The perceptual representation of voice gender

John W. Mullennix
Department of Psychology, Wayne State University, 71 W. Warren Avenue, Detroit, Michigan 48202

Keith A. Johnson
Department of Linguistics, Ohio State University, Columbus, Ohio 43210

Meral Topcu-Durgun and Lynn M. Farnsworth
Department of Psychology, Wayne State University, Detroit, Michigan 48202

(Received 29 November 1993; revised 2 June 1995; accepted 7 June 1995)

The perceptual representation of voice gender was examined with two experimental paradigms: identification/discrimination and selective adaptation. The results from the identification and discrimination of a synthetic male–female voice continuum indicated that voice gender perception was not categorical. In addition, results from selective adaptation experiments with natural and synthetic voice stimuli indicated that the perceptual representation of voice adapted is an auditory-based representation. Overall, these findings suggest that the perceptual representation of voice gender is auditory based and is qualitatively different from the representation of phonetic information. © *1995 Acoustical Society of America.*

PACS numbers: 43.71.Bp

INTRODUCTION

Listeners are able to identify talkers by their voices with a great deal of accuracy and very little trouble. Even when voices are unfamiliar, listeners have very consistent impressions of a talker's gender, height, weight, etc. It is obvious that we have knowledge about voices in long-term memory (LTM) that is used in perceiving voice, as illustrated by the brief mental hesitation we sometimes experience when we answer the telephone and try to identify the voice of the caller. Despite the research on recognition and memory for familiar and unfamiliar voices (Kreiman and Papcun, 1991; Papcun *et al.*, 1989; Van Lancker *et al.*, 1985a, 1985b), we still know very little about how voice is represented during perception and stored in LTM.

The importance of investigating the representation of voice is underscored by recent findings implicating a close relationship between the listener's use of perceived talker voice information and phonetic perception. For instance, Johnson (1990a) found that the perception of vowels is altered when vowels are embedded within carrier phrases whose intonational contours denote different talkers. He concluded that vowel normalization is mediated through processes that rely on contextual talker-related information (see also Ladefoged and Broadbent, 1957). Other studies have shown that variation in talker identity from trial to trial in an experiment interferes with vowel perception (Assmann *et al.*, 1982; Johnson, 1990b; Verbrugge *et al.*, 1976; Weenink, 1986), consonant perception (Fourcin, 1968), and word recognition (Creelman, 1957; Mullennix *et al.*, 1989; Nygaard *et al.*, 1992; Sommers *et al.*, 1992). These interfering effects of voice suggest that the processing of voice and phonetic information are closely tied together. Further evidence for this close relationship was obtained by Mullennix and Pisoni (1990), who found that the processing of voice and phoneme dimensions in a speeded classification task (Garner, 1974) was integral. In their study, the integrality of processing indicated that the perceptual processing of the phonetic information was contingent on the perceptual processing of voice and vice versa.

Given the close relationship between voice perception and phonetic perception, it is important to assess the specific processes and representations relevant to each. Theories of speech perception have concentrated almost exclusively on the perceptual processes involved in processing acoustic–phonetic information (Klatt, 1989). Our view is that to understand speech perception in its entirety, the perceptual representation of voice must also be delineated. Recently, some research has explored how acoustic–phonetic information is stored in memory, with the hypothesis advanced that speech is stored in terms of prototypes (Kuhl, 1991; Miller *et al.*, 1983; Samuel, 1982). A similar hypothesis has been entertained for voices (Papcun *et al.*, 1989). Although the hypothesis that voices are stored in LTM prototypes has merit, there is little empirical evidence currently available in the literature to support this assertion (although, see Kreiman and Papcun, 1991).

In the present study, our goal is to examine the perceptual representation of talker voice. Specifically, we focus on the representation of voice gender. The perception of voice gender is dependent upon a number of acoustic factors, including fundamental frequency, formant frequencies, and breathiness (Klatt and Klatt, 1990). Previous studies have investigated the relative importance of these acoustic factors in contributing to judgements of voice gender (e.g., Lass *et al.*, 1976; Murry and Singh, 1980). The importance of voice gender in talker voice perception is illustrated by the fact that infants are able to categorize voices into male and female categories at a very early age (Miller, 1983; Miller *et al.*, 1982). Other research has indicated that the primary factor underlying the perceived similarity of pairs of normal voices is male–female categorization (Singh and Murry, 1978). Given the importance of gender in the perception of voice, we decided that exploring voice gender was the logical first step in determining how voice in general is perceptually represented. The series of experiments described below was designed to provide specific evidence regarding how voice stimuli varying in gender are represented in memory.

In order to investigate the perceptual representation of voice gender, we used a twofold approach to the problem. First, the issue of categorical perception for voice was examined. Although there are debates about what categorical perception is and how it pertains to speech perception (e.g., see volume edited by Harnad, 1987), we felt it worthwhile to explore categorical perception with voice gender in order to make some preliminary assessments about voice gender representation. Certainly, there is evidence that categorical perception may be a general perceptual ability, rather than a specialized ability restricted only to speech (Harnad, 1987). It is possible that this general ability could exhibit itself with voice-related stimuli that do not vary in acoustic–phonetic dimensions. If, by using standard identification and discrimination paradigms, it is shown that the perception of voice gender is categorical, then discriminations within voice gender categories should be poor. This result would be consistent with the idea that voice information is converted into a reduced representation during perception, with some detailed information about voice gender "lost." On the other hand, if perception of voice gender is not categorical, then this would suggest that, during perception, detailed information relating to voice gender is retained and is available to voice discrimination processes. This latter result would be consistent with recent findings indicating that episodic information about voice is retained in memory (Palmeri *et al.*, 1993).

The second approach is to assess the perceptual representation of voice gender by using selective adaptation techniques. The results from selective adaptation experiments in speech perception have proven useful in examining levels of perceptual processing and representation (Samuel, 1986; Sawusch, 1986). In particular, these findings provide evidence for at least two levels of processing and representation for phonetic perception: an auditory-based level and a higher abstract level (Samuel, 1986; Sawusch, 1986). In the present study, voice gender stimuli were adapted under various conditions in order to determine the level of perceptual representation appropriate for voice: low-level auditory or higher level abstract. Evidence for a higher level perceptual representation would be consistent with the idea that voice is stored in terms of abstract male and female categories in memory. On the other hand, evidence for an auditory representation of voice would suggest that the representation of voice gender is not abstract, but is based on explicit details about the auditory parameters relevant to voice.

In summary, the perceptual representation of voice gender will be investigated by determining whether voice is categorically perceived and whether an auditory representation or a more abstract representation is appropriate.

I. EXPERIMENT 1

In experiment 1, we were interested in determining whether the perception of a synthetic voice gender continuum is categorical. The term categorical perception, as used here, adheres to the standard definition of Liberman *et al.* (1957) for speech. The test of categoricalness for a male/female voice continuum provides information about the degree to which auditory information specific to voice gender is retained during perception. If perception is categorical, this would suggest that some detailed information about voice gender is lost during perception. If perception is not categorical, this would suggest that information about specific auditory attributes related to voice gender is retained and subsequently used for discriminating among voice stimuli within a gender category.

In this experiment, identification and ABX discrimination procedures were used to assess categorical perception. The ABX task is more memory intensive than other discrimination tasks (Pisoni and Lazarus, 1974). However, this is precisely why we chose the task. With the ABX task, the likelihood of obtaining a discrimination peak in the boundary region between male and female voice is maximized due to the additional memory load (Macmillan *et al.*, 1977). If a discrimination peak is not observed with the ABX procedure, we would be confident that the discrimination results reflect an absence of categorical perception, since discrimination peaks would most likely not be found using other low-uncertainty discrimination procedures (e.g., AX, 4IAX, etc.).

The predictions are as follows. If the perception of voice gender is categorical, a steep "voice boundary" identification function should be observed between male and female ends of the continuum along with a discrimination peak in the voice boundary area. In addition, the observed discrimination performance should fit with predicted discrimination data derived from the identification data (Liberman *et al.*, 1957). If perception is not categorical, a gradually sloping identification function with no sharp boundary should be observed along with the absence of a discrimination peak in the obtained discrimination data. These two alternatives represent the two extremes of categorical versus continuous perception. Any pattern of identification and/or discrimination data falling between these extremes would represent an intermediate finding between categorical and continuous perception.

A. Method

1. Subjects

The subjects were 30 volunteers drawn from introductory and upper-level psychology courses at Wayne State University. Subjects received course credit for their participation. All subjects were native speakers of English who reported no history of a speech or hearing disorder.

2. Stimuli

The stimuli were synthesized speech tokens prepared with an updated version of the Klatt (1980) software synthesis package. The stimuli consisted of 250-ms-duration tokens of the steady-state vowel /i/ ranging perceptually from male to female voice. The vowel /i/ was chosen because the formant values for male and female voice tokens of /i/ do not usually overlap with other vowels. In a pilot experiment, the acoustic factors of fundamental frequency ($F0$), formant

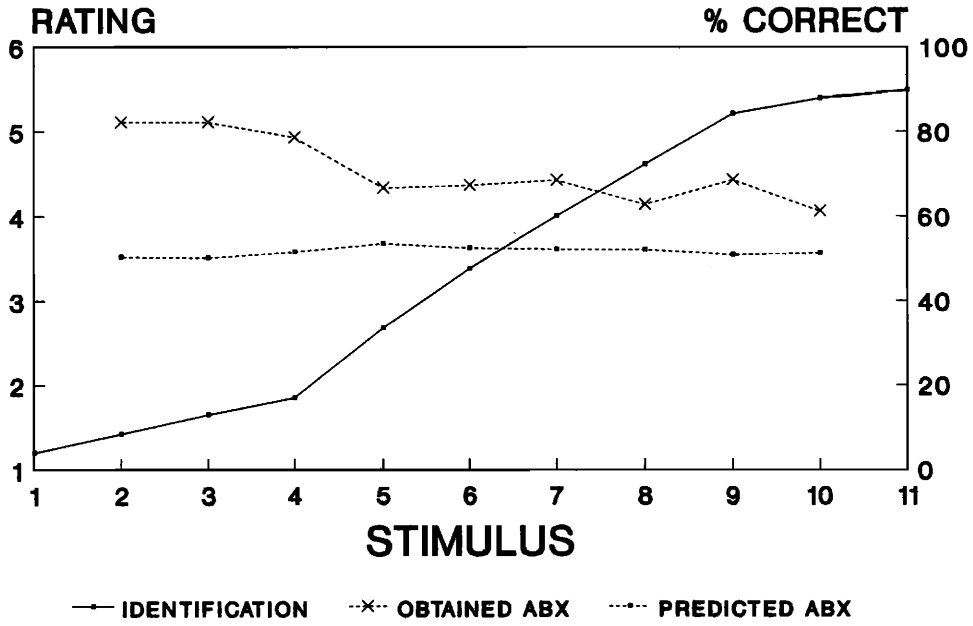

FIG. 1. Identification and ABX discrimination data from experiment 1. The identification data are indexed by the rating scale on the left *Y* axis and the obtained versus predicted discrimination performance in terms of percent correct discrimination is indexed by the percent correct scale on the right *Y* axis.

values ($F1$,$F2$,$F3$), and glottal function (a combination of AH, OQ, and TL) were manipulated to create a set of stimuli presented to 16 subjects for perceptual ratings of voice quality. In the pilot experiment, low $F0$ values and low formant values biased listeners toward "male" voice ratings. However, glottal function values had little effect either way on the classification of stimuli as male or female. Two stimuli corresponding to the most highly rated "male voice" token and the most highly rated "female voice" token were chosen. These two tokens differed in $F0$ and formant values, but the glottal function values were identical and corresponded to the least breathy glottal function. The full synthesis values for these two tokens are listed in Appendix A. These tokens were used as series end points for the synthetic voice continuum used for the experiment. An 11-member synthetic voice continuum varying from male to female voice was generated by incrementing $F0$ and formant values in linear stepwise fashion between the $F0$ and formant values corresponding to the two end points. Glottal function values remained constant across the series.

3. Procedure

The baseline identification trials were presented first. One block of 110 randomized trials (11 continuum stimuli×10 repetitions) was presented. Stimuli were presented one at a time for identification. Listeners were instructed to listen to each stimulus and rate it using a six-point scale from "good male voice" to "good female voice." They were told to use numbers 1 or 6 if the voice was a good exemplar of a male or female voice, numbers 2 or 5 if the voice was clearly male or female but did not sound as good as other voices, and numbers 3 or 4 if they were not sure about the voice gender and had to guess.

After a 2-min break, listeners were presented with the two-step ABX discrimination trials. In the two-step ABX paradigm, two stimuli differing by two places along the continuum (i.e., stimuli 1 and 3, 5 and 7, etc.) are presented as the "A" and "B" stimuli in the trial. The "X" stimulus is identical to either A or B and the listener decides whether the X stimulus matched A or B. 180 randomized trials were presented, with half the trials in the ABA format and half in the ABB format. Each of the three stimulus events on each trial were separated by a 500-ms ISI. For each of the nine possible AB stimulus pairings, the order of stimuli within the pair was counterbalanced. These arrangements produced a total of 20 discrimination responses per stimulus pair. Subjects were given a 1-min break halfway through the trial block. Stimuli were presented on computer using the ONLINE program (Miller, 1990). Stimuli were sampled at a rate of 10 kHz, low-pass filtered at 4.8 kHz, and presented one at a time over AKG K240DF headphones to subjects at a comfortable listening level.

B. Results and discussion

The identification and discrimination data collapsed across subject are displayed in Fig. 1. In experiment 1 and in all subsequent experiments described below, results were also analyzed with a sex of listener factor. Since no effects of sex of listener were observed in any experiment, all results are reported without this analysis variable.

As shown in the figure, the identification function exhibits a gradual slope from male to female voice categories. There is no sharp discontinuity in the boundary region between the male and female ends of the continuum. For the ABX data, two functions are shown: the obtained ABX data from the experiment and the predicted ABX data based on the identification data. The predicted data was obtained by applying the standard formula of Liberman *et al.* (1957) to the identification data for each subject to derive predicted discrimination performance. As seen in the figure, the obtained ABX data consists of high overall discrimination performance with no discrimination peak. This function varies substantially from the near-chance predicted discrimination performance based on the identification data. To verify the differences between obtained and predicted discrimination, a two-way ANOVA with the factors of stimulus pair and data type (observed or predicted) was conducted on the discrimination data to test for differences between observed and predicted discrimination. Significant main effects of stimulus pair [$F(8,232)=3.1$, $p<0.01$] and data type [$F(1,29)=97.4$, $p<0.001$] were obtained. The interaction between pair and data type was also significant [$F(8,232)=5.7$, $p<0.001$]. Newman–Keuls probes of the interaction showed that the obtained data differed significantly from the predicted data for each stimulus pair.

The identification and discrimination results indicate that the perception of the synthetic voice gender continuum is not categorical. There was no sharp discontinuity between male and female categories in the identification data and there was an absence of a discrimination peak in the boundary region. In addition, the discrimination performance was substantially higher than the discrimination performance predicted by applying the Haskins formula to the identification data. The pattern of identification and discrimination performance is typical of many auditory psychophysical functions. The lack of categorical perception indicates that the perception of voice gender is different than the perception of phonetic stimuli, in terms of categoricalness. These results suggest that the perception of voice gender is handled by auditory psychophysical processes. In addition, the results suggest that specific and detailed auditory information about voice gender is retained during perception.

II. EXPERIMENT 2

Although the combined identification and discrimination data from experiment 1 suggest an absence of categorical perception, it is possible that the results were affected by another factor. This factor is that intermediate voice categories could exist in between the male and female end points of the synthetic voice continuum. If such voice categories exist, the presence of these categories could influence the interpretation of the data obtained from experiment 1.

To assess this possibility, a second identification experiment was conducted. In experiment 2, the same stimulus continuum from experiment 1 was used. However, instead of using a six-point male/female rating scale, subjects were given three response alternatives: male, female, or "other." Subjects were told that the other response should be used if they heard a voice that did not fit into the male or female categories. If the number of other identification responses to stimuli in the middle of the continuum turns out to be high, this would suggest that another voice was perceived in the continuum. If the number of other responses to these stimuli is low, this would suggest that subjects in experiment 1 essentially perceived the continuum in terms of the two male/female categories.

A. Method

1. Subjects

The subjects were 22 volunteers drawn from introductory and upper-level psychology courses at Wayne State University. Subjects received course credit for their participation. All subjects were native speakers of English who reported no history of a speech or hearing disorder.

2. Stimuli

The stimuli were the same as used in experiment 1.

3. Procedure

In experiment 2, all aspects of the procedure for presenting the identification trials were identical to experiment 1 except for the response alternatives. Listeners were instructed to listen to each stimulus and respond by pressing a key corresponding to three choices: male, female, or other. Listeners were told that they should use the other response alternative whenever they heard a stimulus that they did not distinctly perceive as a male or female voice.

B. Results and discussion

The identification data collapsed across subject are displayed in Fig. 2. In this figure, the identification data are shown as three separate functions for each response alternative. The pattern of data for the male and female responses was approximately the same as observed for experiment 1. For the other response, this alternative was used primarily to label stimuli from the middle of the continuum. However, when examining the extent to which subjects used this label, identification responses only reached 32% for the stimulus receiving the highest number of other responses (stimulus 6). Also, the overall total number of other responses was 13.1%, compared to 46.9% for the male responses and 39.9% for the female responses. Thus, although there was a tendency for some stimuli to be labeled as an other voice, evidence for the presence of a third voice was weak. Overall, the results from experiment 2 provide little support for the presence of other voices in the stimulus continuum.

III. EXPERIMENT 3

In the next two experiments, the nature of the perceptual representation of voice gender was investigated by using selective adaptation techniques. Selective adaptation in speech perception has a long history (Ades, 1976; Samuel, 1986). Although some researchers have criticized speech adaptation experiments (Diehl, 1981; Diehl *et al.*, 1985), others suggest that the technique is useful in examining the nature of speech perception and representation (Samuel, 1986). The results of adaptation experiments have been used to specify details about the perceptual levels of processing used during speech perception (Sawusch, 1986) and the format of LTM representations of speech (Miller *et al.*, 1983; Samuel, 1982). Here, our assumption about selective adaptation is that the repeated presentation of an adapting stimulus somehow affects or alters the perceptual processing or representation of speech related to the adaptor.

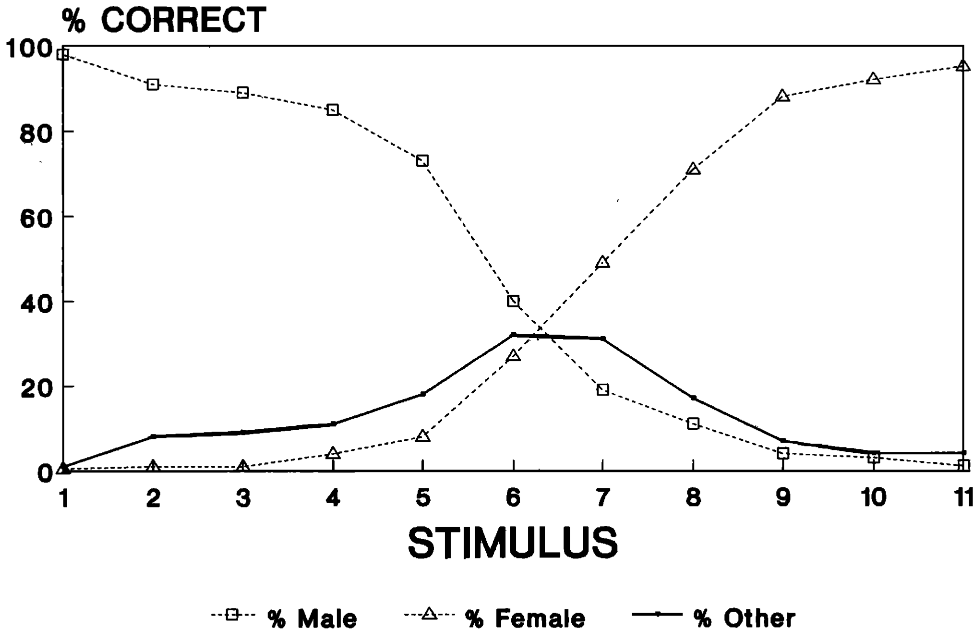

FIG. 2. Identification data from experiment 2. The data are indexed by percent correct labeling on the Y axis for the three response alternatives (male, female, or other).

In experiments 3 and 4, a synthetic male/female voice continuum was used in conjunction with a number of different voice adaptors. The adaptors were varied in terms of their acoustic composition in order to assess whether an auditory representation or a more abstract representation of voice was appropriate. In experiment 3, adaptation with synthetic end-point adaptors was compared to natural voice adaptors that differed slightly from the end points in overall auditory overlap. Overall auditory overlap was defined as a combination of $F0$ and formant value overlap. In experiment 4, synthetic adaptors were used that varied substantially from the end point in one of two acoustic factors, either $F0$ value or formant values. By comparing the results of experiment 3 to experiment 4, two issues can be investigated. The first issue is the effects of degree of auditory overlap on adaptation. The second issue is the relative contribution of each acoustic factor to voice adaptation. By examining adaptation effects in this way, the effects of auditory overlap on voice adaptation can be properly assessed.

In experiment 3, in two conditions, male and female end-point voice stimuli from the continuum were used as adaptors. In two other conditions, naturally produced male and female voice stimuli were used as adaptors. The end-point adaptors were compared to the natural adaptors because we wanted to assess the effects of auditory overlap on voice adaptation. Auditory overlap refers to the degree of acoustic overlap of an adapting stimulus to the continuum end-point stimulus in the continuum being tested. The role of auditory overlap on selective adaptation in speech perception is very important. Some research in speech adaptation has indicated that adaptation does not occur unless there is a substantial auditory overlap of the adaptor with a continuum endpoint (Ades, 1976). These results have been interpreted as evidence for the presence of an auditory level of speech processing and representation (Sawusch and Jusczyk, 1981). Other studies have shown that, in certain situations, adaptation can occur when there is an auditory mismatch between the adaptor and a continuum end point (see Samuel, 1986). These latter results have been interpreted as support for a higher level "abstract" or categorical representation of speech (Samuel, 1986, 1988). Taken together, it appears likely that speech adaptation is related to both auditory-level and higher level processes. In the present study, the auditory overlap of adaptor to continuum is manipulated. The results can provide information about whether the adaptation effects are related to an auditory-based perceptual representation or a more abstract perceptual representation.

In the present experiment, the effects of auditory overlap were assessed for voice adaptation. The male and female synthetic end-point adaptors were congruent to the male and female voice series end points (a 100% auditory overlap of adaptor with end point). But the naturally produced male and female adaptors differed acoustically from the male and female series endpoints. Table I shows the $F0$, $F1$, $F2$, and $F3$ values for the synthetic and natural adaptors. For the

TABLE I. List of frequency values for $F0$ and formants for synthetic and natural adaptors in experiment 3. Difference refers to the raw difference in Hz from the synthetic to the natural adaptors; ratio is equal to the value of the natural parameters divided by the value of the synthetic parameter.

Synthetic adaptors		Natural adaptors	Difference	Ratio
Male	$F0$ 136 Hz	125 Hz	−11 Hz	0.92
	$F1$ 270 Hz	259 Hz	−11 Hz	0.96
	$F2$ 2290 Hz	2074 Hz	−216 Hz	0.91
	$F3$ 3010 Hz	2938 Hz	−72 Hz	0.98
Female	$F0$ 250 Hz	230 Hz	−20 Hz	0.92
	$F1$ 310 Hz	362 Hz	+52 Hz	1.17
	$F2$ 2790 Hz	2506 Hz	−284 Hz	0.90
	$F3$ 3310 Hz	3595 Hz	+285 Hz	1.09

natural adaptors, the auditory overlap from adaptor to series end point varied from 100%. The percent overlap for each acoustic parameter was determined by deriving a ratio value equal to the value of each natural adaptor parameter divided by the value of the corresponding parameter for the appropriate synthetic continuum end point (i.e., male adaptor compared to male end point, female adaptor compared to female end point). This ratio was converted to a percentage overlap value for each separate parameter (see Table I). All parameter values for the natural male adaptor fell below that of the male end point. For the natural female adaptor, the $F0$ and $F2$ values fell below the female end point, while the $F1$ and $F3$ values fell above the end point. Thus the auditory overlap of the natural adaptors to the end points, as defined by a combination of $F0$ and formant values, was close to but not identical to the values for the synthetic end-point adaptors.

The predictions for experiment 3 are as follows. If the synthetic and natural adaptors have the same effect on adaptation, this would indicate that auditory overlap has little effect. This result would be consistent with the hypothesis that an abstract perceptual representation of voice gender was adapted in the experiment. The other alternative is that the amount of adaptation produced by the synthetic and natural adaptors differs. This result would indicate that auditory overlap does affect adaptation, suggesting that an auditory-based representation of voice was adapted.

A. Method

1. Subjects

The subjects were 56 volunteers who participated for course credit. The subjects were drawn from the same pool as the previous experiments.

2. Stimuli

The stimulus continuum was identical to that used for experiment 1. In addition, two naturally produced tokens of the vowel /i/ from a male speaker (JM) and a female speaker (KG) were recorded, digitized, and stored on disk. The duration of the natural male /i/ was 240 ms and the duration of the natural female /i/ was 250 ms. Spectrograms of the natural and synthetic adaptors are shown in Fig. 3.

3. Procedure

There were four separate adaptation conditions: male synthetic adaptor, female synthetic adaptor, male natural adaptor, and female natural adaptor. Fifteen subjects were run in each of the synthetic adaptor conditions and 13 subjects in each of the natural adaptor conditions.

Each listener received stimuli in a baseline condition and in an adaptation condition. The baseline condition was identical to previous experiments. In the adaptation conditions, the same stimuli were presented for identification as in the baseline condition. However, these stimuli alternated with sequences of adaptor repetitions. The adaptor sequence consisted of 50 repetitions of the adaptor, each repetition separated by a 1000-ms ISI. There were ten alternating sequences of trials, with 11 randomized continuum stimuli presented for identification after each adaptor sequence.

For the synthetic male and synthetic female conditions, the adaptors consisted of the male and female end-point stimuli from the continuum, respectively. For the male natural and female natural conditions, the adaptors were naturally produced /i/ vowel tokens spoken by a male and female speaker, respectively.

Each subject received the baseline block of trials first followed by the block of adaptation trials. Each block consisted of 110 trials. Listeners used the same six-point rating scale as before.

B. Results and discussion

The results collapsed across subjects for the four adaptation conditions are displayed in Fig. 4. The figure shows the identification functions before and after adaptation for each of the four adaptor conditions. A visual inspection of the data indicates that the identification function is shifted toward the male end point for the male adaptor conditions and toward the female end point for the female adaptor conditions.

Two three-way ANOVAs with the factors of token (stimulus number in the continuum), condition (baseline or adaptation), and adaptor (synthetic or natural) were run on the combined data from the two male adaptor conditions and the combined data from the two female conditions. This analysis was performed to compare the magnitude of adaptation produced by synthetic versus natural adaptors. Four separate two-way ANOVAs with the factors of token and condition were conducted as followups for each of the four adaptation conditions to assess further the effects of each adaptor.

For the male adaptation data, the three-way analysis on the combined data revealed a significant main effect of token [$F(10,260)=397.1$, $p<0.001$], condition [$F(1,26)=19.5$, $p<0.001$], and adaptor [$F(1,26)=6.8$, $p<0.02$]. A significant interaction of token with adaptor [$F(10,260)=2.8$, $p<0.01$] and token with condition [$F(10,260)=4.8$, $p<0.001$] were observed, but no interaction of condition with adaptor ($F=2.0$, $p<0.17$) and no three-way interaction ($F=1.6$, $p<0.12$) were found. The followup analysis of the synthetic male adaptation condition revealed significant main effects of token [$F(10,140)=188.7$, $p<0.001$] and condition [$F(1,14)=20.5$, $p<0.001$]. The interaction of token with condition was also significant [$F(10,140)=5.3$, $p<0.001$]. Newman–Keuls *post hoc* tests of the interaction showed that differences between baseline and adaptation were reliable for stimuli 1–8 on the continuum.

The followup analysis of the natural male adaptation condition revealed a significant effect of token [$F(10,120)=212.4$, $p<0.001$]. The effects of condition [$F=3.7$, $p<0.08$] and the interaction of token with condition ($F=1.6$, $p<0.11$) were not significant.

For the female adaptation data, the three-way analysis on the combined data revealed a significant main effect of token [$F(10,260)=343.1$, $p<0.001$], condition [$F(1,26)=6.8$, $p<0.02$], and adaptor [$F(1,26)=10.5$, $p<0.01$]. Significant interactions of token with adaptor [$F(10,260)$

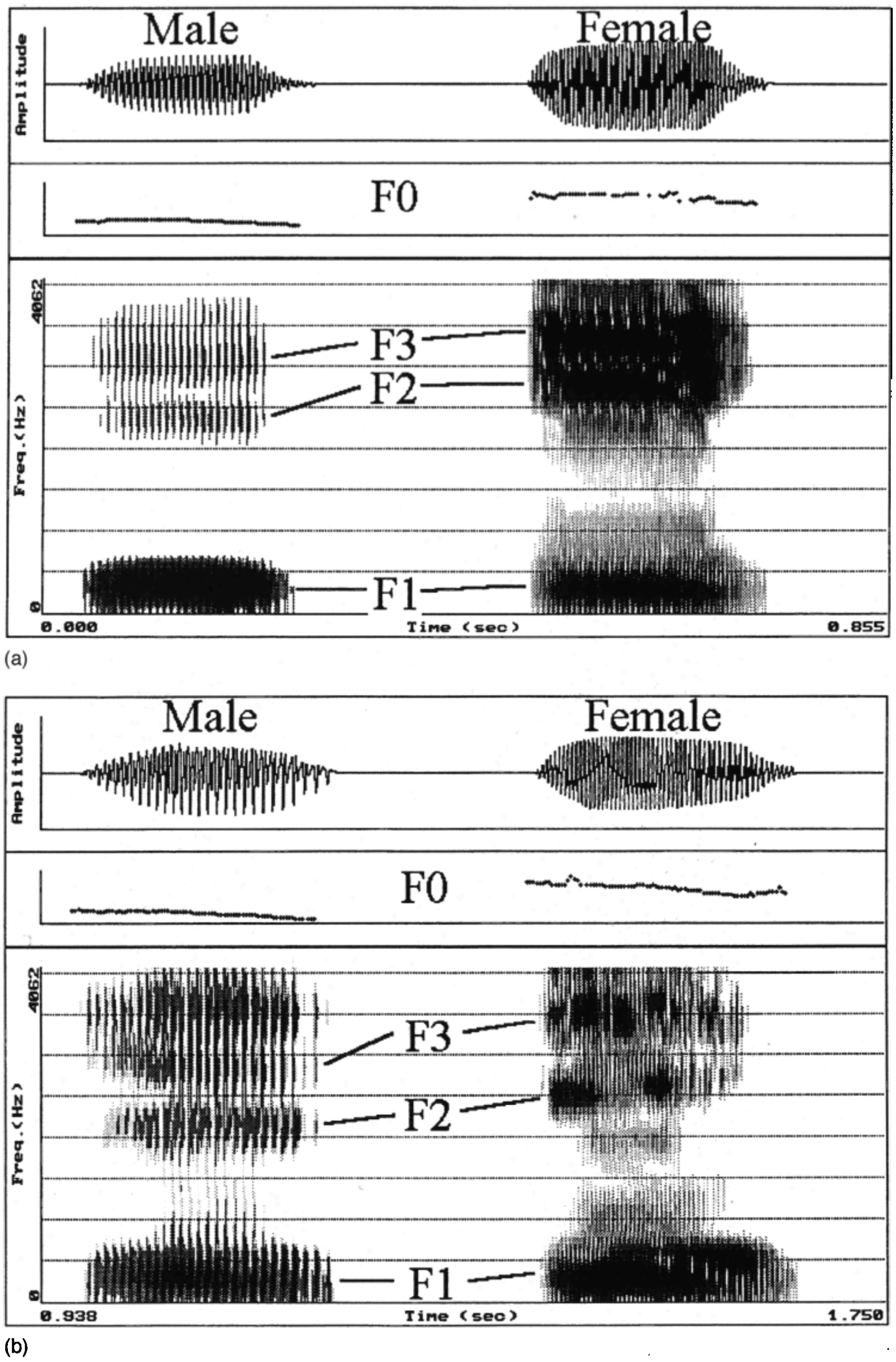

FIG. 3. Spectrograms for the synthetic male and female adaptors. Spectrograms for the synthetic adaptors are shown on top and spectrograms for the natural adaptors are shown at the bottom. Amplitude displays are shown on top for each stimulus, $F0$ values over time in the middle, and spectrographic displays with the center format values for $F1$, $F2$, and $F3$ shown at the bottom.

$=8.1$, $p<0.001$] and token with condition [$F(10,260)=3.5$, $p<0.001$] were observed, but no interaction of condition with adaptor ($F=0.1$, $p<0.79$) and no three-way interaction ($F=1.8$, $p<0.06$) were found. The followup analysis of the synthetic female adaptation condition revealed a significant main effect of token [$F(10,140)=104.7$, $p<0.001$], but not condition ($F=2.8$, $p<0.12$). The interaction of token with condition was significant [$F(10,140)=2.2$, $p<0.03$]. *Post hoc* tests of the interaction showed differences between baseline and adaptation for stimuli 5 and 8–10.

The followup analysis of the natural female adaptation condition revealed a significant main effect of token [$F(10,120)=332.3$, $p<0.001$] and of condition [$F(1,12)=9.1$, $p<0.02$]. The interaction of token with condition was significant [$F(10,120)=3.5$, $p<0.001$]. The *post hoc*

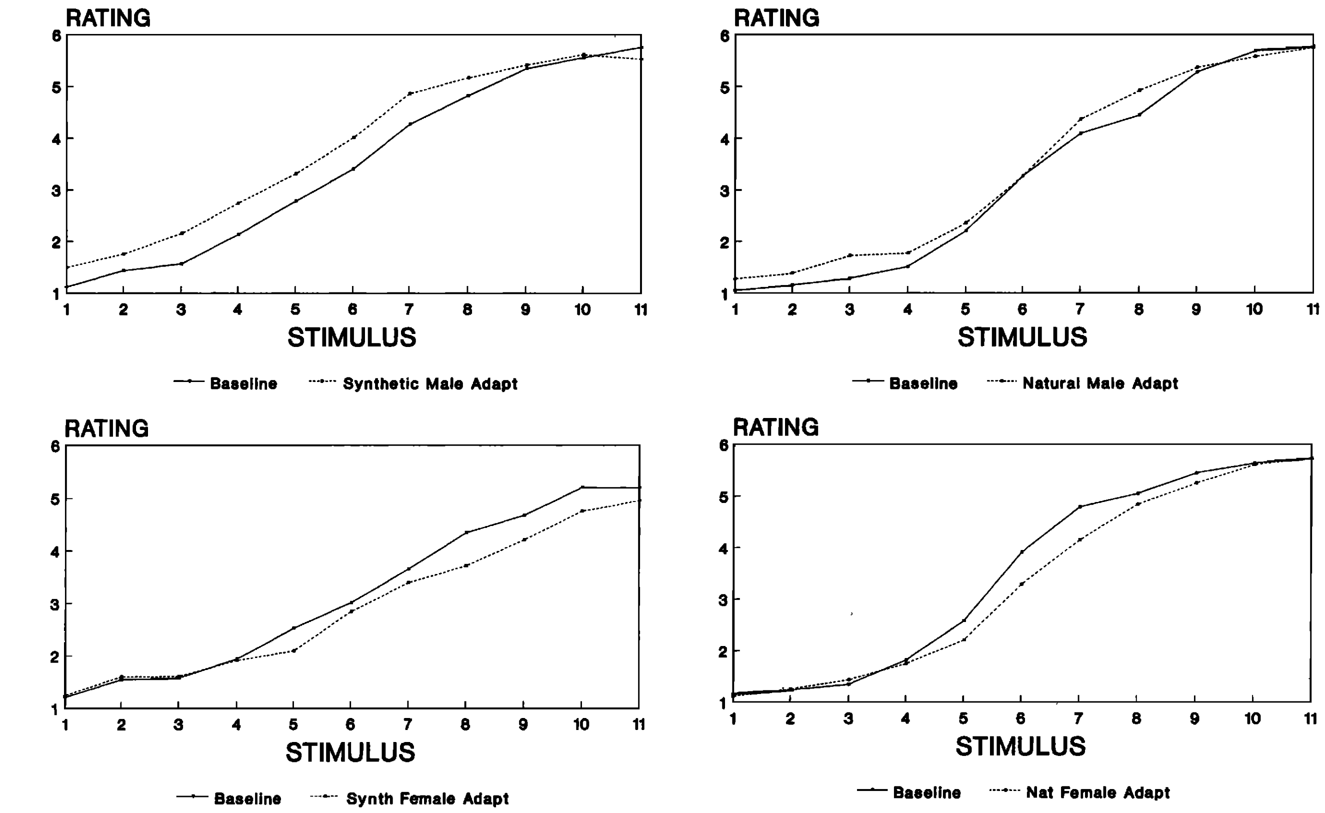

FIG. 4. Baseline and adaptation identification data from experiment 3. The synthetic male and natural male data are shown on the top left and right and the synthetic female and natural female data are shown on the bottom left and right.

tests showed differences between baseline and adaptation for stimuli 5–7.[1]

Overall, the results of experiment 3 indicate that the auditory overlap of adaptor to end point had some effect on adaptation. The first result is that the synthetic adaptors produced significant perceptual changes in the predicted direction toward the end points, as exhibited by the significant token by condition interactions. For the natural adaptors, the male adaptor did not have a significant effect while the female adaptor produced a change similar to the synthetic female adaptor. Although the lack of a condition by adaptor interaction in the overall analyses indicates no differential effects of synthetic versus natural adaptors for male or female adaptation, the individual analyses for each condition show that some differences do exist.

For male voice adaptation, the auditory mismatch of the natural adaptor to the male end point resulted in an absence of adaptation with the natural male adaptor. These results are intriguing when one considers that all values of the acoustic parameters of the natural male adaptor were less than the values of the synthetic male adaptor. One speculative answer is that voice adaptation is tuned to regions of the continuum in a manner similar to that found for phonetic continua (Miller *et al.*, 1983; Samuel, 1982). Under this scenario, even though the male adaptor is unambiguously identified as male, the acoustic parameters of the natural male adaptor may have fallen outside the range of values where it was effective.

The results for the natural female adaptor were different. Although the $F0$ and formant values for the natural female adaptor did not match the female end point, adaptation was still observed. Overall, when considering both the male and female adaptor results, there is weak evidence for the hypothesis that auditory overlap has an effect on adaptation of voice. However, the hypothesis that higher levels of voice processing contribute to adaptation cannot be completely ruled out.

IV. EXPERIMENT 4

Experiment 4 provided a further test of the effects of auditory overlap on adaptation. In experiment 3, auditory overlap was defined as changes in $F0$ and formant values together. However, the adaptation observed could have depended on an auditory match of the adaptor to the continuum end point for the formant values alone, $F0$ alone, or both. In this experiment, two adaptors were used where the formant values and $F0$ values of the adaptors were set up in opposition to one another in terms of values appropriate for male and female series end points. The adaptors in this experiment were designed to test adaptation of the male end of the continuum only. One adaptor possessed formant values identical to the male end point but an $F0$ value identical to the female end point (the "formants adaptor"). The other adaptor possessed an $F0$ value identical to the male end point but formant values identical to the female end point ("$F0$ adaptor"). Thus the voice adaptors used here were ambiguous in

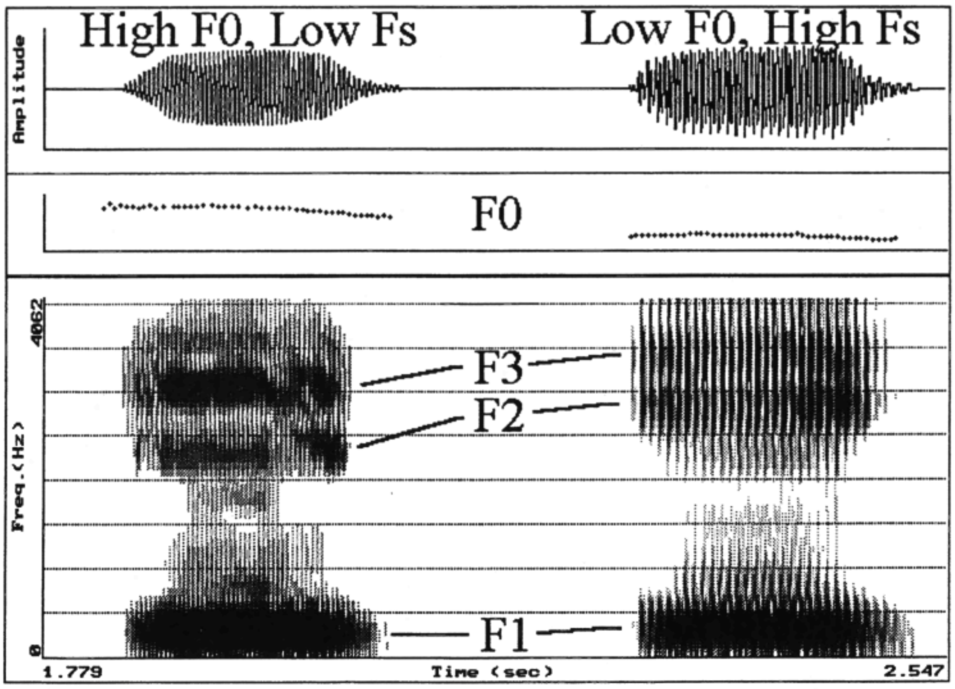

FIG. 5. Spectrograms for the formants and $F0$ adaptors. Amplitude displays are shown on top for each stimulus, $F0$ values over time in the middle, and spectrographic displays with the center formant values for $F1$, $F2$, and $F3$ shown at the bottom.

terms of the major acoustic parameters related to voice gender. However, the perceptual ratings from the pilot experiment described in experiment 1 indicated that there were some differences between the adaptors in terms of voice gender quality. On a five-point rating scale from male to female voice, the $F0$ adaptor received a rating of 1.8 and the formants adaptor received a rating of 3.3. Thus the $F0$ adaptor was rated as more male than the formants adaptor.

The use of these adaptors also allows an assessment of whether perceptual voice quality is responsible for adaptation. If this factor matters, one would expect that a significant amount of adaptation would be obtained with the $F0$ adaptor but not the formants adaptor (although voice information can still be perceived in the absence of $F0$; see Remez *et al.*, 1987). This result would support the involvement of higher-level abstract voice representations in adaptation. On the other hand, the adaptors selected also allow assessments of the acoustic factors that are involved. If formant overlap drives adaptation, then adaptation should be observed with the formants adaptor only. This result would be important because one possibility for the effects of the synthetic adaptors found in the previous experiment was that they were due to simple formant overlap. In the synthetic voice series, formant values changed across the continuum, although all stimuli were identified as /i/. Previous work with auditory/phonetic adaptation has shown that formant overlap or formant pattern overlap results in adaptation of phonetic continua (Ades, 1976; Sawusch, 1977). Since the formant values for the synthetic adaptors completely overlapped with the end points, it is possible that adaptation was driven by an auditory/phonetic formant factor unrelated to voice gender. It is important to assess whether the voice adaptation effects found previously were due to this factor.

Finally, if $F0$ overlap is responsible for adaptation, then adaptation should only be obtained with the $F0$ adaptor. If neither adaptor has an effect, this would indicate that it is the combination of the formant and $F0$ acoustic factors that is important in producing adaptation. This result would strengthen the hypothesis that the perceptual representation of voice gender is based on auditory parameters and not abstract voice representations.

A. Method

1. Subjects

Twenty-eight volunteers from the same pool with the same characteristics as the previous experiments were used.

2. Stimuli

The stimulus continuum was identical to experiment 1. Two adaptors were used that varied in their acoustic and perceptual similarity to the male end-point stimulus. The adaptors were drawn from the pool of stimuli that were rated in experiment 1. Adaptor 1 (the formants adaptor) had formant values identical to the male end point, but $F0$ values identical to the female end point. Adaptor 2 (the $F0$ adaptor) had $F0$ values identical to the male end point, but formant values identical to the female end point (see Appendix B for synthesis values). Spectrograms of both adaptors are shown in Fig. 5. From the pilot experiment, adaptor 1 received a rating of 3.3 on the five-point rating scale from male to female voice and adaptor 2 had a value of 1.8.

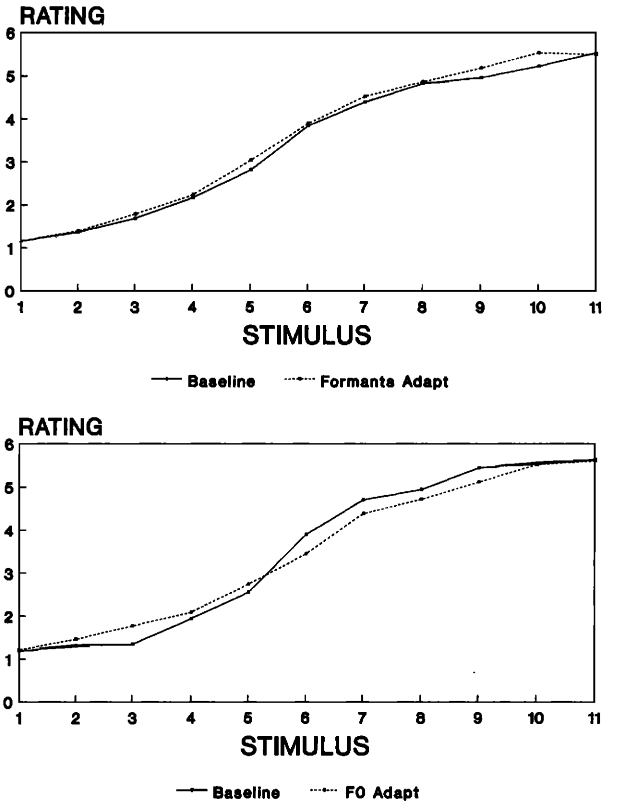

FIG. 6. Baseline and adaptation identification data from experiment 4. The formants adaptation data are shown at the top and the $F0$ adaptation data are shown at the bottom.

3. Procedure

There were two adaptation conditions: formants adaptor and $F0$ adaptor. Fourteen subjects were run in each condition. All other aspects of the procedure were identical to experiment 3.

B. Results and discussion

The results are shown in Fig. 6. At the top of figure are the formants adaptor results and at the bottom of the figure are the $F0$ adaptor results.

For the formants adaptor condition, a significant main effect of token was obtained $[F(10,130)=145.8, p<0.001]$. However, there was no significant effect of condition ($F=1.0$, $p<0.34$) and no significant interaction ($F=0.7, p<0.69$). For the $F0$ adaptor condition, a significant main effect of token was obtained $[F(10,130)=287.8, p<0.001]$, but no effect of condition ($F=0.5$, $p<0.48$). However, a significant interaction of token with condition was observed $[F(10,130)=6.1, p<0.001]$. *Post hoc* tests of the interaction showed differences between baseline and adaptation for stimuli 3 and 6–9.[2]

Overall, there was no evidence for adaptation with either adaptor. Although a significant interaction of token with condition was obtained for the $F0$ adaptor, the identification shift was toward the female end point. If anything, this was an assimilation effect, not adaptation.

There is one possible explanation for the $F0$ adaptor shift toward the female end point. The formant values for the $F0$ adaptor were identical to the formant values for the female end point. If adaptation was driven by formant overlap, then the $F0$ shift could be interpreted as a formant-driven adaptation effect toward the female end of the continuum instead of an assimilation shift. However, this explanation loses strength as one considers that the formants adaptor had no effect. If adaptation was formant driven, both adaptors should have had significant adapting effects in opposite directions. It is unclear why an assimilation effect would occur with the $F0$ adaptor, but the important point is that the adaptation effect was absent. The absence of adaptation effects for either adaptor is important for two reasons: (1) It shows that simple auditory overlap of one isolated voice parameter is insufficient to produce adaptation; and (2) it shows that the perceptual voice quality of the adaptor has little effect on adaptation. The lack of a voice quality effect is consistent with the idea that voice adaptation is related to adaptation of an auditory-based representation of voice. Furthermore, because neither isolated acoustic factor produced adaptation alone, it appears that the perceptual representation of voice gender is one where these acoustic factors are integrated together in an auditory/spectral representation of voice.

V. EXPERIMENT 5

The final experiment was conducted as a further test of the hypothesis that voice adaptation is related to an auditory-based perceptual representation. In experiment 5, a situation was set up where voice adaptors were used that were perceptually ambiguous for each individual subject. Modifying the methods that Sawusch and Pisoni (1976) used for ambiguous adaptors in auditory/phonetic adaptation, the baseline identification performance for each subject was tabulated and a stimulus from the voice continuum closest to the boundary between male and female voice selected as the ambiguous adaptor for that particular subject. Then, each subject was given instructions to bias him/her in terms of the perceptual voice quality of the ambiguous adaptor. Half of the subjects were told that the adaptor was male and half were told the adaptor was female. If voice adaptation is related to how subjects classify voice into categories, then a cognitive instructional manipulation to bias the voice gender of the adaptor should result in adaptation toward the female end point for the female instructions group and adaptation toward the male end point for the male instructions group. However, if voice adaptation is related to auditory-based representations of voice, then the instructions should have no effect and no net difference between baseline and adaptation conditions should be found.

A. Method

1. Subjects

Thirty volunteers from the same pool with the same characteristics as the previous experiments were used.

2. Stimuli

The stimulus continuum was identical to experiment 1. The adaptor for each subject was a stimulus drawn from the continuum.

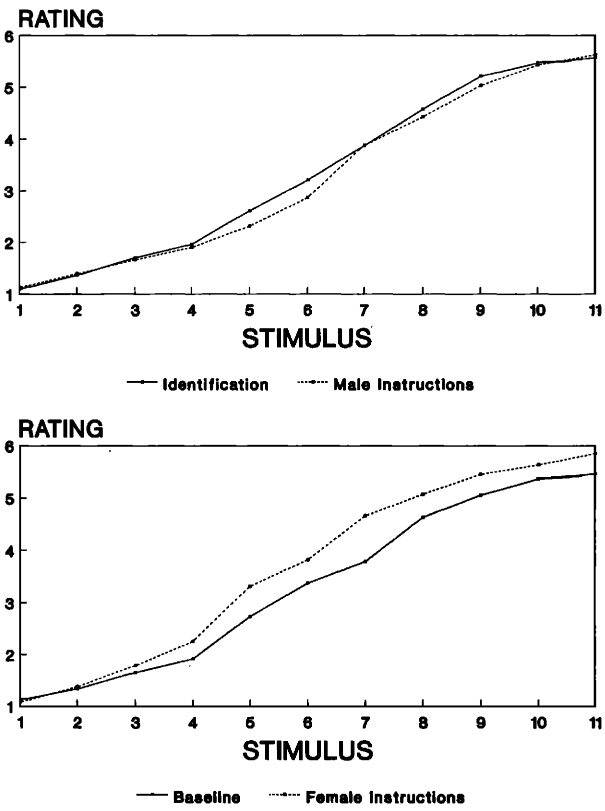

FIG. 7. Baseline and adaptation identification data from experiment 5. The male instructions adaptation data are shown at the top and the female instructions adaptation data are shown at the bottom.

3. Procedure

There were two adaptation conditions: male instructions adaptor and female instructions adaptor. Fifteen subjects were run in each condition. The adaptor for each subject was selected by examining the baseline trials data for each individual subject. The average rating was calculated for each stimulus and the stimulus having a rated value closest to the midpoint value of the male/female rating scale (3.5) was selected as the ambiguous adaptor. Thus, for each individual subject, the adaptor selected was based only on that subject's baseline identification performance. Subjects in the male instructions group were told that they would hear a "male voice" adaptor and subjects in the female instructions group were told they would hear a "female voice" adaptor in the adaptation block. The instructions read by subjects for the adaptation conditions were identical to the instructions used in experiment 4 for the male and female synthetic adaptor conditions. In addition, the experimenters were instructed not to say anything about the ambiguous nature of the adaptor. All other aspects of the procedure were identical to experiment 3.

B. Results and discussion

The results are shown in Fig. 7. At the top of figure are the male instructions group results and at the bottom of the figure are the female instructions group results.

For the male instructions group, a significant main effect of token was obtained $[F(10,140)=350.8,\ p<0.001]$. However, there was no significant effect of condition $(F=2.3,\ p<0.15)$ and no significant interaction of token with condition $(F=1.4,\ p<0.19)$. For the female instructions group, significant main effects of token $[F(10,140)=360.8,\ p<0.001]$ and condition $[F(1,14)=21.3,\ p<0.001]$ were obtained. The interaction of token with condition was also significant $[F(10,140)=3.1,\ p<0.001]$. *Post hoc* tests of the interaction showed reliable differences between baseline and adaptation conditions only for stimuli 4–7, 9, and 11.[3]

The results for the male instructions group were consistent with an auditory-based explanation of voice adaptation. However, the results for the female instructions group were unexpected. When subjects were told that the adaptor was female, they heard the stimuli across the continuum as more female (assimilation). Why would male instructions have no effect but female instructions produce assimilation? One speculative answer is that subjects who received female instructions adjusted their overall criteria to favor female responses to the other stimuli in the series. However, this explanation is *ad hoc*. It is interesting to note that, in experiment 4, assimilation toward the female end point was also found, but was produced by adaptation with the $F0$ adaptor. Although there may be some common mechanism producing this effect in both experiments, the nature of this mechanism remains unclear.

VI. CONCLUSIONS

Overall, the findings from the series of experiments presented here converge to form a preliminary picture of the perceptual representations of voice gender. First, the results from experiment 1 using identification and discrimination procedures indicate that the perception of a synthetic voice gender continuum is not categorical. The absence of categorical perception suggests that the perception of voice gender information in the speech signal is accomplished through auditory psychophysical processes. Furthermore, the high overall discrimination performance indicates that specific auditory information about voice is retained and, at the very least, is available to discrimination processes. Since the perceptual processes that extract voice information from the signal rely on contact with stored representations of speech in LTM, it would appear that voice gender is not represented in memory in terms of abstract representations that contain "reduced" or canonical representations of voice.

Second, the results from the selective adaptation experiments in experiments 3–5 are favorable to the hypothesis that the perceptual representations of voice gender are auditory based. There was some indication in experiment 3 that the auditory overlap of the adaptor to the continuum end point affected adaptation. Interpreting this result within the context of other speech adaptation studies where auditory overlap accounts for adaptation (Ades, 1976; Sawusch, 1986; Sawusch and Jusczyk, 1981), it appears that the perceptual representation of voice gender that was adapted was an auditory-based representation. In experiment 4, further support was given to this idea, as adaptors that shared one acoustic parameter (either formant values or $F0$ value) with a voice end point failed to produce an adapting effect, again suggesting that auditory overlap is important. Finally, in experiment 5, there was no clear-cut evidence in support of a higher level cognitive factor in voice adaptation and by extension no evidence for an abstract representation of voice gender.

When the findings from all experiments are considered together, some differences and similarities between phoneme representations and voice gender representations can be discussed. The most important finding is that voice gender is not stored in abstract male and female voice representations. Instead, voice gender appears to be stored in the form of auditory-based perceptual representations. These representations, in all probability, contain specific auditory information about acoustic voice parameters relevant to gender. The results of experiment 4 suggest that these representations are not based on one isolated parameter like $F0$ or formant frequencies. Instead, the representations are probably an auditory composite of the various acoustic factors relevant to voice gender like $F0$, formant frequencies, breathiness, etc. Although there is a close relationship between phonetic coding and voice coding processes during perception, the representations of phonemes and voices appear to be qualitatively different in that phonetic representations may not be as detailed.

The adaptation effects observed also contrast with studies examining vowel adaptation (Godfrey, 1980; Morse *et al.*, 1976). Vowel adaptation, as assessed in these studies, can occur in some circumstances when the vowel adaptor and vowel end point are spectrally dissimilar. Results of this type can be explained by positing the involvement of either higher level auditory patterns or abstract phonetic representations. However, the lack of adaptation with spectrally dissimilar voice adaptors observed in the present study suggests that voice is tied to lower level auditory representations.

The present results also serve to suggest future directions to pursue concerning the acoustic factors related to voice gender perception (Coleman, 1976; Lass *et al.*, 1976; Murry and Singh, 1980; Singh and Murry, 1978). Instead of focusing an separate individual acoustic factors and how they contribute to gender perception, the results of experiments 3 and 4 suggest that perhaps voice gender should be studied in terms of integrated auditory representations. In addition, the present findings suggest that infants' classification of voice into male and female categories (Miller, 1983; Miller *et al.*, 1982) may be based on heuristics that utilize specific and detailed auditory voice information. Further studies of the type performed in the present study with male-only voices and female-only voices may help to elucidate some of these issues.

Finally, one dimensional issue should be mentioned concerning the present results. All of the experiments reported here used synthetic speech tokens. One possible criticism of this study is that the results found with synthetic speech may not generalize to natural speech. Synthetic voices are reduced stimuli that do not contain the full complement of acoustic information contained in natural voices. There is much evidence indicating that perception and memory for synthetic stimuli is different than for natural stimuli (Ralston *et al.*, 1995). Our reply to this potential discussion is that our findings represent a first step toward defining the nature of the perceptual representation of voice gender and that much of the knowledge we have about acoustic–phonetic perception is based on work with synthetic speech. In addition, synthetic voices have been used by others to determine how listeners judge the perceptual quality of voice (Gerratt *et al.*, 1993). But, we do acknowledge that in order to provide a more definitive examination of voice representation, future studies should compare synthetic speech to natural speech in addressing related issues.

In conclusion, the hypothesis that voice gender is stored in abstract representations in memory received little support. However, the present investigation focused only on a few preliminary aspects of this issue. Future research needs to examine in further detail the prototype hypothesis of storing voices in memory (Papcun *et al.*, 1989) and other details about voice representation not specifically related to voice gender.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant No. R01-DCO 1667 to one of the authors (J. W. M.) We wish to thank Georgianne Baartmans, Renee Dudzinski, Kathy Gorday, and Margaret Webb for running subjects and Lin Zong for programming assistance. Thanks also to Jim Ralston, Jim Sawusch, and one anonymous reviewer for useful comments and critiques.

APPENDIX A

Male end-point synthesis values

SYM	V/C	VAL	SYM	V/C	VAL
sr	C	10 000	*nf*	C	4
du	C	250	*ss*	C	2
ui	C	5	*rs*	C	1
*f*0	V	136	*av*	V	60
*F*1	v	270	*b*1	v	60
*F*2	v	2290	*b*2	v	90
*F*3	v	3010	*b*3	v	150
*F*4	v	3500	*b*4	v	200
*F*5	v	3700	*b*5	v	200
*f*6	v	4990	*b*6	v	500
fz	v	280	*bz*	v	90
fp	v	280	*bp*	v	90
ah	V	35	*oq*	V	75
at	v	0	*t*1	V	20
af	v	0	*sk*	v	0
*a*1	v	0	*p*1	v	80
*a*2	v	0	*p*2	v	200
*a*3	v	0	*p*3	v	350
*a*4	v	0	*p*4	v	500
*a*5	v	0	*p*5	v	600
*a*6	v	0	*p*6	v	800
an	v	0	*ab*	v	0
ap	v	0	*os*	C	0
*g*0	v	64	*dF*	v	0
db	v	0			

Varied parameters

Time	*f*0	*av*	*ah*	*oq*	*t*1
0	136	0	20	25	0
5	136	54	20	28	0
10	136	60	20	31	0
15	136	60	20	34	0
20	136	60	20	37	0
25	136	60	20	40	0
30	136	60	20	43	0
35	136	60	20	46	0
40	136	60	20	50	0
45	136	60	20	50	0
50	136	60	20	50	0
55	136	60	20	50	0
60	136	60	20	50	0
65	136	60	20	50	0
70	136	60	20	50	0
75	136	60	20	50	0
80	136	60	20	50	0
85	136	60	20	50	0
90	136	60	20	50	0
95	136	60	20	50	0
100	136	60	20	50	0
105	136	60	20	50	0
110	136	60	20	50	0
115	136	60	20	50	0
120	136	60	20	50	0
125	136	60	20	50	0
130	135	60	20	50	0
135	134	60	20	50	0
140	133	60	20	50	0
145	132	60	20	50	0
150	131	60	20	50	0
155	130	60	20	50	0
160	129	60	20	50	0
165	128	60	20	50	0
170	127	60	20	50	0
175	126	60	20	50	0
180	125	60	20	50	0
185	124	58	21	50	1
190	123	57	22	51	3
195	122	56	23	52	4
200	121	55	24	53	6
205	120	53	25	53	7
210	119	52	26	54	9
215	118	51	28	55	10
220	118	50	29	56	12
225	117	48	30	56	13
230	116	47	31	57	15
235	115	46	32	58	16
240	114	45	33	59	18
245	113	0	35	60	20

Female end-point synthesis values

SYM	V/C	VAL	SYM	V/C	VAL
sr	C	10 000	*nf*	C	4
du	C	250	*ss*	C	2
ui	C	5	*rs*	C	1
*f*0	V	250	*av*	V	60
*F*1	v	310	*b*1	v	60
*F*2	v	2790	*b*2	v	90
*F*3	v	3310	*b*3	v	150
*F*4	v	4100	*b*4	v	200
*F*5	v	3700	*b*5	v	200
*f*6	v	4990	*b*6	v	500
fz	v	280	*bz*	v	90
fp	v	280	*bp*	v	90
ah	V	35	*oq*	V	75
at	v	0	*t*1	V	20
af	v	0	*sk*	v	0
*a*1	v	0	*p*1	v	80
*a*2	v	0	*p*2	v	200
*a*3	v	0	*p*3	v	350
*a*4	v	0	*p*4	v	500
*a*5	v	0	*p*5	v	600
*a*6	v	0	*p*6	v	800
an	v	0	*ab*	v	0
ap	v	0	*os*	C	0
*g*0	v	64	*dF*	v	0
db	v	0			

Varied parameters

Time	*f*0	*av*	*ah*	*oq*	*t*1
0	250	0	40	45	10
5	250	54	40	48	10
10	250	60	40	51	10
15	250	60	40	54	10
20	250	60	40	57	10
25	250	60	40	60	10
30	250	60	40	63	10
35	250	60	40	66	10
40	250	60	40	70	10
45	250	60	40	70	10
50	250	60	40	70	10
55	250	60	40	70	10
60	250	60	40	70	10
65	250	60	40	70	10
70	250	60	40	70	10
75	250	60	40	70	10
80	250	60	40	70	10
85	250	60	40	70	10
90	250	60	40	70	10
95	250	60	40	70	10
100	250	60	40	70	10
105	250	60	40	70	10
110	250	60	40	70	10
115	250	60	40	70	10
120	250	60	40	70	10
125	250	60	40	70	10
130	248	60	40	70	10
135	246	60	40	70	10
140	244	60	40	70	10
145	242	60	40	70	10
150	241	60	40	70	10
155	239	60	40	70	10
160	237	60	40	70	10
165	235	60	40	70	10
170	234	60	40	70	10
175	232	60	40	70	10
180	230	60	40	70	10

Time	*f*0	*av*	*ah*	*oq*	*t*1
185	229	58	41	70	11
190	227	57	42	71	13
195	225	56	43	72	14
200	224	55	44	73	16
205	222	53	45	73	17
210	220	52	46	74	19
215	219	51	48	75	20
220	217	50	49	76	22
225	216	48	50	76	23
230	214	47	51	77	25
235	212	46	52	78	26
240	211	45	53	79	28
245	209	0	55	80	30

APPENDIX B

"Formants adaptor" synthesis values

SYM	V/C	VAL	SYM	V/C	VAL
sr	C	10 000	*nf*	C	4
du	C	250	*ss*	C	2
ui	C	5	*rs*	C	1
*f*0	V	250	*av*	V	60
*F*1	v	270	*b*1	v	60
*F*2	v	2290	*b*2	v	90
*F*3	v	3010	*b*3	v	150
*F*4	v	3500	*b*4	v	200
*F*5	v	3700	*b*5	v	200
*f*6	v	4990	*b*6	v	500
fz	v	280	*bz*	v	90
fp	v	280	*bp*	v	90
ah	V	35	*oq*	V	75
at	v	0	*t*1	V	20
af	v	0	*sk*	v	0
*a*1	v	0	*p*1	v	80
*a*2	v	0	*p*2	v	200
*a*3	v	0	*p*3	v	350
*a*4	v	0	*p*4	v	500
*a*5	v	0	*p*5	v	600
*a*6	v	0	*p*6	v	800
an	v	0	*ab*	v	0
ap	v	0	*os*	C	0
*g*0	v	64	*dF*	v	0
db	v	0			

Varied parameters

Time	*f*0	*av*	*ah*	*oq*	*t*1
0	250	0	20	25	0
5	250	54	20	28	0
10	250	60	20	31	0
15	250	60	20	34	0
20	250	60	20	37	0
25	250	60	20	40	0
30	250	60	20	43	0
35	250	60	20	46	0
40	250	60	20	50	0
45	250	60	20	50	0
50	250	60	20	50	0
55	250	60	20	50	0
60	250	60	20	50	0
65	250	60	20	50	0
70	250	60	20	50	0
75	250	60	20	50	0
80	250	60	20	50	0
85	250	60	20	50	0
90	250	60	20	50	0
95	250	60	20	50	0
100	250	60	20	50	0
105	250	60	20	50	0
110	250	60	20	50	0
115	250	60	20	50	0
120	250	60	20	50	0
125	250	60	20	50	0
130	248	60	20	50	0
135	246	60	20	50	0
140	244	60	20	50	0
145	242	60	20	50	0
150	241	60	20	50	0
155	239	60	20	50	0
160	237	60	20	50	0
165	235	60	20	50	0
170	234	60	20	50	0
175	232	60	20	50	0
180	230	60	20	50	0
185	229	58	21	50	1
190	227	57	22	51	3
195	225	56	23	52	4
200	224	55	24	53	6
205	222	53	25	53	7
210	220	52	26	54	9
215	219	51	28	55	10
220	217	50	29	56	12
225	216	48	30	56	13
230	214	47	31	57	15
235	212	46	32	58	16
240	211	45	33	59	18
245	209	0	35	60	20

"*F*0 adaptor" synthesis values

SYM	V/C	VAL	SYM	V/C	VAL
sr	C	10 000	*nf*	C	4
du	C	250	*ss*	C	2
ui	C	5	*rs*	C	1
*f*0	V	136	*av*	V	60
*F*1	v	310	*b*1	v	60
*F*2	v	2790	*b*2	v	90
*F*3	v	3310	*b*3	v	150
*F*4	v	4100	*b*4	v	200
*F*5	v	3700	*b*5	v	200
*f*6	v	4990	*b*6	v	500
fz	v	280	*bz*	v	90
fp	v	280	*bp*	v	90
ah	V	35	*oq*	V	75
at	v	0	*t*1	V	20
af	v	0	*sk*	v	0
*a*1	v	0	*p*1	v	80

SYM	V/C	VAL	SYM	V/C	VAL
*a*2	v	0	*p*2	v	200
*a*3	v	0	*p*3	v	350
*a*4	v	0	*p*4	v	500
*a*5	v	0	*p*5	v	600
*a*6	v	0	*p*6	v	800
an	v	0	*ab*	v	0
ap	v	0	*os*	C	0
*g*0	v	64	*dF*	v	0
db	v	0			

Varied parameters

Time	*f*0	*av*	*ah*	*oq*	*t*1
0	136	0	20	25	0
5	136	54	20	28	0
10	136	60	20	31	0
15	136	60	20	34	0
20	136	60	20	37	0
25	136	60	20	40	0
30	136	60	20	43	0
35	136	60	20	46	0
40	136	60	20	50	0
45	136	60	20	50	0
50	136	60	20	50	0
55	136	60	20	50	0
60	136	60	20	50	0
65	136	60	20	50	0
70	136	60	20	50	0
75	136	60	20	50	0
80	136	60	20	50	0
85	136	60	20	50	0
90	136	60	20	50	0
95	136	60	20	50	0
100	136	60	20	50	0
105	136	60	20	50	0
110	136	60	20	50	0
115	136	60	20	50	0
120	136	60	20	50	0
125	136	60	20	50	0
130	135	60	20	50	0
135	134	60	20	50	0
140	133	60	20	50	0
145	132	60	20	50	0
150	131	60	20	50	0
155	130	60	20	50	0
160	129	60	20	50	0
165	128	60	20	50	0
170	127	60	20	50	0
175	126	60	20	50	0
180	125	60	20	50	0
185	124	58	21	50	1
190	123	57	22	51	3
195	122	56	23	52	4
200	121	55	24	53	6
205	120	53	25	53	7
210	119	52	26	54	9
215	118	51	28	55	10
220	118	50	29	56	12
225	117	48	30	56	13
230	116	47	31	57	15
235	115	46	32	58	16
240	114	45	33	59	18
245	113	0	35	60	20

[1] For all four adaptation conditions, the results were also analyzed in terms of category boundary data. Category boundaries were computed for baseline and adaptation conditions for each subject and then analyzed in a one-way ANOVA. The results using this method mirrored the results reported on analyses of raw identification data, in terms of overall effect of condition: $[F(1,14)=6.8,\ p<0.03]$ for synthetic male, $(F=0.8$, n.s.) for natural male $(F=0.4$, n.s.) for synthetic female, and $[F(1,12)=10.7,\ p<0.01]$ for natural female. Thus the effects of adaptation as measured through category boundary movement analyses were approximately the same.

[2] Results for these two conditions were also analyzed using category boundary data. The analyses showed that the effect of condition, using category boundary data, was similar to the analyses based on raw identification data for the formants adaptor $(F=1.3$, n.s.) and for the $F0$ adaptor $(F=3.2$, n.s.).

[3] Results for these two conditions were also analyzed using category boundary data. The results showed that the effect of condition, using category boundary data, was similar to the analyses based on raw identification data for the male instructions condition $(F=2.5$, n.s.) and the female instructions condition $[F(1,14)=9.0,\ p<0.01]$.

Ades, A. E. (**1976**). "Adapting the property detectors for speech perception," in *New Approaches to Language Mechanisms*, edited by R. J. Wales and E. Walker (North-Holland, Amsterdam), pp. 55–107.

Assmann, P. F., Nearey, T. M., and Hogan, J. T. (**1982**). "Vowel identification: Orthographic, perceptual, and acoustic aspects," J. Acoust. Soc. Am. **71**, 975–989.

Coleman, R. O. (**1976**). "A comparison of the contributions of two voice quality characteristics to the perceptions of maleness and femaleness in the voice," J. Speech Hear. Res. **19**, 168–180.

Creelman, C. D. (**1957**). "Case of the unknown talker," J. Acoust. Soc. Am. **29**, 655.

Diehl, R. L. (**1981**). "Feature detectors for speech: A critical reappraisal," Psychol. Bull. **89**, 1–18.

Diehl, R. L., Kleunder, K., and Parker, E. M. (**1985**). "Are selective adaptation and contrast effects really distinct?," J. Exp. Psychol. Hum. Percept. Performance **11**, 209–220.

Fourcin, A. J. (**1968**). "Speech-source interference," IEEE Trans. Audio Electroacoust. **ACC-16**, 65–67.

Garner, W. (**1974**). *The Processing of Information and Structure* (Erlbaum, Hillsdale, NJ).

Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N., and Berke, G. S. (**1993**). "Comparing internal and external standards in voice quality judgements," J. Speech Hear. Res. **36**, 14–20.

Godfrey, J. J. (**1980**). "Comparison of consonantal and vocalic cues in selective adaptation," Percept. Psychophys. **28**, 103–111.

Harnad, S., (Ed.). (**1987**). *Categorical Perception* (Cambridge U. P., Cambridge, England).

Johnson, K. (**1990a**). "The role of perceived speaker identity in *F*0 normalization of vowels," J. Acoust. Soc. Am. **88**, 642–654.

Johnson, K. (**1990b**). "Contrast and normalization in vowel perception," J. Phon. **18**, 229–254.

Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Klatt, D. H. (**1989**). "Review of selected models of speech perception," in *Lexical Representation and Process*, edited by W. D. Marslen-Wilson (MIT, Cambridge, MA), pp. 169–226.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Kreiman, J., and Papcun, G. (**1991**). "Comparing discrimination and recognition of unfamiliar voices," Speech Commun. **10**, 265–275.

Kuhl, P. K. (**1991**). "Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not," Percept. Psychophys. **50**, 93–107.

Ladefoged, P., and Broadbent, D. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., and Bourne, V. T. (**1976**). "Speaker sex identification from voiced, whispered, and filtered isolated vowels," J. Acoust. Soc. Am. **59**, 675–678.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **54**, 358–368.

Macmillan, N. A., Kaplan, H. L., and Creelman, C. D. (**1977**). "The psychophysics of categorical perception," Psychol. Rev. **84**, 452–471.

Miller, C. L. (**1983**). "Developmental changes in male/female classification by infants," Infant Behav. Dev. **6**, 313–330.

Miller, C. L., Younger, B. A., and Morse, P. A. (**1982**). "The categorization of male and female voices in infancy," Infant Behav. Dev. **5**, 143–159.

Miller, C. M. (**1990**). ONLINE: A general-purpose program for the generation and control of laboratory experiments and perceptual tests with acoustic stimuli, for use with microcomputers using MS-DOS [computer program], Laboratory Microsystems, Baton Rouge, LA.

Miller, J. L., Connine, C. M., Schermer, T. M., and Kleunder, K. R. (**1983**). "A possible auditory basis for internal structure of phonetic categories," Percept. Psychophys. **46**, 505–512.

Morse, P. A., Kass, J. E., and Turkienicz, R. (**1976**). "Selective adaptation of vowels," Percept. Psychophys. **19**, 137–143.

Mullennix, J. W., and Pisoni, D. B. (**1990**). "Stimulus variability and processing dependencies in speech perception," Percept. Psychophys. **47**, 379–390.

Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (**1989**). "Some effects of talker variability on spoken word recognition," J. Acoust. Soc. Am. **85**, 365–378.

Murry, T., and Singh, S. (**1980**). "Multidimensional analysis of male and female voices," J. Acoust. Soc. Am. **68**, 1294–1300.

Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1992**). "Effects of stimulus variability on the representation of spoken words in memory," *Research on Spoken Language Processing PR-18* (Indiana University, Bloomington, IN), pp. 163–184.

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (**1993**). "Episodic encoding of voice attributes and recognition memory for spoken words," J. Exp. Psychol. Learn. Memory Cognition **19**, 309–328.

Papcun, G., Kreiman, J., and Davis, A. (**1989**). "Long-term memory for unfamiliar voices," J. Acoust. Soc. Am. **85**, 913–925.

Pisoni, D. B., and Lazarus, J. H. (**1974**). "Categorical and noncategorical modes of speech perception along the voicing continuum," J. Acoust. Soc. Am. **55**, 328–333.

Ralston, J. V., Pisoni, D. B., and Mullennix, J. W. (**1995**). "Perception and comprehension of speech," in *Applied Speech Technology*, edited by A. Syrdal, R. Bennett, and S. Greenspan (CRC, Boca Raton, FL), pp. 233–287.

Remez, R., Rubin, P., Nygaard, L., and Howell, W. (**1987**). "Perceptual normalization of vowels produced by sinusoidal voices," J. Exp. Psychol. Hum. Percept. Performance **13**, 40–61.

Samuel, A. G. (**1982**). "Phonetic prototypes," Percept. Psychophys. **31**, 307–314.

Samuel, A. G. (**1986**). "Red herring detectors and speech perception: in defense of selective adaptation," Cognitive Psychol. **18**, 452–499.

Samuel, A. G. (**1988**). "Central and peripheral representation of whispered and voiced speech," J. Exp. Psychol.: Hum. Percept. Perform. **14**, 379–388.

Sawusch, J. R. (**1977**). "Peripheral and central processes in selective adaptation of place of articulation in stop consonants," J. Acoust. Soc. Am. **62**, 738–750.

Sawusch, J. R. (**1986**). "Auditory and phonetic coding of speech," in *Pattern Recognition by Humans and Machines*, edited by E. C. Schwab and H. C. Nusbaum (Academic, Orlando, FL), Vol. 1, pp. 51–88.

Sawusch, J. R., and Jusczyk, P. W. (**1981**). "Adaptation and contrast in the perception of voicing," J. Exp. Psychol.: Hum. Percept. Perform. **11**, 242–250.

Sawusch, J. R., and Pisoni, D. B. (**1976**). "Response organization in selective adaptation to speech sounds," Percept. Psychophys. **20**, 413–418.

Singh, S., and Murry, T. (**1978**). "Multidimensional classification of normal voice qualities," J. Acoust. Soc. Am. **64**, 81–87.

Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (**1992**). "Stimulus variability and spoken word recognition: effects of variability in speaking rate and overall amplitude," in *Research on Spoken Language Processing PR-18* (Indiana University, Bloomington, IN), pp. 31–52.

Van Lancker, D., Kreiman, J., and Wickens, T. (**1985b**). "Familiar voice recognition: patterns and parameters. Part II: Recognition of rate-altered voice," J. Phon. **13**, 39–52.

Van Lancker, D., Kreiman, J., and Emmorey, K. (**1985a**). "Familiar voice recognition: patterns and parameters. Part I: Recognition of backward voices," J. Phon. **13**, 19–38.

Verbrugge, R. R., Strange, W., Shankweiler, D. P., and Edman, T. R. (**1976**). "What information enables a listener to map a talker's vowel space?," J. Acoust. Soc. Am. **60**, 198–212.

Weenink, D. J. M. (**1986**). "The identification of vowel stimuli from men, women, and children," in *Proceedings 10 from the Institute of Phonetic Sciences of the University of Amsterdam* (University of Amsterdam, Amsterdam, The Netherlands), pp. 41–54.