AUDITORY SPECTRAL INTEGRATION IN NONTRADITIONAL SPEECH CUES IN DIOTIC AND DICHOTIC LISTENING[1,2]

ROBERT ALLEN FOX AND EWA JACEWICZ

Department of Speech and Hearing Science
The Ohio State University

Summary.—Underlying auditory processes in speech perception were explored. Specifically of interest were the stages of auditory processing involved in the integration of dynamic information in nontraditional speech cues such as the virtual formant transitions. These signals utilize intensity ratio cues and changes in spectral center-of-gravity (instead of the actual formant frequency transitions) to produce perceived F3 glides. 6 men and 8 women (*M* age = 24.2 yr., *SD* = 2.1), recruited through posted materials from graduate students at The Ohio State University, participated in two experiments. The results for frequency-based formant transitions (Exp. 1) indicated that spectral cues to syllable identification are combined at more central levels of auditory processing. However, when the components of the virtual formant stimuli were divided between the ears in a dichotic listening task (Exp. 2), the results indicated that auditory spectral integration may occur above the auditory periphery but at stages more intermediate rather than central.

The dynamic time-varying acoustic events in speech have been a focus of speech perception research for several decades. The fact that speech is inherently dynamic poses demands on the auditory system to process efficiently both the segmental and prosodic structures of an utterance. Despite overall progress in understanding the acoustic properties of these complex structures, work still needs to be done to explore the basic underlying auditory processes involved in extracting the relevant information from speech signals.

The present study focuses on the perception of dynamic formant transitions in the syllables /da/-/ga/. Vowel formants represent vowel-specific concentrations of acoustic energy in vowel sounds due to the resonance properties of the vocal tract. The frequencies of these formants are determined by the shape and length of the vocal tract. Because vowels occur in syllables and words and not in isolation, configurations of the vocal tract change appropriately to produce the consonant-vowel-consonant sequences. When moving out of a consonant and into a vowel (or vice versa), vowel formants change, since the vocal tract configuration is chang-

ing, and these transitional portions are termed formant transitions. The important contribution of the time-varying dynamic formant transitions in speech perception has been shown in a number of studies (e.g., Strange, Jenkins, & Johnson, 1983; Nearey, 1989).

The formant transitions in the syllables /da/-/ga/ selected for the present investigation are good examples of dynamic acoustic cues because they reflect the mutual influence of both consonant and vowel on the percept of a higher prosodic unit: the syllable. A number of experiments have shown that the direction of the third formant (F3) transition can signal the perceptual distinction between /da/ and /ga/: a falling F3 transition gives the percept of /da/ and a rising F3 transition gives the percept of /ga/ (e.g., Whalen & Liberman, 1987; Fox, Gokcen, & Wagner, 1997; Gokcen & Fox, 2001). It has also been shown that the /da/-/ga/ distinction is still well perceived when F3 transition is replaced by a gliding tone (i.e., the so-called sine wave replication of speech; Remez, Rubin, Pisoni, & Carell, 1981; Remez, Pardo, Piorkowski, & Rubin, 2001). This indicates that it is the glide and its direction and not the actual formant transition that serve as a cue. Moreover, it has also been shown that listeners can differentiate between /da/ and /ga/ when responding to stimuli with a "virtual F3 transition" (Fox, Jacewicz, & Feth, 2008). The term "virtual" was used to express the fact that the percept of a formant transition was created by changing not the formant frequency but the intensity ratio of a pair of stationary sine waves. Thus, in these virtual F3 transitions, the relevant dynamic frequency information was experimentally produced by modification of the intensities of two stationary-frequency components.

The present paper examines the latter effect in greater detail. As reported in Fox, *et al*. (2008), these nontraditional speech cues—virtual F3 transitions—provide information comparable, although not identical, to that obtained from the actual formant frequency transitions. When presented in isolation, i.e., separated from the syllabic context, these signals were clearly detected as either rising or not rising in frequency, indicating that listeners were able to perceive a dynamic frequency change although the frequencies of the spectral components remained constant. These results were attributed to spectral integration effects in the perception of speech, which were first reported in early research on vowel perception, using primarily a matching task. For example, Delattre, Liberman, Cooper, and Gerstman (1952) showed that listeners could match the phonetic quality of back vowel stimuli containing two closely spaced formants, such as those found in back vowels, with vowel stimuli containing only a single formant. In general, the best matches occurred when the frequency of the single formant was close to the middle of the frequency interval between the pair of formants. Another early proposal was the spectral cen-

ter-of-gravity hypothesis advanced by Chistovich and others (e.g., Bedrov, Chistovich, & Sheikin, 1978; Chistovich & Lublinskaja, 1979; Chistovich, Sheikin, & Lublinskaja, 1979; Chistovich, 1985). A number of experiments have demonstrated that the relative amplitude ratio between two closely spaced formants in back vowels also plays a role in the integration of spectral information. In particular, listeners were able to match single-formant tokens with two-formant tokens whose spectral center-of-gravity (determined by the intensity ratios of the formant pair) corresponded to the frequency of the single formant. This effect occurred only within an integration bandwidth of about 3.5 Bark (the Bark scale is an auditory scale, related to the mel scale, developed by Zwicker, 1961). Again, these results can be interpreted as an indication of spectral integration completed by the auditory system when making vowel identification decisions.

Such integration effects in the perception of speech are believed to occur at more central levels of auditory processing, above the level of the cochlear filtering, and not necessarily in the auditory periphery (e.g., Chistovich, 1985). The auditory mechanism that integrates the spectral information available in two separate vowel formants into one single "perceptual" formant (which is not physically present in the original signal) is understood as a manifestation of more central processing. For example, as shown in Fox, *et al*. (2008), listeners were able to hear frequency glides although there were no frequency changes present in the signal, i.e., unavailable in the auditory periphery. This suggests that they responded to dynamic modifications of the intensities of stationary components (sine waves), which gave rise to a movement of the spectral center-of-gravity over time. However, one methodological issue needs to be addressed in further pursuit of better understanding of spectral integration phenomena in speech. Namely, all experiments to date have presented the experimental stimuli in a diotic condition, so that the listeners responded to the same spectral content delivered to both ears. Although the results suggest involvement of more central auditory processes, it is unclear where in the auditory processing pathway spectral integration is done, at intermediate stages or at the central stage of signal processing.

To shed more light on this question, the present study used a dichotic listening task which requires involvement of central processing in order to combine different spectral cues. The dichotic listening procedure, first introduced into speech perception research by Doreen Kimura (Kimura, 1961), has been used extensively in speech perception research since the early 1970s, following the discoveries that listeners correctly identified syllables when different components of the syllable were presented simultaneously to different ears (Studdert-Kennedy, Shankweiler, & Pisoni, 1972; Cutting, 1976; Hugdahl & Andersson, 1984; Hugdahl, 1988). The ability to

combine the spectral cues and form a syllable percept was considered evidence for central processing. However, as the research progressed, it became apparent that the ability to combine spectral information in dichotic listening is a property of a well-functioning auditory system. Individuals with some form of auditory processing dysfunction will have difficulties in performing a dichotic listening task, which may signal a specific disorder or may come as a function of aging (e.g., Martin & Jerger, 2005).

Because the dichotic listening task is now a widely used experimental paradigm for studying interhemispheric interactions, it needs to be emphasized that the present study is concerned with normal auditory processing in young adults who are free from any form of hearing impairment. The purpose of the present study was to gain a basic understanding of the stages of auditory processing involved in integrating the intensity cues to form a percept of a dynamic F3 change. The syllables /da/ and /ga/ were therefore presented for identification in both diotic (Exp. 1) and dichotic (Exp. 2) listening conditions only to normal-hearing young adults. Based on previous experiments, it was expected that these listeners could easily integrate spectral information (in order to form virtual F3 cues) in the diotic condition. However, it was an empirical question whether they would show a similar performance on a dichotic task.

If integration of spectral information in processing of virtual F3 transitions occurs at the central level of auditory processing, listeners should show good performance in the dichotic task, and the results should be comparable with those from the diotic condition. This is because the dichotic condition constitutes a testing ground for forming a syllable percept at more central levels from partial information delivered to two ears simultaneously. The question arises how "complete" the information must be in order for integration to occur. Given that listeners use amplitude cues to interpret the missing dynamic frequency change, two additional possibilities of the syllable split between the ears were tested, each of which required combining different partials of the virtual F3 transitions across the two channels. The question was whether intensity ratios could still be combined across the channels or if the auditory system utilizes these intensity cues only when they are presented to the same ear, or in other words, whether there is a limit to auditory integration of partial cues across the channels. The two experiments presented below were designed to elucidate these questions.

<div align="center">

EXPERIMENT 1

METHOD

</div>

Participants

Fourteen listeners (6 men and 8 women) served as participants in Exp.

1. One additional listener was initially enrolled but withdrew from the experiment because she was unable to do the task. All remaining participants were undergraduate and graduate students at The Ohio State University and were recruited through posted materials. Their ages ranged from 19 to 28 years ($M = 24.2$, $SD = 2.1$). All of them were native speakers of a Midwestern variety of American English. All individuals had normal hearing and were paid $10.00 per session to participate. The research was approved by the Institutional Review Board Committee at The Ohio State University.

Stimulus Materials

The actual F3 tokens.—The first set of stimuli was a 9-step synthetic three-formant series created with the parallel branch of the HLSYN synthesizer (Sensimetrics Corporation). In this series, the endpoints represented the syllables /da/ and /ga/ and all stimuli contained an actual F3 transition. This series is referred to as the Actual F3. The tokens were 250 msec. in duration and consisted of an initial transition (50 msec.) and a syllable "base" (200 msec.). All three formants in the syllable base were steady-state. All three formants in the steady-state base and the F1 and F2 transitions remained the same for all tokens and only the F3 transition varied (in nine equidistant steps). The frequencies of F1, F2, and F3 in the base were 700, 1220, and 2600 Hz, respectively. The F1 transition rose from 443 to 700 Hz and the F2 transition fell from 1520 to 1220 Hz. The onset of the F3 transition varied in 100-Hz steps from 1800 to 2600 Hz and the offset of each transition was fixed at 2600 Hz. The fundamental frequency was 120 Hz, and the tokens were synthesized using a sampling rate of 11.025 kHz. These tokens are represented schematically in the left panel of Fig. 1.

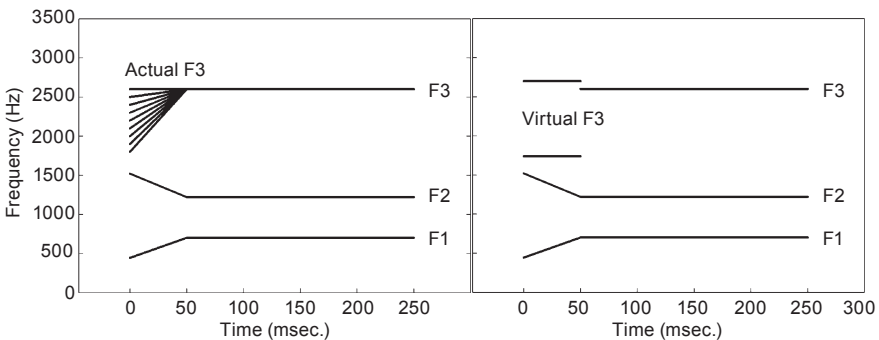*The virtual F3 tokens.*—A second series of 9-step stimuli was created

FIG. 1. Schematic representation of Actual F3 stimulus series (left) and corresponding Virtual F3 series (right) in which the F3 frequency transition is replaced by two sine waves whose intensity ratio changed in steps.

on the basis of the *Actual F3* tokens. The only difference was that the actual F3 transition was replaced by a nontraditional F3 cue which is termed the Virtual F3 (see also Fox, *et al.*, 2008). This "virtual" F3 transition was created in the following way. First, two steady-state 50-msec. sine waves were generated whose frequencies were 1740 Hz and 2700 Hz. These frequencies were chosen as being below and above the frequency of the actual F3 transition in the first series. Next, the relative intensities of these sine waves were changed dynamically across their 50-msec. durations. This was done for each of the nine steps using linear interpolation between the onset and offset intensities of the lower and higher frequency sine waves as listed in Table 1. These particular intensity values were used so that the frequency changes in the spectral center-of-gravity (which was determined by the intensity ratios) matched the actual F3 frequency transition glide in the first series. Thus, the change in the intensities over the 50-msec. sine waves created a percept of a frequency glide which was comparable with the frequency change in the 50-msec. F3 transitions in the Actual F3 tokens. The sine waves with their corresponding intensities were then inserted into the base token (for each of the nine steps) and the mean root-mean-square amplitude of the composite F3 was adjusted to match that of the actual F3 transition.

TABLE 1

Targeted Onset Frequencies of F3 Transitions With Corresponding
Relative Onset Intensities For Virtual F3 Series

Step	F3 Onset Target (Hz)	Relative Intensity at Onset of	
		1740 Hz Sine Wave	2700 Hz Sine Wave
/ga/ 1	1800	0.937	0.063
2	1900	0.833	0.167
3	2000	0.729	0.271
4	2100	0.625	0.375
5	2200	0.521	0.479
6	2300	0.417	0.583
7	2400	0.313	0.687
8	2500	0.208	0.792
/da/ 9	2600	0.104	0.896
F3 offset (vowel onset)	2600	0.104	0.896

Procedure

Each listener was seated in a sound-attenuating booth facing a computer monitor. The experimenter and the computer controlling the experiment were outside the booth. The experimental tokens were presented binaurally over TDH-49 headphones at 70 dB SPL. One interval 2-AFC identification procedure was used with the response choices /da/ and /ga/. Listeners were instructed to identify the syllable and click with a mouse

button on the appropriate box on the screen corresponding to either /da/ or /ga/. The stimuli were presented in two experimental blocks, one containing all Actual F3 tokens and the other all Virtual F3 tokens. The presentation order was counterbalanced across the listeners. In each block, a total of 135 tokens was presented in a random order (15 repetitions of each of the nine steps). A 15-item practice run was administered prior to the experiment to familiarize the listener with the task. No feedback was given in either the experimental block or the practice run.

Results

Listener identification responses are shown in Fig. 2. As can be seen, the slope of the identification function of the Actual F3 tokens is steeper and represents a typical "categorical" response on the /da/-/ga/ continuum. The response pattern to the Virtual F3 tokens was similar, indicating that listeners had no substantial problems in identifying the stimuli as either /da/ or /ga/. However, despite the apparent similarity, there were also differences between these two identification functions. In particular, the number of /da/-responses at the /da/-endpoint in the Virtual F3 set was lower compared to the Actual F3 and was comparatively higher for the first five steps. This pattern yields a shallower identification function for the Virtual F3 series, indicating that the intensity cues gave rise to an increased number of /da/-responses whereas listeners heard /ga/ more often when the actual F3 transition was present.

PROBIT analysis was used to identify the locations of the /da/-/ga/
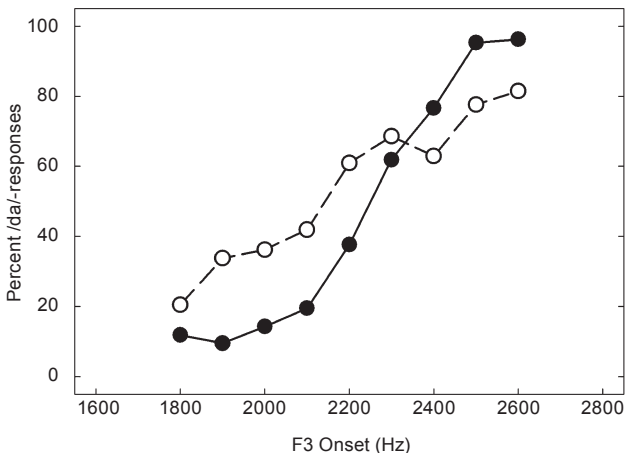

FIG. 2.   Identifications as /da/ and /ga/ in response to Actual F3 and Virtual F3 stimulus series in the diotic listening task. Actual F3 (——●——); Virtual F3 (——○——).

50% cross-over point (representing the category boundary) of the identification functions for both of the stimulus series for each individual listener. A paired-samples t test indicated a significant difference (t_{13}=3.72, p=.003) between the category boundary for the Actual F3 series (M=2229 Hz) and the Virtual F3 series (M=2121 Hz). Cohen's d was 0.994, a large effect (see table in Cohen, 1988, p. 262). Although both types of F3 transition produced generally similar identification functions, replacing the actual F3 with the virtual F3 shifted the location of the /da/-/ga/ category boundary toward the /da/-endpoint by 104 Hz. Next, using linear regression, the slope of each identification function was calculated for all listeners. A paired-samples t test showed that the slope of the identification function for the Virtual F3 series (M=.08 pct/Hz) was significantly shallower (t_{13}=5.45, p<.001) than the slope of the identification function for the Actual F3 series (M=.130 pct/Hz). Cohen's d was 1.48, again indicating a large effect size. These analyses can be interpreted as showing that although the virtual F3 transition can provide the appropriate frequency cue for the distinction of /da/-/ga/, it is not as salient a cue as the actual F3 transition.

EXPERIMENT 2

In Exp. 2, the stimuli were presented dichotically so that different parts of the syllable were played simultaneously to two different ears (in two separate channels). Given that the identification function from the Virtual F3 condition in Exp. 1 was similar to that from the Actual F3 condition despite being shallower, it was predicted that identification function in dichotic listening to these Virtual F3 tokens would follow the same pattern. However, Exp. 2 also explored the integration of intensity cues in syllable parts across the channels to better understand the level of auditory processing involved in interpretation of such syllabic partials.

Method

Participants

The same listeners as in Exp. 1 participated in this experiment. Exp. 2 was administered in a separate session, approximately a week after Exp. 1, depending on listener availability.

Stimulus Materials

The actual F3-1 tokens.—These tokens were a dichotic version of the Actual F3 tokens presented in Exp. 1. In the dichotic task, the syllable was partitioned so that part of the base (consisting of F1 and F2, including the transitions) was played to one ear and the remaining syllable components, i.e., the nine F3 transition steps and the steady-state F3, were played to the other ear. The schematics in Fig. 3 illustrate this partition. All synthesis parameters were exactly the same as in the Actual F3 tokens.

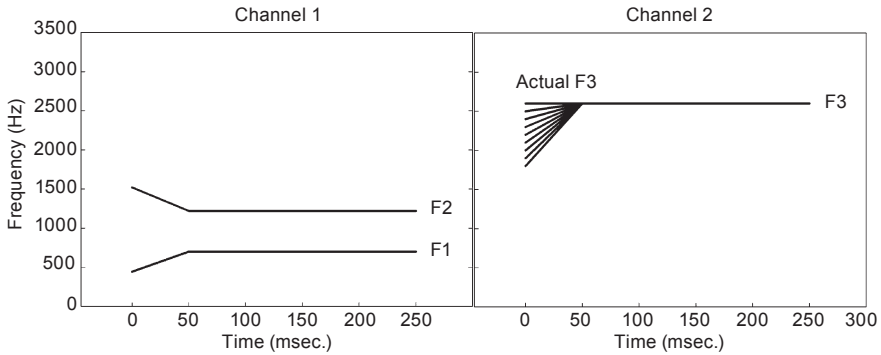*The virtual F3-1 tokens.*—These tokens were also a dichotic version of

FIG. 3.   Actual F3-1 stimulus series presented in the dichotic condition: a base token (left) is presented to one ear and the nine F3 steps (containing both F3 transition and the steady-state F3) to the other ear.

the Virtual F3 tokens presented in Exp. 1. In the dichotic task, the partitioning of the syllable was first done in a manner similar to the Actual F3-1 series, presenting the F1 and F2 base in one channel and the virtual F3 transition steps (consisting of two sine waves with their changing intensity ratio) along with the steady-state F3 in another channel. Fig. 4a illustrates this partition.

The virtual F3-2 and F3-3 tokens.—In addition to the partitioning in the Virtual F3-1 *series*, two other possibilities of the syllable split were tested to assess whether the spectral integration occurs only when the sine waves and their intensity ratios are presented to the same ear or whether these components can be divided between the ears. Fig. 4b illustrates the second possibility, in that the lower sine wave is presented in one channel with the F1 and F2 base and the upper sine wave is presented with the steady-state F3 in the other channel. Note that in this partitioning, the intensity ratios of the sine waves are divided between the ears and have to be combined by the auditory system to form the percept of a glide. Operationally, this type of stimulus split will be called the Virtual F3-2 series. The third possibility of syllable partitioning is presented in Fig. 4c. This time, the upper sine wave is presented with the F1 and F2 base in one channel and the lower sine wave is presented with the steady-state F3 in the other channel. This type of stimulus split will be referred to as Virtual F3-3. In all virtual F3 series in the dichotic task, all synthesis values, the frequencies of the sine waves and the intensity weighting adjustments were exactly as in the *Virtual F3* tokens in Exp. 1.

Procedure

The same procedure was followed as in Exp. 1. Four blocks of stimuli were presented: Actual F3-1, Virtual F3-1, Virtual F3-2, and Virtual F3-
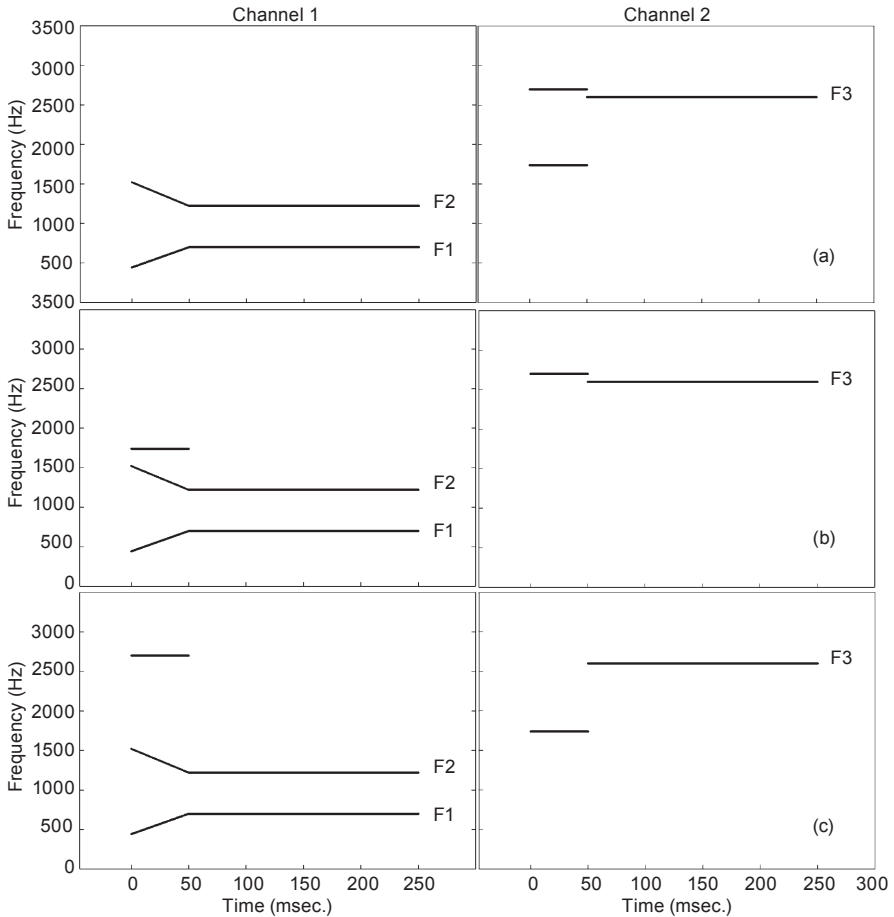
Fig. 4. Partitioning of the two sine waves in Virtual F3 tokens across the channels in dichotic listening: (a) Virtual F3-1, (b) Virtual F3-2, (c) Virtual F3-3. See text for further details.

3. There were 135 stimuli in each block (9 steps × 15 repetitions). The presentation order of the blocks was counterbalanced across the participants. Also counterbalanced was the presentation of the stimuli to different ears so that half the listeners heard Channel 1 in the left ear and half heard it in the right ear. A 15-item practice run was presented at the beginning of each block. Participants were allowed to take breaks during the testing session as needed.

RESULTS

The results for the dichotic listening are displayed in Fig. 5. Responses to the actual (Actual F3-1) and the three virtual series presented in the
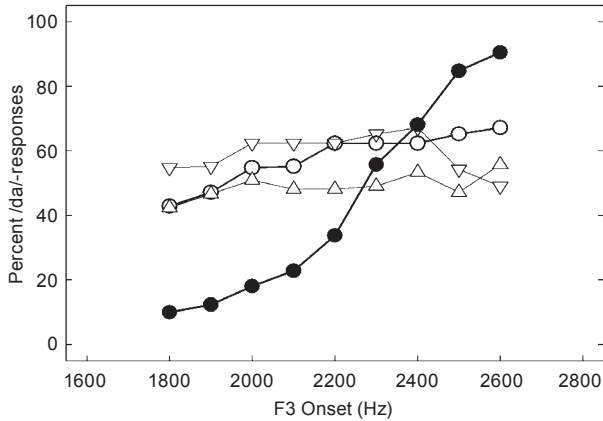
FIG. 5.   Identifications as /da/ and /ga/ in response to the Actual F3 and Virtual F3 stimulus series in dichotic listening. Actual F3-1 (——●——); Virtual F3-1 (——○——); Virtual F3-2 (——▽——); Virtual F3-3 (——△——).

dichotic condition are shown. As can be seen, the identification function for the dichotic Actual F3-1 appears very similar to that in the diotic condition (Actual F3, see Fig. 2). Again, category boundaries of the identification functions for each individual were calculated using PROBIT analysis. A paired-samples t test showed no significant difference in the category boundary between the Actual F3-1 in the dichotic condition ($M = 2263$ Hz) and the Actual F3 in the diotic condition ($t_{13} = -1.65$, ns). Although the mean slopes of the identification functions in the diotic ($M = .126$ pct/Hz) and dichotic ($M = .1126$ pct/Hz) were similar, a paired-samples t test showed a significant difference ($t_{13} = 2.96$, $p = .11$). Cohen's d was .790, indicating an intermediate effect size. These results indicated that the listeners were able to combine effectively the information in F3 transition and the syllable "base" in dichotic listening, although there was a slight decrease in the salience of the F3 cue (as shown by the shallower identification function). However, there was more drastic difference in the identification functions between the diotic and dichotic conditions for the virtual F3 stimuli.

The identification function most closely matching the responses to the Actual F3 stimuli was found in response to the Virtual F3-1 series, in which both sine waves and their changing intensity ratios were presented to the same ear. The mean category boundary for Virtual F3-1 in the dichotic condition ($M = 2206$ Hz) was not significantly different from either the Virtual F3 in the diotic condition ($t_{13} = -.96$, ns) or the Actual F3-1 in the dichotic condition ($t_{13} = .69$, ns). However, the slope of the Virtual F3-1 identification function ($M = .030$ pct/Hz) was significantly shallower

than for either Virtual F3 (t_{13}=6.28, p<.001, Cohen d=1.68) or the Actual F3-1 (t_1=8.55, p<.001, Cohen's d=2.28); both represented a large effect size. Although the identification function was much shallower, its shape generally followed that of the Actual F3-1, indicating that listeners were still able to integrate the spectral information when it was presented to the same ear. However, Virtual F3-1 clearly represents a less salient cue to the /da/-/ga/ distinction.

A very different set of results was obtained when the two sine waves were split between the channels and presented separately to each ear. In each case (Virtual F3-2 and Virtual F3-3), the responses yielded a basically flat identification function. Although the mean slopes for the two identification functions were .009 and .017 pct/Hz, respectively, these values actually overestimate the slopes as the median values were .002 and .001 pct/Hz, respectively (and, for the majority of listeners, the slope was near zero). The median slope for Virtual F3-1 was .027 pct/Hz, close to the mean. These flat slopes indicate that listeners could not make a consistent distinction between /da/ and /ga/ percepts based on the F3 cues in the Virtual F3-2 and Virtual F3-3 stimuli.

Another way to assess whether listeners were able to do the necessary spectral integration in the dichotic condition is to look at the responses to the stimuli near the endpoints. Shown in Fig. 6 are the mean number of
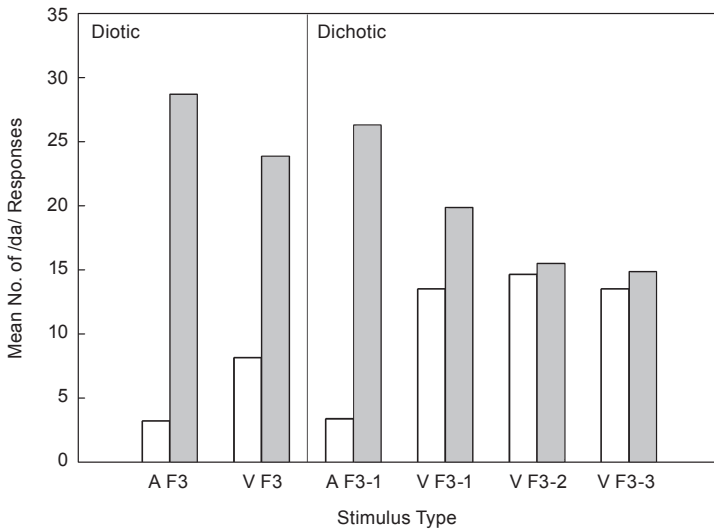

Fig. 6. Mean number of /da/-responses at endpoint stimuli in diotic and dichotic conditions (Steps 1 and 2 correspond to the /ga/-endpoint and Steps 8 and 9 to the /da/-endpoint). Steps 1–2 (□); Steps 8–9 (▣).

/da/ responses to either Steps 1 and 2 (the /ga/-endpoint) or Steps 8 and 9 (the /da/-endpoint) to all stimuli in both the diotic and dichotic conditions. These steps of each continuum should contain the clearest cues to place of articulation. Again, it is clear that the virtual F3 cues were not as salient as the actual F3 cues, and that the virtual F3 cues were less salient in the dichotic condition than the diotic condition. However, there was a significant difference in the number of /da/ responses to Steps 1 and 2 vs Steps 8 and 9 for the Virtual F3-1 continuum ($t_{13} = -4.15$, $p = .001$; Cohen's $d = 1.11$, a large effect size). This difference was not significant for either the Virtual F3-2 ($t_{13} = -.78$, ns) or Virtual F3-3 ($t_{13} = -.79$, ns) continua.

Taken together, these data suggest that spectral integration of the two sine waves does not occur across channels, only within the same channel. This supports the claim that spectral integration of the two sine waves occurs *prior* to being combined with the other cues and thus it is not a central auditory process. Evidence for spectral integration for the F3 cue was found only when both frequency components were presented to the same ear.

DISCUSSION

The purpose of the present studies was to gain a better understanding of the stages of auditory processing involved in integrating the dynamic spectral information in virtual F3 transitions. Examining the baseline performance of the listeners who responded to Actual F3 transition series, their responses were very similar in both the diotic and dichotic listening tasks. The results for dichotic listening verified earlier reports in the literature and provide further support that spectral cues to syllable identification are combined at more central levels of auditory processing prior to phonetic categorization. However, a different set of results was obtained for the Virtual F3 series.

Recall that in the Virtual F3 stimuli, listeners made their identification decisions on the basis of the changing intensity ratio, or a moving spectral center-of-gravity, of stationary sine waves in lieu of the actual F3 transition. Although the identification function in response to the virtual F3 stimuli presented diotically was generally shallower (which was also found in Fox, *et al.*, 2008), listeners were still able to identify the tokens as either /da/ or /ga/ based on the direction of the center-of-gravity movement. However, when the components of the virtual F3 stimuli were divided between the ears in the dichotic listening, accurate identification responses were found only for the Virtual F3-1 series, when the virtual F3 transition was presented in one channel and the base token in the other. The listeners were still able to perceive the Virtual F3 glide and combine the syllabic components across the channels so that the pattern of responses was in the same direction as in the diotic listening. However, the

slope of the identification function was even shallower, which suggests that the virtual F3 cue was not as salient in the dichotic condition. There was no evidence that any spectral integration of the sine waves occurred in the Virtual F3-2 and Virtual F3-3 stimulus sets. The two identification functions showed chance performance, suggesting that listeners were unable to spectrally integrate the two sine waves when they were presented in two separate channels. If listeners cannot perform this spectral integration to produce a changing center-of-gravity, no perceived F3 frequency glide would be expected, and thus no ability to make the /da/-/ga/ identifications.

Although additional experiments are needed to shed more light on this issue, the present results suggest that normal-hearing listeners can perform well in a dichotic task in which the dynamic formant information divided between the ears is frequency-based. Although intensity variations are an integral part of the speech signal and listeners with normal hearing demonstrated ability to integrate intensity cues within a syllabic unit when frequency information in F3 transition was unavailable, the difficulties in processing such signals in a dichotic task were apparent. This indicates that integrating intensity weighting of two spectral components does not occur across the channels and the auditory system may combine these cues only when both components are delivered to the same ear. This finding leads to the conclusion that auditory integration of spectral components occurs prior to combining the spectral cues at central stages of auditory processing. Clearly, spectral integration may occur above the auditory periphery but at stages more intermediate rather than central.

The current set of results needs to be discussed in light of numerous studies which showed that inability to do a dichotic listening task (known as dichotic listening impairment) may signal a form of an auditory processing disorder such as those associated with early onset schizophrenia (Collinson, Mackay, O, James, & Crow, 2009), specific language impairment in children (Cohen, Riccio, & Hynd, 1999), dyslexia (Helland, Asbjørnsen, Hushovd, & Hugdahl, 2007), anterior communicating artery aneurysm (Evitts, Nelson, & McGuire, 2003), or a psychotic disorder (Kaprinis, Nimatoudis, Karavatos, Kandylis, & Kaprinis, 1995). The present results point to the importance of stimulus materials used in the dichotic task. All listeners were normal-hearing and healthy young individuals who had no difficulties with the dichotic task when one type of stimuli was used whereas they were unable to perform similarly with stimuli of a different type. These results clearly indicate that stimulus characteristics and how the spectral material is divided between the ears play an important role in combining information delivered dichotically. This implies that in assessing a dichotic listening impairment, great care needs to

be given to stimulus construction and the levels of signal processing that might be involved in forming the desired percept. In general, more studies are needed to test the potential limits of combining spectral cues in dichotic listening using a greater variety of stimulus materials.

REFERENCES

Bedrov, Y. A., Chistovich, L. A., & Sheikin, R. L. (1978)   Frequency location of the "center of gravity" of formants as a useful feature in vowel perception. *Akusticheskii Zhurnal*, 24, 480-486 (*Soviet Physics–Acoustics*, 24, 275-282).

Chistovich, L. A. (1985)   Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America*, 77, 789-804.

Chistovich, L. A., & Lublinskaja, 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., & 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. Pp. 55-82.

Cohen, J. (1988)   *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cohen, M. J., Riccio, C. A., & Hynd, G. W. (1999)   Children with specific language impairment: quantitative and qualitative analysis of dichotic listening performance. *Developmental Neuropsychology*, 16, 243-252.

Collinson, S. L., Mackay, C. E., O, J., James, A. C., & Crow, T. J. (2009)   Dichotic listening impairments in early onset schizophrenia are associated with reduced left temporal lobe volume. *Schizophrenia Research*, 112, 24-31.

Cutting, J. E. (1976)   Auditory and linguistic processes in speech perception: inferences from six fusions in dichotic listening. *Psychological Review*, 83, 114-140.

Delattre, P., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952)   An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word*, 8, 195-210.

Evitts, P. M., Nelson, L. L., & McGuire, R. A. (2003)   Impairments in dichotic listening in patients presenting anterior communicating artery aneurysm. *Applied Neuropsychology*, 10, 89-95.

Fox, R. A., Gokcen, J., & Wagner, S. (1997)   Evidence for a special speech processing module. In *Proceedings of the 33rdAnnual Meeting of the Chicago Linguistic Society*. Pp. 311-332.

Fox, R. A., Jacewicz, E., & Feth, L. L. (2008)   Spectral integration of dynamic cues in the perception of syllable-initial stops. *Phonetica*, 65, 19-44.

Gokcen, J., & Fox, R. A. (2001)   Evidence for a special speech processing module from electrophysiological data. *Brain & Language*, 78, 241-253.

Helland, R., Asbjørnsen, A. E., Hushovd, A. E., & Hugdahl, K. (2007)   Dichotic listening and school performance in dyslexia. *Dyslexia*, 14, 42-53.

Hugdahl, K. (1988)   *Handbook of dichotic listening: theory, methods, and research*. New York: Wiley.

Hugdahl, K., & Andersson, L. (1984)   A dichotic listening study of differences in cere-
    bral organization in dextral and sinistral subjects. *Cortex*, 20, 135-141.

Kaprinis, G., Nimatoudis, J., Karavatos, A., Kandylis, D., & Kaprinis, S. (1995)   Func-
    tional brain organization in bipolar affective patients during manic phase and
    after recovery: a digit dichotic listening study. *Perceptual and Motor Skills*, 80, 1275-
    1282.

Kimura, D. (1961)   Cerebral dominance and the perception of verbal stimuli. *Canadian
    Journal of Psychology*, 15, 166-171.

Martin, J., & Jerger, J. (2005)   Some effects of aging on central auditory processing.
    Journal of Rehabilitation Research & Development, 42, 25-44.

Nearey, T. M. (1989)   Static, dynamic, and relational properties in vowel perception.
    Journal of the Acoustical Society of America, 85, 2088-2113.

Remez, R. E., Pardo, J. S., Piorkowski, R. L., & Rubin, P. E. (2001)   On the bistability of
    sine-wave analogues of speech. *Psychological Science*, 12, 24-29.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carell, T. D. (1981)   Speech perception with-
    out traditional speech cues. *Science*, 212, 947-950.

Strange, W., Jenkins, J. J., & Johnson, T. L. (1983)   Dynamic specification of coarticu-
    lated vowels. *Journal of the Acoustical Society of America*, 74, 695-705.

Studdert-Kennedy, M., Shankweiler, D., & Pisoni, D. (1972)   Auditory and phonetic
    processes in speech perception: evidence from a dichotic study. *Journal of Cognitive
    Psychology*, 2, 455-466.

Whalen, D. H., & Liberman, A. M. (1987)   Speech perception takes precedence over
    nonspeech perception. *Science*, 237, 169-171.

Zwicker, E. (1961)   Subdivision of the audible frequency range into critical bands.
    Journal of the Acoustical Society of America, 33, 248.