Perceptual adaptation to gender and expressive properties in speech: The role of fundamental frequency

Daniel J. Hubbard[a] and Peter F. Assmann
School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson. Texas 75083-0688

(Received 15 October 2011; revised 26 December 2012; accepted 28 January 2013)

Recent studies have demonstrated perceptual adaptation to nonlinguistic properties of speech involving voice gender and emotional expression. The present study extends this work by examining the contribution of fundamental frequency (F0) to these effects. Voice recordings of vowel-consonant-vowel (VCV) syllables from six talkers were processed using the STRAIGHT vocoder and an auditory morphing technique to synthesize gender (experiment 1) and expressive (experiment 2) speech sound continua ranging from one category endpoint to the other (female to male; angry to happy). Continuum endpoints served as adaptors for *F0 present* and *F0 removed* conditions. *F0 removed* stimuli were created by replacing the periodic excitation source with broadband noise. Confirming previous findings, aftereffects were found in the *F0 present* condition, resulting in a decreased likelihood to identify test stimuli as belonging to the adaptor category. No aftereffects appeared when F0 was removed, highlighting the importance of F0 in adaptation. However, in an identification test listeners were still able to categorize *F0 removed* stimuli at better-than-chance levels, indicating that residual cues for gender and emotion were available even when F0 was not present.

© *2013 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4792145]

PACS number(s): 43.71.Bp, 43.71.An [AA] Pages: 2367–2376

I. INTRODUCTION

The transmission of nonlinguistic information in speech is an important aspect of human communication. Talker attributes such as voice gender and emotional expression are conveyed vocally by dynamic acoustic properties that vary systematically as a result of anatomical differences and sociolinguistic speech production factors (Scherer, 2010; Kreiman and Sidtis, 2011). Much of the research in speech perception has focused on the inherent variability in acoustic properties—for example, fundamental frequency (F0) and formant frequencies—associated with linguistic contrasts (e.g., Peterson and Barney, 1952). However, the acoustic properties that convey nonlinguistic characteristics including voice gender and emotional expression have received less attention. In this paper we examined the representation of gender and emotion using adaptation, a well-documented perceptual phenomenon characterized by decreased response to a stimulus following repeated exposure (for a review, see Webster *et al.*, 2005). When a speech stimulus originates from a continuum spanning two perceptual endpoints (e.g., female to male, or angry to happy), prior exposure to the adaptor may produce a contrastive aftereffect, i.e., a response tendency favoring stimulus parameters opposite those of the adaptor (Schweinberger *et al.*, 2008; Bestelmeyer *et al.*, 2010). As an initial step toward identifying acoustic properties that are important in adaptation of voice gender and expressive speech, the present study addressed the role of F0 in contrastive aftereffects for male/female (experiment 1) and angry/happy (experiment 2) category judgments. We show that the removal of F0, an important acoustic cue for both voice gender and expressive speech perception, had different effects on listener judgments in an adaptation task compared to an identification task.

Contrastive aftereffects demonstrate that perceptual processing related to the adaptor stimulus (or stimulus category) is altered based on recent experience, and this result has important and immediate consequences for perceptual tasks (Webster *et al.*, 2005). The salience of stimulus properties used for ongoing perception is changed, producing a heightened sensitivity to opposing features. Adaptation functions as a calibration mechanism by continuously tuning perceptual decisions related to an adaptor (Webster and MacLeod, 2011). In speech, studies using adaptation have been employed to support distinct auditory and phonetic levels of processing (Eimas and Corbit, 1973; Ades, 1976; Samuel, 1986). More recent speech adaptation studies have shown aftereffects associated with complex nonlinguistic speech categories including voice gender (Mullennix *et al.*, 1995; Schweinberger *et al.*, 2008) and emotional expression (Bestelmeyer *et al.*, 2010).

Mullennix *et al.* (1995) examined the perception of voice gender using a continuum of vowel sounds ranging from male to female. Repeated exposure to male adaptors led to stronger female ratings and vice versa. Schweinberger *et al.* (2008) also explored voice gender perception using multiple adaptor types of different modalities. They demonstrated modality-specific aftereffects following repeated exposure to male or female adaptor syllables. For expressive speech, Bestelmeyer *et al.* (2010) reported aftereffects following adaptation using a continuum of synthesized syllables ranging from anger to fear. Repeated exposure to the original angry adaptors caused synthetic test stimuli along

[a] Author to whom correspondence should be addressed. Electronic mail: dhubbard@utdallas.edu

the continuum from anger to fear to appear more fearful, and vice versa.

Schweinberger *et al.* (2008) also presented listeners with pure tone adaptors corresponding in frequency to average male or female F0, and no aftereffects were found. Based on this result they concluded that aftereffects "clearly do not reflect simple adaptation to pitch." However, there are two ways that these adaptors are not representative of natural speech. First, pure tone adaptors representing a single sinusoid lack the rich harmonic structure of voiced speech. Second, they presented their adaptors as steady-state sounds while F0 varies continuously in natural speech. There are a number of findings that F0 plays a key role in the categorization of voice gender (Coleman, 1976; Assmann *et al.*, 2006; Lass *et al.*, 1976; Gelfer and Mikos, 2005; Hillenbrand and Clark, 2009) and emotional expression (Lieberman and Michaels, 1962; Scherer, 1979, 1986; Murray and Arnott, 1993). For voice gender perception, Lass *et al.* (1976) reported 96% accuracy in an identification task using voiced speech and 75% accuracy with whispered speech. Using a statistical pattern classifier, Hillenbrand and Clark (2009) provided additional evidence that F0 alone distinguished voice gender somewhat better (96%) than mean formant frequencies alone (92%). This provides additional support for the conclusion that F0 makes a significant contribution to gender recognition; however, the fact that voiceless speech produced accuracy rates substantially above chance (Lass *et al.*, 1976) indicates that gender information is contained in voice cues other than F0.

In expressive speech, variation in cues such as F0 mean and range—in conjunction with intensity and speaking rate—are associated with the production of different emotion categories (Scherer, 1979, 1986; Murray and Arnott, 1993). For example, Scherer (1986) reported increases in F0 mean and range, intensity, and speaking rate for elation and fear compared to neutral utterances, and decreases in those parameters for sadness and enjoyment. In a comprehensive review of over 100 expressive speech studies, Juslin and Laukka (2003) reported consistent patterns of acoustic cue changes that vary with the production of different emotions. In their meta-analysis, consistent patterns of F0 mean and F0 variability (high-medium-low) were revealed for major emotion categories across studies. F0 mean and variability typically increased in productions of anger, happiness, and fear, and decreased in productions of sadness (one exception is F0 variability in fear, with mixed results). These findings support the idea that voice gender and expressive speech perception are cued by multiple voice markers but that F0 plays a major role in signaling change in affect.

Some investigators have attempted to determine which aspects of the F0 contour may be used in perceptual judgments of emotional expression. Lieberman and Michaels (1962) found that flattening the F0 contour resulted in significant declines in listener accuracy rates in an emotion identification task. They concluded that much of the emotional information transmitted in speech is carried by the fine structure associated with the F0 contour. Bulut and Narayanan (2008) modified the F0 mean, range, and contour and found that F0 range was the best predictor of emotion judgments, followed by F0 mean. Modifications to the F0 range caused significant changes to emotion category responses. Decreasing the F0 range caused an increase in sad responses and a decrease in happy responses. Collectively, these studies reveal the important contribution of F0 to the perception of voice gender and emotional expression.

In the present study, listeners were exposed to either *F0 present* or *F0 removed* syllables in a selective adaptation task using voice gender (female and male) and emotionally expressive (angry and happy) speech categories. Schweinberger *et al.* (2008) did not find adaptation shifts for voice gender using pure tones corresponding to average male and female F0, and suggested that aftereffects were not the result of adaptation to F0 *per se*. We took a different approach, however, in that listeners were presented with *F0 present* and *F0 removed* syllables in separate adaptation and identification tasks. We sought to determine whether the removal of F0 from natural syllables would systematically alter listeners' judgments in adaptation and identification tasks.

II. METHOD COMMON TO BOTH EXPERIMENTS

A. Recordings

Vowel-consonant-vowel (VCV) syllables were produced by three female and three male talkers. The talkers were college students with experience and were paid a nominal fee for producing the recordings. The recordings were made in a sound-attenuated booth using a Shure SM-94 microphone, Symetrix SX202 dual-microphone pre-amplifier and Tucker-Davis Technologies data acquisition hardware (MA1, RP2.1). Digital waveforms were stored on computer disk at a rate of 48 kHz and 16-bit resolution. Each talker viewed a set of presentation slides displayed on a computer screen to aid in producing the recordings. The slides contained text prompts designed to provide a suitable context to elicit the target emotions (angry or happy). Angry and happy expressive recordings were chosen because they form a representative set of emotional categories extending those used by Bestelmeyer *et al.* (2010). During the recordings, an experimenter was present in the sound booth to assist each talker with the recordings protocol and to encourage the production of appropriately expressive and realistic utterances. The context prompt for angry speech was: "You are on a busy highway in bumper-to-bumper traffic, about to reach your exit, when another driver cuts you off and hits the front of your car, causing an accident." The happy context prompt was: "You purchased a lottery ticket on your way home even though you've never won. You wake up the next morning and turn on the news to learn that your ticket matches the winning numbers and you've won 30 million dollars." Talkers were asked to produce VCV syllables while applying the designated expressive context.

After reading the prompts, the talkers viewed text versions of the following target VCV syllables on separate presentation slides: /aba/, /aga/, /ibi/, /igi/, /obo/, /ogo/, /ubu/, and /ugu/. Each talker was asked to produce 20 repetitions of each syllable in angry, happy and neutral portrayal contexts. Neutral stimuli were recorded first (without prompts), followed by the angry and happy recordings. Twenty repetitions of each syllable for each portrayal context were completed before proceeding to the next syllable. The utterances were recorded in blocks of ten with a short break between blocks. The talkers were encouraged to produce highly expressive angry and happy recordings but were asked to maintain neutral affect during productions of neutral stimuli. Different subsets of the recordings were used for each experiment: pairs of neutral male and female syllables were used for the gender experiment, and the angry and happy syllables were used for the expressive speech experiment.

B. Synthesis

The natural speech recordings described above were analyzed and resynthesized to create the listening test stimuli. Syllable pairs (female/male for experiment 1, angry/happy for experiment 2) with similar overall durations were selected and a waveform editor, WaveSurfer (Sjölander and Beskow, 2000) was used to align temporal onsets and offsets. We used a high-quality vocoder called STRAIGHT (Kawahara *et al.*, 1999) and an auditory morphing technique (Kawahara and Matsui, 2003) based on STRAIGHT to create the synthesized test stimuli. STRAIGHT analyzes input speech into separate components related to the voicing source (F0 and aperiodicity) and the vocal tract transfer function (spectrum envelope). The auditory morphing technique is an automated procedure for interpolating between pairs of time-aligned utterances along a trajectory in the multidimensional space defined by F0, spectrum envelope, and aperiodicity. The procedure utilizes user-specified "anchor points" to systematically interpolate between the two time-aligned utterances. In our application, the anchor points were provided by estimating formant frequencies (F1–F3) at five uniformly spaced time points using an automatic formant tracking algorithm (Nearey *et al.*, 2002). The morphing procedure works by extracting the parameters of each endpoint (F0, aperiodicity spectrogram, and spectrum envelope); then the two endpoints are interpolated at discrete steps to produce an eleven-point continuum of synthesized speech sounds ranging from one endpoint to another in equal acoustic steps.[1]

To measure the contribution of F0 to adaptation aftereffects documented for voice gender and expressive speech, *F0 removed* and *F0 present* adaptor stimuli were created and presented to different listeners preceding the test stimuli. F0 was removed from the synthesized adaptors by replacing the periodic excitation source with broadband noise, and the resulting *F0 removed* adaptors sounded like whispered speech to listeners. Figure 1 displays F0 contours for the expressive *F0 present* test stimuli, showing the systematic variation in pitch as a function of continuum morph level-ranging from the 80% happy (top line in each gender set) to 80% angry (bottom line) morph levels.

C. Procedure

As in Schweinberger *et al.* (2008), the endpoints and the neutral midpoint of each continuum were presented as adaptors to separate listeners in *F0 present* or *F0 removed* adaptation blocks. The adaptors were followed by a common set of voice gender (experiment 1) or expressive (experiment 2) test syllables (with F0 retained) in a forced-choice listening task. The test stimuli consisted of the following continuum morph levels (expressed as percentages): 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, and 20:80. For example, the 80:20 morph level on an angry-to-happy continuum refers to a stimulus in which 80% of the angry endpoint's acoustic characteristics and 20% of the happy endpoint's acoustic characteristics were combined in the auditory morphing process. The test stimuli were presented three times, once per adaptation condition, in separate blocks. For voice gender adaptation the three adaptation blocks were female, neutral and male; for expressive adaptation the three adaptation blocks were angry, neutral, and happy. The presentation order of adaptation blocks was counterbalanced across listeners.

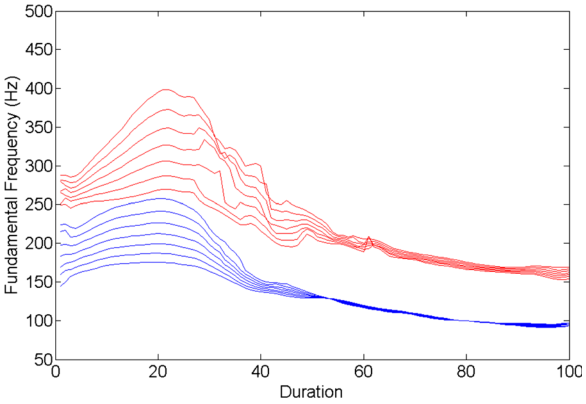

FIG. 1. (Color online) F0 contours for synthesized emotionally expressive test stimuli produced from original recordings of 3 female (top set) and 3 male (bottom set) talkers. Separate lines within each gender category represent average F0 measurements for seven auditory morph proportions (top to bottom from 20:80 to 80:20, angry: happy) along the angry-to-happy expressive speech continua, showing dynamic changes in F0 as a function of morph level.

Per trial, listeners heard four *F0 present* or *F0 removed* adaptor stimuli, followed by a 500 ms pause and one *F0 present* test stimulus. Figure 2 illustrates the experimental procedure and temporal sequence used to study gender and expressive adaptation. The adaptation and test stimuli were presented in random order within each adaptation block. The four adaptor stimuli selected to precede each test stimulus were randomly chosen from a different talker and vowel category so that no talker, syllable, or vowel repetition occurred between the set of adaptors and test stimuli. This was done in both experiments for both F0 conditions to ensure that contrastive aftereffects, if found, were based on high-level perceptual adaptation characteristics rather than low-level stimulus-dependent properties. The adaptation and test stimuli were randomized and presented to listeners using custom MATLAB scripts.

After the test stimulus was presented, listeners selected one of two response buttons in a forced-choice task on a graphic response window using an external mouse. For gender adaptation the two response options were "female" and "male" and for expressive adaptation the two response options were "angry" and "happy." Listeners were asked to listen to the first four (adaptor) stimuli, wait for the pause, listen to the fifth (test) stimulus, and then make a

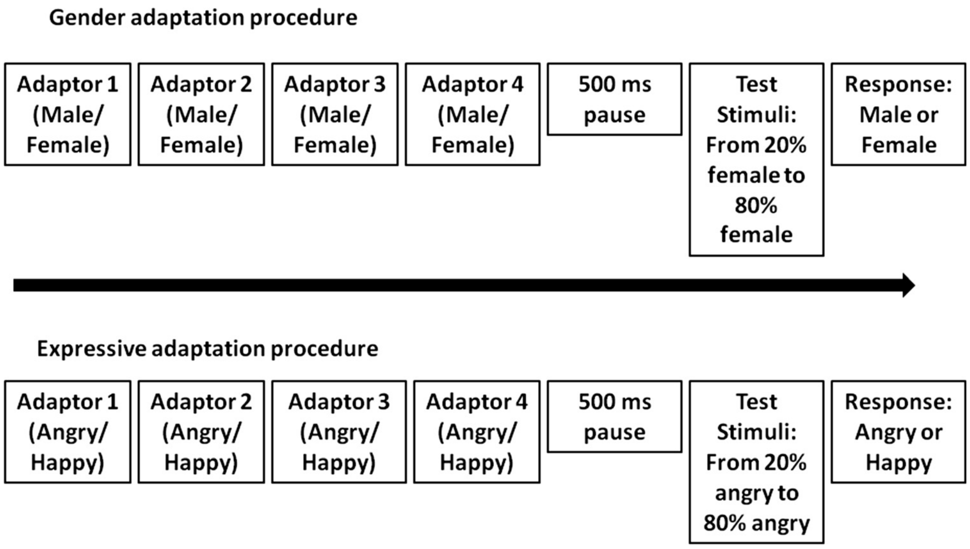

FIG. 2. Adaptation/test trial sequence: Listeners heard four consecutive adaptor stimuli, followed by a 500 ms pause and a single test stimulus. The adaptors originated from different syllables, vowels, and talkers versus the test stimulus.

determination whether the test stimulus sounded female or male (experiment 1), or angry or happy (experiment 2). Participants were told that a new trial would not begin until a selection had been made.

III. EXPERIMENT 1: GENDER ADAPTATION

A. Stimuli and procedure

For synthesis of the gender adaptation test continua, each female talker was matched with a male to create three female/male pairings. Neutral context recordings with matching VCV syllables were used as auditory morphing inputs to create stimuli for the gender listening task. STRAIGHT was utilized to create a female-to-male speech sound continuum for each syllable and female/male pairing. Synthesized adaptors (female and male, plus the neutral midpoint) from eight syllable types for three female/male pairings resulted in 108 potential gender adaptors (36 female, 36 neutral, 36 male) for each F0 condition. For the test stimuli, seven morph levels for each of eight syllable types from three female/male pairings resulted in 168 potential gender test stimuli.

To limit the gender listening task to 1 h of testing, the number of test stimuli was reduced to 84 and the corresponding adaptors were reduced to 36 (12 female, 12 neutral, 12 male). The stimuli were reduced by random selection with the constraints that no female/male talker combinations were eliminated entirely and all eight syllable types were used at least once. The 84 gender test stimuli were each presented three times, once for each adaptation condition in separate blocks, with block order counterbalanced across listeners. This produced a total of 252 trials (84 × 3) for the gender adaptation listening task. See Table I for *F0 present* acoustic measurements for the synthesized female and male endpoint adaptors presented in experiment 1. The table shows F0 and formant frequency (F1–F3) measurements defined as the median of five values surrounding the midpoint of the voiced portion of the first vowel in the VCV syllables. The range is also provided for the voiced portion of the first vowel (in parentheses). F0 was measured using STRAIGHT (Kawahara *et al.*, 1999) and the formant frequencies were estimated using a custom formant tracking algorithm (Nearey *et al.*, 2002). The measurements were verified manually and hand-edited for a few stimuli in which the F0 and/or formant tracker made an error.

B. Participants

Fifteen listeners (11 female and 4 male, native English talkers, mean age: 28.1 years, age range: 18–61 years) participated in the *F0 present* gender adaptation task, and an additional 15 listeners (11 female, 4 male, native English talkers, mean age: 22.2, age range: 18–32 years) participated in the *F0 removed* task.[2] Listeners were recruited using an online research credit sign-up system and were awarded 1 h of required research participation credit as compensation. Participants listened to the adaptation/test trials over headphones and responded to the test stimuli by clicking "female" or "male" in a forced-choice task displayed in a response window on the computer monitor. Participants completed the gender adaptation task with 252 trials in approximately 50 min.

C. Results

The objective of experiment 1 was to measure the contribution of F0 to speech aftereffects for voice gender. Figures 3(a) and 3(b) display overall proportion "female" responses by morph level for each adaptation condition for the *F0 present* adaptors [Fig. 3(a)] and for the *F0 removed* adaptors [Fig. 3(b)]. Female response proportions were well-fit with logistic functions. Figure 3(a) reveals perceptual response shifts due to exposure to *F0 present* adaptors. Repeated presentation of *F0 present* female adaptors decreased female responses, and exposing listeners to male adaptors decreased male responses. The neutral *F0 present* adaptors produced a response pattern that fell between the female and male adaptation conditions, providing a control baseline for comparison [see Fig. 3(a)]. The response shifts following female and male adaptation indicate contrastive aftereffects; detection of stimulus features related to the adaptors was suppressed, and features opposite those of the adaptor were highlighted. For the *F0 removed* condition,

TABLE I. Experiment 1: Fundamental frequency (F0) and formant frequency (F1–F3) midpoint and range (in parentheses) in Hertz (Hz) for *F0 present* voice gender adaptors. Midpoint measurements were defined as the median of the five values surrounding the midpoint of the voiced portion of the first vowel in each syllable. The range was also calculated over the voiced portion of the first vowel in each syllable.

		Female Endpoints							
Talker Pairs		/aba/	/aga/	/ibi/	/igi/	/obo/	/ogo/	/ubu/	/ugu/
T1-T4	F0	200 (37)	207 (7)	212 (13)	218 (27)	218 (11)	210 (19)	230 (42)	222 (22)
	F1	955 (31)	1055 (87)	423 (17)	442 (32)	660 (64)	721 (127)	464 (21)	449 (13)
	F2	1425 (42)	1529 (24)	2893 (42)	2788 (64)	1333 (156)	1478 (32)	1574 (116)	1586 (30)
	F3	2992 (109)	2935 (87)	3197 (128)	3345 (144)	2906 (193)	2904 (134)	3014 (37)	2919 (62)
T2-T6	F0	223 (44)	209 (8)	252 (19)	252 (42)	243 (23)	215 (9)	243 (31)	216 (18)
	F1	941 (70)	928 (75)	454 (6)	380 (80)	721 (31)	660 (106)	501 (32)	459 (42)
	F2	1246 (36)	1270 (169)	2830 (163)	2948 (88)	1476 (81)	1697 (253)	2195 (197)	2169 (74)
	F3	3322 (47)	3232 (102)	3097 (91)	3524 (21)	2924 (24)	2950 (152)	2709 (64)	2643 (121)
T3-T5	F0	183 (11)	187 (11)	195 (8)	199 (8)	193 (4)	184 (22)	187 (9)	187 (9)
	F1	853 (63)	878 (24)	426 (15)	413 (7)	583 (18)	586 (132)	419 (56)	413 (64)
	F2	1447 (29)	1433 (61)	2911 (118)	2951 (103)	1569 (119)	1633 (216)	1777 (144)	2091 (162)
	F3	3043 (87)	2711 (248)	3359 (128)	3455 (203)	3103 (132)	2919 (185)	2750 (80)	2862 (40)
		Male Endpoints							
T1-T4	F0	130 (20)	129 (6)	134 (5)	133 (5)	124 (13)	129 (28)	130 (3)	137 (5)
	F1	793 (46)	805 (73)	309 (23)	317 (27)	520 (58)	578 (37)	412 (19)	426 (6)
	F2	1225 (56)	1313 (49)	2544 (1145)	2405 (28)	1129 (192)	1220 (104)	1477 (125)	1467 (26)
	F3	2714 (135)	2696 (82)	2971 (404)	2931 (83)	2541 (113)	2475 (115)	2184 (93)	2214 (44)
T2-T6	F0	95 (10)	101 (5)	115 (19)	147 (24)	123 (18)	119 (24)	128 (15)	125 (25)
	F1	764 (53)	802 (123)	274 (37)	307 (15)	576 (70)	566 (103)	378 (27)	383 (14)
	F2	1266 (42)	1365 (239)	2276 (56)	2343 (226)	1192 (96)	1193 (162)	1441 (92)	1534 (75)
	F3	2500 (41)	2388 (277)	3312 (166)	3295 (305)	2512 (41)	2460 (48)	2255 (40)	2267 (33)
T3-T5	F0	105 (14)	109 (14)	109 (7)	110 (17)	108 (21)	108 (7)	124 (13)	121 (8)
	F1	653 (51)	658 (76)	283 (26)	258 (20)	425 (39)	499 (53)	373 (23)	355 (20)
	F2	1231 (42)	1225 (100)	2374 (168)	2351 (193)	923 (98)	970 (131)	1007 (76)	1074 (144)
	F3	2425 (525)	2430 (113)	3164 (124)	3154 (204)	2462 (255)	2397 (59)	2340 (62)	2449 (70)

Fig. 3(b) shows that aftereffects were absent when F0 was removed from the adaptor stimuli.

Responses for test stimuli near the endpoints (morph level 2 near the male end of the continuum; morph level 8 near the female end of continuum) after exposure to *F0 present* adaptors for each adaptation condition showed a convergence near the endpoints in that responses were either near 100% female (for the female end of the continuum) or 100% male (for the male end of the continuum), with the largest response differences near the middle of the continuum. For the *F0 removed* condition, Fig. 3(b) shows that aftereffects documented in unaltered adaptors did not persist in the absence of F0. This result suggests that voice gender aftereffects are attributable to F0-related cues, and indicates that residual cues retained in the *F0 removed* adaptors were not adequate to produce aftereffects alone.

Difference scores were calculated by subtracting the proportion of female responses following neutral adaptation

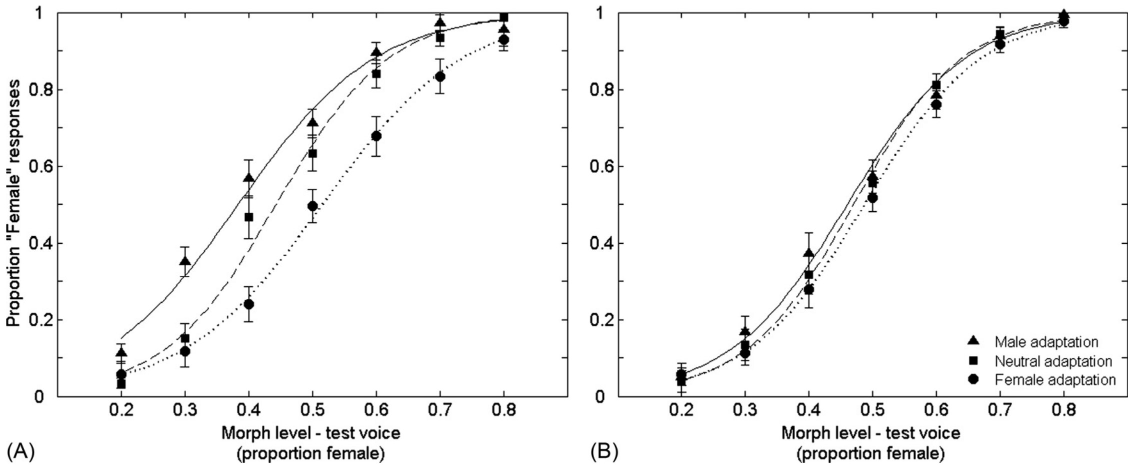

FIG. 3. Experiment 1: Gender adaptation identification functions showing the percentage of “female” responses for each adaptation condition following exposure to (a) *F0 present* and (b) *F0 removed* adaptors. Data from 15 different listeners for each F0 condition are included.

TABLE II. Experiment 1: Gender adaptation pooled *F0 present* and *F0 removed* percent “female” response difference scores comparing each gender and F0 adaptation condition to the neutral control. The difference scores for each talker and morph level were compared using analysis of variance, resulting in the P-values and significance indicators in the columns labeled “*p* value.”[a]

	F0 Present			*F0 Removed*		
Morph level (proportion female)	Female-neutral difference	Male-neutral difference	*p* value	Female-neutral difference	Male-neutral difference	*p* value
0.2	2%	7%	0.2874	2%	1%	0.9152
0.3	−4%	19%	<0.0001***	−2%	3%	0.2874
0.4	−21%	10%	<0.0001***	−4%	6%	0.0709
0.5	−13%	7%	<0.0001***	−4%	2%	0.2874
0.6	−14%	5%	<0.0001***	−5%	−3%	0.6702
0.7	−9%	4%	0.0081**	−3%	−1%	0.6702
0.8	−6%	−3%	0.5946	−1%	1%	0.7494

[a] *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$

from the proportion of female responses following (1) female and (2) male adaptation. This provided two values corresponding to the aggregate response shift for each listener and morph level. To measure whether the response shifts were significant, the difference scores were subjected to analysis of variance. The analysis of variance compared female response proportions for each F0 condition, adaptation condition, and morph level. There was a significant main effect of adaptation condition, F (1, 196) $= 56.41$, p < 0.0001. The mean difference scores for the *F0 present* condition were significantly different at each morph level with the exception of those nearest the endpoints (the 20% and 80% female morph levels). Comparisons of *F0 removed* difference scores at each morph level yielded no differences. See Table II for pooled difference scores for each F0 condition and morph level.

Analysis methods for similar data (Phillips and Hall, 2005; Schweinberger *et al.*, 2008; Bestelmeyer *et al.*, 2010) have included comparison of estimated perceptual boundaries referred to as points of subjective equality (PSEs), or the numeric morph level at which each fitted curve crosses the male/female response boundary. PSE calculations for separate listeners and adaptation conditions were computed as the mean of each fitted curve, or the point along the curve at which responses were 50% female and 50% male. Boundaries were subjected to analysis of variance, comparing the PSE for each adaptation condition. There was a significant main effect of adaptation condition, showing differences in the PSE for *F0 present* adaptors, F (2, 42) $= 8.46$, $p < 0.001$. Planned comparisons for the *F0 present* adaption conditions revealed significant differences in the PSE following female adaptation compared to male adaptation ($p < 0.001$), and following female adaptation compared to neutral adaptation ($p < 0.05$). The difference between male and neutral adaptation was not significant ($p = 0.10$). The PSE analysis for the *F0 removed* adaptors failed to show significant differences between any of the three adaptation conditions.

Following the result that aftereffects did not persist after the removal of F0, a subsequent listening task was conducted to determine if listeners could accurately distinguish female from male voices in the *F0 removed* stimuli.[3] In a two alternative forced-choice listening task, we presented the *F0 removed* gender endpoint adaptors to 13 listeners (7 females and 6 males, native English talkers, mean age: 22.6 years, age range: 19–33 years) five times in random order and calculated identification accuracy rates. Overall identification accuracy rates were 81% for *F0 removed* female stimuli and 93% for the *F0 removed* male stimuli. An analysis of variance confirmed that these rates were significantly above chance (50%) for female talkers, $F(1, 620) = 337.68$, $p < 0.0001$, and male talkers, $F(1, 620) = 644.16$, $p < 0.0001$. This result indicates that residual information in the *F0 removed* adaptors is used in voice gender perception when F0 is absent.

Comparing Fig. 3(a) to Fig. 3(b), an interesting result is that the *F0 present* male adaptation identification function shifted left, indicating an increase in the proportion of female responses. An analysis of variance confirmed that the overall response shift toward the female endpoints was significant, $F(1, 208) = 4.00$, $p < 0.05$. In comparison, the female and neutral identification functions did not shift substantially after the removal of F0. The response shift toward the female endpoint in the identification functions following exposure to male F0-present adaptors is consistent with the results of Coleman (1976), who found that when conflicting cues (gender mismatched F0 and formant frequencies) were presented, stimuli were more likely to be labeled as male. In our results listeners showed greater contrastive response shifts with increased female responses when male gender cues (F0) were present in the adaptors.

IV. EXPERIMENT 2: EXPRESSIVE ADAPTATION

A. Stimuli and procedure

The portrayed expressive VCV syllables served as auditory morphing inputs (endpoints) to create angry-to-happy speech sound continua. Each continuum contained morph levels ranging in equal acoustic steps from the angry to happy synthesized endpoints. Forty-eight continua were constructed, one for each of eight syllable types spoken by six talkers. As in experiment 1, the continuum endpoints and the neutral midpoint were used as adaptors, and morph levels 80:20 (angry: happy) through 20:80 were used as test stimuli. The synthesized adaptors were presented in a pilot identification test to eliminate poorly perceived and misproduced stimuli prior to the adaptation test. In an identification task, 5 female and 4 male native English-speaking listeners with normal hearing were asked to assign emotional category judgments to the synthesized expressive stimuli. Overall accuracy rates were 80% for angry stimuli and 78% for happy stimuli. These accuracy rates are consistent with previous studies on expressive speech categorization in a forced-choice test (Laukka, 2005; Pell *et al.*, 2009).

Eight out of nine listeners (89%) correctly identified both expressive endpoints for 10 of the original 48 continua. Only endpoint adaptors and test stimuli originating from those ten continua were used in the adaptation task. The 89% cutoff was chosen *post hoc* to ensure that the adaptors used were good exemplars of the intended expressive categories. Our goal was to present the most accurately perceived stimuli as adaptors and to limit the task duration to 1 h. Additionally, we would not expect to see contrastive aftereffects following adaptation to stimuli that were poorly identified at the outset. Following the screening, the expressive adaptation listening task consisted of seventy test stimuli (10 continua with 7 morph levels each) and the corresponding 30 adaptors (10 angry and 10 happy endpoints, plus 10 neutral midpoints). The test stimuli were presented three times in separate adaptation condition blocks (neutral, angry, and happy), resulting in 210 trials per F0 condition. As in experiment 1, four randomly selected adaptors from different vowel categories and talkers preceded each test stimulus. Table III displays F0 and F1–F3 acoustic measurements for the synthesized angry and happy endpoint adaptors presented in experiment 2. The midpoint and range is provided for the voiced portion of the first vowel in the VCV syllables. The F0 and formant frequency measurements were collected in the same manner as in experiment 1 and were hand-edited to correct for infrequent tracking errors.

TABLE III. Experiment 2: Fundamental frequency (F0) and formant frequency (F1–F3) midpoint and range (in parentheses) in Hertz (Hz) for *F0 present* expressive adaptors. Midpoint measurements were defined as the median of the five values surrounding the midpoint of the voiced portion of the first vowel in each syllable. The range was also calculated over the voiced portion of the first vowel in each syllable.

	Syllable	Portrayal	F0	F1	F2	F3
			Female Talkers			
T1	/aga/	Angry	225 (28)	880 (89)	1628 (311)	2679 (129)
		Happy	244 (73)	1129 (79)	1808 (82)	3108 (72)
	/obo/	Angry	265 (12)	538 (198)	1231 (318)	2480 (294)
		Happy	343 (76)	1034 (85)	1679 (243)	3843 (532)
	/ogo/	Angry	234 (23)	504 (66)	1398 (126)	2813 (41)
		Happy	367 (128)	935 (367)	1769 (189)	3545 (319)
T3	/ugu/	Angry	213 (7)	430 (60)	1723 (241)	3006 (68)
		Happy	347 (103)	727 (151)	1824 (234)	3160 (112)
			Male Talkers			
T4	/obo/	Angry	160 (28)	532 (46)	1115 (147)	2617 (90)
		Happy	260 (30)	644 (98)	1249 (122)	2572 (61)
	/ogo/	Angry	164 (28)	543 (41)	1167 (127)	2449 (48)
		Happy	261 (47)	554 (129)	1268 (95)	2585 (50)
	/ubu/	Angry	157 (19)	447 (18)	1276 (69)	2195 (359)
		Happy	278 (31)	336 (179)	1459 (110)	2792 (260)
	/ugu/	Angry	167 (8)	462 (84)	1459 (43)	2312 (122)
		Happy	281 (31)	374 (40)	1655 (210)	2750 (300)
T5	/aga/	Angry	114 (13)	765 (131)	1325 (73)	2387 (118)
		Happy	223 (75)	864 (95)	1366 (94)	2565 (121)
T6	/ugu/	Angry	175 (44)	377 (72)	1333 (124)	2295 (79)
		Happy	285 (86)	368 (202)	1316 (164)	2409 (193)

B. Participants

A total of 15 listeners participated in the *F0 present* adaptation task (11 female and 4 male, native English talkers, mean age = 27.6 years, age range = 17–43 years) and an additional 15 listeners (11 female and 4 male, native English talkers, mean age = 23.9 years, age range = 17–48 years) participated in the *F0 removed* task. Listeners were recruited using the same method described in experiment 1 and were awarded research participation credit as compensation. Participants listened to adaptors and test stimuli over circumaural headphones and responded by clicking "angry" or "happy" in a forced-choice task displayed in a response window on the computer monitor. Participants completed the expressive adaptation task with 210 trials in approximately 40 min.

C. Results

The objective of experiment 2 was to measure the contribution of F0 to expressive aftereffects for angry and happy speech. A notable difference between the present study and that of Bestelmeyer *et al.* (2010) is that our expressive continua ranged from angry to happy rather than from anger to fear. Figures 4(a) and 4(b) display listener identification functions showing the overall proportion of "angry" responses per adaptation condition for the *F0 present* (a) and *F0 removed* (b) adaptors. The response data were generally well-fit with logistic functions. For the *F0 present* condition, the presentation of angry adaptors led to fewer angry responses while happy adaptors produced fewer happy responses (i.e., increased angry responses). As in experiment 1, adapting listeners with neutral stimuli produced a response pattern between angry and happy adaptation [see Fig. 4(a)]. No response shifts were found for any of the three F0 removed adaptor types.

Responses for expressive test stimuli near the endpoints for the *F0 present* condition revealed less convergence and reduced accuracy (match between test stimulus category and response) at the endpoints compared to the *F0 present* gender responses in experiment 1. The largest response shifts were observed near the continuum midpoints for angry and happy adaptation. Difference scores for each F0 condition, listener and morph level were calculated using the method described in experiment 1. See Table IV for aggregate angry response proportions and difference scores for each F0 condition and morph level.

The difference scores for each F0 condition, listener, and morph level were subjected to analysis of variance, resulting in a main effect for adaptation condition, $F(1, 196) = 124.75$, $p < 0.0001$. Difference scores for expressive adaptation were significant following adaptation to *F0*

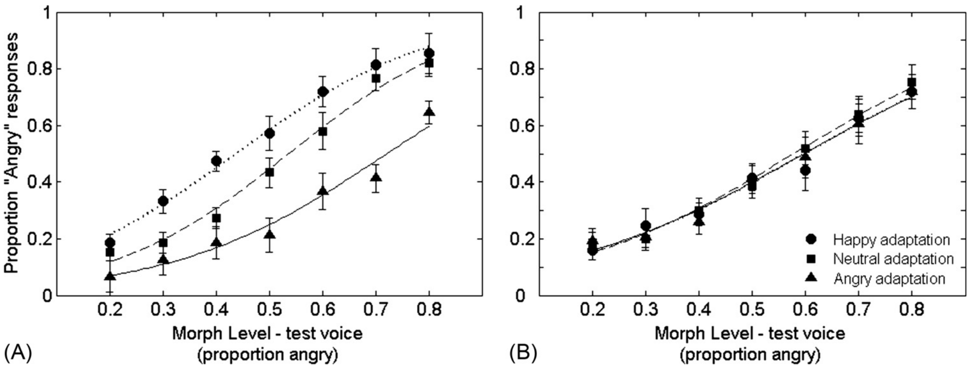

FIG. 4. Experiment 2: Expressive adaptation identification functions showing the percentage of "angry" responses for each adaptation condition following exposure to (a) *F0 present* and (b) *F0 removed* adaptors. Data from 15 different listeners for each F0 condition are included.

present stimuli for each expressive morph level, showing a strong response shift following adaptation to angry and happy adaptors compared to the neutral control (see Table IV). PSE scores representing the mean of the fitted logistic functions were subjected to analysis of variance as well. The data from one participant were eliminated because the PSE could not be accurately estimated. For *F0 present* stimuli there was a significant main effect for adaptation condition, $F\ (2, 42) = 8.46$, $p < 0.001$. Planned comparisons for adaptation condition revealed a significant difference between the proportion of angry responses following angry adaptation compared to those following happy adaptation ($p < 0.0001$). The difference in angry response proportions was also significant for the angry versus neutral adaptation conditions ($p < 0.01$). The happy versus neutral difference was not significant ($p = 0.11$). No significant differences were found between the PSE estimates for the angry, neutral, and happy conditions following adaptation to *F0 removed* stimuli. This suggests that information within F0 is important to adaptation aftereffects in the perception of vocal emotion for angry and happy expressions.

As in experiment 1, a listening task was then conducted to determine if listeners could accurately detect angry and happy expressions from the *F0 removed* stimuli. We presented the *F0 removed* expressive endpoint adaptors to 13 listeners (7 females and 6 males, native English talkers, mean age: 22.6 years, age range: 19–33 years) five times in random order and calculated mean identification accuracy rates. Overall identification accuracy rates were 60% for *F0 removed* angry stimuli and 57% for the *F0 removed* happy stimuli. An analysis of variance performed on the results showed that for angry and happy emotion categories, the *F0 removed* stimuli were identified accurately at a rate above chance level (50%): $F(1, 519) = 9.13$, $p < 0.01$ (angry stimuli); $F(1, 519) = 4.66$, $p < 0.05$ (happy stimuli). This shows that residual expressive cues were used for identification in the absence of F0. The results support the hypothesis that while F0 is a key marker of emotion in speech, other acoustic cues are important carriers as well.

V. GENERAL DISCUSSION

A. The role of F0 in adaptation

To investigate the importance of F0 to perceptual aftereffects in voice gender and emotional expression, we measured listener responses after exposure to *F0 present* and *F0 removed* adaptors using an adaptation paradigm and in an identification task. In experiment 1, listeners exposed to endpoint adaptors from a female-to-male speech sound continuum showed contrastive response shifts following exposure to *F0 present* adaptors. In experiment 2, listeners exposed to endpoint adaptors from an angry-to-happy speech sound continuum showed contrastive response shifts following exposure to *F0 present* adaptors. Aftereffects did not emerge in either experiment following exposure to *F0 removed* stimuli. An interesting result was that listeners were still able to identify voice gender and angry versus happy expressions in the *F0 removed* syllables. Based on these findings, it is clear that F0 plays a prominent role in voice gender and expressive

TABLE IV. Expressive adaptation pooled *F0 present* and *F0 removed* percent "angry" response difference scores comparing each expressive and F0 adaptation condition to the neutral control. The difference scores for each talker and morph level were compared using analysis of variance, resulting in the *p*-values and significance indicators in the columns labeled "*p* value.[a]"

	F0 Present			*F0 Removed*		
Morph level (proportion angry)	Angry - neutral difference	Happy - neutral difference	*p* value	Angry - neutral difference	Happy - neutral difference	*p* value
0.2	−9%	3%	0.0322*	−3%	1%	0.6331
0.3	−6%	15%	0.0002***	5%	1%	0.4036
0.4	−9%	20%	<0.0001***	−1%	−4%	0.8113
0.5	−22%	14%	<0.0001***	3%	3%	0.6331
0.6	−21%	14%	<0.0001***	−8%	−3%	0.1526
0.7	−35%	5%	<0.0001***	−1%	−3%	0.8113
0.8	−17%	3%	0.0002***	−3%	−3%	0.5507

[a] $^*p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

speech adaptation, but that listeners did not rely exclusively on F0 in the identification task.

The research reported here extends the work of Bestelmeyer *et al.* (2010) by replicating speech aftereffects for a new set of emotional expressions, angry, and happy. The results add support for the conclusion that adaptation aftereffects are not confined to linguistic properties of speech but also apply to nonlinguistic features including voice gender and emotion (Schweinberger *et al.*, 2008; Bestelmeyer *et al.*, 2010). Adaptation functions as a recalibration mechanism that continually shapes perception based on experience (Webster *et al.*, 2005). Aftereffects are important not only for understanding how speech is encoded in the auditory system, but also for understanding how perception is altered as a result of stimulus properties to which a listener is exposed. Our results document the contribution of F0 to that process.

Schweinberger *et al.* (2008) concluded that the aftereffects reported in their study did not reflect "simple adaptation to pitch," based on the result that pure tones representing average female and male F0 did not produce adaptation aftereffects. This seems inconsistent with the present finding that the removal of F0 eliminated aftereffects, suggesting that F0 does play an important role. A possible explanation is that aftereffects emerge only when the adaptors convey clear evidence for voice gender or emotion categories. It is unlikely that the pure tone adaptors used in Schweinberger *et al.* (2008) met this criterion.

This account presents an interesting puzzle. Listeners could identify voice gender and angry and happy emotional expressions in the *F0 removed* syllables, yet adaptation using those stimuli did not produce aftereffects. Adaptation to voice gender and angry and happy expressions clearly depends on the presence of F0, but in its absence listeners attended to other cues that yielded better-than-chance identification. Removing F0 lowered identification accuracy, particularly for the expressive stimuli. Research has shown near-perfect accuracy rates for judgments of voice gender using naturally produced syllables containing F0 (Hillenbrand and Clark, 2009). In comparison our *F0 removed* voice gender accuracy rates were 80% and 92% for females and males, respectively. For expressive identification the accuracy rates for *F0 present* stimuli were 80% for angry stimuli and 78% for happy stimuli but dropped to 60% for angry and 57% for happy *F0 removed* stimuli (relative to a chance level at 50%). The weakening of category information may have been responsible for the elimination of aftereffects following exposure to the *F0 removed* adaptors.

Overall, the results show that F0 is a highly salient cue for voice gender and expressive speech, and its salience may be responsible for the induction of adaptation aftereffects. It should be noted, however, that F0 likely operates differently within voice gender and expressive adaptation. Mean F0 is important for distinguishing female from male voices, whereas happy speech is distinguished from angry speech by distinctive patterns of F0 movement (Lieberman and Michaels, 1962). An important next step will be to examine perception of nonlinguistic information in speech using an expanded set of voice and spectrum envelope measures. Future work in expressive speech perception will include additional emotion categories (for example, low-activity level emotions) to similarly measure to what extent perception of those categories is dependent on F0.

B. Comparison of voice gender and expressive adaptation

Comparison of the identification functions for gender (experiment 1) and expressive speech (experiment 2) reveals shallower identification functions for expressive adaptation in which listeners did not reach 100% accuracy for test stimuli near continuum endpoints. These differences indicate greater response ambiguity for the expressive test stimuli and could be related to one of several explanations deserving further study. One possibility is that our gender recordings may have been more distinctively produced than the expressive recordings. Some investigators (e.g., Russell, 1994) have argued that emotions are not as discretely produced as the universal emotions view suggests (e.g., Ekman, 1984). Our test stimuli originated from continua for which both expressive endpoints were accurately identified, however, the relative distinctiveness of voice gender versus emotion categories could account for the flatter response functions and greater response ambiguity in experiment 2. An important consideration related to this idea is that our angry and happy recordings required talkers to manipulate their voices to convey an emotion, whereas the neutral female and male recordings used for gender adaptation did not impose similar demands.

A related possibility is that the voice gender stimuli provided a stronger acoustic contrast than the expressive syllables because the former involved morphing syllables from two separate talkers, while the latter used a pair of syllables from the same talker. The increase in potential category variability for expressive speech and the dichotomous nature of voice gender therefore may have contributed to the flatter expressive adaptation response functions and greater ambiguity for expressive adaptation.

The results reported here offer additional evidence for high-level adaptation to complex stimuli. In everyday speech perception, adaptation functions as a calibration mechanism that has important and immediate consequences for perceptual decisions. The mechanisms responsible for processing streams of incoming speech information are continuously recalibrated based on the acoustic properties of recently heard stimuli. Research has shown that F0 likely performs a calibrating function in speech perception, and is used for comparison to other talkers and to incoming stimuli from the same talker (Nearey and Assmann, 2007; Barreda and Nearey, 2012). We have suggested that F0 plays a major role in perception of voice gender and different emotions in expressive speech, and further work is warranted to better define the nature of perceptual processing for nonlinguistic information in speech.

VI. SUMMARY AND CONCLUSIONS

(1) Contrastive speech adaptation aftereffects were replicated for voice gender and for emotional expression (angry versus happy).

(2) Voice gender and expressive adaptation were eliminated following removal of F0 from adaptation stimuli, indicating that information associated with F0 is critical for adaptation to occur.

(3) Voice gender and angry emotional expressions were reliably labeled in a separate identification task, showing that residual cues in the *F0 removed* stimuli are available for categorization.

ACKNOWLEDGMENTS

The authors would like to thank Hideki Kawahara for providing the STRAIGHT synthesis software, Alice O'Toole for helpful feedback on the research analysis methodology, James Hillenbrand, Drew Rendall, and two anonymous reviewers for helpful comments made on a previous version of the manuscript, and the UT Dallas School of Arts and Humanities for cooperation in posting flyers to recruit talkers. Portions of this work were reported at the 162nd Meeting of the Acoustical Society of America.

[1] Following convention for auditory morphing studies (e.g. Schweinberger *et al.*, 2008) we refer to steps along the continuum as morph levels.

[2] The listeners in the *F0 present* condition of experiment 1 included a 61-year-old participant who was older than the remaining listeners but whose identification functions were comparable to other listeners. One *F0 removed* gender adaptation listener reported tinnitus and partial deafness in the left ear. Following several trials this listener reported no difficulty in hearing adaptation/test trials or in performing the experimental procedure, and therefore the results were included in our analysis. Review of individual results revealed a response pattern similar to that of other listeners.

[3] The authors would like to thank an anonymous reviewer for the suggestion to present the *F0 removed* stimuli to listeners in a subsequent identification task with different listeners to determine if voice gender and emotional expression could be detected from the stimuli used in the adaptation task.

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 Dembling, S. (**2006**). "Effects of frequency shifts on perceived naturalness and gender information in speech," in *Proceedings of the Ninth International Conference on Spoken Language Processing*, Pittsburgh, PA (September, 17–21, 2006), pp. 889–892.

Barreda, S., and Nearey, T. M. (**2012**). "The direct and indirect roles of fundamental frequency in vowel perception," J. Acoust. Soc. Am. **131**, 466–477.

Bestelmeyer, P. E. G., Rouger, J., DeBruine, L. M., and Belin, P. (**2010**). "Auditory adaptation in vocal affect perception," Cognition. **117**, 217–223.

Bulut, M., and Narayanan, S. (**2008**). "On the robustness of overall F0-only modifications to the perception of emotions in speech," J. Acoust. Soc. Am. **123**, 4547–4558.

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

Eimas, P. D., and Corbit, J. D. (**1973**). "Selective adaptation of linguistic feature detectors," Cognit. Psychol. **4**, 99–109.

Ekman, P. (**1984**). "Expression and the nature of emotion," in *Approaches to Emotion*, edited by K. Scherer and P. Ekman (Lawrence Erlbaum, Hillsdale, NJ), pp. 319–344.

Gelfer, M. P., and Mikos, V. A. (**2005**). "The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels," J. Voice **19**, 544–554.

Hillenbrand, J. M., and Clark, M. J. (**2009**). "The role of F0 and formant frequencies in distinguishing the voices of men and women," Atten. Percept. Psychophys. **71**, 1150–1166.

Juslin, P. N., and Laukka, P. (**2003**). "Communication of emotions in vocal expression and music performance: Different channels, same code?," Psychol. Bull. **129**, 770–814.

Kawahara, H., Masuda-Katsuse, I., and de Cheveigné, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction," Speech Comm. **27**, 187–207.

Kawahara, H., and Matsui, H. (**2003**). "Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation," in *Proceedings of the 2003 IEEE Inter. Conf. on Acoust., Speech, and Signal Proc.*, Vol. I, pp. 256–259.

Kreiman, J., and Sidtis, D. (**2011**). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception* (Wiley-Blackwell, West Sussex, UK), pp. 124–341.

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.

Laukka, P. (**2005**). "Categorical perception of vocal emotion expressions," Emotion **5**, 277–295.

Lieberman, P., and Michaels, S. B. (**1962**). "Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech," J. Acoust. Soc. Am. **34**, 922–927.

Mullennix, J. W., Johnson, K. A., Topcu-Durgan, M., and Farnsworth, L. M. (**1995**). "The perceptual representation of voice gender," J. Acoust. Soc. Am. **98**, 3080–3095.

Murray, I. R., and Arnott, J. L. (**1993**). "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion," J. Acoust. Soc. Am. **93**, 1097–1108.

Nearey, T. M., and Assmann, P. F. (**2007**). "Probabilistic 'sliding template' models for indirect vowel normalization," in *Experimental Approaches to Phonology*, edited by M. J. Solé, P. S. Beddor, and M. Ohala (Oxford University Press, Oxford, UK), pp. 246–269.

Nearey, T. M., Assmann, P. F., and Hillenbrand, J. M. (**2002**). "Evaluation of a strategy for automatic formant tracking," J. Acoust. Soc. Am. **112**, 2323.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., and Kotz, S. A. (**2009**). "Factors in the recognition of vocally expressed emotions: A comparison of four languages," J. Phonetics. **37**, 417–435.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.

Phillips, D. P., and Hall, S. E. (**2005**). "Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level," Hear. Res. **202**, 188–199.

Russell, J. A. (**1994**). "Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies," Psychol. Bull. **115**, 102–141.

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

Scherer, K. R. (**1979**). "Nonlinguistic vocal indicators of emotion and psychopathology," in *Emotions in Personality and Psychopathology*, edited by C. E. Izard (Plenum Press, New York), pp. 495–529.

Scherer, K. R. (**1986**). "Vocal affect expression: a review and a model for future research," Psychol. Bull. **99**, 143–165.

Scherer, K. R. (**2010**). "Emotion and emotional competence: conceptual and theoretical issues for modeling agents," in *Blueprint for Affective Computing: A Sourcebook*, edited by K. R. Scherer, T. Bänziger, and E. B. Roesch (Oxford University Press, New York), pp. 3–20.

Schweinberger, S., Casper, C., Hauthal, N., Kaufmann, J., Kawahara, H., Kloth, N., Robertson, D., Simpson, A., and Zaske, R. (**2008**). "Auditory adaptation in voice perception," Current Biology. **18**, 684–688.

Sjölander, K., and Beskow, J. (**2000**). "WaveSurfer—An open source speech tool," in *Proceedings of the Int. Conf. Speech Lang. Proc.*, Vol. IV, pp. 464–467.

Webster, M. A., and MacLeod, D. I. A. (**2011**). "Visual adaptation and face perception," Philos. Trans. R. Soc. B. **366**, 1702–1725.

Webster, M. A., Werner, J. S., and Field, D. J. (**2005**). "Adaptation and the phenomenology of perception," in *Fitting the Mind to the World: Adaptation and Aftereffects in High-Level Vision*, edited by C. E. G. Clifford and G. Rhodes (Oxford University Press, New York), pp. 241–277.