The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness

Katarzyna Pisanski and Drew Rendall[a)]

Department of Psychology, The University of Lethbridge, Lethbridge, Alberta, Canada, T1K 3M4

(Received 26 July 2010; revised 20 December 2010; accepted 4 January 2011)

Key features of the voice—fundamental frequency (F_0) and formant frequencies (Fn)—can vary extensively among individuals. Some of this variation might cue fitness-related, biosocial dimensions of speakers. Three experiments tested the independent, joint and relative effects of F_0 and Fn on listeners' assessments of the body size, masculinity (or femininity), and attractiveness of male and female speakers. Experiment 1 replicated previous findings concerning the joint and independent effects of F_0 and Fn on these assessments. Experiment 2 established frequency discrimination thresholds (or just-noticeable differences, JND's) for both vocal features to use in subsequent tests of their relative salience. JND's for F_0 and Fn were consistent in the range of 5%–6% for each sex. Experiment 3 put the two voice features in conflict by equally discriminable amounts and found that listeners consistently tracked Fn over F_0 in rating all three dimensions. Several non-exclusive possibilities for this outcome are considered, including that voice Fn provides more reliable cues to one or more dimensions and that listeners' assessments of the different dimensions are partially interdependent. Results highlight the value of first establishing JND's for discrimination of specific features of natural voices in future work examining their effects on voice-based social judgments.

[DOI: 10.1121/1.3552866]

PACS number(s): 43.71.Bp, 43.70.Mn, 43.66 Fe [AA] Pages: 2201–2212

I. INTRODUCTION

The human voice is the carrier of language. However, it also conveys cues to many other non-linguistic, indexical dimensions of speakers. For example, it is well-established that listeners can reliably judge the age and sex of speakers based only on hearing their voice, and that such discriminations reflect salient differences between adults and children and between men and women in voice fundamental frequency (F_0) that is associated with the perception of voice pitch and in the pattern of vocal-tract resonances (or formants, Fn) that is associated with percepts of timbre. In turn, these voice acoustic differences reflect maturational and sex-related differences in the length and mass of the vocal folds and in the size of vocal-tract cavities (e.g., Peterson and Barney, 1952; Lieberman and Blumstein, 1988; Childers and Wu, 1991; Titze, 1994; Fitch and Giedd, 1999; Vorperian *et al.*, 1999; Lieberman *et al.*, 2001; Collins and Missing, 2003; Bruckert *et al.*, 2006; Evans *et al.*, 2006; Owren *et al.*, 2007).

Recently, attention has been focused on the extent to which listeners can reliably judge additional indexical dimensions of speakers from voice cues alone. In part, this emphasis reflects efforts to integrate voice research on humans with comparative research on the biological influences shaping systems of animal communication. In animal research, investigators have been interested in how vocal signals might convey cues to important fitness-related, biological, and social dimensions of signalers. Here, Darwin's theories of natural selection and sexual selection predict that, among other possible traits, body size, masculinity (and femininity), and attractiveness will be important dimensions of social discrimination in contexts that involve competing with rivals for access to limited resources or selecting among potential mates (Darwin, 1859, 1871). Subsequent research has confirmed some of these predictions. For example, in many mammals, large-bodied males tend to be more successful in aggressive contests with rival males for access to sexually receptive females (reviewed in Lindenfors *et al.*, 2007). In some taxa, large-bodied males are also selectively chosen by females as mates (Ryan, 1983).

Importantly, some of these biosocial dimensions might also be manifested in the voice due to the integrated nature of anatomical, physiological, and hormonal systems. For example, steroid hormones that affect many aspects of sexual differentiation and body growth in humans and other animals also affect the size, shape, and tension of the vocal folds and thus also their natural vibratory frequency, F_0 (Titze, 1989; Abitol *et al.*, 1999; Dabbs and Mallinger, 1999; Feinberg, 2008; Evans *et al.*, 2008). The result, in humans, is that the sex and body size differences between adult men and women are also associated with a conspicuous, two-fold difference in voice F_0 between them (Titze, 1989). Because steroid hormones also influence ovulation and fertility in women (Alonso and Rosenfield, 2002; Bryant and Haselton, 2009) as well as fertility and immunocompetence in men (Folstad and Karter, 1992), it is possible that variation in voice F_0 in men and women might provide additional cues to speaker quality or attractiveness.

[a)]Author to whom correspondence should be addressed. Electronic mail: d.rendall@uleth.ca

At the same time, voice formant frequencies, Fn, are also predicted to provide reliable cues to speaker body size. This prediction follows a proposal by Fitch (1994, 1997, 2000; Fitch and Giedd, 1999; Fitch and Hauser, 2003) emphasizing that the pattern of voice Fn is determined primarily by the length of the vocal-tract. In turn, vocal-tract length is hypothesized to be constrained by surrounding bony structures (e.g., the length and size of the neck, pharynx, oro-pharynx, and oral cavity) which should in turn be correlated with overall body size (*qua* height). Hence, taller individuals should also have longer vocal-tracts and lower Fn than shorter individuals (Fitch and Giedd, 1999), making voice Fn an inherently reliable cue to body size (Davies and Halliday, 1978; Clutton-Brock, 1979).

Perceptual studies in humans have tended to corroborate these hypotheses in finding that speakers of either sex whose voices have either low F_0 or low Fn are rated as being larger and more masculine and also as being more attractive if male or less attractive if female (Van Dommelen and Moxness, 1995; Oguchi and Kikuchi, 1997; Varosanec-Skaric, 1999; Collins, 2000; Collins and Missing, 2003; Ives *et al*., 2005; Feinberg *et al*., 2005, 2006, 2008; Puts, 2005; Bruckert *et al.*, 2006; Riding *et al*., 2006; Saxton *et al*., 2006; Puts *et al*., 2007). Hence, both F_0 and Fn appear to affect listeners' assessments of speaker size, masculinity or femininity, and attractiveness. What remains unclear is how the two voice features interact in influencing such voice-based social judgments and whether and to what degree one or other vocal feature might be weighted more heavily.

This paper attempts to address these issues. Experiment 1 attempts to replicate and expand on past findings concerning the independent and joint influences of F_0 and Fn on ratings of speaker size, masculinity or femininity, and attractiveness. Experiment 2 then establishes listeners' threshold discrimination abilities for frequency-differences in F_0 and Fn as embedded in naturalistic voices. Finally, Experiment 3 tests the relative salience of voice F_0 and Fn in assessments of speaker size, masculinity or femininity, and attractiveness using naturalistic voices in which the two voice features are put in conflict with one another in equally discriminable amounts based on the values obtained from Experiment 2 for just-noticeable differences (JND's) in F_0 and Fn discrimination.

II. EXPERIMENT 1. INDEPENDENT AND JOINT EFFECTS OF F_0 AND Fn ON ASSESSMENTS OF BODY SIZE, MASCULINITY OR FEMININITY, AND ATTRACTIVENESS

The first step was to replicate previous findings concerning the independent and joint effects that voice F_0 and Fn can have on listeners' ratings of speaker body size, masculinity or femininity, and attractiveness. To that end, an experiment was designed to test how listeners' ratings of these dimensions were influenced by variation in the F_0 and Fn of two sets of voices: one set involved speakers whose voices exemplified the natural range of variation in F_0 and Fn for men and women; and the other set involved speakers whose F_0 and Fn had been systematically manipulated.

A. Methods

1. Voice recording

The voices used as stimuli in this experiment were derived from a speaker database that was collected 2 years earlier to minimize the likelihood that participants might recognize the voices of the particular speakers. The speaker database included 57 males and 57 females who lacked strong regional accents and whose native language was Canadian English. Speakers' voices were recorded in a sound-controlled room in the Laboratory of Comparative Communication and Cognition at the University of Lethbridge using procedures identical to those used in previous research (for additional details, see Rendall *et al*., 2005, 2007). Briefly, speech samples were collected using an adjustable head-mounted microphone [AKG C420 (Vienna, Austria)] connected to a preamplifier and computer through a Butterworth antialias filter [Frequency Devices 900/9L8B (Ottawa, IL)]. The speech material recorded included a list of single-syllable bVt words (e.g., *bit*, *bet*, *bat*, *bait*, *butt*, and *boot*). Prior to recording, each speaker was given time to familiarize themselves with the list and to get comfortable saying the words aloud while wearing the headset microphone, while the recordist titrated appropriate signal recording levels. They were subsequently asked to say each item on the list slowly and clearly in a natural speaking voice while a formal recording was made.

2. Voice stimulus selection and manipulation

To facilitate the selection of voice stimuli, values of F_0 and Fn were obtained for each speaker in the recording database. Measurements were performed using PRAAT software (version 4.6, Boersma, 2001) and were taken from the central, steady-state portion of the vowel in each of the words. Pitch values were obtained using PRAAT's pitch-tracking function with a range setting of 75–300 Hz. Formant values were obtained using PRAAT's formant-tracking function and interactive modification of the number of formants identified (i.e., 4 ± 2), depending on the speaker's sex and visual evaluation of the fit of putative formant tracks to manifest peaks in the spectrum of vowels. The resulting values for F_0 and Fn for different vowels were then averaged within speakers to establish baseline F_0 and Fn values for each speaker. The mean baseline F_0 for this sample of male speakers was 120 Hz (range: 114–127 Hz) with mean values for $F1$–$F4$ of 353, 1635, 2690, and 3522 Hz, respectively. The mean baseline F_0 for female speakers was 207 Hz (range: 172–242 Hz) with mean Fn values of 443, 1914, 3006, and 4068 Hz. These values agree well with those of previous samples of speakers of American English (e.g., Stevens, 1998; Bachorowski and Owren, 1999; Rendall *et al*., 2005).

From this speaker database, four speakers (two males, two females) were selected whose natural voice F_0 and Fn values were either relatively low or relatively high for their sex (see Table I). This provided a natural opportunity to test the extent to which inherent variation among speakers representing the extremes of F_0 and Fn for each sex might influence listener ratings of size, masculinity, and attractiveness. Four additional speakers (two males, two females) with intermediate values of F_0 and Fn were selected to create a larger speaker sample. Four independent manipulations of F_0 and Fn were then performed for all eight speakers as a further test of listener sensitivity to these two voice features. The manipulations involved either raising or lowering F_0 by 20% while holding Fn constant and either raising or lowering all formants by 10% while holding F_0 constant. The magnitude of the frequency manipulations was chosen to parallel those used in previous studies (e.g., Feinberg *et al*., 2005; Rendall *et al*., 2007). In this way, a set of five different voice stimuli was created for each of the eight speakers. Each voice stimulus consisted of five bVt words from the original speech recording (e.g., *bit*, *bet*, *boot*, *bait*, and *bat*) with each word separated by 50-ms of silence and standardized to 65 dB [sound pressure level (SPL)].

3. Participants

Thirty-one females and 30 males completed the experiment. All participants were recruited from the undergraduate community at the University of Lethbridge and received partial course credit. Participants provided informed consent and all but one self-identified as heterosexual. Because notions of masculinity/femininity and attractiveness might vary appreciably between heterosexual and homosexual individuals, the data from the one self-identified homosexual participant were omitted prior to analysis.

4. Experimental procedure

Participants completed the experiment privately in a sound-controlled room via a custom computer interface that was designed in RUNTIME REVOLUTION (version 2.8.1) to implement the experiment and to collect and archive the data. Before testing began, the program presented the participant with a set of instructions outlining the experimental task and procedures. Participants were instructed that they would hear a series of individual voices played to them one at a time. They heard the voices through Sennheiser HD 280 professional headphones (Germany) at a comfortable pre-set volume. Their task was to rate each voice on one of the three dimensions (either size or masculinity/femininity or attractiveness) using a six-point scale. The scale was represented on-screen by a set of six unlabeled radio buttons, three on either side of a mid-point marker. For each dimension, the left and right endpoints of the scale were anchored with text labels which were, respectively, either *small* and *large*; or *feminine* and *masculine*; or *unattractive* and *attractive* depending on the dimension being evaluated.

The experiment then commenced and the program presented one voice stimulus per trial with the order of voice stimuli randomized. The dimension to be rated on a given trial was also randomized within the constraint that each voice stimulus ultimately had to be rated three times, once for each of the three dimensions. Each participant thus completed 120 trials (five voice stimuli × eight different speakers × three different biosocial dimensions). Participants received three scheduled rest-breaks at equally spaced intervals.

5. Statistical analysis

The effects of F_0 and Fn on participants' ratings of each of the three dimensions were tested using repeated measures analysis of variance (rmANOVA) with voice condition as a within-subjects factor and participant sex as a between-subjects factor. Separate analyses were conducted for male and female speakers. Statistical analyses were performed using NCSS version 5.1 (Hintze, 1989) using an alpha level of 0.05. *Post-hoc* analyses used the Tukey-Kramer multiple comparison test.

B. Results and discussion

The first tests involved examining listener ratings of size, masculinity, and attractiveness for the set of four speakers who exemplified naturally low and high extremes in both F_0 and Fn. These tests revealed main effects of voice condition on ratings of all three dimensions (Fig. 1). Thus, for both male and female raters, speakers of both sexes whose voice had naturally low F_0 and Fn were rated as larger [male speakers (m): $F_{1,59} = 39.56$, $P < 0.001$; female speakers (f): $F_{1,59} = 72.89$, $P < 0.001$] and more masculine (m: $F_{1,59} = 130.19$, $P < 0.001$; f: $F_{1,59} = 267.88$, $P < 0.001$) than speakers with naturally high F_0 and Fn. There were also predictable differences in the average absolute value of these ratings as a function of speaker sex: both male and female listeners rated male speakers as being larger ($M = 4.34$) and more masculine ($M = 4.75$) than female speakers ($M = 2.91$, 2.64).

Ratings of attractiveness also differed as a function of F_0 and Fn, but here there were also predictable interactions with the sex of both the speaker and the listener. Thus,

TABLE I. Values of F_0 and Fn for speaker sample used in experiment 1.

	Male speakers				Female speakers			
Baseline frequency (Hz)	1 Low F_0Fn	2 High F_0Fn	3 Mixed F_0Fn	4 Mixed F_0Fn	1 Low F_0Fn	2 High F_0Fn	3 Mixed F_0Fn	4 Mixed F_0Fn
F_0	93	138	82	134	132	234	165	178
$F1$	412	410	399	472	452	535	483	526
$F2$	1471	1922	1892	1697	1633	2122	2055	2051
$F3$	2467	2772	2597	2562	2612	3054	3067	3034
$F4$	3433	3814	3494	3560	3587	4157	4224	3872

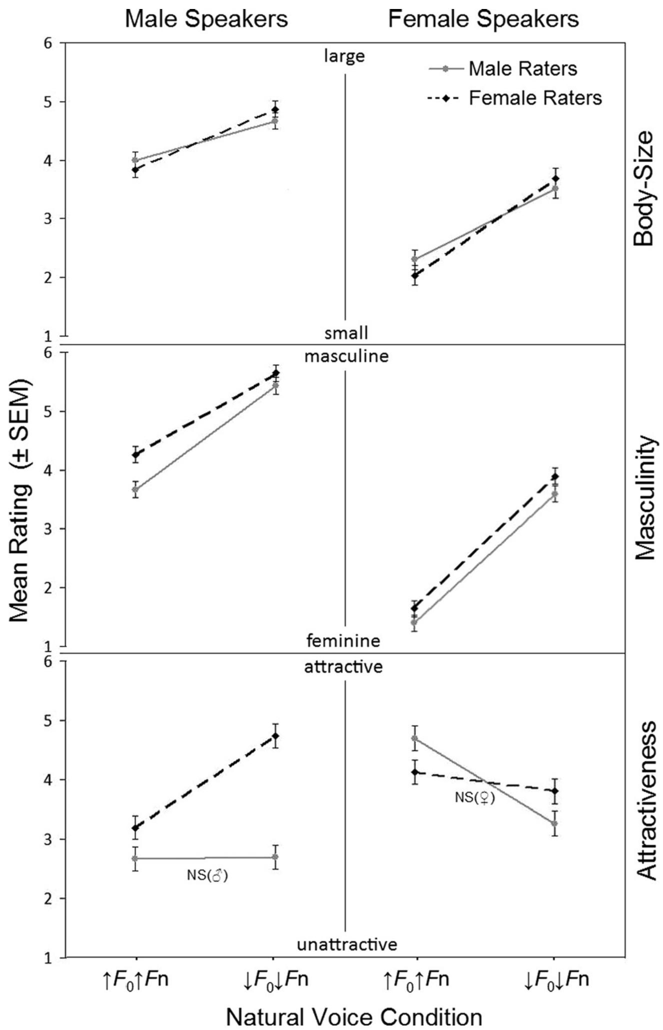

FIG. 1. Mean [±standard error mean (SEM)] ratings of body size (top panels), masculinity or femininity (middle panels), and attractiveness (bottom panels) for male and female speakers with naturally high F_0 and Fn compared to those with naturally low F_0 and Fn ($\uparrow F_0 \uparrow F$n vs $\downarrow F_0 \downarrow F$n). Not significant (NS): $P > 0.05$.

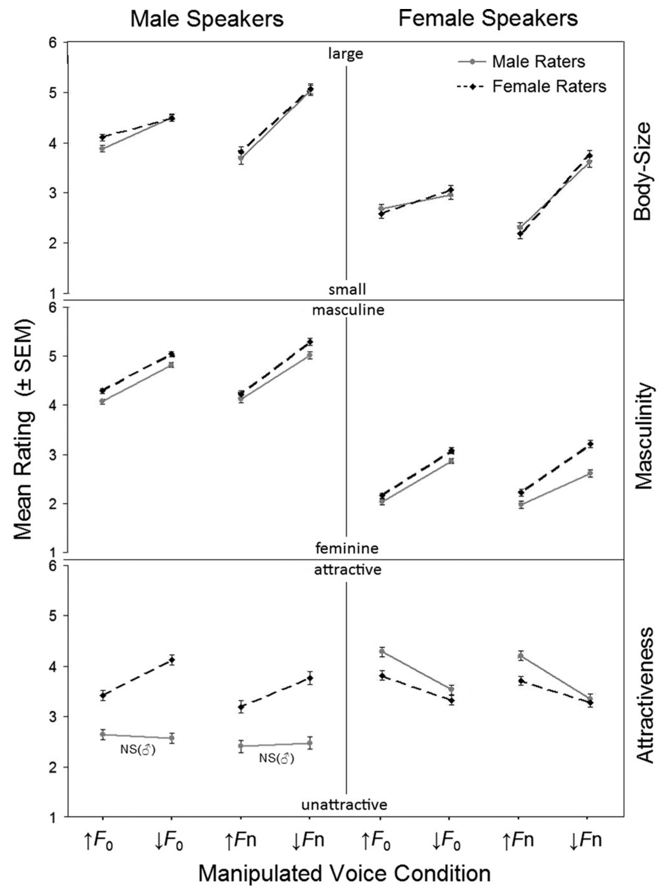

FIG. 2. Mean (±SEM) ratings of body size (top panels), masculinity or femininity (middle panels), and attractiveness (bottom panels) for male and female speakers with manipulated high vs low F_0 ($\uparrow F_0$ vs $\downarrow F_0$) or high vs low Fn ($\uparrow F$n vs $\downarrow F$n). NS: $P > 0.05$.

speakers with naturally low F_0 and Fn were rated as more attractive if they were male ($F_{1,59} = 15.35$, $P < 0.001$) and less attractive if they were female ($F_{1,59} = 17.76$, $P < 0.001$), and these effects held differentially for male and female listeners. Thus, only female listeners rated male speakers with low F_0 and Fn as more attractive than those with high F_0 and Fn ($F_{1,59} = 14.08$, $P < 0.001$), while only male listeners rated female speakers with low F_0 and Fn as less attractive than those with high F_0 and Fn ($F_{1,59} = 7.11$, $P = 0.009$). In other words, for both male and female listeners, ratings of the attractiveness of same-sex speakers did not vary as a function of voice F_0 and Fn, perhaps because rating the attractiveness of same-sex speakers is a more difficult or unnatural task.

These outcomes for the joint effects of either low or high F_0 and Fn in natural voices were corroborated in analyses of listener ratings of manipulated voices where F_0 or Fn were systematically increased or decreased (Fig. 2). Thus, male and female speakers whose voices had been lowered in F_0 were rated by both sexes as larger (m: $F_{1,59} = 53.48$, $P < 0.001$; f: $F_{1,59} = 18.86$, $P < 0.001$) and more masculine (m: $F_{1,59} = 233.8$, $P < 0.001$; f: $F_{1,59} = 246.38$, $P < 0.001$) compared to speakers whose F_0 had been increased. Similarly, male and female speakers whose voices had been lowered in Fn were rated as larger (m: $F_{1,59} = 158.92$, $P < 0.001$; f: $F_{1,59} = 193.75$, $P < 0.001$) and more masculine (m: $F_{1,59} = 180.43$, $P < 0.001$; f: $F_{1,59} = 120.14$, $P < 0.001$) than speakers with raised Fn, and this was true for both male and female listeners. Once again, both male and female listeners rated male speakers as being, on average, larger ($M = 4.32$) and more masculine ($M = 4.61$) than female speakers ($M = 2.87$, 2.56).

Ratings of attractiveness also differed as a function of the independent manipulations of F_0 or Fn and in ways that again interacted with the sex of both speaker and listener. Thus, speakers whose F_0 or Fn had been lowered, compared to those whose F_0 or Fn had been raised, were rated as more attractive if they were male (F_0: $F_{1,59} = 10.12$, $P = 0.002$; Fn: $F_{1,59} = 6.62$, $P = 0.013$) and less attractive if they were female (F_0: $F_{1,59} = 45.21$, $P < 0.001$; Fn: $F_{1,59} = 47.64$, $P < 0.001$). As before, these effects of F_0 and Fn on attractiveness ratings held only for female (and not male) listeners rating male speakers (F_0: $F_{1,59} = 14.75$, $P < 0.001$; Fn: $F_{1,59} = 4.15$, $P = 0.046$) but held for both male and female listeners rating female speakers.

Taken together, listeners consistently judged speakers with either low F_0 or low Fn, or both, as larger and more masculine, and either attractive if the speaker was male or unattractive if female. These patterns generally replicate outcomes reported in previous studies as reviewed earlier (see Introduction). One notable exception is that many previous studies have focused on listener ratings of male voices and often have not included female voices in the rating task. Here, our results suggest that the same voice features of relatively low F_0 or low Fn that elicit ratings of larger size and greater masculinity when they occur in male voices elicit the same ratings of larger size and greater masculinity (lower femininity) when they occur in female voices. This finding suggests that the algorithm listeners use to assess body size and masculinity (or femininity) from the voice may be a general one that is not differentiated by the sex of the speaker.

In addition, two previous studies have failed to find an effect of voice Fn on ratings of the attractiveness (Feinberg *et al.,* 2005) or pleasantness (Bruckert *et al.*, 2006) of male voices, but possibly for very sensible reasons. Thus, Feinberg *et al.* used a 5% shift in Fn relative to baseline and so it is possible that many of the formant contrasts actually fell below listeners' threshold discrimination abilities (see Sec. III). The study by Bruckert *et al.* involved assessments of "pleasantness" as opposed to "attractiveness" *per se* and the two constructs might be meaningfully different. At the same time, Bruckert *et al.* used naturalistic speech sequences in which the more protracted F_0 intonation patterns of different speakers were allowed to vary (i.e., involve variably rising or falling F_0 contours across words). Ultimately, these variable F_0 intonation patterns were found to influence assessments of vocal pleasantness more than did the absolute frequencies of either F_0 or Fn.

Although largely consistent, the result patterns reported to date do not yet allow us to say whether one or the other of voice F_0 or Fn exerts a greater influence on listeners' voice-based social judgments. This is an important issue for two reasons. First, according to conventional theories of speech production, voice F_0 and Fn are largely independent of one another (Müller, 1848; Fant, 1960). Empirical studies, and everyday experience, confirm this independence such that there can be considerable independent variation in F_0 and Fn both within and across utterances by the same individual and in the average values of these features in different individuals. For example, there can be individuals with relatively low F_0 but high Fn (and vice versa) as well as individuals exemplifying all possible combinations of F_0 and Fn between the extremes. As a result, it cannot simply be assumed that, for any given speaker, voice F_0 and Fn are necessarily highly correlated and thus largely redundant cues to size, masculinity, and attractiveness.

At the same time, it cannot simply be assumed that voice F_0 and Fn provide equally reliable cues to these biosocial dimensions, even in principle. For example, although variation in voice F_0 correlates with gross differences in body size (height and weight) between adults and children and between men and women, it does not appear to correlate very well with variation in body size within adults of either sex (Künzel, 1989; Van Dommelen and Moxness, 1995; Collins, 2000; Rendall *et al.*, 2005). Instead, variation in voice Fn appears to correlate more highly with body size variation at this level (i.e., within age-sex classes: Greisbach, 1999; González, 2004; Rendall *et al.*, 2005; Bruckert *et al.*, 2006; Evans *et al.*, 2006).

The corollary of both points is that listeners might naturally weight voice F_0 and Fn differentially in making voice-based social assessments of others. This possibility is examined in Secs. III and IV.

III. EXPERIMENT 2. FREQUENCY DISCRIMINATION THRESHOLDS FOR F_0 AND Fn

To test which of voice F_0 or Fn listeners might weight more heavily when making voice-based social judgments, an experiment was designed that pitted the two voice features against one another (see Experiment 3). Put simply, experimental voice stimuli were created to mimic natural speakers whose F_0 and Fn features provided conflicting cues to size, masculinity, and attractiveness because they combined relatively low F_0 with relatively high Fn (or vice versa). To conduct this experiment properly, however, the discordance between the two voice features must not be biased inadvertently in favor of one or the other as might occur if the differences in one feature were simply more easily discriminated than were differences in the other.

As a result, an important preliminary step was to establish discrimination thresholds for frequency-differences in F_0 and Fn. Although there is a substantial literature on frequency discrimination thresholds (or JND's) in humans, much of it focuses on the discrimination of either pure-tone stimuli (Wier *et al.*, 1977; Klinge and Klump, 2009) or broad-band stimuli with emphasized center-frequencies to simulate formants, rather than naturalistic speech material *per se* (Mermelstein, 1978; Kewley-Port and Watson, 1994). There is comparatively little research on discrimination thresholds of either F_0 or Fn in naturalistic speech (cf. Puts *et al.*, 2007). Because our interest is ultimately to determine how voice-based social judgments are variably affected by F_0 and Fn as they are embedded in naturalistic speech, first it was necessary to establish frequency discrimination thresholds for these two features and for the particular voice samples being used.

A. Methods

1. Voice stimulus selection and manipulation

The voices used as stimuli in this experiment were derived from the same speaker database and involved the same word material (single-syllable words in bVt context) as described for the previous experiment. From this database, eight new speakers (four males, four females) were selected and an identical set of four words was used for each one. Experimental stimuli involved a pairing of two sets of the same four bVt words (*boat*, *beat*, *book*, and *bait*) spoken by the same individual with a 600-ms silent interval between each word set and a 50-ms silent interval between each word within a set. In test-trials, each stimulus contained the original, unmanipulated recording (baseline condition) of the four-word set by a given speaker followed by a repetition of the same word set by the same speaker but with either F_0 or Fn increased by 1%–10% relative to that speaker's mean baseline values (manipulated conditions). In catch-trials, each stimulus contained two presentations of the same word set by a given speaker, either two presentations of the baseline condition or two presentations of one of the manipulated conditions (1%–10%). Frequency manipulations were performed using PRAAT either by specifying a new absolute pitch median for each word (in the case of F_0 modifications) or a proportional shift applied to all formants (in the case of Fn modifications). All experimental stimuli were standardized for length (at 5-s) and amplitude (65 dB).

2. Participants

Thirty-seven females and 25 males completed the experiment. All participants were recruited from the University of Lethbridge undergraduate community and received partial course credit.

3. Experimental procedure

Before testing began, participants were assigned randomly to one of four testing groups that involved making frequency discriminations in either F_0 ($n = 31$) or in Fn ($n = 31$) and in the voices of one or the other of two sets of four speakers (set A or B, each containing two male and two female speakers). Otherwise, the general procedures for this experiment were the same as those for the previous experiment. Participants once again completed the experiment in a sound-controlled room via a custom computer interface. The program first presented the participant with a set of instructions outlining the two-alternative, forced-choice frequency discrimination task. Participants were instructed that they would hear a series of voice-comparisons that would involve two presentations of the same set of four words spoken by the same person. They were informed that the two repetitions might be identical to one another in frequency or slightly different; their task was simply to indicate whether the repetitions were the *same* or *different* by clicking on the appropriate button on the computer screen bearing that label. Participants were able to watch a demonstration-trial before beginning the formal experiment. There were no time restrictions and they were free to replay any given stimulus one time using a *replay* button on the computer screen. The experiment then commenced. Each participant received 224 trials subdivided into four blocks. Each block contained ten test-trials and four catch-trials for each of the four speakers ($n = 56$ trials/block), with the order of these trials randomized.

4. Statistical analysis

Repeated measures ANOVA tests were used to examine variation in listeners' discrimination performance as a function of participant sex, speaker set (A or B), and frequency-difference condition (1%–10%). Separate tests were run for discriminations of F_0 and Fn and for male and female speakers. All ANOVAs were a mixed-design with listener sex and speaker set included as between-subjects factors and frequency-difference condition as a within-subjects factor. Statistical tests used an alpha level of 0.05.

In addition, a signal detection paradigm was used to calculate d-prime (d') scores for each incremental frequency change in F_0 or Fn from 1% to 10%. d' is a standard metric used to assess listener sensitivity to stimulus variation while controlling for individual response biases (Macmillan and Creelman, 2005). It can be estimated as the standardized difference between the means of the distributions of listeners' correct responses on test-trials (hits) vs their incorrect responses on catch-trials (false alarms). A d' value of 1 was used as a criterion for recognizing a reliable degree of discrimination performance (a JND).

B. Results and discussion

Overall, there were significant main effects of the frequency-difference conditions on the proportion of correct discriminations made by listeners for both F_0 and Fn in both male and female speakers (male F_0: $F_{9,243} = 38.25$, $P < 0.001$; female F_0: $F_{9,243} = 42.67$, $P < 0.001$; male Fn: $F_{9,243} = 65.80$, $P < 0.001$; female Fn: $F_{9,243} = 70.19$, $P < 0.001$). There were no effects of listener sex on discrimination performance (male F_0: $F_{1,27} = 1.75$, $P = 0.20$; female F_0: $F_{1,27} = 0.02$, $P = 0.88$; male Fn: $F_{1,27} = 0.08$, $P = 0.79$; female Fn: $F_{1,27} = 0.79$, $P = 0.38$). Thus, men and women performed equally well when discriminating F_0 and Fn in speakers of the same or the opposite-sex.

In F_0 discriminations for male speakers, there was also a significant main effect of speaker set ($F_{1,27} = 10.02$, $P = 0.004$) and an interaction between speaker set and frequency dimension ($F_{9,243} = 4.25$, $P < 0.001$) where differences in F_0 were more easily discriminated in male speakers in set B than in set A. This outcome suggests that the discriminability of voice features might vary slightly from speaker-to-speaker (see Table II). There were no other significant effects.

Predictably, listener discrimination of frequency-differences in F_0 and Fn improved steadily as the magnitude of the frequency-difference between the baseline and test voice stimulus increased from 1% to 10%. This pattern is shown in Table II where listeners' discrimination performance for each speaker and for all frequency-difference conditions is summarized as d' scores. Inspection of these scores confirms variation in the discriminability of F_0 and Fn differences from speaker-to-speaker. However, it also shows that d' scores generally increased steadily as the magnitude of the frequency-difference increased. Moreover, there was some consistency in the point at which d' scores exceeded and remained above the criterion value of 1. For both F_0 and Fn, that point was generally at frequency-differences of 4% to 7%.

Figure 3 displays these patterns graphically, where d' scores for each frequency-difference condition are calculated from the pooled responses of all listeners and plotted separately for male and female speakers. The pattern of steadily increasing discrimination performance is made clearer as is the consistency in the point at which this exceeds and

TABLE II. Summary of d' scores for the discrimination of variation in F_0 and Fn for each speaker in experiment 2.

Speaker		n[b]	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
		d' scores for F_0[a]										
Males	**1**	15	−0.39	−0.02	0.29	0.93	0.80	0.93	0.84	1.54	1.63	1.40
	2	15	−0.04	−0.27	0.25	−0.25	0.39	0.18	0.58	0.58	0.81	0.70
	3	16	0.44	0.66	0.63	1.76	1.61	1.57	2.18	1.91	1.92	2.33
	4	16	0.44	0.67	0.48	0.85	1.27	1.37	1.78	1.57	1.95	2.48
Females	**5**	15	0.46	1.39[c]	0.96	1.37	1.45	1.94	2.58	1.89	1.84	1.84
	6	15	0.22	0.56	0.90	1.42	0.64	1.23	1.00	1.53	0.83	1.54
	7	16	0.37	0.79	0.37	1.29	0.77	1.37	1.52	2.46	2.47	2.09
	8	16	−0.12	−0.33	0.62	0.23	0.56	1.24	0.86	1.07	1.83	1.77
		d' scores for Fn[a]										
Speaker		n[b]	**1%**	**2%**	**3%**	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
Males	**1**	16	−0.10	−0.27	−0.22	0.40	0.69	0.96	1.02	1.50	2.16	2.66
	2	16	−0.21	0.00	0.39	0.20	0.67	1.36	1.66	2.30	3.03	2.68
	3	15	0.58	0.54	0.90	1.07	1.28	2.08	1.66	1.91	2.99	3.00
	4	15	−0.09	−0.35	−0.09	0.15	0.14	0.43	1.07	1.40	1.93	2.19
Females	**5**	16	−0.03	0.20	0.14	0.32	1.02	1.69	2.06	2.00	2.44	2.49
	6	16	0.37	0.56	0.54	0.79[c]	1.21	1.15	1.89[c]	2.15	3.06	2.67
	7	15	0.00	−0.30	0.38	0.67	1.35	1.61	1.67	1.78	2.45	2.65
	8	15	−0.46	0.03	0.02	0.32	0.98	0.88	1.49	2.12	1.92	2.49

[a] d' scores were calculated for each incremental increase in frequency from 1% to 10% according to the formula: $d' = Z(\text{hits}) - Z(\text{false alarms})$.
[b] n = number of listeners discriminating the corresponding speaker.
[c] False alarm rate (FA) = 0; d' scores calculated by assigning FA = 1 (Stanislaw and Todorow, 1999).

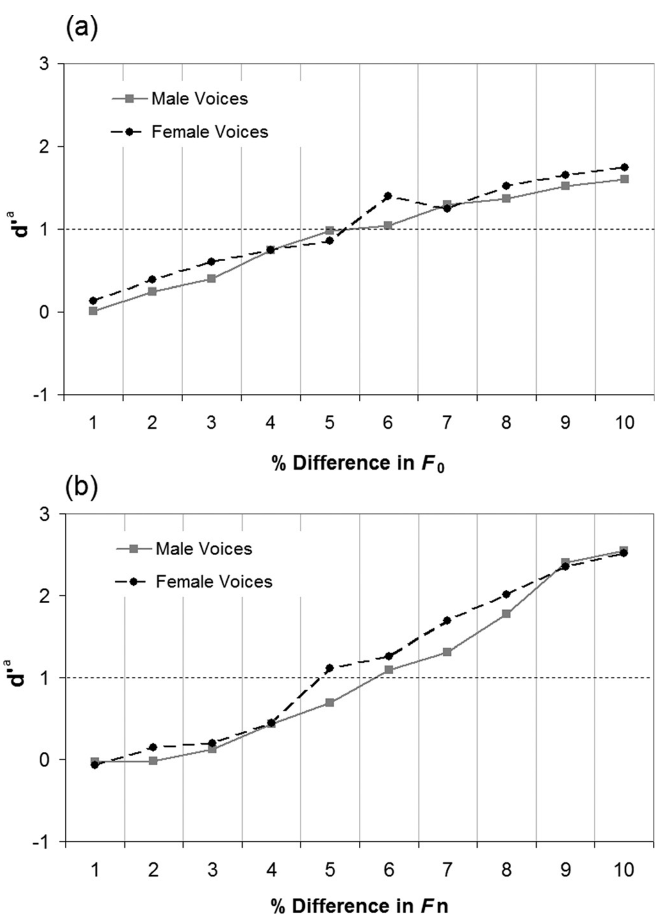

FIG. 3. Mean d' scores[1] for the discrimination of frequency-differences in F_0 (a) or Fn (b) as a function of speaker sex.

remains above the criterion. From these outcomes, reliable discrimination performance (1 JND) was delineated for F_0 differences in both male and female speakers at 6% and for Fn differences at 6% for male speakers and at 5% for female speakers.

As already noted, the discrimination task using naturalistic speech materials in Experiment 2 is not directly comparable to many of the past studies. Hence, it is difficult to compare outcomes directly. For example, previous studies using pure-tones or tone complexes (Moore, 1973; Wier *et al.*, 1977; Sinnott *et al.*, 1987, 1992; Klinge and Klump, 2009) have generally reported JND's lower (1%–2%) than those observed in Experiment 2 for the discrimination of F_0 in naturalistic speech. This difference might prove robust and point to higher discrimination thresholds generally for F_0 variation in naturalistic speech compared to synthetic tones or tone complexes. Stimuli used in Experiment 2 also involved a sequence of multiple words, rather than only a single sound, and thus included some variation among words in F_0 (as well as in Fn). This additional variation might also have contributed to higher discrimination thresholds in Experiment 2 compared to the discrimination of pure-tones in previous studies. Because the stimuli used for discrimination were sufficiently different in Experiment 2 compared to most past studies, it is difficult to fully reconcile the differences.

Other comparisons are more appropriate and promising. Thus, previous studies testing formant discrimination thresholds using synthesized vowel sounds have reported JND's more comparable to those observed in Experiment 2, in the range of 1.5%–9% (reviewed in Kewley-Port *et al.*, 1996) with variation among studies attributed to many subtle methodological differences between them including: whether or not participants were trained prior to testing (Kewley-Port and Watson, 1994; Kewley-Port, 1995); whether vowels were presented in isolation or in consonant context (Mermelstein, 1978; Kewley-Port and Watson, 1994; Ives *et al.*, 2005); whether only one or many different speakers were used and whether they were male or female and involved either natural or synthesized voices (Kewley-Port and Watson, 1994; Kewley-Port, 1995; Smith *et al.*, 2005). Although none of these studies involved short stretches of naturalistic speech or such a large sample of both speakers and listeners, the results are relatively consistent and encompass the discrimination thresholds of 5%–6% that we report.

The most comparable previous study to ours was by Puts *et al*. (2007) who reported comparable JND's in F_0 and Fn of approximately 7% and 4%, respectively, for full-sentence stimuli from male speakers.

IV. EXPERIMENT 3. THE RELATIVE SALIENCE OF F_0 VS Fn IN VOICE-BASED SOCIAL JUDGMENTS

Having established JND's for discrimination of F_0 and Fn in naturalistic speech samples, it was possible to develop an experiment to test listeners' weighting of the two voice features in voice-based social judgments. This was done by creating experimental stimuli that put the two voice features in direct conflict with one another by amounts that were equally perceptually discriminable. Once again listeners were asked to rate these voices for their relative size, masculinity or femininity, and attractiveness.

If listeners attended to and weighed F_0 and Fn cues equally in making such judgments, then there should be no consistent variation in their ratings of such stimuli because the conflicting effects of the F_0 and Fn cues would cancel each other. In contrast, if listeners attended to and weighed one voice feature more than the other, then there should be consistent variation in their ratings with the direction of the effects indicating which of the two features were more salient.

A. Methods

1. Voice stimulus selection and manipulation

The voices used as stimuli in this experiment were once again derived from the same speaker database and involved the same word material as described for the previous two experiments. From this database, 20 new speakers (ten males, ten females) were selected and for each speaker an identical set of five words was used. For male speakers used in this sample, the mean F_0 was 98 Hz (range: 82–117 Hz) and the mean values for $F1$–$F4$ were 469, 1619, 2584, and 3511 Hz, respectively. For female speakers, the mean F_0 was 204 Hz (range: 178–236 Hz) and the mean $F1$–$F4$ values were 586, 1949, 2969, and 4047 Hz, respectively.

Experimental stimuli for this study involved the same sample of single-syllable words used previously, this time specifically the words *bet*, *butt*, *bite*, *beat*, and *book* with each word separated by a 50-ms silent interval and all standardized to 65 dB. The natural (unmanipulated) sample of these words for each speaker was used as a baseline condition and then four additional experimental conditions were developed that involved modifications of F_0 and Fn. These modifications entailed putting pitch and formant cues in conflict with each other by raising F_0 while lowering Fn or by lowering F_0 while raising Fn. Modifications to F_0 and Fn were performed in increments designed to be equally perceptually discriminable to listeners and hence were done in JND increments guided by the results of the previous experiment. Those results indicated that, for both male and female speakers, the JND for F_0 was 6% while the JND for Fn was also 6% for male speakers and 5% for female speakers. In previous studies of F_0 and Fn effects, subtle frequency manipulations in the range of one JND to one or other feature have failed to produce consistent results (Feinberg *et al.*, 2005) compared to shifts of two or more JND's (Puts, 2005) and our own pilot testing confirmed this pattern. Hence, experimental stimuli were created using F_0 and Fn shifts of two and three JND's. Such shifts would be more perceivable to listeners but still result in parameters for both voice features within the normal range for each sex. For example, modifications to baseline F_0 values for the male speakers by two and three JND's yielded F_0 values of 67–138 Hz, while modifications of two and three JND's to baseline F_0 values for the female speakers yielded F_0 values of 145–279 Hz.

The four experimental conditions thus entailed stimuli in which either the F_0 was raised and Fn lowered by two or three JND's (i.e., $\uparrow 2F_0 \downarrow 2F$n; $\uparrow 3F_0 \downarrow 3F$n) or the F_0 was lowered and Fn raised by two or three JND's (i.e., $\downarrow 2F_0 \uparrow 2F$n; $\downarrow 3F_0 \uparrow 3F$n). Manipulations were performed using PRAAT by delineating a new absolute pitch median for each word and applying an equal proportional shift to all formants.

2. Participants

Thirty-two females and 36 males completed this experiment. All participants were recruited from the University of Lethbridge undergraduate community and received partial course credit. All participants self-reported as heterosexual.

3. Experimental procedure

Before testing began, participants for this experiment were assigned randomly to one of two versions of the experiment. In version A, voice stimuli and rating dimensions were completely randomized, within the constraint that each voice stimulus had to be rated three times, once for each of the three dimensions. In version B, rating dimensions were blocked, with each block of trials involving the same dimension (either size, or masculinity/femininity or attractiveness). The order of these blocks was randomized across participants. The order of voice stimuli presented within each block was also randomized within and between participants. In this experiment, listeners only rated speakers of the opposite-sex. Otherwise, both versions of the experiment included the complete speaker set (n = ten speakers of one or the other sex), the same set of baseline and experimental conditions for each speaker (n = five conditions per speaker), and the same three rating dimensions, for a total of 150 trials per participant.

All other procedures for this experiment were identical to those used in the first experiment. On each trial, participants were presented with a single voice stimulus of the opposite-sex with the opportunity to replay the stimulus once. Their task was to rate each voice on one of the three dimensions using the same six-point scale with no mid-point described for the first experiment. Subjects were given unlimited time to complete the experiment. Rest-breaks were scheduled at 50-trial intervals.

4. Statistical analysis

Statistical testing paralleled that for the first experiment except that experimental version was included as a between-subjects factor and both rating dimension and experimental frequency condition were within-subjects factors in rmANOVA tests. Trials on which participant response time was less than 50-ms were removed prior to analysis.

B. Results and discussion

Results are shown in Fig. 4. For female listeners responding to male speakers, there were significant main effects of experimental frequency condition ($F_{4,120} = 160.6$, $P < 0.001$) and rating dimension ($F_{2,60} = 90.17$, $P < 0.001$) as well as a significant interaction between the two ($F_{8,240} = 85.76$, $P < 0.001$). *Post-hoc* analyses showed that females' ratings of all three dimensions were significantly greater for the voice conditions where Fn was lowered while F_0 was raised either by two JND's or by three JND's compared to the corresponding voice conditions where F_0 was lowered and Fn was raised (size: $\uparrow 2F_0 \downarrow 2F$n, $M = 4.88$; $\downarrow 2F_0 \uparrow 2F$n, $M = 3.05$; $\uparrow 3F_0 \downarrow 3F$n, $M = 5.53$; $\downarrow 3F_0 \uparrow 3F$n, $M = 2.44$; masculinity: $\uparrow 2F_0 \downarrow 2F$n, $M = 5$; $\downarrow 2F_0 \uparrow 2F$n, $M = 3.44$; $\uparrow 3F_0 \downarrow 3F$n, $M = 5.47$; $\downarrow 3F_0 \uparrow 3F$n, $M = 3.02$; attractiveness: $\uparrow 2F_0 \downarrow 2F$n, $M = 3.58$; $\downarrow 2F_0 \uparrow 2F$n, $M = 2.93$; $\uparrow 3F_0 \downarrow 3F$n, $M = 2.53$; $\downarrow 3F_0 \uparrow 3F$n, $M = 1.86$). There was no effect of version on female's ratings ($F_{1,30} = 2.18$, $P = 0.15$), indicating that blocking trials by rating dimension did not affect ratings.

Similarly, for male listeners responding to female speakers, there were significant effects of experimental frequency condition ($F_{4,136} = 46.72$, $P < 0.001$), rating dimension ($F_{2,68} = 6.72$, $P = 0.002$), and a significant interaction between the two ($F_{8,272} = 112.72$, $P < 0.001$). Likewise, *post-hoc* analyses showed that males' ratings of size and masculinity were significantly greater for the voice conditions where Fn was lowered while F_0 was raised either by two JND's or by three JND's compared to the corresponding voice conditions where F_0 was lowered and Fn was raised (size: $\uparrow 2F_0 \downarrow 2F$n, $M = 4.05$; $\downarrow 2F_0 \uparrow 2F$n, $M = 2.68$; $\uparrow 3F_0 \downarrow 3F$n, $M = 4.86$; $\downarrow 3F_0 \uparrow 3F$n, $M = 2.47$; masculinity: $\uparrow 2F_0 \downarrow 2F$n, $M = 3.34$; $\downarrow 2F_0 \uparrow 2F$n, $M = 2.7$; $\uparrow 3F_0 \downarrow 3F$n, $M = 4.14$; $\downarrow 3F_0 \uparrow 3F$n, $M = 2.88$). Male ratings of female attractiveness were the reverse of this pattern where their ratings were significantly greater for the voice conditions where Fn was raised while F_0 was lowered either by two JND's or three JND's compared to the corresponding voice conditions where F_0 was raised while Fn was lowered (attractiveness: $\downarrow 2F_0 \uparrow 2F$n, $M = 3.85$; $\uparrow 2F_0 \downarrow 2F$n, $M = 3.15$; $\downarrow 3F_0 \uparrow 3F$n, $M = 3.1$; $\uparrow 3F_0 \downarrow 3F$n, $M = 2.05$). There was no effect of version on male's ratings ($F_{1,34} = 0.74$, $P = 0.4$).

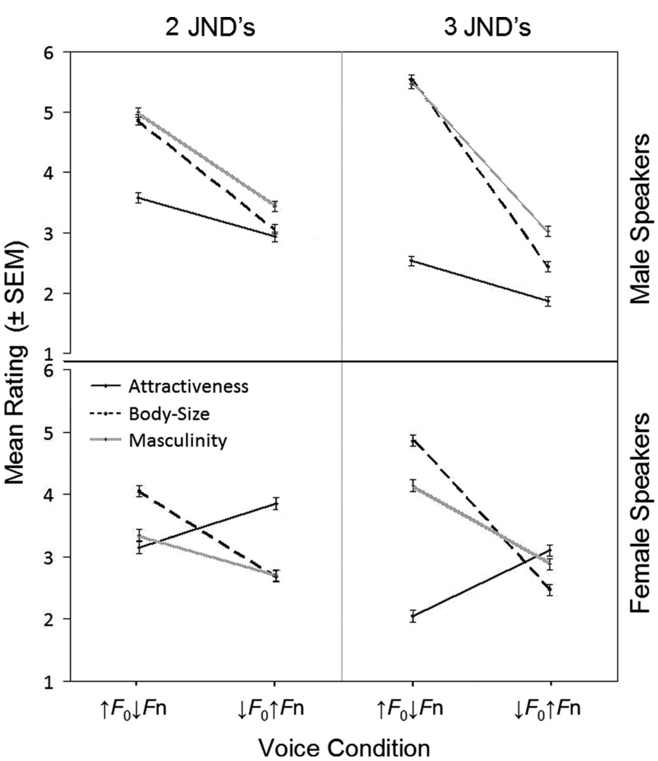

FIG. 4. Mean (±SEM) ratings of body size, masculinity (femininity), and attractiveness for male speakers (top panels) and female speakers (bottom panels) with manipulated high F_0 and low Fn ($\uparrow F_0$ vs $\downarrow F$n) compared to those with manipulated low F_0 and high Fn ($\downarrow F_0$ vs $\uparrow F$n). Separate panels are shown for manipulations involving either two JND's (left panels) or three JND's (right panels). Ratings: 1 = small, feminine, or unattractive; 6 = large, masculine, or attractive. All effects shown are significant at $P < 0.05$.

Taken together, listeners of both sexes consistently rated speakers of the opposite-sex as larger and more masculine (or less feminine) in experimental conditions in which Fn had been lowered (while F_0 had been raised). Speakers with lowered Fn (but raised F_0) were also rated more attractive if they were male speakers and less attractive if they were female speakers. Because ratings of larger size and greater masculinity (and greater attractiveness in male voices but lower attractiveness in female voices) are generally associated with lower-frequency voices, as reviewed earlier, these outcomes are consistent in pointing to listener ratings preferentially tracking the Fn cues over the F_0 cues in this experiment.

V. GENERAL DISCUSSION

Briefly summarizing, the results of Experiment 1 were consistent with those of most previous studies in finding that speakers of either sex with relatively low voice F_0 and/or low Fn elicited ratings of larger size and greater masculinity (or lower femininity). In addition, speakers with relatively low F_0 and/or low Fn were rated as more attractive if male and less attractive if female (Figs. 1 and 2). After quantifying JND's for the discrimination of F_0 and Fn in Experiment 2 (Fig. 3), the possibility that listeners might differentially weight the two features was tested in Experiment 3 that put the two voice features in conflict. Results showed that, in assessing all three social dimensions, listeners appeared to weight Fn cues more heavily than F_0 cues (Fig. 4).

This latter outcome is notable in as much as previous studies that have attempted to test the relative salience of F_0 and Fn have generally reported stronger effects for F_0 in ratings of speaker size (Collins, 2000), speaker sex (Coleman, 1976) or masculinity (Collins, 2000), and speaker attractiveness (Collins, 2000; Collins and Missing, 2003; Bruckert *et al*., 2006). This pre-eminence of voice F_0 in such social ratings is certainly intuitive (Ohala, 1983). The voice differences typically associated with these distinctions in the natural classes of speakers that exemplify them (i.e., men and women) are much larger for F_0 than for Fn: there is a nearly two-fold difference in voice F_0 between men and women, while the Fn differences between the sexes are approximately 15% (Petersen and Barney, 1952; Bachorowski and Owren, 1999; Rendall *et al.,* 2005). Hence, using F_0 more than Fn cues to judge speaker sex, and the closely associated dimensions of masculinity/femininity and possibly also body size, might simply be taking advantage of the much greater natural variation in F_0 compared to Fn that mark these distinctions between males and females.

At the same time, this point highlights the value of controlling the magnitude and relative discriminability of F_0 and Fn differences used in experimental tests of their relative salience. Previous studies have not made this a priority but it was the explicit goal of Experiments 2 and 3. By first establishing equally discriminable differences in F_0 and Fn and then putting the two cues in conflict by equally discriminable amounts, it was possible to more definitively test their relative effects. And, at least under these conditions, results suggest that listeners weight Fn cues more than F_0 cues in voice-based social judgments.

Ultimately, this outcome does not actually contradict the results of previous studies so much as complement them in showing that, under controlled conditions in which differences in the two voice features are made equally perceptually discriminable, the typical priority of F_0 compared to Fn cues obtained for many naturalistic speaker contrasts can be reversed. Notably, one other recent study established and used JND's for F_0 and Fn in the synthesis of experimental voice stimuli and found that, while both low F_0 and low Fn elicited high ratings of social and physical dominance in male speakers, the effects were greater for Fn than for F_0 (Puts *et al*., 2007).

If true and generalizable in future work, the apparent priority of Fn over F_0 in contexts like those tested in Experiment 3 might occur for one of several reasons. First, it might be that the Fn variation in Experiment 3 was simply more discriminable than in previous studies. Care was taken to avoid this possibility by the precautions of first establishing independent discrimination thresholds for F_0 and Fn and then using stimuli that put the two features in conflict by equivalent JND increments. Nevertheless, it is possible that thresholds for discriminating variation in Fn and F_0 might differ slightly in the context of complex stimuli containing variation in both features simultaneously compared to those observed for each feature in isolation. Indeed, Kewley-Port and Watson (1994) reported an effect like this where discrimination thresholds for Fn differences varied slightly as a function of variation in the F_0 of the stimuli in which they were jointly embedded. It is not clear exactly how or why such a shifting discrimination threshold effect might consistently favor Fn over F_0 cues. However, in the broader context of examining potential nonlinearities in the auditory processing of speech signals, this possibility might be worth considering in future research.

A second and related possibility is that the variation in Fn was not more discriminable *per se* but rather more salient in the degree to which it indexed variation in the biosocial dimensions being rated. For example, there is a two-fold difference in average voice F_0 between men and women but a difference of only about 15% in voice Fn; hence, a 6% difference (or one JND) in voice Fn might have a greater proportional effect on perceptions of the relative masculinity (or femininity) of a given voice than would a 6% difference (one JND) in voice F_0. Thus, although the 6% shifts in Fn and F_0 used were specifically designed to be equally *discriminable*, they may not have been equally *salient* in terms of the proportional difference they represented in each rating dimension. Again, it is unclear how this kind of effect might consistently favor Fn over F_0 in all three social assessments (but see below).

An additional possibility for why listeners might have tracked Fn over F_0 is that Fn is actually a more reliable cue to the biosocial dimensions being evaluated. There is certainly some reason to think this might be true *vis a vis* body size. Thus, Fitch (1994, 1997, 2000) proposed that voice Fn should correlate better than voice F_0 with body size variation because vocal-tract length is constrained by the size of the bony anatomy that comprises it which in turn should be determined by developmental programs regulating growth of bony structures and overall body size. In contrast, vocal fold length is comparatively unconstrained because the larynx in humans is positioned low in the vocal-tract and largely outside the bony confines of the skull; hence, vocal fold development should be less tied to growth programs determining overall body size. There is some evidence that such emancipation of vocal-fold length holds in other mammals as well, even without a descended larynx (Fitch and Hauser, 1995; Fitch, 2000).

Subsequent empirical work has generally confirmed this proposal for several mammalian species, reporting reliable three-way correlations between body size, vocal-tract length, and Fn (Fitch, 1997; Fitch and Giedd, 1999; Riede and Fitch, 1999; Reby and McComb, 2003). Studies of humans have also generally reported similar correlations between body size and either vocal-tract length or Fn across age-sex classes as well as within both adult male (Fitch and Giedd, 1999; Greisbach, 1999; González, 2004; Rendall *et al.,* 2005; Evans *et al.,* 2006) and adult female speakers (Griesbach, 1999; Collins and Missing, 2003; González, 2004), although a small number have failed to find correlations within one or the other sex (Van Dommelen and Moxness, 1995; Rendall *et al.,* 2005; Bruckert *et al.,* 2006). In contrast, although variation in F_0 tracks gross differences in body size between adults and children and between men and women, research is consistent in finding that it is not strongly correlated with body size variation within adults of either sex (Hollien and Jackson, 1972; Majewski *et al.*, 1972; Künzel, 1989; Hollien *et al.*, 1994; Van Dommelen and Moxness, 1995; Collins and Missing, 2003; González, 2004; Rendall *et al.,* 2005; Evans *et al.,* 2006). Hence, listeners in the experiments reported here might have tracked *F*n over F_0 in their assessments of speaker size because *F*n is in fact the more reliable cue, at least within sexes.

It is not so apparent why listeners also should have weighed *F*n cues more heavily in assessments of masculinity (and femininity) and attractiveness, if they did not do so for reasons outlined in the second possibility above, because neither rating dimension seems to have any privileged connection to *F*n variation *per se*. Indeed, masculinity and femininity, which are closely tied to the difference in biological sex, are most conspicuously cued by the sizeable difference in voice F_0 between men and women. And, in as much as assessments of attractiveness are rooted in assessments of relative masculinity and femininity, it would seem that attractiveness ratings might then also be rooted in the more conspicuous F_0 difference between men and women.

At the same time, there is some evidence to suggest that *F*n can predict testosterone levels in men (Bruckert *et al.,* 2006). It is also the case that the changes in male vocal anatomy at puberty include a small secondary descent of the larynx, yielding a subtle additional, and size-independent, increase in overall vocal-tract length in adult men compared to adult women (Lieberman *et al*., 2001; Fitch and Giedd, 1999). Hence, listeners might interpret lower *F*n cues as indicating greater masculinity and thus also greater attractiveness in male voices.

At the same time, it is possible that the prioritization of *F*n cues in assessments of both masculinity (or femininity) and attractiveness simply followed their prioritization as more reliable cues in assessments of body size and that they did so because of the inherent size-related associations among all three rating dimensions. Men and women differ consistently in size, at least on average; hence, notions of masculinity and femininity and the notions of attractiveness that might flow from them could be commonly anchored by basic perceptions of size, with "larger voices" epitomizing masculinity and attractiveness in males and "smaller voices" epitomizing femininity and attractiveness in females.

If true, an important corollary of the latter two possibilities is that size, masculinity/femininity, and attractiveness might not actually represent independent rating dimensions in voice-based social judgments, even though investigators typically analyze them independently and generally assume that listeners can evaluate them independently. Instead, they might be intrinsically linked to one another and tied to basic perceptions of size which can be a relevant dimension of perceptual evaluation not just in the perception of human voices but in the perception of, and adaptive responding to, many other environmental sounds (for further discussion of this possibility, see Grassi, 2005; Rendall *et al.,* 2007).

As a first step toward discriminating among these possibilities, additional experiments are being conducted to test the degree to which ratings of speaker size, masculinity, and attractiveness inherently overlap or can be independent of one another in voice-based social judgments.

ACKNOWLEDGMENTS

We are grateful to Sandeep Mishra for collecting and making available the speaker recordings used in experiments. We also thank John Vokey and Louise Barrett for discussions related to experimental design and analysis. Research was approved by the Human Subjects Research Committee of the University of Lethbridge and was supported by operating grants from the Natural Sciences and Engineering Research Council of Canada.

[1]Mean d′ scores for each frequency-difference condition were calculated separately for *F*0 and *F*n and for male and female speakers using discrimination performance data pooled from all raters (*F*0: $n = 31$; *F*n: $n = 31$).

Abitol, J., Abitol, P., and Abitol, B. (**1999**). "Sex hormones and the female voice," J. Voice. **13**, 424–446.

Alonso, L. C., and Rosenfield, R. L. (**2002**). "Oestrogens and puberty," Best Pract. Res. Clin. Endocrinol. Metab. **16**, 13–30.

Bachorowski, J.-A., and Owren, M. J. (**1999**). "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech," J. Acoust. Soc. Am. **106**, 1054–1063.

Boersma, P. P. G. (**2001**). "Pratt, a system for doing phonetics by computer," Glot Internat. **5**, 341–345.

Bruckert, L., Liénard, J.-S., Lacrois, A., Kreutzer, M., and Leboucher, G. (**2006**). "Women use voice parameters to assess men's characteristics," Proc. R. Soc. B. **273**, 83–89.

Bryant, G. A., and Haselton, M. G. (**2009**). "Vocal cues of ovulation in human females," Biol. Lett. **5**, 12–15.

Childers, D., and Wu, K. (**1991**). "Gender recognition from speech. Part II: Fine analysis," J. Acoust. Soc. Am. **90**, 1841–1856.

Clutton-Brock, T. H., and Albon, S. D. (**1979**). "The roaring of red deer and the evolution of honest advertising," Behaviour **69**, 145–170.

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 Hear. Res. **19**, 168–180.

Collins, S. A. (**2000**). "Men's voices and women's choices," Anim. Behav. **60**, 773–780.

Collins, S. A., and Missing, C. (**2003**). "Vocal and visual attractiveness are related in women," Anim. Behav. **65**, 997–1004.

Dabbs, J. M., Jr., and Mallinger, A. (**1999**). "High testosterone levels predict low voice pitch among men," Pers. Individ. Differ. **27**, 801–804.

Darwin, C. (**1859**). *The Origin of Species by Means of Natural Selection* (John Murray, London).

Darwin, C. (**1871**). *The Descent of Man, and Selection in Relation to Sex* (John Murray, London).

Davies, N. B., and Halliday, T. R. (**1978**). "Deep croaks and fighting assessment in toads Bufo bufo," Nature **274**, 683–685.

Evans, S. N., Neave, N., and Wakelin, D. (**2006**). "Relationship between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice," Biol. Psychol. **72**, 160–163.

Evans, S. N., Neave, N., Wakelin, D., and Hamilton, C. (**2008**). "The relationship between testosterone and vocal frequencies in human males," Physiol. Behav. **93**, 783–788.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague, Holland).

Feinberg, D. R. (**2008**). "Are human faces and voices ornaments signaling common underlying cues to mate value?" Evol. Anthropol. **17**, 112–118.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., and Perrett, D. I. (**2008**). "The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices," Perception **37**, 615–623.

Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., Hillier, S. G., and Perrett, D. I. (**2006**). "Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice," Horm. Behav. **49**, 215–222.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., and Perrett, D. I. (**2005**). "Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices," Anim. Behav. **69**, 561–568.

Fitch, W. T. (**1994**). "Vocal tract length perception and the evolution of language" (Unpublished doctoral dissertation, Brown University, Providence, Rhode Island).

Fitch, W. T. (**1997**). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," J. Acoust. Soc. Am. **102**, 1213–1222.

Fitch, W. T. (**2000**). "The evolution of speech: A comparative review," Trends Cogn. Sci. **4**, 258–267.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract. A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Fitch, W. T., and Hauser, M. D. (**2003**). "Unpacking 'honesty': Vertebrate vocal production and the evolution of acoustic signals," in *Acoustic Communication*, edited by A. M. Simmons, R. R. Fay, and A. N. Popper (Springer, New York), pp. 65–137.

Folstad, I., and Karter, A. J. (**1992**). "Parasites, bright males, and the immunocompetence handicap," Am. Nat. **139**, 603–622.

González, J. (**2004**). "Formant frequencies and body size of a speaker: A weak relationship in adults humans," J. Phonetics. **32**, 277–287.

Grassi, M. (**2005**). "Do we hear size or sound? Balls dropped on plates," Percept. Psychophys. **67**, 274–284.

Greisbach, R. (**1999**). "Estimation of speaker height from formant frequencies," Forensic Linguist. **6**, 265–277.

Hintze, J. L. (**1989**). *NCSS Version 5.1* (Kayeswille, Utah).

Hollien, H., Green, R., and Massey, K. (**1994**). "Longitudinal research on adolescent voice changes in males," J. Acoust. Soc. Am. **96**, 2646–2654.

Hollien, H., and Jackson, B. (**1973**). "Normative data on the speaking fundamental frequency characteristics of young adult males," J. Phonetics. **1**, 117–120.

Ives, D. T., Smith, D. R. R., and Patterson, R. D. (**2005**). "Discrimination of speaker size from syllable phrases," J. Acoust. Soc. Am. **118**, 3816–3822.

Kewley-Port, D. (**1995**). "Thresholds for formant-frequency discrimination of vowels in consonantal context," J. Acoust. Soc. Am. **97**, 3139–3146.

Kewley-Port, D., Li, X., Zheng, Y., and Neel, A. T. (**1996**). "Fundamental frequency effects on thresholds for vowel formant discrimination," J. Acoust. Soc. Am. **100**, 2462–2470.

Kewley-Port, D., and Watson, C. S. (**1994**). "Formant frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.

Klinge, A., and Klump, G. M. (**2009**). "Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans," J. Acoust. Soc. Am. **125**, 304–314.

Künzel, H. J. (**1989**). "How well does average fundamental frequency correlate with speaker height and weight," Phonetica **46**, 117–125.

Lieberman, P., and Blumstein, S. E. (**1988**). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge University Press, Cambridge, MA).

Lieberman, D. E., McCarthy, R. C., Hiiemae, K., and Palmer, J. B. (**2001**). "Ontogeny of postnatal hyoid and larynx descent in humans," Arch. Oral Biol. **46**, 117–128.

Lindenfors P., Gittleman J. L., and Jones K. E. (**2007**). "Sexual size dimorphism in mammals," in *Sex, Size and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism,* edited by D. J. Fairbairn, W. U. Blanckenhorn, and T. Szekely (Oxford University Press, Oxford), pp. 19–26.

Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge, MA).

Majewski, W., Hollien, H., and Zalewski, J. (**1972**). "Speaking fundamental frequency of Polish adult males," Phonetica **25**, 119–125.

Mermelstein, P. (**1978**). "Difference limens for formant frequencies of steady-state and consonant-bound vowels," J. Acoust. Soc. Am. **63**, 572–580.

Moore, B. C. J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.

Müller, J. (**1848**). *The Physiology of the Senses, Voice and Muscular Motion with the Mental Faculties* (Walton & Maberly, London).

Oguchi, T., and Kikuchi, H. (**1997**). "Voice and interpersonal attraction," Jpn. Psychol. Res. **39**, 56–61.

Ohala, J. J. (**1983**). "Cross-language use of pitch: An ethological view," Phonetica **40**, 1–18.

Owren, M. J., Berkowitz, M., and Bachorowski, J.-A. (**2007**). "Listeners judge talker sex more efficiently from male than female vowels," Percept. Psychophys. **69**, 930–941.

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

Puts, D. A. (**2005**). "Mating context and menstrual phase affect women's preferences for male voice pitch," Evol. Hum. Behav. **26**, 388–397.

Puts, D. A., Hodges, C. R., Cardenas, R. A., and Gaulin, S. J. C. (**2007**). "Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men," Evol. Hum. Behav. **28**, 340–344.

Reby, D., and McComb, K. (**2003**). "Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags," Anim. Behav. **65**, 519–530.

Rendall, D., Kollias, S., Ney, C., and Lloyd, P. (**2005**). "Pitch (F_0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," J. Acoust. Soc. Am. **117**, 944–955.

Rendall, D., Vokey, J. R., and Nemeth, C. (**2007**). "Lifting the curtain on the Wizard of Oz: Biased voice-based impressions of speaker size," J. Exp. Psychol. Hum. Percept. Perform. **33**, 1208–1219.

Riding, D., Lonsdale, D., and Brown, B. (**2006**). "The effects of average fundamental frequency and variance of fundamental frequency on male vocal attractiveness to women," J. Nonverbal Behav. **30**, 55–61.

Riede, T., and Fitch, W. T. (**1999**). "Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*)," J. Exp. Biol. **202**, 2859–2869.

Ryan, M. J. (**1983**). "Sexual selection and communication in a neotropical frog, *Physalaemus pustulosus*," Evolution **37**, 261–272.

Saxton, T. K., Caryl, P. G., and Roberts, S. C. (**2006**). "Vocal and facial attractiveness judgments of children, adolescents and adults: The ontogeny of mate choice," Ethology **112**, 1179–1185.

Sinnott, J. M., Brown, C. H., and Brown, F. E. (**1992**). "Frequency and intensity discrimination in Mongolian gerbils, African monkeys and humans," Hear. Res. **59**, 205–212.

Sinnott, J. M., Owren, M. J., and Peterson, M. R. (**1987**). "Auditory frequency discrimination in primates: Species differences (*Cercopithecus, Macaca, Homo*)," J. Comp. Psychol. **101**, 126–131.

Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (**2005**). "The processing and perception of size information in speech sounds," J. Acoust. Soc. Am. **117**, 305–318.

Stevens, K. N. (**1998**). *Acoustic Phonetics.* (MIT Press, Cambridge, MA).

Titze, I. R. (**1989**). "Physiological and acoustic differences between male and female voices," J. Acoust. Am. Soc. **85**, 1699–1707.

Titze, I. R. (**1994**). *Principles of Voice Production.* (Prentice Hall, Englewood Cliffs, NJ).

Van Dommelen, W. A., and Moxness, B. H. (**1995**). "Acoustic parameters in speaker height and weight identification: Sex-specific behaviour," Lang. Speech **38**, 267–287.

Varosanec-Skaric, G. (**1999**). "Relation between voice pleasantness and distribution of the spectral energy," in *Proceedings of the XIV Conference in Phonetic Sciences*, San Francisco, pp. 1013–1016.

Vorperian, H. K., Kent, R. D., Gentry, L. R., and Yandell, B. S. (**1999**). "Magnetic resonance imaging procedures to study the concurrent anatomic development of the vocal tract structures: Preliminary results," Int. J. Pediatr. Otorhinolaryngol. **49**, 197–206.

Wier, C. C., Jesteadt, W., and Green, D. M. (**1977**). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Am. Soc. **61**, 178–184.