Spectral integration in vowel perception: Matching and discrimination studies

Keith Johnson
Marisa Fernandez
Michael Henninger
Jim Sandstrum

Introduction

This paper is about the **spectral integration** of formants in vowel perception. Chistovich and her colleagues demonstrated in the late 70's that if you change the relative amplitudes of two closely spaced formants in a two-formant vowel, the formant frequency of the best matching one-formant vowel will change, provided the formants are "closely spaced". Chistovich & Lublinskaja (1979) reported that the critical distance for this center of gravity effect is about 3.5 Bark. This estimate of the critical distance and the hypothesis that vowel formants are percepually merged has been often used but seldom studied. A fact which is all the more remarkable given that the authors served as their own listeners. We will present 3 experiments which were designed to investigate the spectral integration of vowel formants.

The experiments test the hypothesis that the center of gravity effect is caused by spectral integration, and so when two formants are within some critical distance they merge into a single perceptual formant.

Experiment 1

The first experiment was a matching study. Listeners heard synthetic stimuli presented in pairs. The first stimulus in each pair had two formants, the second had only one. The 5 naive listeners were asked to adjust the frequency and bandwidth of the formant in the one-formant stimulus, until the two stimuli sounded as similar as possible.

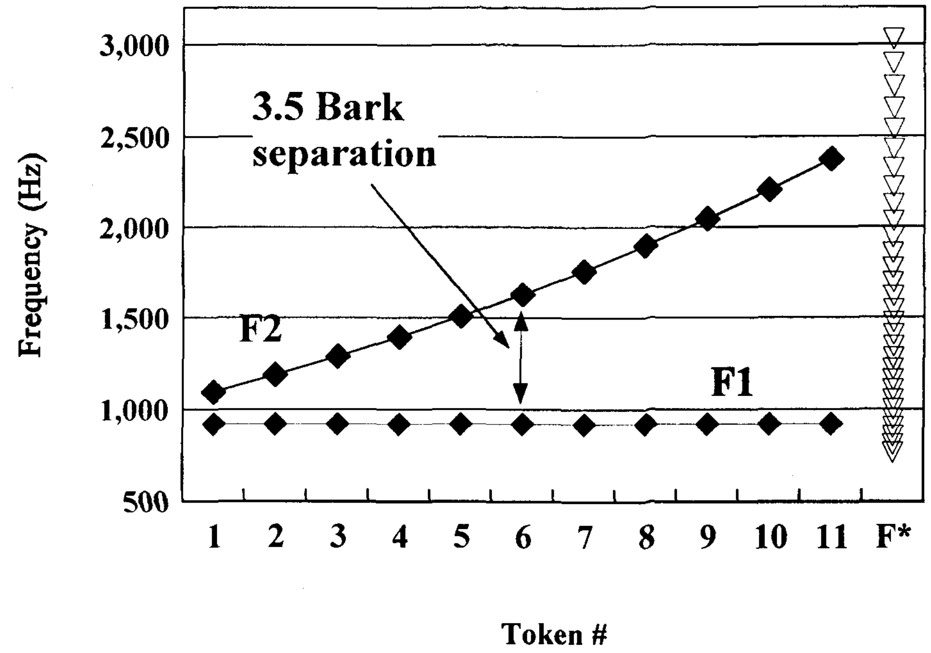

Figure 1. Formant frequencies of the synthetic stimuli used in Experiment 1. The solid symbols show the frequencies of F1 and F2 of the two-formant stimuli, and the open symbols on the right show the frequencies of F* in the one-formant stimuli. As indicated, F1 and F2 are separated by at least 3.5 Bark in two-formant stimuli 6-11.

The two-formant stimuli formed an 11-step continuum (shown in Figure 1) with F1 fixed and F2 varied across the members of the continuum. When the F2 was low, these stimuli sounded like the vowel in "odd", and when the F2 was high, they sounded like the vowel in "add". The distance between F1 and F2 was 3.5 Bark in token 6.

The one-formant stimuli also formed a continuum, labelled F* in Figure 1. Their frequencies ranged from 764Hz to 3027Hz in 0.3 Bark steps. We synthesized 8 stimuli at each frequency step spanning a range of bandwidth values from 50Hz to 400Hz in 50Hz steps. This manipulation was added because we noticed that single formant stimuli (especially those with high F*) sounded more natural with wider bandwidths.

The spectral integration hypothesis leads to a prediction about this matching experiment (illustrated with some hypothetical data in Figure 2). If closely spaced formants are integrated into a single perceptual formant, we expect that, when the formants of the two-formant standard are close to each other, listeners will choose F* values which are between the formants of the two-formant standard, but when they are not close to each other, listeners will tend to adjust F* so that it matches either the F1 or the F2 of the standard. Chistovich et al. (1979) stated that they expected F* to match the F2 of the standard. We can see no principled motivation for this expectation. Note that this pattern of results gives us a definition of "closely spaced" (the critical distance between formants), as well as a test of the spectral integration hypothesis.

As predicted, listeners did tend to choose F* values which were between F1 and F2, early in the continuum (see Figure 3). The boxes in Figure 3 enclose 50% of the responses, the notches enclose 33% of the responses, and the horizontal line marks the median response. As can be seen in the figure, there was greater variability in the listener's responses as F1 and F2 separated. However, we are hard pressed to identify any point along the continuum where there is a sudden shift in the response pattern. Also, contrary to our expectation, listeners chose F* values in between the formants even when F1 and F2 were widely separated.

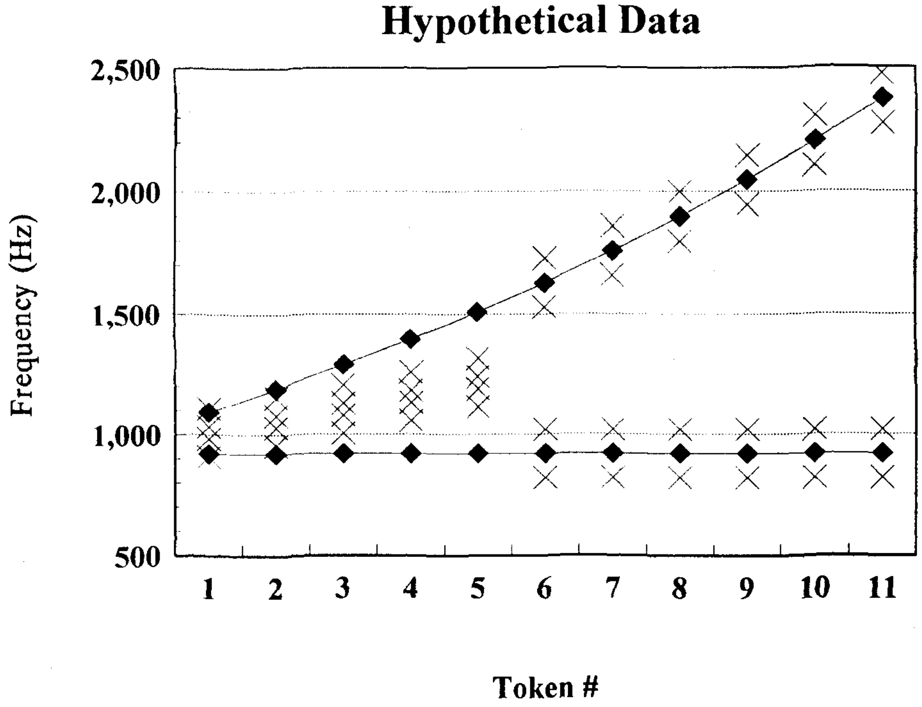

Figure 2. Pattern of matching responses predicted by the spectral integration hypothesis. Frequencies of F1 and F2 of the two-formant stimuli are indicated by the filled diamonds, and hypothetical best matching one-formant F* values are marked with x.

Our decision to allow the listeners to adjust the bandwidth of the one-formant stimuli, may have clouded the results of this experiment. As Figure 4 shows, the listeners tended to choose tokens with wide bandwidths (that is, the stimuli which sounded to us less harsh or nonspeech-like), so, the F* choices may have been corrupted by the less sharply defined spectral envelopes of these wide bandwidth tokens.

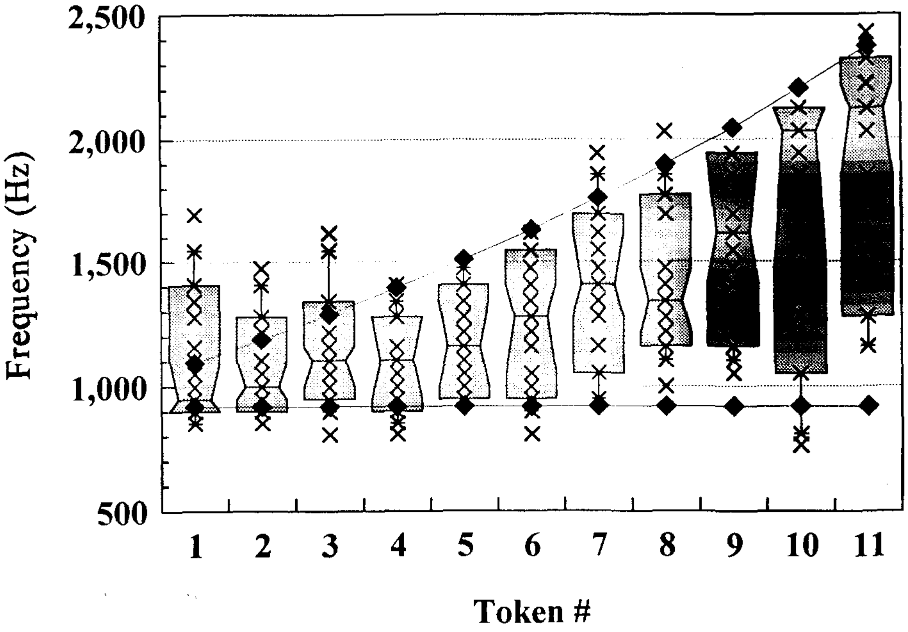

Figure 3. Results of Experiment 1. Frequencies of F1 and F2 are marked by filled diamonds and the observed F* values are marked with x. The overlaid boxes show the median response and the thirds and quartiles of the response distributions. The error bars include 90% of the responses.

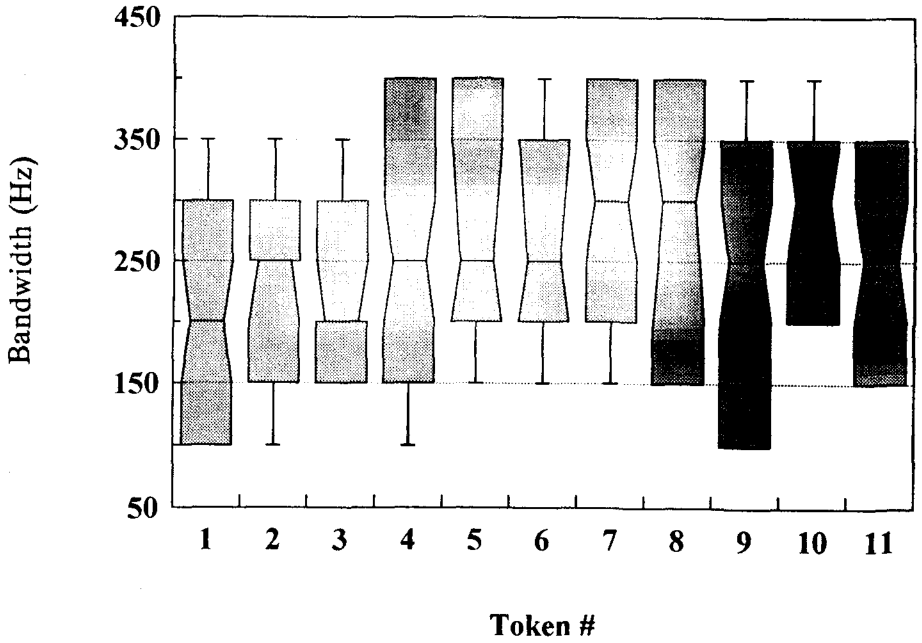

Figure 4. Results of Experiment 1. Bandwidth values selected by the listeners for each two-formant token are plotted by token number. Box and whisker display as in Figure 3.

Experiment 2

Therefore, in Experiment 2 we tested the effect of changing the bandwidth of the F* in a matching experiment. We also manipulated the overall amplitude of the two-formant stimuli, on the assumption that higher amplitude might result in some frequency smearing which would affect the results. Unfortunately, our amplitude manipulation was not very large, and no effects of the manipulation were observed. Therefore this report will focus on the bandwidth manipulation.

One group of 3 listeners performed the matching task as in Experiment 1, with narrow-bandwidth F* stimuli, while another group of 3 listeners performed the task with wide bandwidth stimuli. The bandwidths were 150Hz and 250Hz respectively.

Responses in the wide bandwidth condition (Figure 5) were very scattered. There are a few observable trends, such as the steady increase in the median frequency of F* across the continuum, but on the whole the data in this condition were quite messy. Thus, given the relatively wide bandwidths chosen by listeners in Experiment 1, it seems likely that the data in that experiment included some of the variability we see in Figure 5.

Data from the narrow bandwidth condition (Figure 6) showed much less scatter. However, this box-and-whisker display leads to the false impression that listeners were equally probable to chose any of the stimuli surrounded by the boxes. The distribution of responses is clarified in a spectral plot of the same data (Figure 7). In this figure, responses to the stimuli are plotted by token number and F* value with the number of responses coded as a shade of gray. The darker the shading, the greater the number of responses at that F* frequency. Note how the distribution of responses goes from unimodal for tokens 1 through 5 (that is, the responses fall near one frequency) to bimodal for tokens 6 through 11. This is pooled data, so it is important to note that the distributions were bimodal for each listener. Recall that this was the pattern of data we predicted given the spectral integration hypothesis. Chistovich et al.'s prediction that F* would follow F2 was not confirmed by these data, but the location of the shift from unimodal to bimodal performance corresponds with Chistovich & Lublinskaya's (1979) estimate of the critical distance.

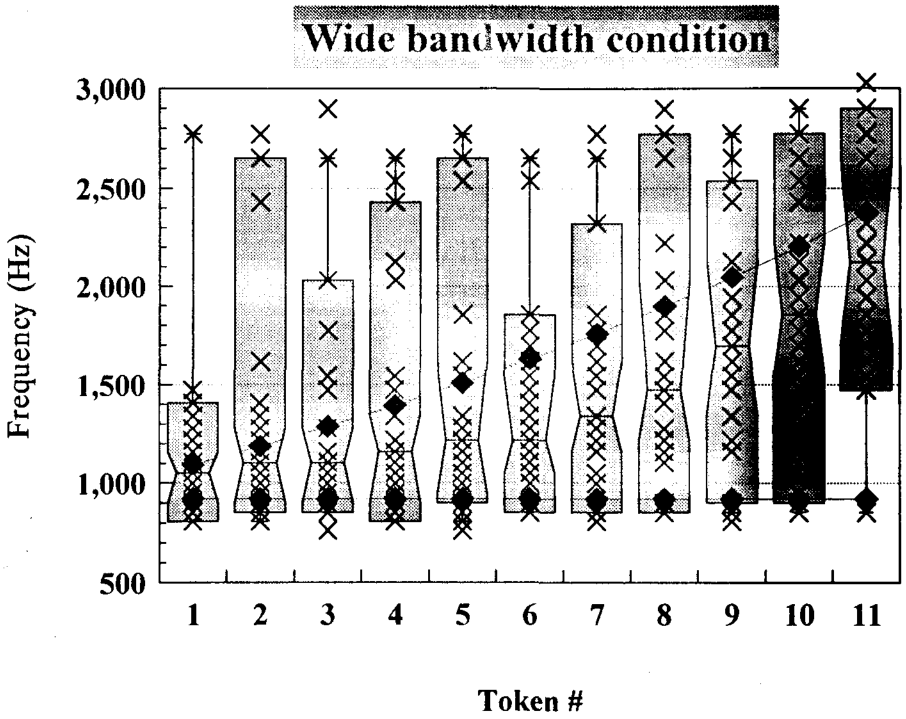

Figure 5. Results of Experiment 2. Data from the wide bandwidth condition.

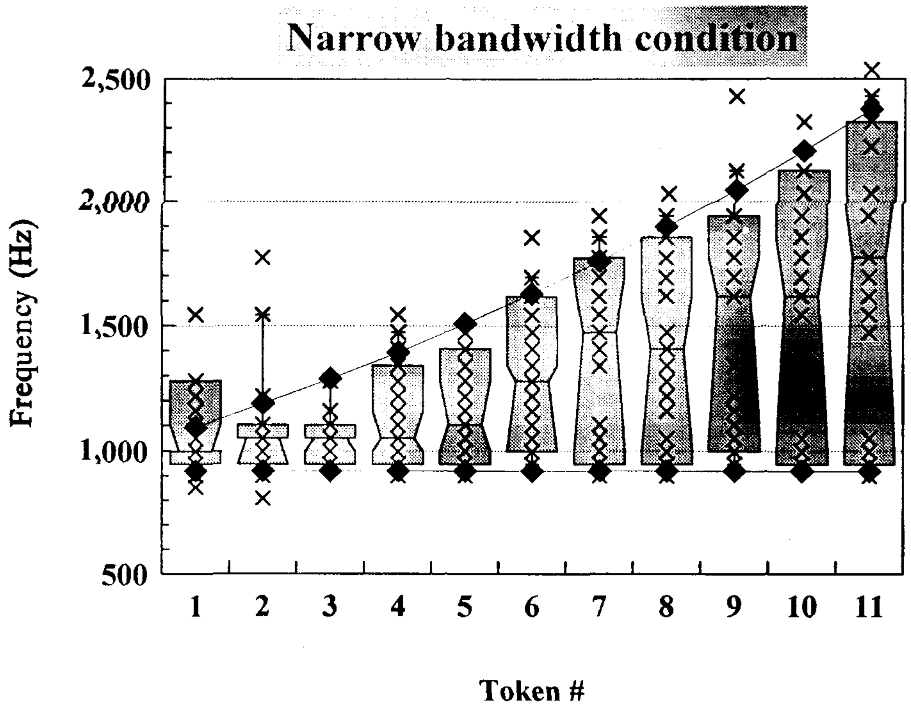

Figure 6. Results of Experiment 2. Data from the narrow bandwidth condition.

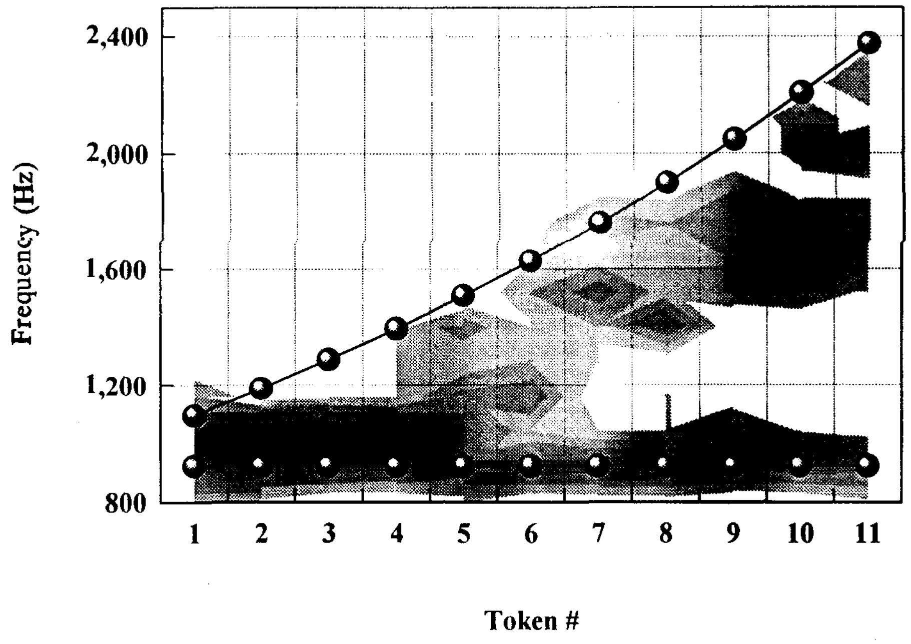

Figure 7. Results of Experiment 2. Data from the narrow bandwidth condition shown in a spectral plot. The number of responses to each two-formant token having a particular F* value is indicated as a shade of gray; the darker the shading, the greater the number of responses.

Experiment 3

Earlier we stated the spectral integration hypothesis this way: "When two formants are within some critical distance they merge into a single perceptual formant". Experiment 2 found evidence in favor of this hypothesis. Experiment 3 was designed to further investigate this hypothetical spectral integration.

The experiment was a very simple discrimination task. We saught to discover whether listeners can detect the difference between two-formant stimuli and their best-matching one-formant counterparts. For stimuli in which the two formants are far apart we can't really talk about the best-matching one-formant stimulus because the distribution of best-matches is bimodal, and even when you compare the most common matching tokens they sound nothing like the two-formant stimuli. However, when the two-formants are close to each other, the one- and two-formant stimuli do sound surprizingly similar (to our ears). Klatt (1985) and Hanson & Javkin(1990) asked listeners to rate the similarity of tokens which had the same calculated spectral center of gravity, but different formant frequencies and formant amplitudes. They found that sounds with the same center of gravity were judged to be less similar than acoustically identical sounds. Hanson & Javkin (1990) attributed the difference between matching results and similarity judgements to differences in the tasks. They described the matching task as "linguistic" because it calls for a linguistic judgement. The distinction that Hanson & Javkin (1990) make between linguistic judgements and other sorts of judgements is important, but not very well developed. Our understanding of this distinction is somewhat different from theirs.

When thinking about the relevance of spectral integration in vowel perception it is important to distinguish between (1) information used to identify sounds, and (2) ways that sounds can be similar. We assume that any aspect of a sound in its auditory representation may be used for identification, while judgements of sound similarity may disregard auditory differences (or ignore some of the available information). Therefore, the question we are posing in Experiment 3 can be stated this way: Does spectral integration affect the information available for vowel identification, or does it affect sound similarity only? Many researchers have assumed that 3-Bark integration is relevant for vowel identification, because they have assumed that the merged 'perceptual formant' is an **auditory** property of sounds with close formants. When two formants are closer than 3.5 Bark they become auditorily one and the listener has no record or representation of their former two-formantness. We will call this the **auditory hypothesis**. An alternative hypothesis, which we will call the **perceptual hypothesis**, is that spectral integration data (such as Experiment 2 above) reflects the existence of a dimension of perceptual similarity. However, this perceptual aspect is built up in addition to the auditory representation, and does not replace the more detailed spectral information in the auditory representation. To quote Chistovich et al. (1979), "information about spectrum shape is not exhausted by the centre of gravity location".

These two hypotheses about 3 Bark spectral integration make different predictions about the discrimination of one- and two-formant vowels. The auditory hypothesis predicts that listeners will not be able to discriminate one- and two-formant vowels when the two formants are within the critical distance, while the perceptual hypothesis predicts that listeners **will** be able to distinguish one- and two-formant vowels.

For purposes of this experiment we defined "best-matching" as the average F* value found in the narrow bandwidth condition in Experiment 2. In a pilot experiment, with the authors as listeners, we used long vowels, in the main experiment we used short vowels. We also manipulated the interstimulus interval in the pilot and the main experiment.

Results of the pilot experiment (Figure 8) showed that regardless of "spectral integration" the listeners could easily detect the differences between the one- and two-formant stimuli. This result is consistent with the perceptual hypothesis rather than the auditory hypothesis. But what happens when we bring this discrimination function down from ceiling? We made the task more difficult by reducing the durations of the stimuli from 225ms to 60ms. Discrimination performance for naive listeners at long ISI's was near chance on stimuli 1-3 (Figure 9). It is interesting that the discrimination function is gradual, there is no sudden shift in performance as might be predicted by the auditory hypothesis. Rather, this gradual shift in discriminability seems to be more indicative of a gradual decrease in <u>perceptual</u> similarity.

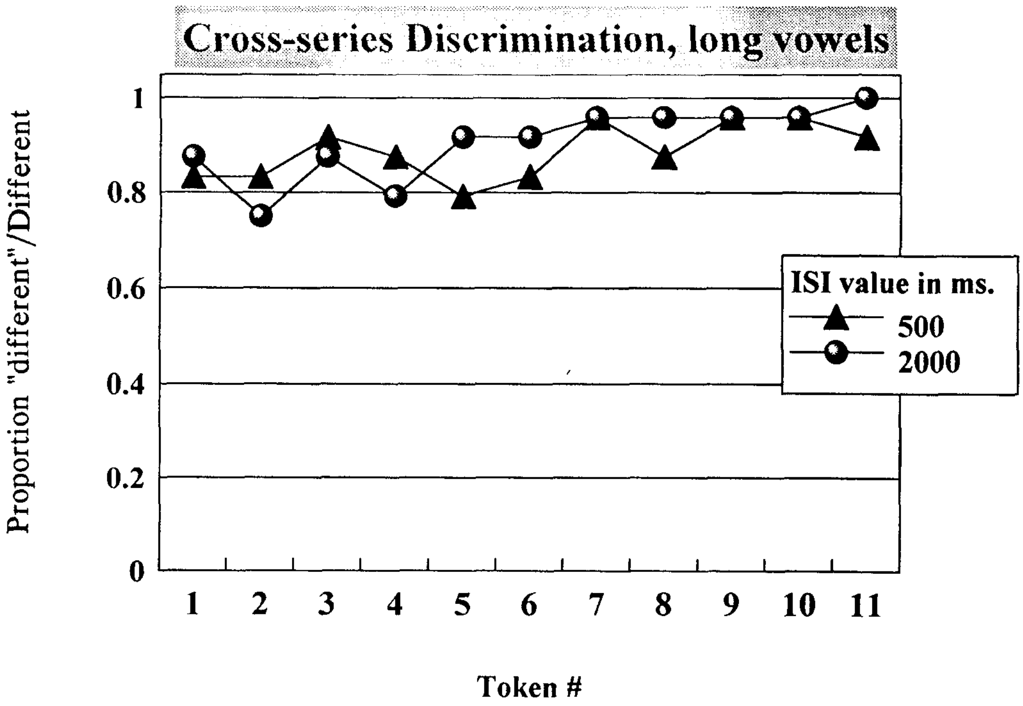

Figure 8. Results of the preliminary cross-series discrimination experiment. The proportion "different" responses for trials in which the stimuli were actually different is plotted by token number for short (triangles) and long (balls) interstimulus intervals.

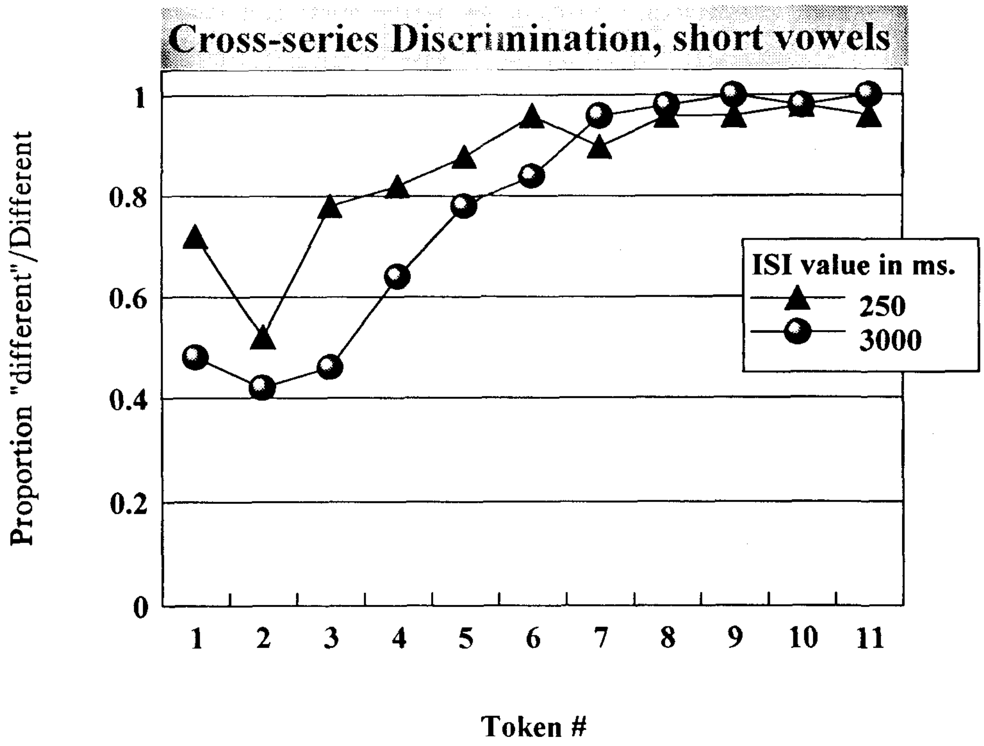

Figure 9. Results of Experiment 3. The proportion "different" responses for trials in which the stimuli were actually different is plotted by token number for short (triangles) and long (balls) interstimulus intervals.

Conclusion

We found evidence in a matching experiment to support the spectral integration hypothesis. The data also confirmed Chistovich et al.'s estimate of the critical distance between formants. Additionally, the results of Experiment 3 suggested that spectral integration affects perceptual similarity, but not auditory representation.

Note

Johnson is now at: Dept of Biocommunication, UAB, Birmingham, AL 35294-0019. The experiments were carried out as undergraduate honors projects (Fernandez - Experiment 1, Henninger - Experiment 2, and Sandstrum - Experiment 3) at UCLA, Spring, 1992. Presented to the 124th meeting of the Acoustical Society of America, November 3rd, 1992.

References

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. *Hearing Research* **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 (B. Lindblom & S. Öhman, Eds.) *Frontiers of speech communication research.* London: Academic Press.

Hanson, B.A. & Javkin, H.R. (1990) Evidence for the three bark integration interval. *STL Research Reports* **2**, 2-1 -- 2-7, Santa Barbara, CA: Speech Technology Laboratory.

Klatt, D.H. (1985) A shift in formant frequencies is not the same as a shift in the center of gravity of a multi-formant energy concentration. *J. Acoust. Soc. Am.* **77**, S7.