20 dB). When the separation between $F1$ and $F2$ exceeded 3.5 bark, there was no agreement as to the frequency of matching. One listener matched the single formant to one of the two actual formants and the second performed highly unreliable matches to intermediate frequencies between the two. Chistovich and Lublinskaja (1979) concluded that spectral integration fails for more than critical formant separation, i.e., for more than 3.5 bark.

Chistovich (1985) proposed a method for predicting listeners' matches to a two-formant reference vowel. The perceptual formant was defined as the centroid of the vowel spectrum weighted on the bark scale rather than absolute frequency (in hertz). However, this centroid measurement was only good for predicting results for signals with closely spaced formants. Chistovich and Lublinskaja's COG and critical distance effects were well manifested when listeners matched a single-formant signal of variable frequency to a two-formant reference. Results contrary to the COG effect were obtained, subsequently, in which multiformant vowels were used as both reference and variable signals (Beddor and Hawkins, 1990; Assmann, 1991).

Testing the interaction of frequency and amplitude in low-frequency formants predicted by the COG effect, Assmann (1991) carried out two experiments. Using additive harmonic synthesis to control for altering the levels of adjacent harmonics in the region of the formant, he manipulated the amplitude of the first two closely spaced formants in back vowels. The experimental tokens (both the matching and the reference stimuli) included six formants to achieve more natural-sounding vowels. In one experiment, listeners adjusted the frequency of both $F1$ and $F2$ in the target tokens to match the reference tokens with modified amplitudes of $F1$ and $F2$. In the second task, listeners identified the tokens with modified amplitude ratios as instances of five selected English vowels. The stimuli were presented in two conditions, with low f_0 (125 Hz) and high f_0 (250 Hz). Results from the first experiment failed to replicate the findings from the single-formant matching experiments as in Chistovich and Lublinskaja (1979). Support for the COG effect was also weak from the results of the second experiment as some shifts in vowel quality were only obtained in the high f_0 condition.

An important finding from Beddor and Hawkins' study (1990) was that the influence of the spectral envelope on perceived vowel quality was greater when spectral peaks in the low-frequency region were less pronounced. That is, no COG effect was observed for oral vowels with well-defined peaks, whose quality was perceptually determined by the frequency of $F1$. For nasal vowels with less-visible peaks and valleys, the matches were more consistent with the COG effect and listeners relied less on the frequency of $F1$. Their third experiment addressed the issue of the effect of well- versus poorly defined formants on the perceived vowel quality by manipulating formant bandwidths. The results indicated that, in perceiving vowel quality, listeners relied on formant frequency for vowels with well-defined peaks (the narrow-bandwidth condition, BW=45 Hz), whereas the combination of both frequency and bandwidth influenced the perception of vowels with poorly defined peaks (the wide-bandwidth condition, BW=150 Hz).

In summary, the results from experiments in which a single formant was adjusted to match reference signals supported the COG effect. However, in the experiment where two formants were adjusted simultaneously to match the reference signal, the results were incompatible with the pre-

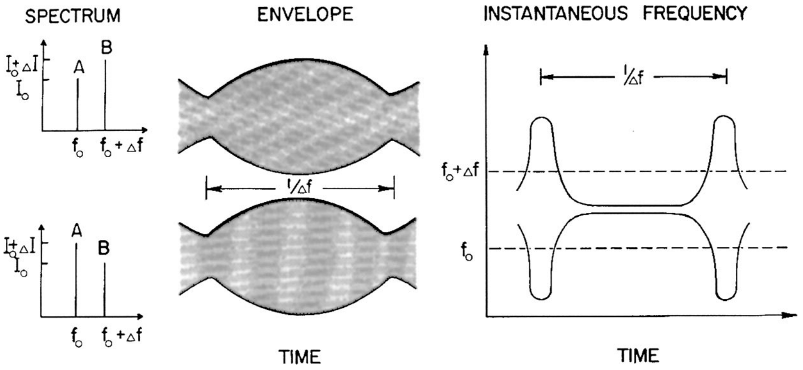

FIG. 1. Long-term spectrum, envelope, and instantaneous frequency functions for both members of the two-component complex tones that make up the simplest set of common envelope signals, as defined by Voelcker (1966a). From Feth and O'Malley (1977) with permission.

dicted pattern. This suggests that the COG effect may not directly contribute to the phonetic perception of a full vowel spectrum, playing a different role in speech signal processing. Consequently, listeners may use different strategies in responding to different types of signals in matching tasks.

B. Two-tone complex signals and the 3.5-bark resolution bandwidth

Independently of the development of both the COG effect and the spectral centroid model by the Leningrad group, Feth (1974) and Feth and O'Malley (1977) studied spectral integration using psychoacoustic signals. This research investigated the pitch of two-tone complexes, following Voelcker's unified theory of modulation (1966a, 1966b). Voelcker began with the envelope–fine-structure representation of a signal

$$S(t)=E(t)*\cos(\varphi(t)), \tag{1}$$

where $E(t)$ is the envelope of the signal, and $\varphi(t)$ is the angle. The envelope function represents slow variations in the amplitude of the signal and is often characterized as the result of amplitude modulation. The first time derivative of the angle function is often defined as the instantaneous frequency, $f(t)$. Voelcker defined sets of common envelope signals that have identical envelope functions but different fine structure. The simplest set of common envelope signals is shown in Fig. 1. For each signal, two sinusoids are separated by a small difference in frequency, ΔF, and their amplitudes differ by a small amount, ΔI. For one signal, the lower-frequency component is more intense; for the complementary signal the higher-frequency component has the higher intensity. It can be shown that these two-component signals have identical envelope functions, as shown in the middle panel of Fig. 1. The instantaneous frequency functions change in opposite directions. When the lower component is the more intense, $F(t)$ moves to lower frequencies as the envelope moves toward its minimum value. When the higher-frequency component is more intense, the frequency modulation moves toward higher frequencies. Listeners hear these fine-structure differences as small changes in the spectral pitch of the signal (Helmholtz, 1954; Jeffress, 1968).

Feth and O'Malley (1977) used the discriminability of complementary, two-component complex tones suggested by Voelcker to investigate the spectral resolving power of the auditory system. They reported that the percentage of correct responses, $P(C)$, in a 2IFC task increased from chance to approach 100% for moderate separations (1 to 3 bark) of the two-component frequencies. However, further separation of the components led to decreased discriminability when the components were apparently resolved by the peripheral auditory filter. The frequency separation for which the Voelcker signals become indiscriminable was suggested as a psychophysical estimate of auditory spectral resolving power. For each center frequency tested, this estimate is approximately 3.5 bark. Thus, in two different experimental paradigms, i.e., vowel matching and two-tone discrimination, the 3.5-bark critical separation plays a role. The critical distance can be viewed as a limit to the range of spectral integration that the auditory system can perform.

In an earlier study, Feth (1974) had proposed that the perceived pitch of a two-tone complex with components of unequal amplitude should correspond to the envelope-weighted average of the instantaneous frequency (EWAIF) of the waveform. According to the EWAIF model, the instantaneous frequency occurring where the envelope was large would contribute more to the perceived pitch of the two-tone complex. The EWAIF model accounted for experimental data using narrow-bandwidth complex tones. Its later version, the intensity-weighted average of instantaneous frequency model (IWAIF) (Anantharaman *et al.*, 1993), is a simple mathematical variation of the EWAIF that eliminates the computational difficulty of the EWAIF calculation. Anantharaman *et al.* (1993) showed that the IWAIF formulation can be transformed into the frequency domain where the IWAIF value corresponds to the COG of the signal spectrum. Given that speech signals are essentially dynamic and complex, changing in both frequency and amplitude over time, the IWAIF model can be viewed as a candidate for testing the COG effect in vowels. IWAIF model predictions were verified for the stationary two-formant speech signals used in Chistovich and Lublinskaja (1979) by Anantharaman (1998).

C. Modeling the COG effect

Hermansky (1990) proposed the perceptual linear prediction model (PLP) to account for spectral integration in vowel-like signals. The model is consistent with both the $F1-F2'$ concept and with the 3.5-bark auditory integration theory. A speech signal is first processed to produce the so-called auditory spectrum by using three processes derived from psychoacoustics: critical band frequency resolution, equal loudness compensation, and intensity-loudness compression. This auditory spectrum is then approximated by an autoregressive all-pole filter, just as in traditional LP modeling, to produce a low-dimensional representation of the speech-signal spectrum. Hermansky suggested that a fifth-order PLP model could extract at most two major peaks from the auditory spectrum, removing the minor spectral peaks. Experimental results from natural and synthetic speech signals confirmed model predictions. When two vowel formants were far apart, the PLP estimated two distinct spectral peaks. With smaller separation of formants of about 3.5 bark, these peaks merged into one peak.

In this paper, we use a modified PLP model to predict and test the COG effect, taking into account the amplitude ratio of two closely spaced formants. In so doing, we investigate whether the COG effect disappears with the 3.5-bark separation of the formants, which would be predictable by the model. Since we manipulate the amplitude ratio in testing the COG effect, we introduce two more stages as our modification to Hermansky's original model, i.e., a peak detection (stage 6) and a decision maker (stage 7). An additional modification is the use of a fourth-order LP model spectrum instead of the fifth-order LP filter because our stimuli do not have more than two formants. Consequently, the number of detected peaks at stage 6 is likely to be either one or two depending on the separation of the two formants and their relative levels. We use the modified PLP model to make predictions of the perceptual formant (F_v) as a function of formant frequency separation and amplitude ratio of a two-formant synthetic signal. We hypothesize that if only one local peak is detected, listeners match it at the F_v. If two peaks are detected, the decision about F_v is based on the frequency-level vector of two local peaks returned from the stage of peak detection. A more detailed description of the model is presented in Xu (1997).

In testing the COG effect in experiment 1, the modified PLP model is used to predict listeners' performance in matching an adjustable, single-formant vowel signal to two-formant vowel signals. The model predicts that

(1) listeners will match single-formant variable signals to the COG of two-formant reference vowels (i.e., F_v changes systematically with respect to the relative level $A2/A1$) only when the distance between $F1$ and $F2$ is smaller than the critical distance of 3.5 bark; and
(2) when the distance between $F1$ and $F2$ exceeds 3.5 bark, listeners match F_v either closer to $F1$ or $F2$ or exhibit chance performance.

Experiment 1 also investigates the effect of spectral shape (poorly defined spectral prominence versus well-defined spectral prominence) on the matching values of F_v; and how the type of excitation source (noise versus pulse train) affects the matching values of F_v.

II. EXPERIMENT 1: COG EFFECT IN TWO-FORMANT SYNTHETIC VOWELS

A. Methods

1. Listeners

Three young adults, two female and one male, participated. Two listeners were volunteer graduate students in speech and hearing science and one was an undergraduate student paid for her participation. All listeners showed normal hearing at all audiometric frequencies on two screening tests: a tone-threshold test in quiet and a frequency difference limen (DLF) test. All listeners were given extensive practice before participating in the experiment. Each participant listened 2 h per day, 5 days per week, for a period of several weeks. All experimental data were collected after their response patterns in the adaptive-tracking procedure were convergent and stable.

TABLE I. Synthesis parameters for two-formant reference vowels used in experiment 1.

Vowel	$F1$ (Hz)	$F2$ (Hz)	$F2-F1$ (bark)	BW1 (Hz)	BW2 (Hz)	Source
1	800	1300	2.5	80	80	Impulse
2	800	1300	2.5	80	80	Noise
3	800	1300	2.5	45	45	Noise
4	800	1300	2.5	150	150	Noise
5	800	1400	3.0	80	80	Noise
6	700	1400	3.6	80	80	Noise
7	700	1500	4.0	80	80	Noise
8	600	1700	5.3	80	80	Noise

2. Stimuli

Two-formant vowel-like signals (reference vowels) were generated using a formant synthesizer in parallel configuration. The excitation source was either an impulse train ($f_0 = 100$ Hz) or broadband, Gaussian noise. The $F2-F1$ distance was varied in five conditions: 1300−800, 1400−800, 1400−700, 1500−700, and 1700−600 Hz, equivalent to 2.5, 3.0, 3.6, 4.0, and 5.3 bark, respectively. For the 2.5-bark separation, four reference vowels were created with either a different excitation source or different bandwidths. For the remaining four conditions, only the formant frequency separation varied and the other parameters remained constant. A summary of the parameters is given in Table I. The relative level $A2/A1$ of the two formants was manipulated from −20 dB to +20 dB in steps of 10 dB. This resulted in five relative amplitude conditions for each reference vowel.

The output sequence of the variable single-formant signals was obtained by passing a source through a second-order IIR filter with the required $F1$ and BW1. In order to generate a sequence with two formants, the source was passed through a second IIR filter ($F2$, BW2). The output of the second filter was then scaled and added to that from the first formant. The scale factor ($A2/A1$) denotes the level of the second formant peak relative to that of the first.

Both the reference (two-formant) and the variable single-formant target signals were generated by a TDT system II D/A board controlled by a laboratory PC (166-MHz Pentium) at a sampling frequency of 20 kHz. Signal duration was 300 ms. After being shaped by a 5-ms cosine rise−fall window, the signals were attenuated by a TDT programmable attenuator and filtered by a passive analog filter (TTE), with cutoff frequency at 8 kHz. Signals were played, monaurally, through the right channel of a Sennheiser HD-414-SL headset at 65 dB SL.

The use of noise-excited signals was motivated by the following two reasons. First, when synthetic vowels are impulse-train excited, the interaction between source harmonics and formant frequencies on the effective bandwidth may become significant. The target spectral shapes of both reference and variable signals may be distorted, especially when the frequencies of formant peaks are not in alignment with source harmonics and the bandwidths of those formants are relatively narrow. This suggests that noise-excited synthetic vowels might be better suited for this kind of vowel-matching task. A second advantage of using noise excitation is to encourage subjects to focus on the formant frequency comparison rather than on the fundamental frequency.

3. Model predictions

The modified PLP model was applied to each two-formant reference vowel. Selected results for vowel # 1 (2.5-bark separation, pulse-train-excited) are displayed in Figs. 2(a)–(c), which shows spectra for three relative formant levels $A2/A1 = -20$, 0, $+20$ dB, respectively). The acoustic spectrum (dashed line), auditory spectrum (dashed-dotted line), and PLP-model spectrum (solid line) are plotted together. The maximum level for each plot is aligned at 0 dB for comparison. Essentially, the auditory spectrum reduces the contrast between the peaks and valleys of the acoustic spectrum (note, however, that the two peaks are still visible). The model smoothes the auditory spectrum further so that in each condition, there is only one peak. In (a) ($A2/A1 = -20\,\mathrm{dB}$), the single peaks of the model spectra fall in the vicinity of $F1\,(800\,\mathrm{Hz})$, while in (c) ($A2/A1 = +20\,\mathrm{dB}$), they are close to $F2\,(1300\,\mathrm{Hz})$. In (b) ($A2/A1 = 0\,\mathrm{dB}$), the F_{v} prediction is between $F1$ and $F2$ but is skewed toward the lower frequency.

The simulation results for the noise-excited vowel # 2 are similar to those for vowel # 1. The only difference is that the model spectra have even broader peaks for the noise-excited source than for the pulse-excited source (see Xu, 1997, for more details). In simulations with supracritical separation of formants (e.g., vowel # 7 with 4.0-bark separation), two peaks are detected in the 0- and +10-dB conditions. The difference between the two peaks is greater than 1 dB. Thus, the predicted value of F_{v} is the higher peak near $F2$ in both conditions. In the remaining three $A2/A1$ conditions (i.e., −20, −10, and +20 dB), only one peak is detected, and the predicted F_{v} is the frequency at that peak for each condition.

Note that the model spectrum more likely results in two peaks whenever the two formants of the original signal are of equal level (i.e., $A2/A1 = 0\,\mathrm{dB}$). For a comparison across frequency separations, the results for vowels # 2, # 6, and # 8 each with $A2/A1 = 0$ dB are shown in Fig. 3. For the subcritical separation shown in (a) (2.5 bark, vowel # 2), only one peak it is detected in the model spectrum. For the two supracritical separations depicted in (b) (3.6 bark, vowel # 6), and (c) (5.3 bark, vowel # 8), two peaks are detected successfully, although the second peak in (b) is not separated by a valley as it is in (c). In the latter plot, the second formant peak is more than 1 dB higher in level than the first. The trend observed with the increasing distance between the formants reveals that (1) the single peak in the model spectrum becomes broader as separation increases, and (2) two peaks are detected when separation exceeds the critical distance. The two peaks become sharper as the formant distance increases.

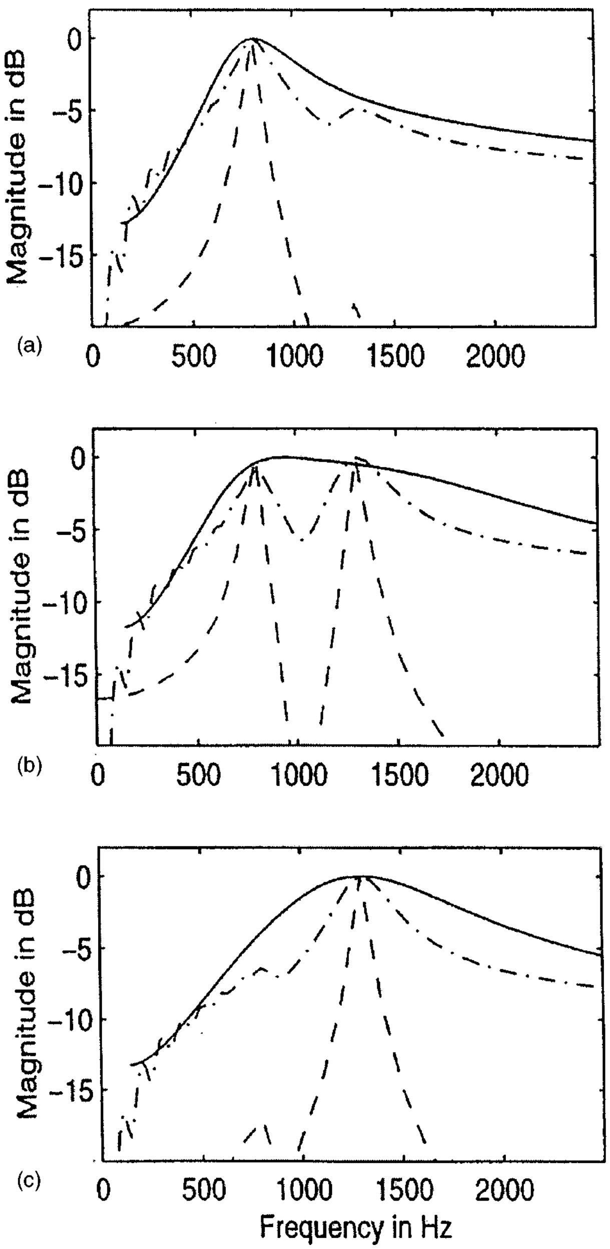

FIG. 2. Spectra derived from the proposed PLP model for a two-formant vowel # 1 (2.5-bark formant separation, pulse-train excited). Relative level ($A2/A1$) ranges from (a) −20 dB, (b) to 0 dB, (c) to +20 dB. Acoustic spectra (dashed line), auditory spectra (dashed-dotted line), and PLP model spectra (solid line) are plotted for each $A2/A1$ ratio.

These simulation results applied to a set of two-formant vowel-like signals indicate that the modified PLP model can account for the COG effect reasonably well.

4. Procedures

The listeners were asked to indicate whether the spectral pitch of the variable single-formant vowel was higher or lower than that of the two-formant reference vowel using a “double-staircase” adaptive tracking procedure (Jesteadt, 1980). The two-formant reference vowel was always pre-

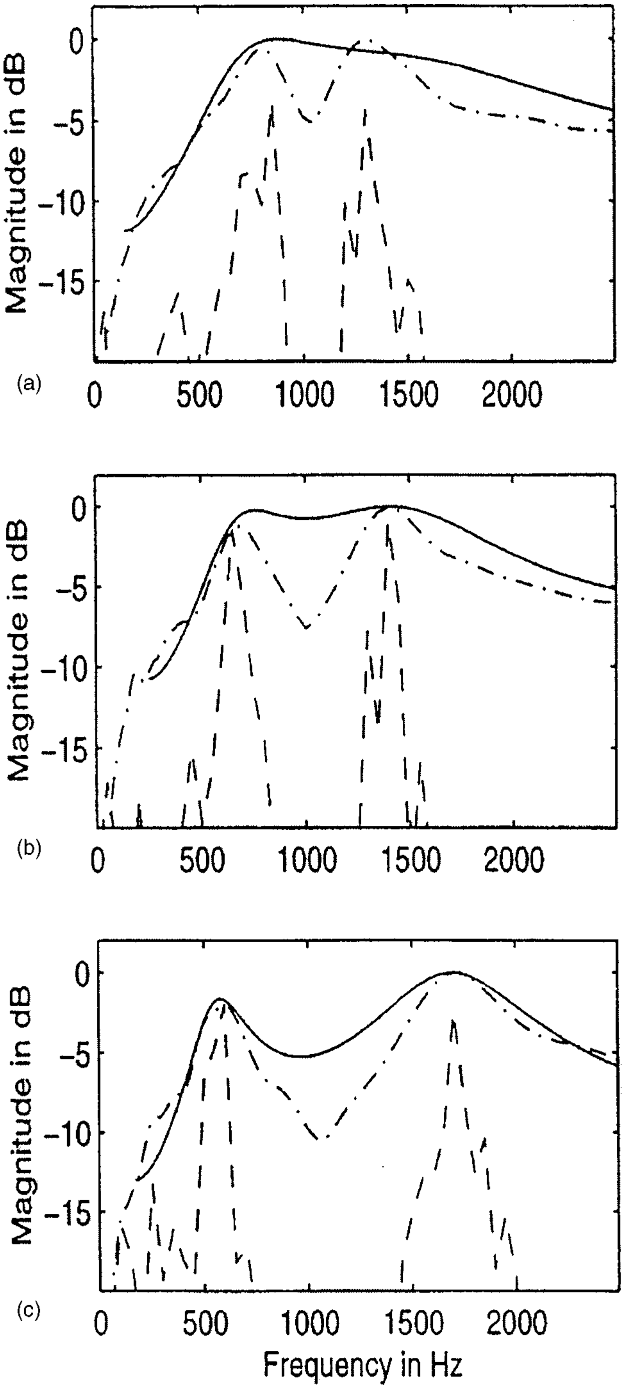

FIG. 3. Spectra derived from the proposed PLP model for two-formant vowels for three different $F2-F1$ separations and at a constant level $A2/A1=0$ dB. (a) vowel # 2 (2.5 bark); (b) vowel # 6 (3.6 bark); (c) vowel # 8 (5.3 bark). Acoustic spectra (dashed line), auditory spectra (dashed-dotted line), and model spectra (solid line) are plotted for each vowel.

sented first and the variable single-formant target vowel second, after a 600-ms silent interval. The listeners were instructed to push the right button of the mouse whenever they judged the variable signal to be higher in pitch than the reference signal. When they judged it to be lower in pitch than the reference, they pushed the left button. The double-staircase procedure maintains two adaptive tracking rules simultaneously. The "two-up, one-down" rule (Levitt, 1971) was applied to both the ascending and descending tracking sequences. Thus, the F_v estimate for the descending sequence was at a level for which the variable signal was judged higher than the reference signal on 71% of the trials. For the ascending sequence, F_v was judged lower on 29% of the trials.

In order to obtain the value of F_v at a given $A2/A1$ ratio, each of these estimates was based on eight consecutive 70-trial blocks of the double-staircase procedure. For each block, the F_v estimates for the descending and ascending sequences were first obtained by averaging the reversal points within each sequence. Then, the point-of-subjective-equality, PSE, was estimated by simply averaging the F_v estimates from the two sequences.

The use of the double-staircase adaptive procedure in lieu of the traditional matching tasks, as in early experiments on COG, was dictated by preliminary results from Lester (1996), who tested the COG hypothesis in a direct matching experiment. The listeners heard one of the two signals, alternatively. They were asked to adjust the second variable vowel to match the first vowel for quality. The results revealed considerable variability in matching data. The matches were tightly grouped around the lower formant frequency when $A2/A1=-20$ dB, but more scattered when $A2/A1=+20$ dB. Lester noted that direct matching to the "perceptual formant" was difficult even for well-trained subjects of the study. In related work (Feth *et al.*, 1996), the double-staircase adaptive procedure was used. The results showed that the dispersion of matching values at the higher frequency formant disappeared. This indicates that improvement in performance is related to eliminating inherently subjective judgments and including an objective criterion to measure the correctness of responses. Thus, the decision rules no longer continually select signals near the PSE but focus instead on points above and below it. In this case, the selection of signals by the adaptive procedure is controlled only by the observer's use of the two subjective response categories and not by how the responses relate to the objective properties of the stimuli. Consequently, the difficulty of the subjective tasks is greatly reduced.

III. RESULTS AND DISCUSSION

A. Overall results

The overall results of the experiment are shown in Figs. 4 and 5. All three listeners showed similar F_v matching values for all conditions for vowels # 1 through # 5 (2.5- and 3.0-bark separations), and # 8 (5.3 bark). The data for these conditions are therefore collapsed across the listeners. However, the results showed individual differences for vowels # 6 and # 7 (3.6 and 4.0 bark, respectively), which have their frequency separations near the 3.5-bark critical distance. These data were plotted for each subject individually in Figs. 5(b) and (c). Overall, the results indicate that frequency separation between the two formant peaks of the reference signals along with their relative levels have an effect on F_v matching values. When the frequency separation between the two formants does not exceed 3.5 bark, listeners show similar performance. The same is true for the supracritical separation of 5.3 bark. A substantial difference is observed in

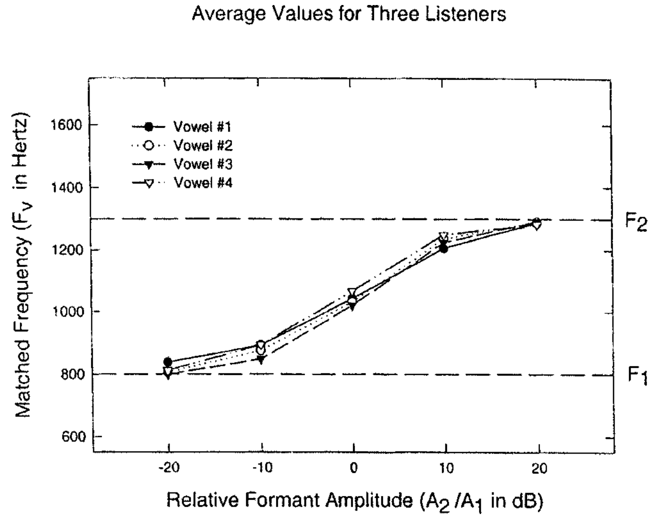

FIG. 4. Mean F_v matching values for vowels # 1 through # 4 (2.5-bark separation, variable source, variable bandwidth) with changing amplitude ratio ($A2/A1$). Dashed lines mark the locations of $F1$ and $F2$.

listeners' performance when the frequency separation approaches the "critical region" near 3.5 bark. For both 3.6- and 4.0-bark separations, variable patterns of responses were obtained.

The data were subjected to a two-way analysis of variance (ANOVA) with subject and relative level $A2/A1$ as factors. The relative level was significant for all vowel conditions ($p<0.001$). With $df=(4, 105)$, the largest F-ratio of 11 952.49 was obtained for vowel # 8 (5.3-bark condition) and the smallest F ratio of $F=492.68$ for vowel # 1 (pulse-train, 2.5-bark condition). The results of this ANOVA confirmed that, for each vowel condition, the level $A2/A1$ had a significant effect on F_v. Listener differences were not significant for vowels # 1 through # 5, and # 8, but they were significant for vowels # 6 and # 7 [$F(2,105)=31.78$, $p<0.001$ and $F(2,105)=22.51$, $p<0.001$, respectively]. This confirms the observed variability in listeners' performance for both 3.6- and 4.0-bark separations.

B. The COG effect

When the distance between the two formants of a reference vowel is smaller than the critical distance (3.5 bark), F_v values are similar to those reported by Chistovich and Lublinskaja (1979). For example, in Fig. 4, the data show a continuous relationship between F_v and $A2/A1$ for vowels # 1 through # 4 ($F2-F1=2.5$ bark). Consequently, all match-

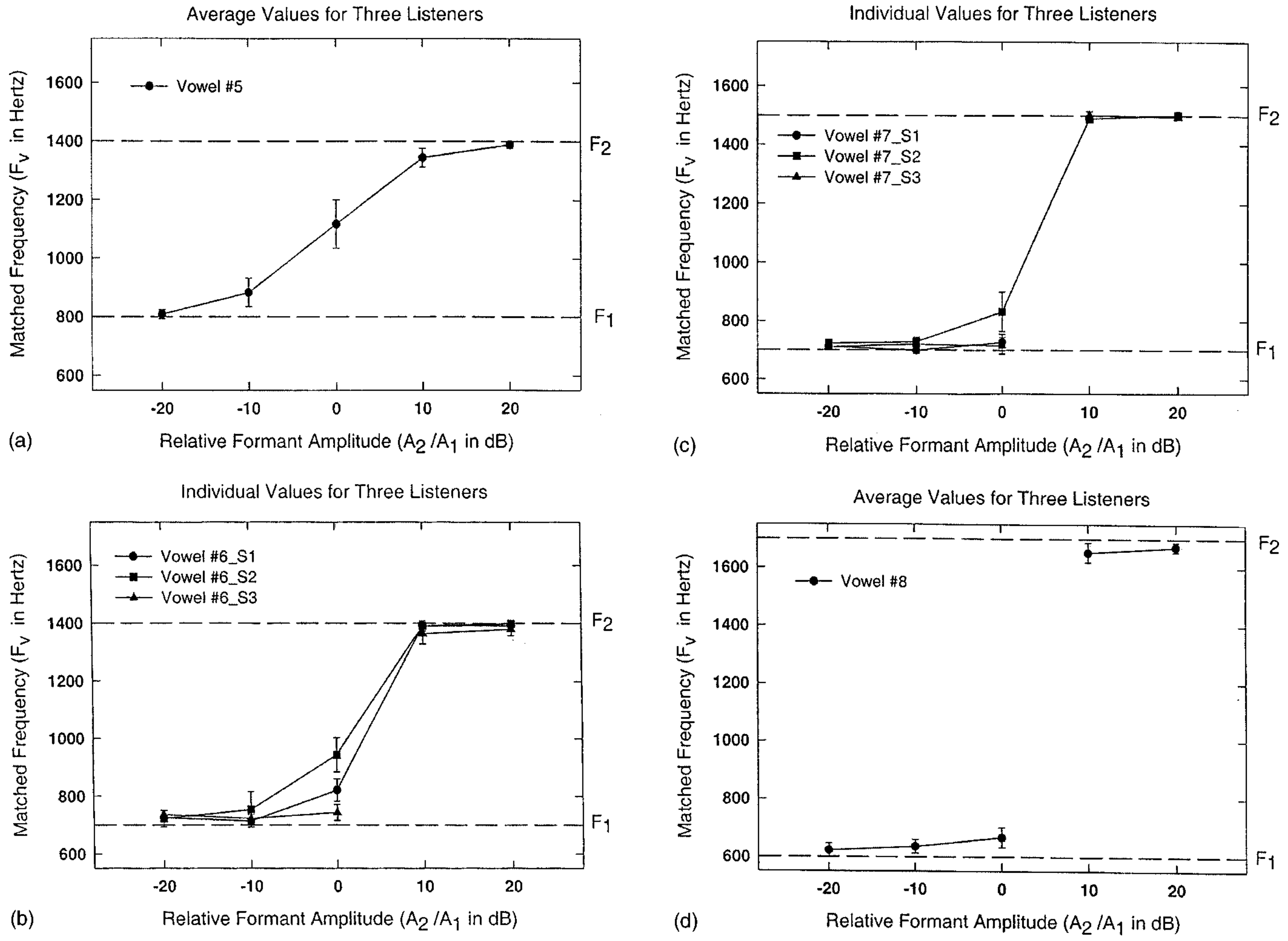

FIG. 5. F_v matching values for vowels # 5 through # 8 with changing $A2/A1$ ratio. (a) mean F_v values for three listeners for vowel # 5 (3.0 bark); (b) individual F_v values for each listener for vowel # 6 (3.6 bark); (c) individual F_v values for each listener for vowel # 7 (4.0 bark); (d) mean F_v values for three listeners for vowel # 8 (5.3 bark). Error bars at each data point represent 2 standard deviations. Horizontal dashed lines mark the locations of $F1$ and $F2$.

ing F_v points fall between $F1$ and $F2$. However, the data show a different pattern of variance with relative level. That is, the variance in the middle area ($A2/A1 = 0$ dB) is much greater than toward the ends ($A2/A1 = \pm 20$ dB), regardless of the excitation source or formant bandwidth. Standard deviation for the -20-dB condition ranges from 5.01–19.53 and from 13.80–24.46 for the 20-dB condition, whereas the range for 0 dB is from 48.03–76.37. This might indicate that listeners' uncertainty was greater when the auditory system integrated the spectra of two equal-level formants than when one of the two formants dominated. Another observation is that the range of $A2/A1$ over which F_v changed as a function of $A2/A1$ is about 40 dB (-20 to $+20$ dB). Figure 5(a) shows a similar result for 3.0-bark separation. These results confirm the existence of spectral COG effect for two-formant vowels when both formants are spaced closely enough to display the predicted variation in the matching frequency of the single-formant vowel.

C. The critical distance

Formant separations in the reference signals were set to 2.5, 3.0, 3.6, 4.0, and 5.3 bark to verify the existence of a critical distance. For all three listeners, a continuous relationship between the F_v values and $A2/A1$ was observed for frequency separations of up to 3 bark. However, when the separation between the formants of the reference vowel increased, this continuous relationship gradually disappeared. It is interesting to note that each listener showed a different breakdown point. For 3.6-bark separation [Fig. 5(b)], listener # 3 first shows a discontinuous relationship between F_v and $A2/A1$, while the other two maintain the continuous relationship. For 4.0-bark separation [Fig. 5(c)], listener # 1 demonstrates the breakdown, and only results from listener # 2 remain continuous. Finally, for 5.3-bark separation [Fig. 5(d)], the continuity disappeared in the performance of all three listeners. This indicates that the three subjects had different estimates of critical distance. This distance may be somewhere less than 3.6 bark for listener # 3, between 3.6 and 4.0 bark for listener # 1, and somewhat larger than 4.0 bark for listener # 2. A comparable pattern of responses was obtained by Chistovich and Lublinskaja (1979) for the supracritical separation. Performance of both their listeners differed considerably in that one listener showed a clear discontinuous relationship between F_v and $A2/A1$ and the continuity did not disappear in the performance of the second listener despite the large variation of the values of F_v. For the latter listener, no clear breakdown point was obtained.

Because the experimental procedure was already very time consuming, we did not explore additional values of $A2/A1$ to determine where exactly the F_v values shifted abruptly from $F1$ to $F2$. We can only determine that all fell within $A2/A1$ values ranging from 0 to $+10$ dB. This differs from results of Chistovich and Lublinskaja, where the shift from $F1$ to $F2$ occurred when $A2/A1$ was about -10 dB. This discrepancy in results for one experimental condition is most likely due to differences in the procedures used. Increased variability in the data collected in a direct matching task in Chistovich and Lublinskaja's study may have contributed to this result as well.

The data in Fig. 5 show that, for each listener, the variance of F_v is smaller within the supracritical separation than within the subcritical separation. This is true especially for the two formants with equal levels. This is similar to Chistovich and Lublinskaja's results, in which most values for the supracritical distance showed less variation than for subcritical separation. One explanation is that when the two formants of the reference signal are resolved, listeners simply pick either $F1$ or $F2$ for F_v match. However, the degree of uncertainty in making judgments may be considerably reduced in the double-staircase procedure than in the traditional adjustment task.

D. Formant bandwidths

For the subcritical distance of 2.5 bark, four reference vowels were generated with variable bandwidths. The results for the noise-excited vowels # 2, # 3, and # 4 (see Fig. 4) show that F_v values for vowel # 3 (the narrowest formant bandwidth) approximated closely the frequency values of either $F1$ or $F2$, whereas those for vowel # 4 (the widest bandwidth) were closer to the COG point. The only exception occurred when $A2/A1 = 10$ dB. To investigate the effect of formant bandwidth, a mixed-design ANOVA was performed on F_v data for vowel # 2 (BW=80 Hz), vowel # 3 (BW=45 Hz), and vowel # 4 (BW=150 Hz) with subject and formant level $A2/A1$ as two between-subject factors, and formant bandwidth as the within-subject factor. Formant bandwidth was significant [$F(2,210) = 17.172$, $p < 0.001$] and so were the interactions between bandwidth and $A2/A1$ [$F(8,210) = 3.621$, $p < 0.001$], and bandwidth and subject [$F(4,210) = 4.074$, $p < 0.004$]. The effect of subject was not significant [$F(2,105) = 0.323$, $p < 0.725$] and the interaction bandwidth×subject×$A2/A1$ was not significant [$F(16,210) = 1.651$, $p < 0.059$].

These results are in accord with Beddor and Hawkins' (1990) findings for their three bandwidth conditions: narrow (BW=45 Hz), medium (BW=75 Hz), and wide (BW=150 Hz), thus supporting the hypothesis that the relative contribution of formant frequency and amplitude might depend on the spectral characteristics of signals. Beddor and Hawkins propose that, in the matching task, formant frequency may be more important than spectral shape for signals with well-defined spectral peaks. Formant amplitudes and spectral shape might be more important for signals with poorly defined spectral peaks. In the former case, the F_v might lie closer to the formant frequency values. In our study, the vowel with narrower-than-normal formant bandwidth (45 Hz) had a well-defined spectral peak, while the vowel with wider-than-normal bandwidth (150 Hz) had a poorly defined spectral peak. Our results agree with Beddor and Hawkins' hypothesis except when $A2/A1 = 10$ dB.

E. Excitation source

Figure 4 shows the matching F_v values for the pulse-train-excited vowel # 1 and the noise-excited vowel # 2 with the same frequency separation (2.5 bark) and formant bandwidth (80 Hz). Generally, listeners showed similar response patterns for the two reference signals. F_v values for vowel # 1 were lower than those for vowel # 2 when $A2/A1$ ranged from −20 to 0 dB, and higher when $A2/A1$ was 10 or 20 dB. The variances of the matching values were also comparable at each $A2/A1$ level for the pulse-train and noise-excited signals. For the two vowels, a mixed-design ANOVA was performed on the F_v data with subject and relative level $A2/A1$ as two between-subject factors, and excitation source type as a within-subject factor. Source type was not significant $[F(2,105)=1.281,\ p=0.260]$ and subject effect was not significant $[F(2,105)=4.554,\ p=0.013]$; however, the interaction between source type and relative level $A2/A1$ was significant $[F(8,105)=5.352,\ p=0.001]$. This justifies the use of Gaussian noise excitation in generating all reference vowels to exclude potential effects of a harmonic complex tone source. The interaction between source type and subject was not significant $[F(4,105)=1.731,\ p=0.182]$.

Overall, these data suggest that listeners are able to perform a spectral integration which depends on frequency separation between two formants. The results confirm the existence of a critical frequency region, in which the listeners' behavior changes rather drastically. This region, identified as "critical distance" of about 3.5 bark in early matching experiments by Chistovich and Lublinskaja (1979) and verified with various degrees of success by subsequent research, may indicate a limit to the range of spectral integration that the auditory system can perform. Thus, the observed COG effect may reflect a more broadly defined auditory behavior which is not peculiar to the perception of speech sounds. The second experiment was conducted to verify this possibility. The striking similarity of the two-tone resolution results of Feth and O'Malley (1977) to the critical distance observed in vowel matching tasks substantiates the claim that both the critical region and the complex-tone discriminability estimates may be related to a common mechanism, i.e., to an auditory spectral resolving power.

IV. EXPERIMENT 2: TWO-TONE SPECTRAL RESOLUTION BANDWIDTH

A. Methods

The same three trained listeners from experiment 1 participated. As in the synthetic vowel task, all listeners were given an extensive practice on complex signals before participating in the second experiment. Data were collected after there were no further improvements in their performance.

The stimuli used in experiment 2 were two-tone complex signals, which were geometrically centered around a frequency, f_c (1000 Hz) so that $f_c=(f_2 f_1)^{1/2}$; $\Delta f=f_2-f_1$. In the target signal, the intensity difference $(\Delta I=L_2-L_1)$ between the two components was +1 dB, while in the reference standard signal ΔI was −1 dB. The level of a given component at f_i is designated by L_i. All signals were generated by the laboratory PC using the TDT system II with D/A board. Each signal had duration of 300 ms and was shaped by a 5-ms cosine rise–fall window. The silent interval between signals was 300 ms. The signals were played at a 20-kHz sampling rate. The smoothing filter was the same analog low-pass filter with a cutoff frequency at 8 kHz, as used in experiment 1. Sound level was controlled by a programmable attenuator. All signals were presented monaurally through the HD-414-SL earphones at 50 dB SL.

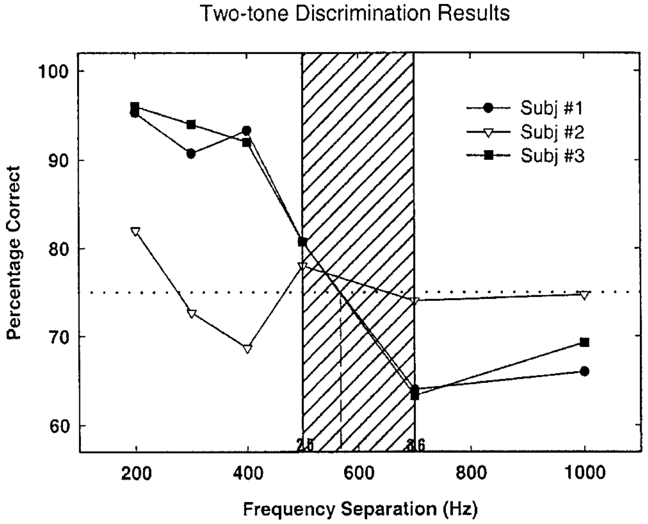

FIG. 6. Percentage of correct discrimination in a 2Q-2AFC task as a function of frequency separation, Δf, of the two components in each complex-tone pair. Each data point is the average of 150 trials. Individual psychometric functions are shown for each of the three listeners. The horizontal dotted line indicates the jnd at 75%. The shaded region delimits the frequency separation from 2.5 to 3.6 bark.

The two-cue, two-alternative, four-interval forced-choice paradigm was used. The target signal was presented with equal probability in either the second or the third interval. The reference signal was presented in the remaining three intervals. For the target signal, the higher-frequency component had the higher intensity, i.e., $\Delta I=+1$ dB. As in experiment 1, the listeners were asked to indicate whether the pitch of the variable signal was higher or lower than that of the reference signal.

At first, a two-up, one-down adaptive procedure was used with the initial value of the frequency separation, Δf, set at 100 Hz around the center frequency of 1000 Hz. However, the spectral pitch of both target and standard signals changed whenever Δf was changed from one value to another, and all three subjects reported that the task was too difficult to perform in this adaptive procedure. Therefore, a fixed standard procedure as in the Feth and O'Malley study (1977), was used instead. The Δf was fixed for a block of 50 trials and the percentage of correct discrimination, $P(C)$, was recorded. Each data point at a given Δf value was obtained from 150 trials. The psychometric function for each subject was generated by plotting $P(C)$ for Δf values of 200, 300, 400, 500, 700, and 1000 Hz.

B. Results and discussion

Figure 6 displays psychometric functions for each of the three participants. Percentage of correct responses, $P(C)$, is plotted as a function of the frequency difference (Δf) between the two tones. Listeners # 1 (●) and # 3 (■) show performance similar to that in Feth and O'Malley (1977), while the performance of listener # 2 (△) is clearly different.

For listeners # 1 and # 3, the $P(C)$ values are high for moderate Δf values near 200 Hz. Discrimination performance drops when Δf exceeds 400 Hz. The Δf at which the listener's $P(C)$ falls to 75% is used to estimate the spectral resolution bandwidth. For these two listeners, the estimate is 580 Hz.

Recall the results from experiment 1. We estimated that the critical distance for listener # 1 was somewhere between 3.6 and 4.0 bark, while it was less than 3.6 bark for listener # 3. The same tendency in performance is observed in experiment 2. The shaded area in Fig. 6 extends from a frequency separation equal to 2.5 bark to one equal to 3.6 bark. The ability of these two listeners to distinguish between the two-tone complexes drops below 75% within that same critical region. In the vowel task, when the formant separation was greater than the critical distance, we concluded that the auditory system failed to integrate across the two formants (i.e., the two formants were resolved). In the two-tone discrimination task, we assume that the discriminability of complementary two-tone pairs dropped below 75% when the two components were resolved by the auditory system. Thus, we conclude that both the critical distance between formants and the complex-tone discriminability limit are dependent upon the same auditory spectral resolving power.

The performance of listener # 2 in the discrimination task is very different from that of the other two listeners. The $P(C)$ values fluctuate across the whole Δf range as seen in Fig. 6. Even for moderate Δf values, $P(C)$ does not approach 100%. Furthermore, discriminability does not drop with larger Δf values. We have no explanation for why this listener performed so differently. Although this listener was given much more practice than the other two, she was unable to do the task and her performance is clearly different. Thus, we cannot determine the spectral resolution bandwidth for listener # 2. We infer from experiment 1 that her estimated critical distance is greater than 4.0 bark. Based on this value alone, we might conclude that her spectral resolution bandwidth is larger than 4.0 bark. This would lead us to predict that her ''-75%'' estimate is greater than 800 Hz in the two-tone task. However, her performance in experiment 2 was never regular enough to give us a solid basis to reach this conclusion.

V. GENERAL DISCUSSION AND CONCLUSIONS

As outlined in the Introduction, the phenomenon of spectral integration has been studied from two perspectives, i.e., as a perceptual averaging of a formant cluster in approximating phonetic quality of vowels ($F2'$ or center of gravity of the cluster), and as a frequency-displacement effect resulting from the interaction of frequency and relative amplitude ratio of two closely spaced formants (the COG effect). A common underlying approach for the two lines of research was to verify the early observations that spacing between two formants plays an important role in making perceptual decisions about vowel quality.

Although linking the 3.5-bark limit of spectral integration with $F2'$ brought consistent experimental results to confirm the importance of COG in estimating vowel quality in two-formant models, formalizing the COG effect encountered difficulties. The question of adequate methodology to study the COG effect arose as a consequence of attempts to formally account for its manifestation. Assmann's study (1991) showed no support for a COG effect using the first two formants of multiformant back vowels. His results from the vowel identification paradigm were also inconclusive. The use of matching tasks advocated by the Leningrad group has also been questioned as the only, and perhaps not very reliable procedure, to measure the COG effect. Considerable variability in the data from Chistovich and Lublinskaja (1979) and Lester (1996) provide additional evidence that a refinement in the experimental procedure to study the COG effect was necessary. It is worthwhile to note that both the ''double-staircase'' adaptive procedure used in Feth *et al.* (1996) and the two-cue, two-alternative forced-choice paradigm, as in Feth and O'Malley (1977), reduced the variability in the data.

A more reliable methodology gave rise to the question of the role of the COG effect and the formant amplitude ratio in estimating vowel quality. That is, for two closely spaced formants, listeners performance is predictable. When the frequency separation between the formants exceeds the 3.5-bark distance, relative levels of the formants do not contribute to otherwise predictable listening behavior: decisions about vowel identity are based entirely on the frequency of the formants. The experimental data on Russian vowels show that, for a subcritical separation of 3.5 bark, a change in $A2/A1$ ratio in two-formant vowels is perceptually equivalent to the frequency change which determines vowel quality.

In this study, we examined whether the COG effect plays role other than approximating phonetic vowel quality. Consequently, our listeners responded to the changes in vowel pitch, not vowel quality. In this respect, the task in experiment 1 was similar to that in experiment 2, in which the same listeners responded to differences in pitch in two-tone signals. In testing the COG effect in experiment 1, we sought to verify the role of the critical separation of 3.5 bark in spectral integration, using both a more reliable experimental procedure and model predictions incorporating the interaction between formant frequencies and their relative amplitudes. In this endeavor, we re-examined some aspects of Chistovich and Lublinskaja's results (1979) and introduced further modifications to the methods such as source type (Gaussian noise) and bandwidth manipulations, as in Beddor and Hawkins' study (1990). We used well-trained listeners as subjects of our study because we were interested in the optimal performance of the human auditory system in testing the COG effect.

The results of experiment 1 confirm that the COG effect occurs within the limit of spectral integration of 3.5 bark. Listeners' responses to changes in vowel pitch indicate a match in frequency according to the COG mechanism: when $F2$ is much weaker than $F1$ ($A2/A1=-20\,\mathrm{dB}$), the matches fall in the vicinity of the stronger $F1$ and, conversely, listeners match the frequency closer to $F2$ when $F1$ is weaker ($A2/A1=+20\,\mathrm{dB}$). However, when both formants are of equal strength ($A2/A1=0$ dB), the matches fall somewhere between the two formants, and an increased variability is observed in the data. Crucially, this tendency is maintained for all conditions within the subcritical formant separation for vowels # 2 through # 5.

In designing experiment 1, we hypothesized that spectral shape (poorly defined spectral prominence versus well-defined spectral prominence) affects the matching values of F_v. Accordingly, three two-formant vowels with variable bandwidth were generated with subcritical separation of 2.5 bark (vowels # 2 through # 4). The vowel with narrower-than-normal formant bandwidth of 45 Hz had a well-defined spectral peak (vowel # 3), and the vowel with wider-than-normal bandwidth (150 Hz) had a poorly defined spectral peak (vowel # 4). Model predictions were verified in the experimental data. Listeners approximated the frequency of either $F1$ or $F2$ for the narrow-bandwidth vowel # 3, thus selecting one of the ‘‘stronger’’ peaks. For vowel # 4, the matches fell closer to a frequency between the formants, which is in accord with the COG effect. Our results and model predictions support Beddor and Hawkins’ hypothesis with one exception, i.e., when $A2/A1 = 10$ dB.

Experiment 2 verified earlier findings by Feth and O’Malley (1977) that spectral integration in two-tone complex signals occurs within the limit of the 3.5-bark resolution bandwidth. Given that the responses in experiment 2 came from the same participants as in experiment 1, we conclude that the complex-tone discriminability and the spectral integration limits reflect the same auditory spectral resolving power. This further suggests that the auditory processing of complex signals at the intermediate stage, i.e., before higher-level decision processes apply, is the same for speech and nonspeech signals.

As discussed in the Introduction, several models were previously proposed for predicting the performance of listeners asked to match an adjustable one-formant to a fixed two-formant signal. Early work suggested that a simple calculation of the spectral centroid of the two-formant signal would serve this purpose (Chistovich *et al.*, 1979). We verified the predictions of the spectral centroid model and also used the IWAIF model for a comparison with the present results. Figure 7 displays a summary of predictions of our modified PLP model, the IWAIF model, and the spectral centroid model, along with the group data for vowel # 2 (2.5-bark separation, noise-excited) averaged across three listeners. To implement the IWAIF and spectral centroid models, we assumed that signal energy outside the critical distance was not available for the calculation.

Clearly, the spectral centroid model performs most poorly in this comparison. The predicted relationship between F_v and relative formant levels does not fit the experimental data. The endpoints are overestimated for $F1$ peak and underestimated for $F2$ peak, and only the value at the center of the $A2/A1$ range falls within the error bars. For both the PLP model and the IWAIF model, there is much better agreement with the data. The PLP model predicts well the performance at the endpoints (i.e., both extreme $A2/A1$ values) but underestimates listeners’ matches by more than 2 standard deviations at $A2/A1 = 0$ dB. The IWAIF model, on the other hand, predicts listeners’ matches that fall just above those obtained in the experiment. An exception to this tendency is the $F2$ region where the predicted values are lower than those in the experimental data. At this stage, it is still premature to determine which of the two models, the PLP model or the IWAIF model, is better suited to predicting COG effects with greater accuracy. Further refining of the IWAIF model is probably necessary, since we had to assume an arbitrary limitation on signal bandwidth to produce a reasonable prediction of performance.

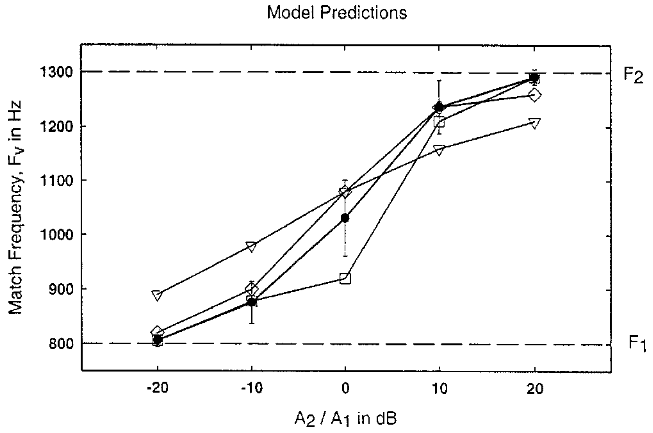

FIG. 7. Model predictions and experimental data for vowel # 2 (2.5 bark): (□) predictions by the proposed PLP model; (◇) by the IWAIF model, (▽) by the spectral centroid model with linear amplitude weighting, (●) the experimental data for vowel # 2; each error bar represents 2 standard deviations. The 3.5-bark rectangular window is used to select the spectra for the IWAIF and the centroid models. Horizontal dashed lines mark the locations of $F1$ and $F2$.

In conclusion, the two experiments reported here have shown that the 3.5-bark limit on spectral integration appears to reflect a property of auditory processing rather than some special processing unique to speech sounds. This study, and that of Chistovich and Lublinskaja (1979), used a small number of very well-trained listeners, but some evidence for individual differences in the size of the critical separation is apparent in both. The effects of formant bandwidth and excitation source were predictable. Finally, a preliminary comparison of model predictions of COG performance did not favor one model strongly; however, the simple spectral centroid model was clearly the poorest at predicting listener matches.

The signals used in the current study were spectrally static. That is, the parameters of both the adjustable and the reference signals remained constant for the entire duration of the sound. Real speech sounds are typically dynamic; that is their formant frequencies and amplitudes change, sometimes rapidly, over time. Lublinskaja (1996) has recently reported that the COG effect can be demonstrated dynamically by amplitude modulation of formant amplitudes. Further work on modeling of the COG effects therefore should take into account these dynamic effects.

ACKNOWLEDGMENTS

This work was supported by a grant from AFOSR. The authors wish to thank Douglas O’Shaughnessy and several anonymous reviewers of the manuscript for many helpful comments.

Anantharaman, J. N., Krishnamurthy, A. K., and Feth, L. L. (**1993**). ''Intensity weighted average of instantaneous frequency as a model for frequency discrimination,'' J. Acoust. Soc. Am. **94**, 723–729.

Anantharaman, J. N. (**1998**). ''A perceptual auditory spectral centroid model,'' Ph.D. thesis, The Ohio State University, Columbus, OH.

Assmann, P. F. (**1991**). ''The perception of back vowels: Centre of gravity hypothesis,'' Q. J. Exp. Psychol. **43A**, 423–448.

Beddor, P. S., and Hawkins, S. (**1990**). ''The influence of spectral prominence on perceived vowel quality,'' J. Acoust. Soc. Am. **87**, 2684–2704.

Bedrov, Y. A., Chistovich, L. A., and Sheikin, R. L. (**1978**). ''Frequency location of the 'center of gravity' of the formants as a useful parameter in vowel perception,'' Akust. Zh. **24**, 480–486 (Sov. Phys. Acoust. **24**, 275–282).

Bladon, A. (**1983**). ''Two-formant models of vowel perception: Shortcomings and enhancements,'' Speech Commun. **2**, 305–313.

Bladon, A., and Fant, G. (**1978**). ''A two-formant model and the cardinal vowels,'' Royal Inst. Tech., Stockholm, STL-QPRS **1**, 1–8.

Carlson, R., Granström, B., and Fant, G. (**1970**). ''Some studies concerning perception of isolated vowels,'' Royal Inst. Tech., Stockholm, STL-QPRS **2–3**, 19–35.

Carlson, R., Fant, G., and Granström, B. (**1975**). ''Two-formant models, pitch and vowel perception,'' in *Auditory Analysis and Perception of Speech*, edited by C. G. M. Fant and M. A. Tatham (Academic, New York and London), pp. 58–82.

Chistovich, L. A. (**1985**). ''Central auditory processing of peripheral vowel spectra,'' J. Acoust. Soc. Am. **77**, 789–804.

Chistovich, L. A., and Lublinskaja, V. V. (**1979**). ''The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli,'' Hear. Res. **1**, 185–195.

Chistovich, L. A., Sheikin, R. L., and Lublinskaja, V. V. (**1979**). '''Centres of gravity' and spectral peaks as the determinants of vowel quality,'' in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, London), pp. 55–82.

Delattre, P., Liberman, A., Cooper, F., and Gerstman, L. (**1952**). ''An experimental study of the acoustic determinants of vowel color,'' Word **8**, 195–210.

Escudier, P., Schwartz, J.-L., and Boulogne, M. (**1985**). ''Perception of stationary vowels: Internal representation of the formants in the auditory system and two-formant models,'' Franco–Swedish Seminar, SFA, Grenoble, 143–174.

Fant, G. (**1959**). ''Acoustic analysis and synthesis of speech with applications to Swedish,'' Ericsson Tech. **1**, 3–108.

Feth, L. L. (**1974**). ''Frequency discrimination of complex periodic tones,'' Percept. Psychophys. **15**, 375–378.

Feth, L. L., Lester, J., Xu, Q., and Ross, J. (**1996**). ''Testing the 'center of gravity' effect for vowel-like complex sounds,'' J. Acoust. Soc. Am. **100**, 2626(A).

Feth, L. L., and O'Malley, H. (**1977**). ''Two-tone auditory spectral resolution,'' J. Acoust. Soc. Am. **62**, 940–947.

Helmholtz, H. L. F. (**1954**). *On the Perception of Tone*, 2nd English ed. (Dover, New York).

Hermansky, H. (**1990**). ''Perceptual linear predictive (PLP) analysis of speech,'' J. Acoust. Soc. Am. **87**, 1738–1752.

Jeffress, L. A. (**1968**). ''Beating sinusoids and pitch changes,'' J. Acoust. Soc. Am. **43**, 1964.

Jesteadt, W. (**1980**). ''An adaptive procedure for subjective judgments,'' Percept. Psychophys. **28**, 85–88.

Lester, J. (**1996**). ''Spectrographic analysis of sound vs. auditory perception,'' Senior Honors thesis, The Ohio State University, Columbus, OH.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lublinskaja, V. V. (**1996**). ''The 'center of gravity' effect in dynamics,'' in *Proceedings of the Workshop on the Auditory Basis of Speech Production*, edited by W. Ainsworth and S. Greenberg, pp. 102–105, ESCA.

Mantakas, M., Schwartz, J.-L., and Escudier, P. (**1988**). ''Vowel spectrum processing and the large-scale integration concept,'' 7th FASE Congress, Edinburgh.

Schwartz, J.-L., and Escudier, P. (**1989**). ''A strong evidence for the existence of a large-scale integrated spectral representation in vowel perception,'' Speech Commun. **8**, 235–259.

Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (**1997**). ''The dispersion-focalization theory of vowel systems,'' J. Phonetics **25**, 255–286.

Voelcker, H. B. (**1966a**). ''Toward a unified theory of modulation I. Phase-envelope relationship,'' Proc. IEEE **54**, 340–353.

Voelcker, H. B. (**1966b**). ''Toward a unified theory of modulation II. Zero manipulation,'' Proc. IEEE **54**, 735–755.

Xu, Q. (**1997**). ''A signal processing model for spectral integration of vowel sounds,'' M.A. thesis, The Ohio State University, Columbus, OH.