Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Phonetics 34 (2006) 202–240

Journal of Phonetics

www.elsevier.com/locate/phonetics

The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech

Benjamin Munson*, Elizabeth C. McDonald, Nancy L. DeBoe, Aubrey R. White

Department of Speech-Language-Hearing Sciences, University of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr., SE, Minneapolis, MN 55455, USA

Received 7 January 2005; received in revised form 25 April 2005; accepted 3 May 2005

Abstract

Three experiments examined acoustic and perceptual characteristics of the speech of gay; Lesbian; Bisexual (GLB) and heterosexual people. Experiment 1 examined the acoustic characteristics of single words produced by both men and women who identified as either GLB or heterosexual. The largest differences between GLB and heterosexual women were in the F1 frequency of /ɛ/ and the F2 frequency of /oʊ/. The largest differences between the groups of men were in the F1 frequency of /ɛ/ and /æ/, and the spectral skewness of the fricative /s/. Experiment 2 showed that listeners' judgments of perceived sexual orientation were related to the acoustic parameters found to differ in Experiment 1: Listeners showed greater sensitivity to differences in men's sexual orientation when listening to words containing low front vowels and than when listening to words containing back vowels. Moreover, Regression analyses showed that judgments of men's sexual orientation were influenced by /s/ skewness, the F1 frequency of low front vowels, and the F2 frequency of back vowels. Judgments of women's sexual orientation were predicted most strongly by the F1 frequency of low front vowels and the F2 frequency of back vowels. Experiment 3 showed that the judgments of perceived sexual orientation collected in Experiment 2 were strongly related to judgments of perceived height and perceived speech clarity made by independent groups of listeners. Taken together, the results provide a more comprehensive picture of the acoustic and perceptual characteristics of GLB speech styles than has been provided by previous research. Moreover, the results of Experiment 3 suggest that listeners' percepts of GLB speech styles may be related to their perception of other speech characteristics.

*Corresponding author. Tel.: +612 624 0304; fax: +612 624 7586.
E-mail address: Munso005@umn.edu (B. Munson).

0095-4470/$ - see front matter
doi:10.1016/j.wocn.2005.05.003

1. Introduction

1.1. Sexual orientation and speech

Speech communicates multiple types of information in parallel. A single spoken CVC word, such as an utterance of the word *man*, can convey semantic information (the spectral patterns characteristic of /m/, /æ/, and /n/ would indicate that the talker is referring to an adult human male); the talker's pragmatic intent (in some dialects of English, a rising f0 contour would indicate that the talker intends the listener to interpret the utterance with respect to upcoming information in discourse; Pierrehumbert & Hirschberg, 1990); the talker's emotional state (a relatively flat f0 contour might indicate the talker's indifference, boredom, or grief; Scherer, 1986); the talker's regional dialect (i.e., the articulation of /æ/ vowel might indicate that the talker speaks a Northern Cities dialect; Labov, 1994); and idiosyncratic acoustic characteristics that would allow a familiar listener to identify the talker.

The focus of this investigation is talkers' use of the speech signal to convey and perceive social-group membership. Specifically, this investigation examines talkers' manipulation of phonetic detail to mark whether they are gay/lesbian/bisexual (henceforth GLB) or heterosexual. It has often been speculated that a subset of GLB people use a distinctive speech style that allows naïve listeners to identify their sexual orientation in the absence of an overt disclosure. Anecdotal reports characterize gay/bisexual (henceforth G/B) men's speech style as having 'dynamic intonation', 'lisping', or a higher-pitched voice (see reviews in Gaudio (1994) and Zwicky (1997)). Descriptions of lesbian/bisexual (henceforth L/B) women's speech patterns are less common, but generally indicate more monotone speech patterns than heterosexual women (see reviews in Moonwoman-Baird (1997) and Waksler (2001)).

One of the major challenges in studying GLB speech styles is that anecdotal descriptions of this style are often cast in pejorative terms. Descriptions of G/B male speech styles often have accompanied broad characterizations of G/B men's behavior as effeminate; descriptions of L/B women's speech styles typically accompany descriptions of behavior as being inappropriately masculine. Moreover, popular-culture descriptions and discussions of the G/B male speech style often use the word 'lisp', suggesting the style to be indicative of pathology. These descriptions may come from members of the GLB community itself. Sedaris (2001) describes his experience receiving speech therapy for /s/ misarticulation in childhood. He makes the observation that the other boys receiving therapy for /s/ concurrently appeared less masculine than their peers not receiving therapy. Thompson and Bellini (1998) present a fictionalized autobiography of a stereotypically gay character who describes himself as lisping. These examples are only two of many, and show that the popular-culture stereotypes of GLB speech styles often involve negative evaluations.

Studies of GLB speech styles are additionally complicated by the perception of many lay people that research on this topic confirms these negative stereotypes. That is, it is thought that this research aims only to reinforce the negative, pathologized stereotypes described above. Finally, this area of research—like research on stylistic variants in many social groups—takes place in the context of a GLB community that is both large and culturally diverse. Not surprisingly, not all people in this diverse community speak with an identifiably GLB speech style. Critiques of research on this topic inevitably invoke the heterogeneity of the speech styles used by GLB people when questioning the generality of research findings on this topic (e.g., Murray, 1999). As emphasized by Zwicky (1997), these and other issues make research on linguistic variation in GLB people challenging.

Despite this, a small number of empirical studies have suggested that a distinctive GLB speech style does exist. Most of these studies have examined men only. Gaudio (1994) reports on acoustic and perceptual analyses of a small number of self-identified gay and heterosexual male speakers from the San Francisco bay area ($n = 4$ in both groups). Talkers read two types of texts, described as *technical* and *dramatic*. A variety of measures of mean fundamental frequency and variability in fundamental frequency were taken. In addition, 13 listeners rated the samples on four scales (*straight–gay*, *effeminate–masculine*, *reserved–emotional*, *affected–ordinary*). Self-identified gay men were rated as sounding more gay and more effeminate than self-identified heterosexual men in both texts. Measures of perceived sexual orientation and perceived masculinity/effeminacy were highly correlated. Group differences in the other two perceptual dimensions were not found for both text types. Measures of pitch range and pitch variability were moderately correlated with measures of perceived sexual orientation and perceived masculinity/effeminacy (Pearson's r for all comparisons was approximately 0.65). These did not achieve statistical significance, perhaps due to the small sample size. This suggests that one of the cues that listeners use when perceiving sexual orientation is modulation of fundamental frequency.

A similar study was presented by Linville (1998). Linville examined acoustic differences between a small sample of Milwaukee, Wisconsin-area self-identified G/B and heterosexual men's speech ($n = 5$ and 4, respectively). Linville collected readings of a dramatic text. She analyzed a variety of acoustic measures, including measures of the acoustic characteristics of /s/ and measures of the long-term average spectrum. Gay men produced /s/ with higher-frequency spectral peaks and longer durations than heterosexual men; no differences were found in long-term spectra. Judgments of perceived sexual orientation were gathered from a group of listeners. Acoustic measures of /s/ spectra and duration predicted listener judgments. As with Gaudio (1994), these results suggest that acoustic cues to sexual orientation are available in read speech.

One of the most comprehensive studies on sexual orientation and men's speech to date is presented by Smyth, Jacobs, and Rogers (2003). Smyth et al. (2003) report on the development of a database of 25 Toronto-area male talkers who vary in the extent to which naïve listeners judge their voice as gay-sounding. Smyth et al. examined listener ratings for three speech samples: a reading of a scientific text, a reading of a dramatic text, and spontaneous speech. The influence of talkers' self-stated sexual orientation and the type of speech sample on ratings of perceived sexual orientation and perceived masculinity/femininity were examined. Smyth et al. found significant main effects of talker sexual orientation and passage type on ratings of perceived sexual orientation. Self-identified gay men were rated as more gay-sounding than self-identified heterosexual men. There was an interaction between talker sexual orientation and text type: ratings of self-identified gay men were similar across the three types of speech materials. In contrast, self-identified heterosexual men were rated as more-gay sounding when reading the scientific text than in the other two conditions. This finding suggests that one of the parameters that listeners use when judging sexual orientation is formality of speech style, as the heterosexual men were rated as more gay-sounding in the more formal reading condition.

Smyth et al. also examined various measure of vocal pitch, none of which were correlated with ratings of perceived sexual orientation. As in Gaudio (1994), correlations between perceived sexual orientation and perceived masculinity/femininity were highly correlated; however, absolute values differed for the two types of measures. The discrepancy between the two types of measures may have been due to the different phonetic cues that listeners used to judge them; men with low-pitched voices were rarely rated as sounding feminine, but were sometimes judged to sound gay.

Smyth, Rogers, and colleagues have presented a parallel series of descriptive reports of the acoustic characteristics of the speech samples described in Smyth et al. (2003). These studies have documented many acoustic differences between more- and less-gay sounding men, and the relationships between these acoustic measures and listeners' ratings of the talkers. Rogers and Smyth (2003) showed that the more-gay sounding men were more likely to produce vowels closer to the periphery of the vowel space than less-gay sounding men. Smyth and Rogers (2002) report that more-gay sounding men produce stop consonants with longer voice-onset times, longer sibilant fricatives with higher peak frequencies, and more-alveolar variants of /l/ than less-gay sounding men. Rogers and Smyth (2003) found that, while mean f0 and f0 variability did not predict gayness ratings, gayness ratings were strongly correlated with independently made judgments of perceived intonational variability. That is, the voices that one group rated to sound gay were rated by an independent group of listeners to sound as if they had greater f0 modulation. Listeners were more likely to falsely judge a voice to be produced with greater f0 modulation if that voice had been judged by an independent group to sound gay.

Relatively few studies have examined acoustic characteristics of lesbian/bisexual women's speech. Moonwoman-Baird (1997) presented anecdotal evidence suggesting that lesbian women produced conversational speech with more restricted pitch ranges than heterosexual women. More recently, Waksler (2001) examined pitch range in San Francisco bay area women's read speech ($n = 24$), and found no differences between self-identified heterosexual and L/B women. Neither Moonwoman-Baird nor Waksler collected listener judgments.

Pierrehumbert, Bent, Munson, Bradlow, and Bailey (2004) examined short samples of read speech from a large group ($n = 103$) of Chicago-area GLB and heterosexual men and women. Pierrehumbert et al. examined five vowels: /i/, /ɛ/, /æ/, /ɑ/, and /u/. The duration, F1, and F2 of each vowel was measured. Average F1 and F2 were calculated. Vowel-space dispersion was measured, using the technique presented in Bradlow, Toretta, and Pisoni (1996), as the mean Euclidian distance from the center of the vowel space. GLB people produced hyperarticulated vowel spaces relative to their same-sex heterosexual peers. For G/B men, this appeared to be due to an overall hyperarticulation of the vowel space. For women, this effect appeared to be driven by the lesbian/bisexual women producing more-back variants of /u/ and /ɑ/. A separate unpublished study (Bailey, 2003) examined listener judgments of perceived sexual orientation using the speech materials analyzed by Pierrehumbert et al. In that study, it was found that listeners were successful in judging talkers' sexual orientation at greater-than-chance levels from these materials.

In general, the small number of published studies on sexual orientation and speech provide a mixed picture. While some studies have found that listeners are able to detect men's sexual orientation through speech, there is little consensus on the specific acoustic parameters that cue these judgments. There is essentially no published evidence that listeners are able to judge women's sexual orientation through speech. Moreover, only one study, Pierrehumbert et al. (2004), has examined cues for men and women's sexual orientation using the same speech materials and measures.

Given the paucity of data, it is not surprising that few researchers have offered explanations for differences between GLB and heterosexual people's speech. This stands in sharp contrast to research on male-female differences in pronunciation. In that area, a wealth of explanations have been offered for sex differences in production, appealing to social (Eckert & McConnell-Ginet, 1999; Gordon, 1997), cultural (Van Bezooijen, 1995), anatomic (Simpson, 2001, 2002), and perceptual (Diehl, Lindblom, Hoemeke, & Fahey, 1996) factors.

1.2. Purpose

The purpose of this study is to examine the acoustic and perceptual correlates of sexual orientation in women and men's speech. This study contains three experiments. Experiment 1 explores the extent to which GLB and heterosexual men and women differ in their production of vowels, /s/, and /ʃ/ in read speech. Experiment 2 assess whether naïve listeners can perceive sexual orientation when presented with a subset of these words as stimuli. Analyses for Experiment 2 focus on both the influence of talker sex and sexual orientation on mean ratings, as well as the relationship between acoustic measures and ratings of perceived sexual orientation across talkers.

Experiment 3 examines the relationship between measures of perceived sexual orientation and two other perceptual measures: perceived height, and perceived speech clarity. The purpose of Experiment 3 is twofold. First, shows the extent to which judgments of sexual orientation are related to other perceptual judgments made for the same speech stimuli. Second, it determines the extent to which measures of perceived sexual orientation can be predicted by judgments of perceptual characteristics that are more robustly coded in the speech signal.

This investigation differs from previous studies on sexual orientation and speech in two key ways. First, it examines both men and women using a consistent set of speech materials and acoustic analyses. Only one previous study (Pierrehumbert et al., 2004) examined both men and women in a consistent design, but it did not collect listener judgments. This allows us to compare G/B men's and L/B women's speech to both heterosexual men and to women. Thus, this study expands considerably on previous research by examining actual and perceived sexual orientation in men and women in a consistent design.

The second way in which this study differs from previous research is that the primary speech materials used in acoustic and perceptual analyses are words selected for their phonetic composition. That is, this investigation differs from other studies (Pierrehumbert et al., 2004; Smyth et al., 2003; etc.) in that it primarily examined single-word productions, rather than read sentences or connected speech. There are obvious benefits to investigating connected speech. Connected speech is used in daily social communication; consequently, measures from connected speech have an ecological validity that the same measures made from single words do not have. It is plausible to posit that certain speech styles would only be used in natural connected speech and not in the formal speech style that tends to accompany reading tasks using single words. However, numerous factors complicate the interpretation of acoustic measures made from sounds in sentences. It is well established that the prosodic structure of a sentence influences the articulatory and acoustic characteristics of the sounds in it (e.g., Cho & Keating, 2001; Herman, 2000). For example, voiceless stop consonants have longer voice onset times at the left edge of a prosodic phrase, and vowels and consonants are hyperarticulated when they occur in pitch-accented syllables. Research that examines the acoustics of sounds in sentences—as Pierrehumbert et al.

did—may be looking not at differences in articulatory targets for sounds per se, but in differences in prosodic structure. Consider, for example, Pierrehumbert et al.'s finding that G/B men hyperarticulated their vowel space relative to heterosexual men. This finding may reflect more-peripheral vowel targets in G/B men relative to heterosexual men. Conversely, it may reflect differences in prosodic structure: the G/B male speech style might be characterized by greater prosodic variation than that of the heterosexual men, leading indirectly to more-expanded vowel spaces.

The use of sentences in perception tests is similarly problematic. Listeners are sensitive to the prosodic structure of sentences; indeed, prosodic structure is used to infer to the pragmatic function of sentences in an ongoing discourse (Herman, 2000). Perception of sexual orientation in sentences may reflect listeners' attention to prosodic structure rather than to the acoustic characteristics of the sounds they contain. Moreover, the effects of prosodic structure extend over multiple sounds and words within a sentence, and thus may lead acoustic characteristics of sounds in sentences to be correlated. For example, consider the effect of placing contrastive emphasis on multiple words within a sentence. The result of this might be that vowels would be hyperarticulated; voiceless stop consonants would be produced with longer voice-onset times; and /s/ would be produced with greater tongue–palate contact, leading to higher-energy frication noise. Regression analyses predicting listener judgments from any of these acoustic measures might find a predictive relationship, but the predictive relationship would be due primarily to role of prosodic structure in mediating these relationships. Indeed, two studies which have examined listener sensitivity to social-indexical information using materials that varied systematically in their linguistic complexity (Lass, Tecca, Mancuso, & Black, 1979; Van Bezooijen & Gooskens, 1999) found that listeners were increasingly more accurate in identifying social-indexical categories (race in Lass et al. (1979), and regional dialect in Van Bezooijen & Gooskens (1999)) with linguistically more complex stimuli.

The acoustic and perceptual measures taken from single words in this study do not have these confounds. By using a variety of controlled speech samples in both perception and production tasks, we are able to determine the specific segmental characteristics that influence actual and perceived sexual orientation without the confounding influence of higher-level prosodic structure. It is important to point out that this methodology is meant to complement, rather than replace, experiments utilizing sentences and other connected-speech measures. A full characterization of GLB speech styles must describe distinctive segmental and prosodic structures at multiple levels of linguistic complexity and in multiple speech tasks.

This study attempts to integrate its findings with those in past research to form a more cohesive explanation for GLB speech styles than has been offered previously. In general, past studies of GLB speech styles have focused on establishing this phenomenon as something that exists outside of popular culture stereotypes. For example, Gaudio (1994) and Waksler (2001) state that the primary goal of their studies is simply to examine whether the popular-culture notion of G/B men as producing more-dynamic intonation or L/B women as producing less-dynamic intonation has any reality in a controlled experiment. This investigation attempts to go beyond merely documenting the existence (or lack thereof) of GLB speech styles. By integrating the results of production measures (Experiment 1) with perception measures (Experiments 2 and 3), we hope to make a stronger statement about the origin and functions of GLB speech styles than has been offered by previous research.

2. Experiment 1: single-word reading

The purpose of Experiment 1 was to examine the acoustic characteristics of vowels and sibilant fricatives in single words read by GLB and heterosexual men and women.

2.1. Participants

Participants consisted of 44 talkers (11 each L/B women, heterosexual women, G/B men, and heterosexual men). Talkers were recruited from a variety of locations, both from the University of Minnesota, and from the St. Paul/Minneapolis metropolitan area. Fliers were circulated on campus and in the community advertising a study examining people's ability to perceive 'personal characteristics' in speech. No reference to sexual orientation was made in the flier. To insure that a sufficient number of GLB people enrolled in the study, fliers were posted strategically in GLB social and professional groups on campus and in the community. The study's principal focus on sexual orientation and speech was not disclosed when securing informed consent. After data were collected, participants were informed the focus of the study. They were then given the opportunity to withdraw their consent and have their data destroyed. None opted to do so.

All of the talkers were between 18 and 40 years of age. Men and women differed significantly in age ($M_{women} = 29$ years, $SD_{women} = 7.2$ years; $M_{men} = 24.5$, $SD_{men} = 4.5$ years; $t(42) = -3.5$, $p<0.01$). This difference was likely due to differences between the populations from which the two sexes were drawn. Many of the women were from the St. Paul/Minneapolis community, or were students in the University of Minnesota's professional programs in speech-language pathology and audiology. These programs include students outside of traditional college age. In contrast, many of the men were from the general undergraduate student population at the University of Minnesota. However, GLB and heterosexual people did not differ in age, nor did age and sexual orientation interact. Men and women differed significantly in height ($M_{women} = 168.7$ cm, $SD_{women} = 7.4$ cm; $M_{men} = 181.4$ cm, $SD_{men} = 6.6$ cm; $t(42) = 7.0$, $p<0.01$). Height did not differ as a function of sexual orientation, nor did sex and sexual orientation interact. Kolmogorov–Smirnoff tests of normality showed that the distribution of height in men and women did not differ significantly from a normal distribution.

Participants reported no history of speech, language, or hearing disorders. All reported being native speakers of English. All participants were from Minnesota, Wisconsin, Iowa, North Dakota, or South Dakota, corresponding to the North Central dialect described by Labov, Ash, and Boberg (in press). All reported having spent the majority of their post-teenage lives in those areas. This minimized the extent to which differences in regional dialect might have affected results. All of the participants had graduated high school, and were either enrolled in or had completed an undergraduate university degree. All participants passed a pure-tone hearing screening at 0.5, 1, 2, and 4 kHz at 20 dB HL (ANSI, 1989) bilaterally, or reported a normal audiometric evaluation within the past 3 months. All had normal or corrected-to-normal vision. Participants were paid $5.00 for completing the experiment.

2.2. Stimuli

Stimuli for Experiment 1 were 32 CVC words, shown in Table 1. Stimuli were chosen to contain a variety of vowels (/i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ɔɪ/, /oʊ/, and /u/) and to contain words both with and without /s/ and /ʃ/. The sibilant fricatives were chosen for this investigation because their production is closely related to popular-culture stereotypes of G/B men's speech (e.g., Sedaris, 2001), and because previous research had suggested that they might differ between G/B and heterosexual men (Linville, 1998). All of the words had an average familiarity rating (Pisoni, Nusbaum, Luce, & Slowiacek, 1985) of 6.0 or greater, indicating that they would be familiar to most participants. As Table 1 shows, there were more words containing /æ/ than the other vowels. One of the goals of Experiment 1 was to examine voice quality, by measuring the difference in amplitude between the first and second harmonics (H2−H1) of the glottal waveform. The H2−H1 measure has the potential to be affected by F1 resonance, particularly for high vowels (which have low F1 frequencies) spoken by people with high f0. The inclusion of many /æ/ words insured that we would have a large enough sample to measure H2−H1 reliably.

2.3. Procedures

During this task, participants were presented with a list in which the words were typed in 14-point Times New Roman font. Different quasi-randomized lists were created. Each talker read through the list three times. Participants were given no special instructions for this task. They were not told that the stimuli would be used in perception experiments until a post-experiment debriefing (see above). This task was embedded in a larger protocol consisting of four speech-production tasks, only one of which is discussed in this article. Task order was randomized, and equal numbers of participants were assigned to the different task orders.

Table 1
Stimuli for the single-word production task

Orthography	Phonetic transcription	Orthography	Phonetic transcription
Bell	[bɛl]	Sack	[sæk]
Face	[feɪs]	Sad	[sæd]
Fade	[feɪd]	Safe	[seɪf]
<u>**Gas**</u>	[<u>gæs</u>]	**Said**	[sɛd]
Guess	[gɛs]	**Same**	[seɪm]
Half	[<u>hæf</u>]	Seed	[sid]
Hoop	[hup]	Shack	[ʃæk]
Joke	[dʒoʊk]	Shape	[ʃeɪp]
Lake	[leɪk]	Sheep	[ʃip]
Less	[lɛs]	Ship	[ʃɪp]
Loose	[lus]	Sit	[sɪt]
Miss	[mɪs]	<u>**Soap**</u>	[<u>soʊp</u>]
<u>**Note**</u>	[<u>noʊt</u>]	**Soon**	[sun]
Pack	[pæk]	**Tooth**	[tuθ]
<u>**Path**</u>	[<u>pæθ</u>]	Voice	[vɔɪs]
Peace	[pis]	Wish	[wɪʃ]

Stimuli that were used in the perceived sexual orientation task (Experiment 2) and the perceived height task (Experiment 3) are bolded. Stimuli used in the perceived speech clarity task (Experiment 3) are underlined.

The data were recorded on a Marantz CDW300 CD recorder, at a sampling rate of 44.1 kHz, with 16-bit quantization, and an anti-aliasing filter with a cutoff frequency of 22.05 kHz; they were then down-sampled to 22.05 kHz and passed through an 11.025 kHz anti-aliasing filter. Participants wore an AKG-C420 head-mounted condenser microphone, attached to a Rolls phantom power source. Data were recorded in a sound-treated room.

2.4. Analysis

2.4.1. General

Acoustic measures were made using the Praat signal-processing program (Boersma & Weenink, 2003). The onset and offset of each phoneme in each word was marked in Praat by a coder who was blind to the talker's sexual orientation. All acoustic analyses were done automatically in Praat using custom-written scripts; all of these made reference to these labels. Reliability was judged by having the first author re-mark ~3% of words (3 per talker); reliability coding was also done blindly. The range in duration between these measures and the original measures was −11 to 13 ms; the average absolute difference was 7 ms. Prior to conducting acoustic analyses, tokens containing extraneous noise, disfluencies, or reading errors were removed. 3.2% of the data were lost, for a total of 4090 usable tokens.

2.4.2. Filter characteristics: vowel formant frequencies and duration

The duration, F1 and F2 of each vowel were measured. Duration was measured automatically in Praat. F1 and F2 were extracted from LPC formant analyses with eight coefficients; these were taken at vowel midpoint. As in previous research (Pierrehumbert et al., 2004), these were expressed in Bark units (Zwicker & Ternhardt, 1980). Formant values were hand-measured if they appeared to have been mistracked by the LPC algorithm. It was assumed that formant values had been mistracked if they were greater than 2 standard deviations away from the mean values reported by Munson and Solomon (2004), who studied speakers of the same dialect as those in this study.

For each talker, average F1 and F2 were calculated for individual vowels. These values were averaged to compute an overall average F1 and F2. This average was taken over the eight vowel types rather than the 32 word types so that the overall average formant values were not unduly influenced by the preponderance of low front vowels in the stimulus set. In addition, average vowel-space expansion was computed for each subject. As in previous research (Bradlow et al., 1996; Pierrehumbert et al., 2004) this was measured as the average Euclidian distance from the center of the subject's vowel space. This measure is related to clarity in speech production (Bradlow et al., 1996), and was shown by Pierrehumbert et al. (2004) to differ between G/B and heterosexual men.

2.4.3. Source characteristics: fundamental frequency and spectral tilt

Three measures of the voicing source were taken. All of these measures were made based on the voiced portions of vowels only. The first of these was average fundamental frequency at vowel midpoint. As with average F1 and F2, these values were calculated for individual vowels, and the average values were used to calculate each talker's overall f0. These values were expressed in Equivalent Rectangular Bandwidths (ERBs). This measure has been shown previously to correlate more strongly with perceptual judgments of vocal pitch than traditional linear or logarithmic measures (Hermes & van Gestel, 1991). The second voicing-source measure was f0 range. This was calculated by subtracting the lowest f0 in the vowel from the highest f0. The first and last 20 ms of voicing were excluded from this analysis to minimize the effect of consonant perturbations on f0. These values were calculated for individual vowels, and the average f0 range was calculated based on the averages for each vowel.

The final measure was the amplitude of the first harmonic frequency minus the amplitude of the second harmonic frequency (H1−H2), henceforth called *spectral tilt*. This measure indexes voice quality. Larger values indicate a sharper roll-off in the amplitude of the voice source waveform, and are related to breathy voiced qualities (Hanson, 1997; Klatt & Klatt, 1990). This measure was taken for the vowel /æ/ only, for reasons described in Section 2.2. The measurement of H1 and H2 was made in Praat using a method from A.C.L. Remijsen (pers. comm.) as follows: First, Praat was used to determine the f0 at the midpoint of the vowel. Second, it determined a window 3.5 times the period of the f0, centered at the vowel's midpoint. This window of data was extracted, and its spectrum was calculated. The intensity maximum in a frequency range of 60 Hz was then logged, centered on the f0. This was the H1 amplitude. The intensity maximum in a frequency range of 60 Hz was likewise logged, centered on twice the f0; this was the amplitude of H2. The spectral tilt was measured as the H1 amplitude minus the H2 amplitude.

2.4.4. Sibilant fricative spectra

The final set of measures taken in Experiment 1 were of the spectra of /s/ and /ʃ/. Previous research has shown that the acoustic characteristics of these sounds differ between G/B and heterosexual men (Linville, 1998; Smyth & Rogers, 2002). Four measures were taken. The first two measures were the first spectral moments of /s/ and /ʃ/ (Forrest, Weismer, & Milenkovic, & Dougall, 1988). In addition, the third spectral moment, skewness, was calculated for /s/. These were calculated from the entire frequency range, from 0 Hz to 11.025 kHz. The final measure was based on Munson (2004). Briefly, Munson (2004) examined within-speaker variability in fricative spectra in an attempt to understand age-related changes in precision in speech production. Munson described a novel measure of within-speaker variability in fricative production. This measure is calculated by first computing the first spectral moment for each nonoverlapping 10 ms window of frication noise in multiple tokens of fricatives produced by a talker in a single phonetic context. As above, these are calculated from the entire frequency range, from 0 Hz to 11.025 kHz. These values are plotted relative to position in fricative, expressed as a proportion of the total duration of the fricative. The resulting scatter-plots are similar to that shown in Fig. 1.

Polynomial regressions are used to predict the first spectral moments from their position in the fricative. The measure of scatter around the regression line is used as an index of within-speaker variability, weighted to the number of data-points in the regression. Munson (2004) referred to this measure as the *Weighted Sum of Average Residuals*. In the results section, this is referred to as the *precision* of /s/. Larger WSAR values indicate greater trial-to-trial variation (i.e., less precision) than lower ones. These measures were calculated separately by vowel context and averaged across vowel contexts.

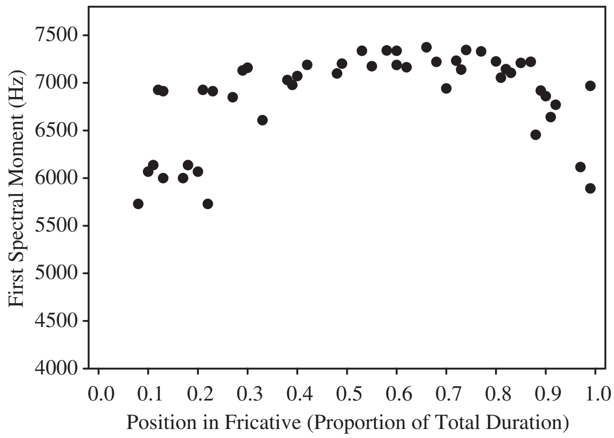


Fig. 1. Scatterplot of the first spectral moment in five productions of the fricative /s/ in the word *sack,* produced by a self-identified gay man, illustrating the fricative precision measure used in this study. See text for details.

2.5. Results

2.5.1. General

The analysis section is divided into three sets of analyses, paralleling the measures described in Sections 2.4.2–4. For each set of dependent measures, a two-factor multivariate analysis of variance (MANOVA) was used. In each MANOVA, talker sex and sexual orientation were the between-subjects factors. Across these MANOVAs, no corrections were made for multiple comparisons. These analyses were exploratory; hence, the use of a less-stringent α-level is justified. For each significant main effect and interaction, a measure of effect size, partial η^2, is reported. Means and standard deviations for all dependent measures are given in Table 2, separated by talker sex and talker sexual orientation. These values are not replicated in the text below.

2.5.2. Filter characteristics: vowel formant frequencies and duration

The results of the MANOVA examining average F1 and F2 showed strong, significant effects of sex on both F1 and F2 ($F[1,40] = 35.7$, $p<0.01$, partial $\eta^2 = 0.47$ for F1; $F[1,40] = 120.1$, $p<0.01$, partial $\eta^2 = 0.75$ for F2). Formants frequencies were higher for women than for men. This finding is somewhat unremarkable, given that the group differences in vocal-tract length that exist between men and women in this study and in the general population (Fitch & Giedd, 1999) have predictable effects on vocal-tract resonant frequencies (Fant, 1966, 1975; Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952). However, there was no significant effect of sexual orientation on either F1 or F2. Sex and sexual orientation did not interact significantly.

Vowel-space expansion differed significantly as a function of sex, $F(1,40) = 10.1$, $p<0.01$, partial $\eta^2 = 0.20$. Pooled across sexual orientations, women produced vowel spaces with greater dispersion than men. This is in agreement with previous research (Bradlow et al., 1996; Pierrehumbert et al., 2004). There was no significant main effect of sexual orientation on vowel-space expansion, nor did the two factors interact. The latter finding is in contrast with the work of Pierrehumbert et al. and Rogers and Smyth (2003), both of whom found vowel-dispersion differences as a function of men's sexual orientation.

Table 2
Summary of acoustic analyses, Experiment 1

Measure	Women				Men			
	L/B		Heterosexual		G/B		Heterosexual	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
F1 (Bark)	5.25	(0.54)	5.41	(0.28)	4.79	(0.27)	4.56	(0.30)
F2 (Bark)	12.33	(0.28)	12.52	(0.28)	11.50	(0.33)	11.47	(0.24)
Expansion (Bark)	2.29	(0.15)	2.46	(0.30)	2.14	(0.28)	2.08	(0.34)
Vowel duration (ms)	208.28	(29.84)	190.42	(23.44)	199.28	(22.81)	190.03	(11.82)
Mean f0 (ERB)	4.87	(0.43)	4.84	(0.26)	3.26	(0.33)	3.11	(0.45)
f0 range (ERB)	0.63	(0.26)	0.77	(0.38)	0.43	(0.19)	0.38	(0.17)
/s/ COG (Hz)	7477.21	(632.19)	7630.90	(521.89)	6455.96	(512.69)	6517.45	(610.01)
/ʃ/ COG (Hz)	4243.86	(246.53)	4345.11	(580.54)	3607.29	(444.94)	3617.99	(752.95)
/s/ skewness	−1.32	(0.92)	−1.65	(0.65)	−0.85	(0.59)	−0.32	(0.62)
/s/ precision	29.79	(13.56)	26.68	(7.64)	35.02	(14.24)	44.64	(14.24)
H1−H2 (dB)	6.64	(3.55)	4.97	(4.13)	1.64	(1.63)	2.61	(3.46)

The F1 by F2 vowel spaces for women and men are shown in Figs. 2 and 3, respectively. As Fig. 1 shows, the largest raw differences between the L/B and heterosexual women were in the F1 of the vowel /ɛ/, and the F2 of the vowel /oʊ/. Indeed, these differences were found to be significant in separate single-factor ANOVAs, $F(1,20) = 5.05$, $p < 0.05$ for the F1 of /ɛ/; $F(1,20) = 4.8$, $p < 0.05$ for the F2 of /oʊ/. For the men, the largest raw differences were in the F1 of /ɛ/ and /æ/. Again, these differences were found to be significant in separate single-factor ANOVAs, $F(1,20) = 8.3$, $p < 0.01$ for the F1 of /æ/; $F(1,20) = 4.7$, $p < 0.05$ for the F1 of /ɛ/.

2.5.3. Source characteristics: fundamental frequency and spectral tilt

The three summary measures of the voice source (mean f0, mean f0 range, and mean amplitude of H2−H1 of /æ/) were submitted to a fully MANOVA examining the influence of sex and sexual orientation. For mean f0, there was a significant main effect of sex, $F[1,40] = 214.1$, $p < 0.01$, partial $\eta^2 = 0.84$. Unremarkably, women had average F0's that were higher than men. There was also a significant main effect of sex on f0 range, $F(1,40) = 13.5$, $p < 0.01$, partial $\eta^2 = 0.25$. Women demonstrated larger f0 ranges than men. Finally, there was a significant main effect of sex on /æ/ H2−H1 values, $F(1,40) = 13.4$, $p < 0.01$, partial $\eta^2 = 0.25$. Women's voices showed a sharper spectral tilt than men's, indicating a breathier voice quality. Sexual orientation did not have a significant effect on any of the dependent measures, nor did sex and sexual orientation interact.

2.5.4. Sibilant fricative spectra

Four summary measures of sibilant fricative spectra (center of gravity of /s/, center of gravity of /ʃ/, skewness of /s/, and precision of /s/) were submitted to a fully MANOVA examining the influence of sex and sexual orientation. There was a significant main effect of sex on all of these measures, $F(1,40) = 38.4$, $p < 0.01$ for /s/ center of gravity, $F(1,40) = 17.7$, $p < 0.01$ for /ʃ/ center

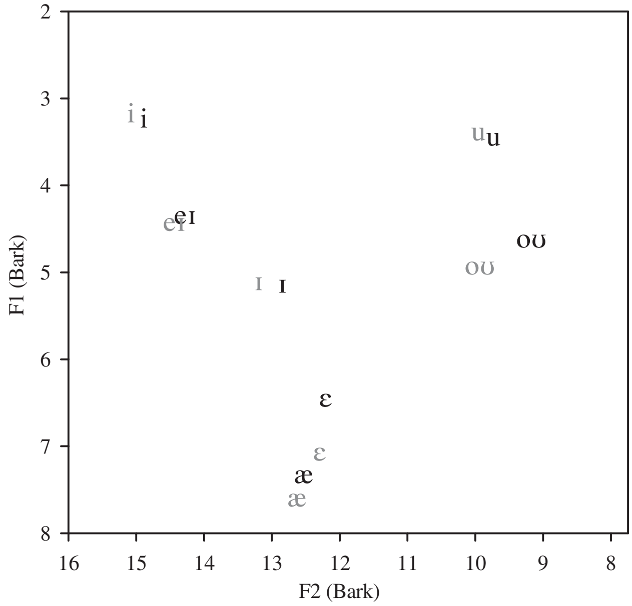


Fig. 2. Average Bark-scaled F1/F2 values for vowels produced by lesbian/bisexual (black) and heterosexual (gray) women. (Symbols are shown in bold font to prevent confusion with Fig. 2.)

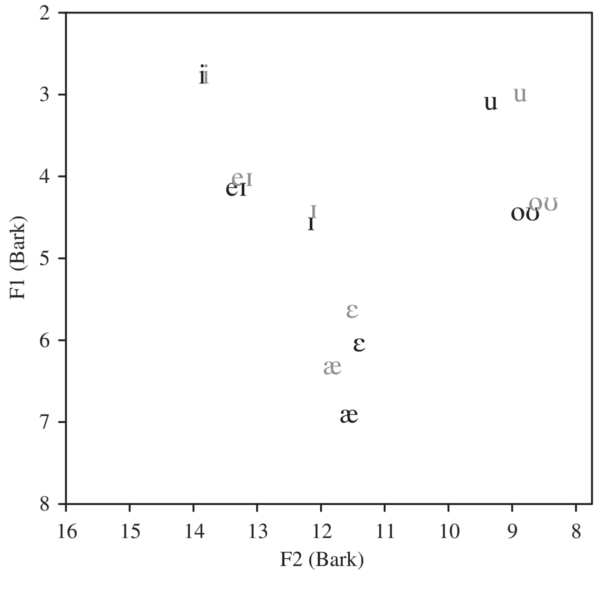


Fig. 3. Average Bark-scaled F1/F2 values for vowels produced by gay/bisexual (black) and heterosexual (gray) men. (Symbols are shown in unbolded font to prevent confusion with Fig. 1.)

of gravity, $F(1,40) = 17.6$, $p < 0.01$ for /s/ skewness, and $F(1,40) = 7.7$ for /s/ precision. These differences were in the expected directions: women produced both fricatives with higher-frequency spectra than men; women produced /s/ with a more negatively skewed spectrum than men, indicating more spectral energy above the spectral mean than in men; and produced /s/ with more precision (i.e., less trial-to-trial variability) than men.

Sexual orientation did not have a significant main effect on any of these variables. Sex and sexual orientation interacted significantly for /s/ skewness, $F(1,40) = 4.1$, $p < 0.05$. This interaction arose because the spectra of G/B men's /s/ was significantly more-negatively skewed than that of heterosexual men ($M_{G/B\ men} = -0.85$, $M_{het.\ men} = -0.32$, $t(20) = -2.1$, $p < 0.05$). In contrast, the L/B and heterosexual women did not differ significantly in this parameter ($M_{L/B\ women} = -1.32$, $M_{het.\ women} = -1.65$, $t(20) < 1$, $p > 0.05$). Sample fricatives produced by a G/B man and a heterosexual men exemplifying these skewness differences are given in Figs. 4 and 5.

2.6. Discussion of Experiment 1

The results of this experiment are consistent with much previous research regarding sex-based differences in speech acoustics. Women produced vowels with higher and more-variable f0's than men, with breathier voice qualities than men, and with higher F1 and F2 frequencies than men. Women produced the fricatives /s/ and /ʃ/ with higher-frequency energy than men. In addition, women produced /s/ with greater precision (i.e., less trial-to-trial variability) and more energy above the median frequency (i.e., with more negatively skewed spectra) than men. Each of these findings is consistent with a past finding from the literature on male–female differences in speech production.

In contrast, the acoustic characteristics of GLB and heterosexual people differed only subtly. Differences between G/B and heterosexual men were limited to the vowels /æ/ and /ɛ/, and the skewness of /s/ spectra. Differences between L/B and heterosexual women were even more subtle, and appeared to be limited to the formants of /ɛ/ and /oʊ/. The findings of Experiment 1 stand at odds with some previous research on sexual orientation and speech. For example, Pierrehumbert et al. (2004) and Rogers and Smyth (2003) both found that hyperarticulated vowel spaces were associated with G/B men's speech. Gaudio (1994) presented evidence suggesting that more gay-sounding men produced speech with greater pitch ranges than less gay-sounding men. Neither of those findings was replicated in this experiment. The results of Experiment 1 are consistent with

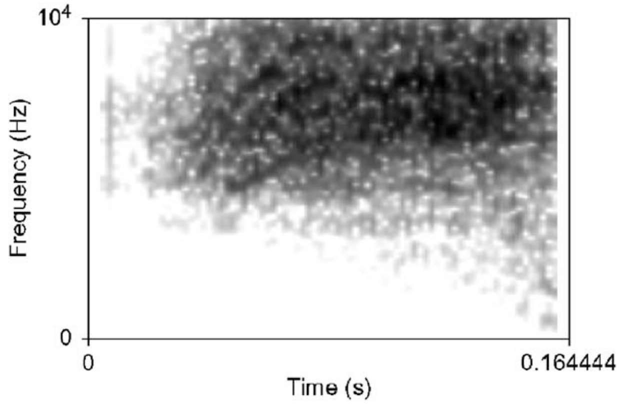


Fig. 4. Spectrogram (made with a 20 ms Hanning window) of the fricative /s/ in the word *sack* produced by a self-identified gay man. Skewness = −1.67 Hz.

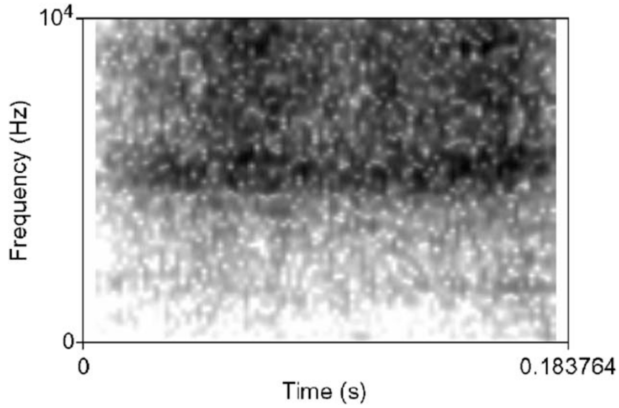


Fig. 5. Spectrogram (made with a 20 ms Hanning window) of the fricative /s/ in the word *sack* produced by a self-identified heterosexual man. Skewness = −0.03 Hz.

Waksler (2001), in that they did not find any differences in pitch range as a function of women's sexual orientation. However, both of these differences must be qualified by the differences in the types of speech stimuli used. Previous studies have measured f0 variation in sentences and connected speech, while the current study used single words. It is reasonable to assume that there would be a much larger f0 range overall in sentences and connected speech than in single-word readings.

The specific set of results did not support the conjecture that GLB speech patterns represent a simple scaling towards the values associated with the opposite sex. Evidence against this conjecture can be found in two sets of measures. First, there were no significant differences between GLB and heterosexual people in acoustic characteristics of the voice source. That is, there was no evidence that G/B men produce a higher f0 or a more modulated f0 than heterosexual men, or that L/B women produced a lower, less modulated f0 than heterosexual women. For the latter finding, it is notable to mention that the hypothesis that women produce larger f0 ranges than men is qualified by the specific units used to measure this (Henton, 1989). The specific set of findings in this study may be related to the use of the psychophysically motivated ERB scale rather than the linear Hertz scale. There were no group differences in breathiness of the voice during /æ/ production, estimated as the amplitude difference between the first and second harmonics. Second, average F1 and F2 measures did not suggest that GLB people were scaling their articulation of the vowel space uniformly towards the values associated with the opposite sex. Instead, the differences in speech production were limited to specific phonemes. The specific differences that were significant, however, do suggest a selective adoption of some speech patterns characteristic of the opposite sex. G/B men produced a higher F1 in /ɛ/ and /æ/, and a more-skewed /s/. All of these patterns are similar to those produced by heterosexual women. L/B women produced a lower F1 in /ɛ/ and a lower F2 in /oʊ/. Again, these patterns are similar to those produced by heterosexual men. The specific vowel differences appear to be learned ways of speaking that convey a talker's sexual orientation. The specific origin of these differences is not clear from Experiment 1 alone. They will be discussed further in Section 5, in conjunction with the results from the perception experiments to be presented in Sections 3 and 4.

3. Experiment 2: perception of sexual orientation in read speech

The purpose of Experiment 2 was to examine whether the differences found in Experiment 1 were available to listeners as cues to the perception of sexual orientation from read speech. Experiment 2 had two goals. The first was to examine whether listeners could reliably rate the 44 talkers differently as a function of their self-stated sexual orientation. The second goal was to use multiple regression to examine relationships between average perceived sexual orientation and the acoustic measures reported in Section 2 above. In this way, it paralleled other investigations (e.g., Gaudio, 1994; Linville, 1998; Smyth et al., 2003) in that it examined the perceptual correlates of both self-identified and listener-perceived sexual orientation.

Experiment 2 differed from previous research in two key ways. First, it examined the perceptual correlates of women and men's sexual orientation in the same group of listeners with similar speech stimuli. Second, unlike many previous studies, it examined the specificity of different perceptual correlates by collecting multiple judgments per talker with speech stimuli that varied in their segmental composition. That is, most previous studies have examined the specific parameters that listeners use in judging sexual orientation by collecting listener ratings and then using multiple regression to examine predictors of these judgments. This approach has provided fruitful data regarding possible correlates of perceived sexual orientation. However, this method has limitations. Multiple regression cannot detect whether two predictor measures contribute independently to judgments of perceived sexual orientation if those predictors are highly correlated. Experiment 2 reduced this confound by collecting multiple judgments per talker for stimuli whose phonemic composition was carefully controlled and whose prosodic structure was nearly identical. This allows us to make more definitive conclusions regarding the extent to which specific acoustic-phonetic parameters cue judgments of sexual orientation than have been offered by previous research.

3.1. Participants

Listeners were 40 individuals from the University of Minnesota community. They were solicited through fliers advertising the study. The fliers did not make explicit reference to sexual orientation. All of the listeners were between 18 and 40 years of age and had no self-reported history of speech, language, or hearing disorders. All were native speakers of North American English. Unlike some previous studies (Carahaly, 2000; Smyth et al., 2003) no attempt was made to stratify listeners by their sexual orientation. Rather, we presumed that the distribution of sexual orientation in our listener sample would follow that in the population.

3.2. Stimuli

Stimuli were 12 words each from the 44 talkers. These are shown as bolded and italicized in Table 1. A subset was used rather than the entire list of 32 words so that the experiment would be reasonable in length. As this list shows, the 12 words included three words with front vowels and sibilant fricatives (*gas*, *said*, *same*); three words with front vowels and no sibilant fricatives (*bell*, *fade*, *path*); three words with back-round vowels and sibilant fricatives (*loose*, *soap*, *soon*), and three words with back-round vowels and no sibilant fricatives (*hoop*, *note*, *tooth*). For each talker, one token of each of these words was chosen at random. These had a 22.05 kHz sampling rate with 16-bit quantization, and had been processed through an 11.025 kHz anti-aliasing filter. The 528 stimuli were normalized for amplitude such that the peak amplitude of each stimulus was equal.

3.3. Procedures

This experiment took place in a double-walled sound-proof booth in the first author's laboratory. The experiment was designed and executed using the *E-prime* experiment-management software (Schneider, Eschman, & Zuccolotto, 2002). On each trial, three words were played. The words varied in the backness of the vowels they contained, and whether or not they contained a sibilant fricative, as described in Section 3.2. The choice to present three words at once was designed to keep the experiment to a reasonable length. An orthographic display of the words was presented on a 17″ computer monitor concurrent with their presentation in 36-point courier font. After each trial, listeners rated the talkers' sexual orientation on an five-point equal-interval scale, where 1 indicated *definitely sounds heterosexual*, 3 indicated *sounds neither heterosexual nor GLB* and 5 indicated *definitely sounds GLB*. The numbers 2 and 4 indicated intermediate values. Cards with the above wording were placed above buttons on a button-box. Participants responded by pressing buttons; their responses were logged automatically. Words were presented over a powered speaker (Roland DS-90A) located 0° azimuth from the listener's head at a level of approximately 65 dB HL, as calibrated prior to the experiment using the slow, dB A-scale of a sound-level meter positioned at the approximate location of the listener's head during the experiment. Experimental blocks were preceded by a practice block containing two talkers not used in the experiment. There were 6 trials during the practice block, using two words that did not appear in the experiment. Experimental stimuli were presented in fully randomized order.

This experiment was embedded in a larger protocol consisting of four perception tasks. This experiment was always the fourth tasks. In the first tasks, listeners made speeded judgments of the talkers' sex. This task was included to ensure that all of the talkers were reliably identified as male or female. In the second task, listeners rated their perception of the talkers' height; the results of this task are presented in Section 4. The third task asked listeners to rate talkers' perceived clarity on a five-point scale. This task yielded no interpretable data, and is not analyzed in this paper. The entire experimental protocol in which this task was embedded took approximately 30 min to complete.

In designing these experiments, the 44 talkers were divided into 4 groups of 11. Each group contained approximately equal numbers of GLB and heterosexual men and women. Within each of the four groups, the different categories (G/B man, L/B woman, etc.) contained talkers whose perceived sexual orientations varied, based on the first author's ratings of perceived sexual orientation. That is, the groups were constructed so that there would not be one group that contained all of the most-GLB sounding gay men and the most-heterosexual sounding heterosexual women. In each of the four perception tasks described above, listeners rated only 11 talkers. Listeners never provided different ratings for the same talkers. Across the four experiments, each of the 40 listeners listened to all 44 talkers, but never made more than one type of rating (sex, height, clarity, and sexual orientation) for a talker. It has been noted in previous research that perceptual judgments about talker characteristics are correlated (e.g., Gaudio, 1994). The decision not to have listeners rate the same talker twice was made so that listeners' judgments about one characteristic of a talker would not affect their judgments about another characteristic. It is well established that listeners can reliably encode, identify, and remember differences among talkers (e.g., Goldinger, 1997). It is possible, then, that perceptual judgments may be correlated simply because listeners remember judgments of one set of characteristics made about a talker (i.e., that the male talker sounds atypically masculine) and give a judgment of a different perceptual parameter for that talker that is consistent with the earlier judgment (i.e., that the same talker is unlikely to be G/B). Indeed, Gaudio (1994) speculated that the high correlations between judgments of masculinity–effeminacy and judgments of perceived sexual orientation might be due to this phenomenon. By having listeners rate only one perceptual correlate per talker, we reduced this potential confound. Of course, it is possible that merely being asked to make a judgment of one perceptual parameter (i.e., perceived height) might affect a subsequent judgment of another perceptual parameter (i.e., perceived sexual orientation), regardless of whether the same talker was being rated for both parameters. Nonetheless, the methodology we chose likely attenuated the influence of earlier perceptual judgments on later ones, given that no talker was ever rated in more than one perception tasks.

3.4. Analysis

Prior to analysis, responses occurring more than 3 s after the presentation of the stimuli were excluded. This was designed to minimize the influence of listener inattention on the perception data. These constituted less than 0.5% of the data. Two sets of summary statistics were calculated. The first was average perceived sexual orientation per talker. This was used in an ANOVA by items (Section 3.5.1) and as the dependent measure for a series of multiple regressions (Section 3.5.2). The second was the average perceived sexual orientation ratings per listener. These were calculated separately for judgments of GLB and heterosexual men and women based on front-vowel and back-vowel words with and without a fricative, for a total of 16 data points per listener. These were used as the dependent measures in an ANOVA by subjects (Section 3.5.1).

3.5. Results

Prior to analyzing data on perceived sexual orientation, accuracy in judging talker sex was analyzed. The male talkers' sex was judged accurately 99% of the time. Values for individual male talkers ranged from 97% to 100%, with a mode of 100%. Women's sex was judged accurately 97% of the time. Values for individual talkers ranged from 90% to 100%, with a mode of 100%. Only one talker was identified at 90% accuracy; the remaining talkers were identified with at least 95% accuracy. The high accuracy rate with which talker sex was judged suggests that results from the other perception experiments presented in this article are not confounded by listeners systematically misapprehending the talkers' sex.

3.5.1. Analyses of variance

A four-factor ANOVA was used to examine the influence of talker sexual orientation, talker sex, vowel backness, and presence/absence of a fricative on listeners' ratings of perceived sexual orientation. Separate ANOVAs were calculated by listeners (a fully within-subjects ANOVA by subjects) and talkers (a mixed-model ANOVA by items). For each significant main effect and interaction, a measure of effect size, partial η^2, is reported. A separate series of ANOVAs was conducted using talker group (discussed in Section 3.3) as a factor. This factor was not significant; thus, the ANOVA by subjects reported here does not include this factor.

There was a significant main effect of sex, sexual orientation, and vowel backness in both the subjects and items ANOVAs (for sex: $F1(1,38) = 19.6$, $p<0.01$, partial $\eta^2 = 0.34$, $F2(1,40) = 4.3$, $p<0.05$, partial $\eta^2 = 0.10$; for sexual orientation: $F1(1,38) = 74.9$, $p<0.01$, partial $\eta^2 = 0.66$, $F2[1,40] = 14.9$, $p<0.01$, partial $\eta^2 = 0.27$; for vowel: $F1(1,38) = 32.9$, $p<0.01$, partial $\eta^2 = 0.46$, $F2(1,40) = 19.4$, $p<0.01$, partial $\eta^2 = 0.33$). Three interactions were significant in both ANOVA types: sex by vowel ($F1(1,38) = 23.3$, $p<0.01$, partial $\eta^2 = 0.38$, $F2(1,40) = 14.6$, $p<0.01$, partial $\eta^2 = 0.22$), sexual orientation by vowel ($F1(1,38) = 16.4$, $p<0.01$, partial $\eta^2 = 0.30$, $F2(1,40) = 11.7$, $p<0.01$, partial $\eta^2 = 0.23$), and sex by vowel by fricative ($F1(1,38) = 6.0$, $p<0.01$, partial $\eta^2 = 0.14$, $F2(1,40) = 4.1$, $p<0.05$, partial $\eta^2 = 0.09$). The sex by fricative interaction was significant only in the ANOVA by subjects ($F1(1,38) = 4.5$, $p<0.05$, partial $\eta^2 = 0.11$, $F2(1,40) = 1.8$, $p>0.05$).

The complex set of interactions was examined through post-hoc tests of significant main effects. Men and women were examined separately. When women were examined, only sexual orientation affected ratings significantly, $F(1,20) = 5.1$, $p<0.05$. L/B women were rated as more-GLB sounding than heterosexual women. For men, significant main effects of sexual orientation and vowel backness were found ($F(1,20) = 10.5$, $p<0.01$ for sexual orientation; $F(1,20) = 30.2$, $p<0.01$ for vowel backness); these interacted significantly ($F(1,20) = 11.3$, $p<0.01$). No other significant interactions were found.

These data are shown in Figs. 6 and 7. Fig. 6 shows data for women; Fig. 7 shows data for men. As these figures show, GLB people were rated as sounding more GLB than heterosexual people, across the eight comparisons. For women, this difference was approximately equal in all four conditions. For men, the difference in ratings between G/B and heterosexual men was exaggerated in the condition in which listeners rated single words containing front vowels. Moreover, the interaction between vowel backness and presence/absence of a fricative appeared to be due to fricatives influencing judgments made over back-vowel words, but not front-vowel words.

3.5.2. Regressions

The next set of analyses examined the relationship between acoustic measures, collected in Experiment 1, and the average perceived sexual orientation for the 44 talkers. These were examined using a series of hierarchical multiple regressions, with average perceived sexual orientation as the dependent measure. The independent measures were the summary acoustic measures shown in Table 2, with the exception of /ʃ/ center of gravity and /s/ precision. These were not entered into the regression because none of the perception stimuli contained /ʃ/, nor were there enough different tokens of /s/ to expect the listeners to be able to gauge perceptually the talkers' trial-to-trial variability in producing that sound.

Table 3 shows correlations (Pearson's r) among the acoustic and perceptual measures, separated by talker sex. As this table shows, there were a number of significant correlations for men and for women. Interestingly, the correlations were generally in the opposite directions for men and for women. The only two exceptions were the correlations between perceived sexual orientation and average vowel duration, and between average perceived sexual orientation and /s/

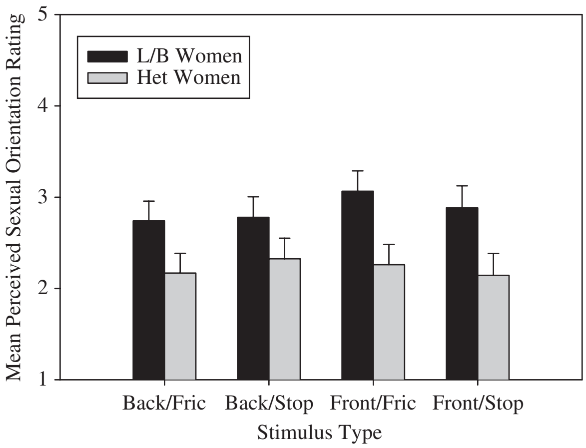


Fig. 6. Mean judgments of women's sexual orientation, separated by vowel place of articulation and presence or absence of a sibilant fricative in the stimulus set. Error bars represent one standard error of measurement.

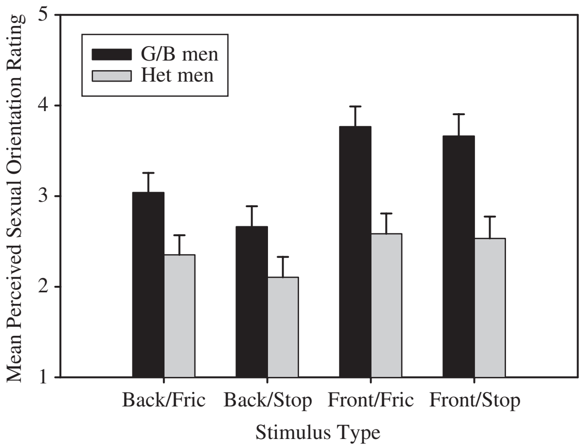


Fig. 7. Mean judgments of men's sexual orientation, separated by vowel place of articulation and presence or absence of a sibilant fricative in the stimulus set. Error bars represent one standard error of measurement.

center of gravity. In both of these cases, the correlation coefficient for one sex was very close to zero. That is, the correlations broadly support the conjecture that the perceived sexual orientation represents a listener's tacit measurement of sex-typicality: a correlation in one direction for women implies a correlation in the other direction for men.

Four hierarchical multiple regressions were computed. These examined judgments made over front-vowel and back-vowel words (averaged across those that contained /s/ and those that did not) separately for men and for women. The independent measures were entered into the regression stepwise if they accounted for a significant proportion of variance ($\alpha < 0.05$) beyond what was accounted for on the previous step. These regressions are summarized in Tables 4 and 5.

As these tables show, all of the regressions were significant, indicating that a significant proportion of variance in perceived sexual orientation could be accounted for by the acoustic characteristics of the stimuli.

For women, the strongest predictor of judgments of perceived sexual orientation made from front-vowel words was F1 frequency; this accounted for 40% of the variance. The negative β-weight associated with this factor indicates that women with higher F1 values were more likely to be judged as sounding heterosexual sounding than women with lower F1 values. An additional 14% of the variance was accounted for by F2 frequency. Again, the negative β-weight indicates that women with higher F2 frequencies were more likely to be judged as heterosexual sounding than women with lower F2 frequencies. Finally, 10% of the variance was accounted for by overall vowel-space dispersion. The negative β-weight suggests that women with more-dispersed vowel spaces were more likely to be rated as heterosexual sounding than women with less-dispersed vowel spaces. This contribution was statistically independent from the contribution of the group differences in F1 frequency. In the regressions predicting judgments from back-vowel words, F2 frequency accounted for the greatest proportion of variance. The negative β-weight indicates that women with low F2 frequencies were more likely to be judged to sound GLB than women with high F2 frequencies. Again, vowel-space dispersion predicted a significant proportion of variance, such that women with overall more-contracted vowel spaces were more likely to be judged to sound GLB than women with overall more-expanded vowel spaces.

For judgments of men's sexual orientation made from front-vowel words, the largest proportion of variance (25%) was accounted for by F1 frequency. The positive β-weight associated with this predictor indicates that men with higher F1 values were more likely to be perceived as GLB-sounding than men with lower F1 values. This was the opposite pattern than was found with women's voices. In addition, skewness of /s/ predicted 21% of the variance in perceived sexual orientation judgments. Men with more negatively skewed /s/ were more likely to be rated as sounding GLB than men with less negatively skewed /s/. For judgments made over back-vowel words, the primary predictor of perceived sexual orientation was F2 frequency, accounting for 26% of the variance. Men with higher average F2 frequencies were more likely to

Table 3
Correlations between perceived sexual orientation from Experiment 2 and selected acoustic measures from Experiment 1

	Perceived sexual orientation	
	Women	Men
Average vowel duration	0.03	0.42*
Average F1	−0.65**	0.49**
Average F2	−0.65**	0.45*
Vowel-space expansion	−0.57**	0.24
Average f0 at vowel midpoint	−0.39	0.34
Average f0 range	−0.35	0.12
Average H1−H2	−0.14	0.02
Average /s/ center of gravity	−0.41*	0.08
Average /s/ skewness	0.43*	−0.38*

*$p<0.05$, ** $p<0.01$.

Table 4
Stepwise multiple regression predicting perceived sexual orientation from selected acoustic measures, women only

	Step	Variable	ΔR^{2a}	B^b	SE B^b	β^b
Front-vowel words	1	Average F1	0.40**	−0.70	0.30	−0.38*
	2	Average F2	0.14*	−0.98	0.42	−0.36*
	3	Vowel-space expansion	0.10*	−1.03	0.48	−0.33*
Back-vowel words	1	Average F2	0.42**	−1.35	0.37	−0.56**
	2	Vowel-space expansion	0.17**	−1.18	0.43	−0.42**

Regressions calculated separately for ratings made from front-vowel words and back-vowel words. $F(3, 18) = 10.2$, $p<0.01$, $R^2 = 0.64$ for the full model based on the ratings of front-vowel words; $F(2, 19) = 13.5$, $p<0.01$, $R^2 = 0.59$ for the full model based on the ratings of back-vowel words.
* $p<0.05$, ** $p<0.01$.
[a]Increase in R^2 over the model containing all previous steps.
[b]Coefficients for the full model.

Table 5
Stepwise multiple regression predicting perceived sexual orientation from selected acoustic measures, men only

	Step	Variable	ΔR^{2a}	B^b	SE B^b	β^b
Front-vowel words	1	Average F1	0.25**	1.74	0.51	0.59**
	2	Average /s/ skewness	0.21**	−0.63	0.24	−0.46*
Back-vowel words	1	Average F2	0.26**	1.40	0.37	0.57**
	2	Average f0	0.16*	0.75	0.26	0.43**
	3	Average /s/ skewness	0.17**	−0.44	0.16	−0.41*

Regressions calculated separately for ratings made from front-vowel words and back-vowel words. $F(2, 19) = 8.24$, $p<0.01$, $R^2 = 0.46$ for the full model based on the ratings of front-vowel words; $F(3, 18) = 8.8$, $p<0.01$, $R^2 = 0.60$ for the full model based on the ratings of back-vowel words.
* $p<0.05$, ** $p<0.01$.
[a]Increase in R^2 over the model containing all previous steps.
[b]Coefficients for the full model.

be rated as GLB-sounding than men with lower F2 frequencies. The skewness of /s/ also accounted for 17% of the variance in of perceived sexual orientation. As with the regression for front-vowel words, men who produced a more-negatively skewed /s/ were more likely to be rated as sounding GLB than men with less-negatively skewed /s/. Average f0 also accounted for 16% of the variance in back-vowel words: men with higher f0 were more likely to be judged as GLB sounding than men with low f0.

3.6. Discussion of Experiment 2

The results of Experiment 2 suggest that listeners are able to discriminate between GLB and heterosexual women and men's readings of single words. As a group, G/B men and L/B women were rated as sounding more GLB than heterosexual people. It is noteworthy to mention that the groups were not completely separated in their perceptual ratings. For example, the two men who were rated as most GLB sounding (both of whom received a mean rating of 4.3, averaged across the four conditions) included one self-identified heterosexual man and one self-identified G/B man. The ratings for the self-identified L/B women ranged from 1.83 to 3.67. The most-GLB sounding heterosexual woman received an average rating of 3.4, which was close to the upper end of the range for L/B women.

For both men and women, the strongest predictors of perceived sexual orientation were found to be measures of vowel formant frequencies. For low-front vowels, F1 frequency was the most salient predictor. A high F1 frequency was associated with more-heterosexual sounding women's voices, and more-GLB sounding men's voices. For back vowels, F2 frequency was the most salient predictor. A high F2 frequency cued listeners to judge women's voices as heterosexual sounding, and men's voices as GLB sounding. In addition, /s/ spectra were associated with judgments of sexual orientation in men, and overall vowel-space dispersion was associated with judgments of women's sexual orientation. In sum, the cues that listeners used when judging sexual orientation from speech were phoneme-specific. This was particularly true for men. Listeners showed greater sensitivity to men's sexual orientation when listening to words that containing front vowels than when listening to words containing back vowels. Sensitivity to women's sexual orientation was approximately equal across conditions.

Some of the results of Experiment 2 are consistent with previous research on sexual orientation and speech. For example, the finding that /s/ skewness predicted a significant proportion of variance in judgments of men's sexual orientation is broadly consistent with Linville's (1998) investigation and with that of Rogers and Smyth (2003). Other findings, however, contradicted previous work. For example, regression analyses failed to find an influence of vowel-space expansion on judgments of men's sexual orientation. Previously, Rogers and Smyth (2001) had posited an association between these two variables, using slightly different measures. Our analyses also failed to support Moonwoman-Baird's (1997) hypothesis that f0 range was associated with women's sexual orientation. Although this parameter was correlated with ratings of women's sexual orientation, it did not predict a significant, independent portion of variance in this measure beyond what was accounted for by other acoustic measures.

It is possible that the differences between this investigation and earlier ones are due to the linguistic complexity of the stimuli being used. Other investigations have elicited judgments using connected speech. As mentioned earlier, the use of connected speech in perception tasks allows listeners to use broader prosodic features in addition to or instead of low-level acoustic–phonetic cues. In this experiment, listeners were forced to attend to segment-level detail.

The acoustic differences in Experiment 1 did not support the conjecture that GLB people's speech was the result of global scaling towards opposite-sex values. The results of Experiment 2 complement this finding, in that they show that listeners were not listening for global indices of masculinity and femininity when making judgments of perceived sexual orientation. This is best illustrated by the fact that acoustic measures of the voice source (mean f0, f0 range, and spectral tilt of the voicing source for /æ/) generally did not predict a significant proportion of variance in perceived sexual orientation in the regression analyses. In fact, these measures were not correlated with average sexual orientation, although the correlation for between mean f0 and perceived sexual orientation in women approached significance ($p = 0.10$).

Perhaps the most interesting finding of Experiment 2 concerns the asymmetry in predictors of men and women's perceived sexual orientation. These asymmetries were reflected in two findings.

First, when analyses were completed with respect to the talkers' self-stated sexual orientation, stimulus composition affected judgments of men's sexual orientation, but not judgments of women. This finding might indicate a greater familiarity with G/B men's speech styles. That is, the listeners' greater familiarity with G/B men's speech may have lead them to attend to certain speech features (i.e., the height of /æ/ and /ɛ/, the skewness of /s/) over others.

Second, the variance in women's perceived sexual orientation that was accounted for by acoustic measures was higher than that for men. One possible explanation of this asymmetry is that women might have demonstrated a wider range of scores than men. That is, there simply might have been more variance available to be accounted for in women's voices than in men's. Examination of the data would suggest otherwise. Indeed, the opposite was true: there was a greater range of perceived sexual orientation scores within the group of men than in women. This was particularly true for judgments made from words containing front vowels.

4. Experiment 3: perceived height and perceived clarity

The results of Experiments 1 and 2 demonstrate that there are subtle differences between GLB and heterosexual people's speech, and that these subtle differences are perceptible to listeners. However, it is not clear *how* listeners are able to accurately rate sexual orientation from speech, in this investigation and in other past investigations. One potential explanation is that the speech of G/B men is globally more feminine than that of heterosexual men, and the speech of L/B women is more masculine than that of heterosexual women. If that were true, people might judge the sexual orientation of a talker by gauging the extent to which his or her voice and speech resembles that of the opposite sex. Indeed, previous studies show that cues to talker gender are present even in short speech samples, and even when these samples have been modified to filter out most of the speech signal (Lass, Almerino, Jordan, & Walsh, 1980). One other possibility is that judgments of sexual orientation are not judgments of sexual orientation per se. Rather, it is possible that these judgments reflect judgments of some other parameter that is robustly coded in short segments of speech, and which listeners associate either tacitly or overtly with sexual orientation.

Experiment 3 explores the extent to which the judgments of perceived sexual orientation from Experiment 2 are correlated with other judgments made from the same stimuli. The first parameter that we examine is speech clarity. Speech clarity is typically defined by measures of intelligibility: speech is considered to be clear if it is easily understood by naïve listeners. Within and between individuals, speech differs in how clearly it is produced. Individual talkers increase their speech clarity in response to a variety of talker- and listener-specific characteristics, such as the level of ambient noise and the perceived needs of the listener (Bradlow, 2002; Picheny, Durlach, & Braida, 1985). Talkers also differ from one another in how clearly they speak: Ferguson (2004) showed that speakers differ systematically in the clarity of their speech, and in the extent to which they can modify their speech style to enhance clarity. Hazan and Markham (2004) show that clarity differences are salient to both children and adults. Clarity differences result from acoustic modifications that talkers make to individual sounds and words, and can be seen both in single-word productions and in connected speech.

The choice of clarity as a possible perceptual parameter related to perceived sexual orientation was motivated by two previous findings. The first was Smyth et al.'s (2003) finding that men's voices were rated as being more-gay sounding in a more formal speaking task (reading) than in a less-formal task (spontaneous speech). Reading tasks elicit clearer, more hyperarticulate signals than spontaneous speech; hence, this finding may reflect an association between clarity of speech production and ratings of sexual orientation. Second, many of the acoustic characteristics that have been found to differ between GLB and heterosexual people arguably reflect differences in speech clarity. The vowel-production differences found by Rogers and Smyth (2003) and Pierrehumbert et al. (2004) show that G/B men produced more-expanded vowel spaces than heterosexual men; more-expanded vowel spaces are associated with clearer speech (Bradlow et al., 1996). Moreover, analyses presented in Section 3.5.2 show a relationship between vowel-space dispersion and ratings of women's sexual orientation. Listeners in Experiment 2 might have been estimating talker clarity from the speech signal, and applying beliefs about how GLB and heterosexual people might differ in this parameter. If so, we would expect strong correlations between independent judgments of perceived sexual orientation and perceived speech clarity.

The second parameter that we examined is perceived height. The acoustic characteristics of speech sounds are related, albeit weakly, to the size of the vocal tract. As documented by Fitch and Giedd (1999), women's vocal tracts are smaller than men's. This has a predictable effect on vocal-tract resonances: Peterson and Barney (1952), Fant (1966, 1975), and Hillenbrand et al. (1995) all show that adult women produce vowels with higher formants than adult men; children produce higher formants than adults of either sex. The influence of vocal-tract size on speech acoustics can be observed both in connected speech and in single words (González, 2004).

Despite the semi-regular relationship between height and speech acoustics, listeners do not appear able to gauge height from speech signals reliably (van Dommelen, 1993; van Dommelen & Moxness, 1995). The relationship between measures of body size and perceived stature is complicated by the fact that people can volitionally manipulate the size of the vocal tract to give the illusion of a larger or smaller vocal tract. Lowering the larynx and protruding the lips lengthens the vocal tract, and would serve to lower all formants. Indeed, the specific acoustic modifications that were made to vowels in this study may have been chosen to convey differences in vocal-tract size and, perhaps, differences in overall stature or height. Consider the influence of /oʊ/ fronting and /æ/ lowering on the perception of vocal-tract size. Fronted /oʊ/ is characterized by a high F2, and lowered /æ/ is characterized by a high F1. Together, these give the impression of a smaller vocal tract. Perhaps the predictive relationship between those parameters and perceived sexual orientation seen in Section 3.5.2 is due to their cueing the perception of height (taller for GLB-sounding women; shorter for GLB-sounding men). That is, listeners might be estimating talker height from the speech signal, and applying beliefs about how the stature of GLB and heterosexual people might differ. If so, we would expect strong correlations between independent judgments of perceived sexual orientation and perceived height.

A crucial fact in this line of argumentation is that both clarity and height can be estimated from short speech samples. González (2004) showed that relationships between height and speech acoustics were relatively weak, but were equally weak for short samples as they were for connected speech. Clarity transformations affect most, if not all, of the sounds in a word. They can be seen even in short speech samples (Ferguson, 2004).

Thus, the purpose of Experiment 3 is to examine relationships between measures of perceived sexual orientation, taken from Experiment 2, and measures of perceived speech clarity and perceived height. A finding that measures of sexual orientation are strongly predicted by the other two measures might suggest that the listeners in Experiment 2 were successful in judging sexual orientation from speech because they were judging a different parameter, either height or clarity, from the speech stimuli.

4.1. Participants

Participants in the perceived height experiment were the same as those described in Section 3.1. Participants in the perceived clarity experiment ($n = 10$) were native speakers of English between 18 and 40 years of age with no reported history of speech, language, or hearing disorders. They were recruited from fliers on the University of Minnesota campus. Seven women and three men participated. None was aware of the purposes of the experiment. Listeners in the perceived clarity experiment were paid $10.00 for their participation.

4.2. Stimuli

Stimuli for the perceived height experiment were 12 words each from the 44 talkers. These were the same stimuli used in Experiment 2, as described in Section 3.2 and shown in Table 2. As this list shows, these stimuli included words both with back vowels and with front vowels, and both with and without a sibilant fricative. A subset of these words was used in the perceived clarity experiment. The four words chosen for this experiment (*gas*, *path*, *soap*, *note*) included one word each from the four categories used in Experiment 2. The choice of only four stimuli for the perceived clarity experiment was motivated by time constraints. As described below, the method used to elicit clarity judgments was very time-consuming, and a larger set of stimuli would have made the experiment unacceptably long.

4.3. Procedures

4.3.1. Perceived height

The procedures for collecting perceived height judgments were similar to those used to measure perceived sexual orientation in Experiment 2. On each trial, three words were played. The words varied in the backness of the vowels they contained, and whether or not they contained a sibilant fricative, as described in Section 4.2. After each trial, listeners rated the talkers' perceived height on an 5-point equal-interval scale, where 1 indicated *taller than average*, 3 indicated *of average height* and 5 indicated *shorter than average*. The numbers 2 and 4 indicated intermediate values. Cards with the above wording were placed above buttons on a button-box. All other details about this experiment are identical to the perception experiment described in Section 3.3.

4.3.2. Perceived speech clarity

Previous studies examining speech clarity have typically used sentence intelligibility measures (Bradlow et al., 1996). Typical stimuli in sentence-intelligibility experiments are syntactically well-formed but semantically unpredictable sentences, often played in the presence of background noise. These are played to listeners, who repeat what the talker said. Percent words correctly repeated scores are used to rank talkers by intelligibility. These methods require talkers to produce a large number of sentences (such as the IEEE sentences, IEEE, 1969) to avoid the confounds that would occur if listeners rated the same sentence twice. Our perceived speech clarity experiment was conceived after the data for Experiment 1 had been collected. We did not collect sentence productions, nor did we collect enough different word types to be able to use a methodology analogous to ones used in previous intelligibility studies.

Given these constraints, we measured perceived speech clarity using paired comparisons. On each trial, eight words were played. First, four words from one talker were presented through a speaker, concurrent with an orthographic display of the words presented on a 17″ computer monitor in 36-point courier font. After a 0.5 s pause, the same four words produced by a different talker were presented concurrent with their orthographic display. Listeners were instructed to press one button if they thought that the first talker had produced the words more clearly, and a different button if they thought the second talker spoke more clearly. These responses were logged automatically. As with the other perception experiments, words were presented over a powered speaker (Roland DS-90A) located 0° azimuth from the listener's head at a level of approximately 65 dB HL, as calibrated prior to the experiment using the slow, dB A-scale of a sound-level meter positioned at the approximate location of the listener's head during the experiment.

Unlike the other perception experiments, stimuli were presented in the presence of speech-shaped background noise at +10 dB signal-to-noise ratio (SNR). The long-term spectrum of the entire set of 176 stimuli was calculated using the Praat signal-processing program (Boersma & Weenink, 2003). A digital file of noise matching this spectrum was then created and mixed with each stimulus at the appropriate SNR prior to presentation. The decision to present stimuli in noise was motivated by the assumption that clarity differences among talkers would not be evident in a quiet presentation, in which all talkers would likely be rated as very clear. The experiment was blocked by sex; listeners only judged pairs of speakers of the same sex. Equal numbers of listeners participated in the two experiment orders. Each block contained 462 responses. The total listening time for this experiment was approximately 90 min.

4.4. Results

4.4.1. Analyses of variance

The ANOVAs in this section report on analyses by items. That is, this section reports on differences among the 44 talkers, rather than among the 50 listeners. As described below, the only summary statistic available for the perceived clarity experiment was for the 44 talkers, rather than the 10 listeners. Moreover, the purpose of these experiments was to examine the extent to which judgments of talkers' perceived sexual orientation were correlated with other judgments made from the same speech samples. Hence, we examined statistics by talkers because these data were comparable to the measures of the 44 talkers' perceived sexual orientation.

4.4.1.1. Perceived height. Average perceived height was calculated separately for each of the 44 talkers. Prior to calculating averages per talker, responses occurring more than 3 s after the stimuli were presented were excluded. These represented less than 0.5% of the total responses. Data were submitted to a four-factor mixed-model ANOVA, with vowel backness and presence/absence of fricative as the within-subjects factors, and talker sex and sexual orientation as the between-subjects factors. The only significant main effect was for sex, $F[1,40] = 6.95$, $p<0.05$, partial $\eta^2 = 0.15$. Averaged across stimulus type and sexual orientation, women were rated to sound shorter than men ($M_{women} = 3.02$, $SD_{women} = 0.65$, $M_{men} = 2.48$, $SD_{men} = 0.75$). There was no main effect of sexual orientation, nor did sex interact with sexual orientation. Stimulus composition did not affect judgments, nor did it interact with any of the other factors.

4.4.1.2. Perceived speech clarity. The paired comparison data were summed across the 10 listeners, to derive a matrix showing for each pair the number of times the first talker was selected as the clearer talker. Perceived clarity was analyzed by fitting a Bradley–Terry model (Linacre, 1995) to the summed paired-comparison matrix. The result of this model was a score for each of the talkers indicating the likelihood that the talker was chosen in the paired comparisons. Lower values indicated talkers who were more likely to be chosen as the clearer talker. Separate Bradley Models were calculated for men and for women. Bradley–Terry scores for men and women were normalized separately prior to analysis.

Two separate ANOVAs examined the influence of sexual orientation on normalized Bradley–Terry scores of perceived speech clarity. A significant effect of sexual orientation was found for men ($F(1,20) = 9.03$, $p<0.01$) and women ($F(1,20) = 5.7$, $p<0.05$); however, the difference was in the opposite direction for the two sexes. G/B men were rated as producing speech more clearly than heterosexual men ($M_{het.\ men} = 0.54$, $SD_{het.\ men} = 1.1$, $M_{G/B\ men} = -0.54$, $SD_{G/B\ men} = 0.49$). L/B women were rated as producing speech less clearly than heterosexual women ($M_{het.\ women} = -0.46$, $SD_{het.\ women} = 0.94$, $M_{L/B\ women} = 0.46$, $SD_{L/B\ women} = 0.86$).

4.4.2. Regression

The next set of analyses used multiple regression to examine relationships among the measures of perceived height and perceived clarity gathered in Experiment 3, and the measures of perceived sexual orientation gathered in Experiment 2. Perceived height and perceived speech clarity were significantly correlated in women ($r = -0.57$, $p<0.01$) but not in men ($r = -0.30$, $p>0.05$). These findings on perceived height parallel the findings of González (2004) on actual height. González found stronger correlations between actual height and formant frequencies in women than in men.

The results of the regression analyses are shown in Tables 6 and 7. These regressions were calculated separately for front-vowel words and back-vowel words. Average perceived sexual orientation was the dependent measure; average perceived height (as described in Section 4.4.1.1) and normalized Bradley–Terry scores of perceived clarity (as described in Section 4.4.1.2) were the independent measures. Women and men were examined separately. The results for the regressions on women are shown in Table 6. The R^2 in these regressions were 0.81 and 0.73, indicating that 81% and 73% of the variance in women's perceived sexual orientation was predicted by perceived height and perceived clarity; this regression was significant. In both regressions, the standardized β-weights associated with perceived height were negative, indicating that women who were rated as sounding shorter than average were also likely to be rated as sounding more heterosexual. The standardized β-weights associated with perceived clarity were positive, indicating that women who were judged to produce speech less clearly were also likely to be judged as more-GLB sounding. Two post hoc stepwise multiple regressions showed that a considerably greater proportion of variance in perceived sexual orientation was accounted for by perceived height than by perceived clarity in both of the regressions. These relationships are shown in Figs. 8a and 9a.

The R^2 for the regressions on male talkers was 0.48 and 0.55, indicating that 48% and 53% of the variance in perceived sexual orientation was predicted by perceived height and perceived clarity. Both regressions were significant. The standardized β-weights associated with perceived clarity were negative. This indicates that men who were judged to produce speech more clearly were also likely to be judged as more-GLB sounding. The standardized β-weights associated with perceived height were not significant. These relationships are shown in Figs. 8b and 9b.

A separate series of analyses were conducted in which actual height was entered into the regressions. This variable did not predict a significant proportion of variance in perceived sexual orientation for men or women beyond what was accounted for by perceived height and perceived speech clarity in a fully stepwise regression. When actual height was forced as the first variable in a hierarchical regression, it did not predict a significant proportion of variance in men's sexual orientation. It did predict a small but significant proportion of variance in women's sexual orientation; however, perceived height and perceived clarity both predicted a significant proportion of variance beyond what was accounted for by actual height, and the β-weight associated with actual height was never significant in the full regression model.

The results of these regressions suggest that a significant proportion of variance in perceived sexual orientation can be predicted by measures of perceived height and perceived speech clarity.

Table 6
Simultaneous multiple regression predicting perceived sexual orientation from perceived sexual orientation and perceived speech clarity measures, women only

	Variable	*B*	SE *B*	β
Front-vowel words	Perceived height, front-vowel words	−0.77	0.11	−0.71**
	Perceived speech clarity	0.30	0.08	0.38**
Back-vowel words	Perceived height, back-vowel words	−0.72	0.14	−0.65**
	Perceived speech clarity	0.28	0.09	0.39**

Regressions calculated separately for ratings made from front-vowel words and back-vowel words. $F(2,19) = 39.4$, $p<0.01$, $R^2 = 0.81$ for the full model based on the ratings of front-vowel words; $F(2,19) = 20.1$, $p<0.01$, $R^2 = 0.73$ for the full model based on the ratings of back-vowel words.
* $p<0.05$, ** $p<0.01$.

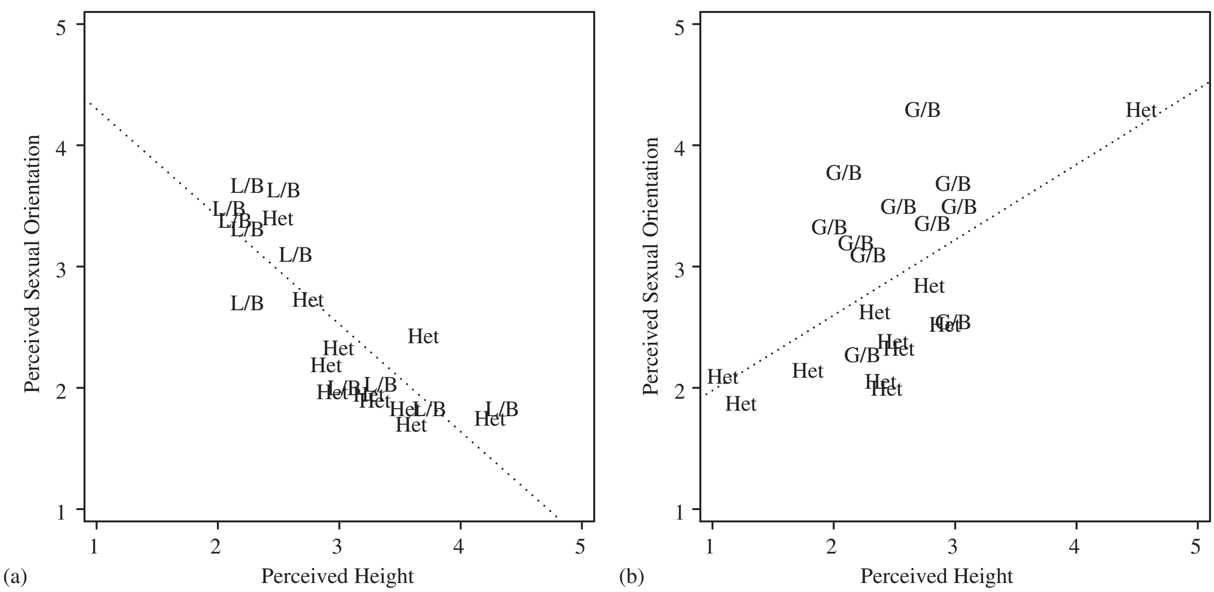
Table 7
Simultaneous multiple regression predicting perceived sexual orientation from perceived sexual orientation and perceived speech clarity measures, men only

	Variable	*B*	SE *B*	β
Front-vowel words	Perceived height, front-vowel words	0.15	0.24	0.12
	Perceived speech clarity	−0.59	0.17	−0.66**
Back-vowel words	Perceived height, back-vowel words	0.22	0.13	0.31
	Perceived speech clarity	−0.47	0.19	−0.48*

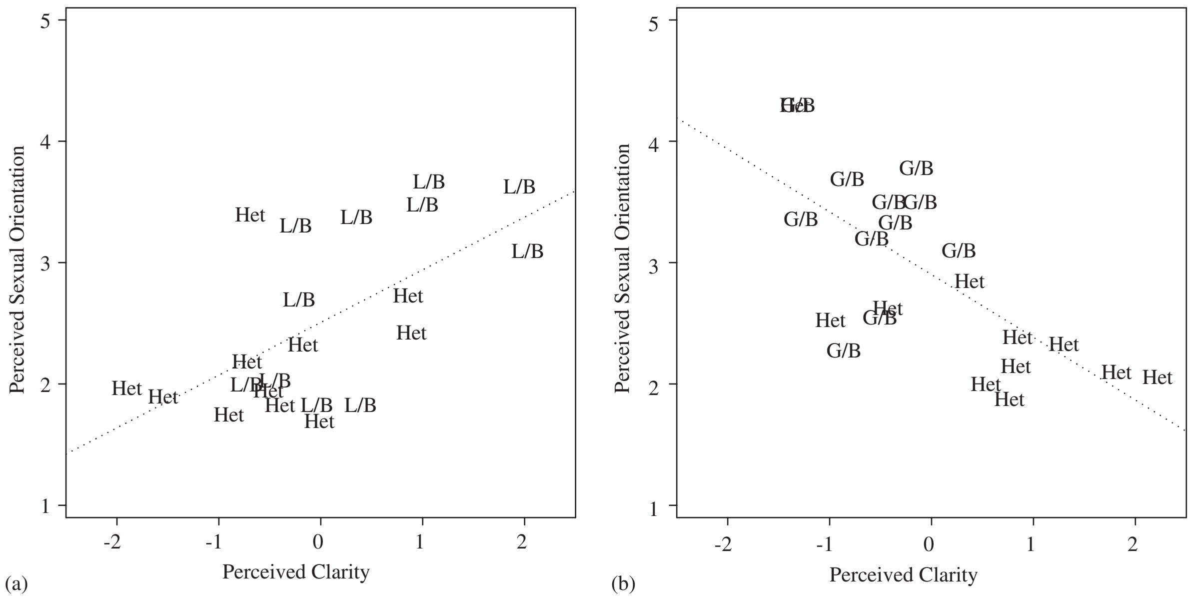
Regressions calculated separately for ratings made from front-vowel words and back-vowel words. $F(2,19) = 8.6$, $p<0.01$, $R^2 = 0.48$ for the full model based on the ratings of front-vowel words; $F(2,19) = 11.7$, $p<0.01$, $R^2 = 0.55$ for the full model based on the ratings of back-vowel words.
* $p<0.05$, ** $p<0.01$.

The last set of regression analyses examined whether acoustic measures continued to predict a significant proportion of variance in perceived sexual orientation when perceived height and clarity were controlled. As with the other regression analyses reported in this section, these were done separately for ratings of women and men's sexual orientation based on front-vowel and back-vowel words, for a total of four separate regressions. In the first step of the regression, average perceived height and perceived speech clarity measures were entered into the regression. On the second step, the same acoustic measures used in the regressions described in Section 3.5.2 (mean F1, mean F2, vowel-space dispersion, mean vowel duration, mean f0, f0 range, /s/ center of gravity, and /s/ skewness) were entered stepwise if they accounted for an additional proportion of variance in the dependent measure ($\alpha < 0.05$) beyond what was accounted for by the previously entered variables.

When ratings of women's sexual orientation were examined, the results were identical to those previously described in this section; the acoustic measures did not predict any variance beyond what was accounted for by perceived height and perceived speech clarity. This contrasted with the results of the regressions predicting men's sexual orientation. In the regression predicting ratings made from back-vowel words, two acoustic measures predicted a significant proportion of variance beyond what was accounted for by perceived height and perceived speech clarity. These were average F2, which accounted for an additional 9.0% of variance, and /s/ skewness, which predicted an additional 9.5% of variance. The standardized β-value for F2 was 0.367; the value for /s/ skewness was -0.387; these were both significant at the $\alpha < 0.05$ level. These values parallel those found in the regressions presented in Table 5, and show that men with higher F2



Figs. 8. (a,b) Scatterplots showing the relationship between perceived height (1 = *taller than average*, 3 = *of average height*, 5 = *shorter than average*) and perceived sexual orientation (1 = *definitely sounds heterosexual*, 3 = *sounds neither GLB nor heterosexual*, 5 = *definitely sounds GLB*) for lesbian/bisexual (*L/B*) and heterosexual (*Het*) women (Fig. 8a) and gay/bisexual (*G/B*) and heterosexual (*Het*) men (Fig. 8b).



Figs. 9. (a,b) Scatterplots showing the relationship between perceived speech clarity (where larger values indicate voices less likely to be chosen as the clearer voice in the paired comparison) and perceived sexual orientation (1 = *definitely sounds heterosexual*, 3 = *sounds neither GLB nor heterosexual*, 5 = *definitely sounds GLB*) for lesbian/bisexual (*L/B*) and heterosexual (*Het*) women (Fig. 9a) and gay/bisexual (*G/B*) and heterosexual (*Het*) men (Fig. 9b).

and more-negatively skewed /s/ were more likely to be rated as GLB-sounding than men with lower F2 and less-negatively skewed /s/. In the ratings for the front-vowel words, /s/ skewness and F1 frequency predicted a significant proportion of variance in the dependent measure. F1 frequency predicted 7.6% of the variance and had a standardized β-weight of 0.329. /s/ skewness predicted 8.0% of the variance and had a standardized β-weight of -0.366. Both β-weights were significant at the $\alpha < 0.05$ level. These variables predicted perceived sexual orientation in a manner similar to that in the regressions in Table 5: men with more negatively skewed /s/ were rated as more-GLB sounding than men with less-negatively skewed /s/; men with lower F1 were rated as more-heterosexual sounding than men with higher F1.

4.5. Discussion of Experiment 3

This experiment found that perceived sexual orientation was strongly predicted by measures of perceived height and perceived speech clarity. This relationship was particularly strong for ratings of women's voices: over 70% of the variance in judgments of perceived sexual orientation was accounted for by variance in the other two measures. The variance in perceived male sexual orientation that was accounted for was slightly lower, and was due to the influence of perceived speech clarity only. The fact that perceived height did not predict a significant proportion of variance in men's sexual orientation might have been due the strong correlation between perceived height and perceived speech clarity in men's voices.

These findings suggest that judgments of sexual orientation from single words may not be judgments of sexual orientation per se, but judgments of another parameter that is robustly coded in short samples of speech. These parameters, height and clarity, were chosen because they are present even in very short speech signals. A listener would be able to make a stable judgment of height or clarity from a short signal like a single read word. This study did not take measures of whether listeners consciously associate clarity of speech production with G/B men and taller stature with L/B women. However, the finding that GLB-sounding men's speech is rated as clearer-sounding is broadly consistent with at least one parameter that is part of the popular-culture stereotype, namely, that G/B men speak with more 'dynamic' intonation. It has been shown previously that larger f0 excursions are associated with more-intelligible speech (Laures & Weismer, 1999).

The results of this investigation suggest a mechanism for understanding why GLB speech styles are learned and maintained. As mentioned earlier, there does not seem to be a clear functional benefit for GLB people to convey their sexual orientation outside of certain social situations (i.e., conveying their sexual orientation to show group solidarity). If, however, GLB speech styles are conveyed via manipulations that have a plausible benefit (producing speech more clearly; giving the illusion of being larger in stature), then it is easier to understand why they would be learned, and maintained across different speaking situations. These are discussed further in the general discussion. A cautionary note should be offered that the direction of causality cannot be determined definitively using regressions methods. It may be the case that judgments of speech clarity or height are driven by listeners' perception of sexual orientation. We consider this unlikely, given that height and clarity, unlike sexual orientation, are intrinsically associated with specific acoustic parameters. However, we cannot rule out this possibility altogether, and future research will have to address this topic directly.

5. General discussion

5.1. Summary of Experiments 1–3

This investigation examined the acoustic and perceptual correlates of judgments of women and men's sexual orientation from read speech. The most salient findings are as follows. First, as expected, acoustic characteristics varied considerably as a function of talker sex. However, differences as a function of talkers' self-identified sexual orientation were very subtle. For men, group differences were most found in the first formant frequency of the vowels /ɛ/ and /æ/ (self-identified G/B men showed higher F1 frequencies than self-identified heterosexual men), and the skewness of /s/ spectra (G/B men produced /s/ with a more negatively skewed spectrum than heterosexual men). For women, group differences were found in the first formant frequency of /ɛ/ (L/B women produced a lower F1 frequency than heterosexual women) and the second formant frequency of /oʊ/ (L/B produced a lower F2 frequency than heterosexual women). No differences between GLB and heterosexual people were found in voicing-source measures.

In Experiment 2, listeners rated self-identified GLB people as more GLB sounding than heterosexual people. For women's voices, these ratings were not affected by the phonetic content of the stimuli over which the ratings were made. For men's voice, ratings were affected strongly by the vowels that the stimuli contained: G/B men were rated as more-GLB sounding when ratings were made from words containing low-front vowels than for words containing back vowels. Regression analyses predicting ratings of perceived sexual orientation of men's voices from acoustic measures showed that ratings were related to the parameters that differed as a function of actual sexual orientation in Experiment 1. Men with higher F1 in low vowels, higher F2 in back vowels, and more-negatively skewed /s/ were more likely to be rated as GLB sounding than men whose speech demonstrated the opposite characteristics. Regression analyses on ratings of women's voices showed that listeners were likely to rate a woman's voice as GLB sounding if she produced a lower F1 in low vowels, a lower F2 in back vowels, and a less-expanded vowel space overall.

Finally, Experiment 3 showed that ratings of perceived sexual orientation were related to measures of perceived height and perceived speech clarity. For women's voices, measures of perceived sexual orientation were strongly related to these two parameters, and acoustic measures did not predict any of the variance in perceived sexual orientation beyond what was accounted for by the two other perceptual measures. For men's voices, approximately 50% of the variance in perceived sexual orientation was accounted for by perceived speech clarity, and none was predicted by perceived height. Moreover, acoustic measures continued to predict a significant proportion of variance in perceived sexual orientation beyond what was accounted for by perceived height; the predictors were similar to those found in Experiment 2.

5.2. Accounting for the observed differences

This section considers possible explanations for the differences between GLB and heterosexual people seen in Experiments 1–3. As a point of reference, consider the multiplicity of explanations that have been offered for male–female differences more generally. Plausible explanations for differences can be constructed that appeal to social, motoric, perceptual, and cultural differences among the sexes. Differences between GLB and heterosexual people are no less subject to this multiplicity of explanations. One possible but unlikely explanation for these differences is that they reflect globally more-feminine speech in G/B men and globally more-masculine speech in L/B women. Indeed, this conjecture is consistent with much popular-culture speculation regarding this speech style. This conjecture is clearly not supported by the data presented in Experiment 1. G/B men did not show an overall higher scaling of vowel formants or f0, as would be expected if their speech were globally more feminine; L/B women did not produce speech with a lower f0 and overall lower formant frequencies. Though the speech of GLB isn't globally sex atypical, the parameters that did differ between the groups might indicate a selective adoption of some speech characteristics of the opposite sex. For example, G/B men's low F1 in /æ/ and /ɛ/, as well as their more negatively skewed /s/, are approximations of the speech produced by heterosexual women. Interestingly, the G/B-heterosexual male vowel-production differences found in this study mirror the male–female differences found by Fant (1996, 1975). Fant showed the largest differences between men and women to be in low vowels than in high vowels. However, we can conclude definitively from Experiment 1 that GLB speech styles are not a whole-scale adoption of sex-opposite speech patterns.

For men, the G/B speech style that we observed may have been related to two factors. First, it may reflect a habitual clear-speech style. This is supported by two pieces of data. First, the talkers that listeners identified as GLB-sounding in Experiment 2 were also likely to have been chosen as the clearer talker in the paired comparisons of speech clarity measured in Experiment 3. The strong relationship between these two variables suggests that the G/B speech style observed in this investigation involves many of the same modifications that talkers make when attempting to increase their clarity of speech production. Second, many of the acoustic characteristics of the speech of G/B men reported in this study and in previous research are similar to characteristics associated with clear speech production in other studies. For example, Pierrehumbert et al. (2004) found that G/B men produced more-expanded vowel spaces than heterosexual men; previous research has established that hyperarticulated vowel spaces are associated with more intelligible talkers (Bradlow et al., 1996). Gaudio (1994) reported a moderate but statistically nonsignificant relationship between listeners' judgments of sexual orientation and pitch variability; previous research has argued that greater pitch variability is associated with increased speech intelligibility (Laures & Weismer, 1999). In the current experiment, we were unable to replicate these results, but we did demonstrate that G/B men produced more-extreme low-front vowels /æ/ and /ɛ/ than heterosexual men. Thus, at least the front-vowel space demonstrated greater expansion for G/B men than for heterosexual men.

The origin of the more-negatively skewed /s/ may also relate to the habitual clear-speech style. The acoustic consequence of a strongly negatively skewed /s/ is a concentration of spectral energy above the mean. Much ambient environmental noise has a spectrum with a concentration of energy in the low frequencies (e.g., Busch, Hodgson, & Wakefield, 2003). Consequently, a token of /s/ with more energy in the high frequencies would be easier to perceive in the presence of typical background noise than a token of /s/ with a more diffuse spectrum. It would also make the /s/ acoustically more distinct than the sound that it is acoustically closest to, /ʃ/. The fact that /s/ skewness predicts a proportion of variance in men's perceived sexual orientation ratings beyond that accounted for by perceived clarity may mean that this variant of /s/ is an exaggeration of the clear-speech style. Future research should further examine the relationship between perceived sexual orientation and speech clarity two ways. First, it should utilize more-traditional sentence-intelligibility measures of speech clarity. Second, it should examine whether intentionally clear speech elicits systematically different ratings of perceived sexual orientation than more-natural conversational speech.

Differences in speech clarity did not explain all of the differences between G/B- and heterosexual-sounding male voices. This was shown in Section 4.4.2, in which we demonstrated that acoustic measures predicted a significant proportion of variance in perceived sexual orientation of men's voices even when measures of perceived clarity were controlled. Specifically, men who produced an extremely low /æ/, more-fronted /oʊ/ and /u/, and more-negatively skewed /s/ were more likely to be rated as GLB sounding than men whose speech demonstrated the opposite characteristics. We hypothesize that the vocalic features that distinguish G/B- and heterosexual-sounding men's voices may be related to differential engagement in ongoing sound changes. The dialect of English spoken in Minnesota is typically described as having extreme backness and rounding on the back vowels /u/ and /oʊ/, and a sharply falling F2 trajectory and low F1 associated with /æ/ (Labov et al., in press). This contrasts with many other dialects of English, in which /æ/ is produced with a high F1 and a flat F2, and /u/ and /oʊ/ are produced as more fronted and less rounded than in Minnesota. It is plausible to conjecture that the speech community in Minnesota is currently engaging in sound change in progress, in which /æ/, /oʊ/, and /u/ are becoming progressively more like the variants seen in other American English dialects. In this scenario, the differences between G/B and heterosexual men may relate to differential engagement in this sound change: G/B men might be using the newer variant more frequently than heterosexual men. Indeed, previous research has suggested that sound changes in progress are likely to be initiated with young women (Labov, 2001); G/B men may be following this pattern. This hypothesis could be tested with a larger-scale study of GLB-heterosexual speech-production differences in different-aged speakers.

Differences between L/B- and heterosexual-sounding women are less clear. One possible explanation arises from the finding of a strong relationship between ratings of perceived sexual orientation and measures of perceived height for women's voices: women who were rated as GLB-sounding were likely to be rated by an independent group of listeners to sound taller than average. The heterosexual-sounding women's speech style may reflect an attempt to convey a stature that is smaller than average. This explanation is consistent with Van Bezooijen's (1995) analysis of cross-cultural differences in vocal pitch, in which it was shown that Japanese women speak with a habitually higher vocal pitch than Dutch women, and that higher-pitched voices are associated with more positive affective qualities in Japan than in the Netherlands. An analogous situation may be at play in these data: in the culture studied in this investigation, positive affective qualities associated may be associated with speech that sounds as if it was produced by a woman of below-average stature, just as positive affective qualities are associated with high-pitched female voices in Japan. Future research should examine this question more directly by examining whether speech that has been manipulated acoustically to show the characteristics of taller people elicits systematically different ratings of perceived sexual orientation and other affective qualities than unmodified speech.

The hypothesis that GLB speech styles arise from modifications made for another benefit (speaking more clearly, sounding taller) might also explain why there is an imperfect correlation between actual sexual orientation and listener-identified sexual orientation. As noted earlier, there was an imperfect overlap between talkers' self-stated sexual orientation and listener-identified sexual orientation. The two men who were rated as most-GLB sounding included one self-identified gay man and one self-identified heterosexual man. Large overlap was also noted between the two groups of women. It may be that the heterosexual male talker who listeners rated as very GLB sounding was simply someone who strove to produce especially clear speech without concern with or knowledge of the effect that it would have on the perception of sexual orientation in his speech. The L/B women talkers who were rated as sounding shorter than average may have valued conforming to social expectations of the diminutivity of women's voices over expressing their sexual orientation through speech.

5.3. Future research

One clear area for future research is to expand the type and complexity of the speech materials used in production and perception experiments of GLB speech styles. The experiments in this article utilized read single words. It is unlikely that the talkers are able to convey a GLB speech style fully in such a constrained speaking task. Rather, the group differences seen in this investigation are likely to be a very pared-down version of the GLB speech style that is evident in more naturalistic speech. Indeed, previous studies on this topic have all utilized connected or conversational speech. Thus, future research on this topic should examine the instantiation of sexual orientation in speech using multiple levels of linguistic complexity, including single words and conversational speech. Such a study could shed further light on the extent to which sentence- and discourse-level prosody convey sexual orientation. It may be that differences between GLB and heterosexual people are stronger in conversational tasks than in single-word reading. Indeed, Lass et al. (1979) and Van Bezooijen and Gooskens (1999) both found that linguistically more-complex speech stimuli were needed for listeners to accurately and reliably identify social-indexical characteristics of speech (race in Lass et al., and regional variety in Van Bezooijen and Gooskens). Future studies may find that the GLB style in conversational speech to be an exaggeration of the style noted in single words, which we have argued to be related to intentional modifications to effect changes in speech clarity and perceived stature. Moreover, future research should also examine naturalistic speech of GLB people in social interactions, to examine the extent to which the GLB speech styles observed in controlled settings like single-word reading are used in natural social interactions.

More importantly, however, future research should examine the development of this speech style. If this style indeed reflects social-group membership (as proposed by Linville, 1998, and others), then we would expect that it would not be evident in people who have not yet identified themselves as GLB. If, however, this style reflects engagement in ongoing sound changes, a habitually clear speech style, or a style that conveys stature, then we might expect to see evidence of this style prior to a person overtly identifying as GLB. This research question poses logistic challenges, as there is no clear method for identifying all of the children, adolescents, and young adults who are likely to adopt a GLB identity. However, it is possible to identify a subset of the children who are likely to adopt a GLB identity as an adult, namely, children who demonstrate extremely gender nonconforming behavior. These children are sometimes given the label Gender Identity Disorder (GID, Zucker & Bradley, 2000). They demonstrate a variety of behaviors different from their gender-conforming peers, including the sex composition of their chosen peer group, avocations and interests, and, in some cases, overt gender dysphoria. Longitudinal studies of children with GID suggest that they are more likely than their gender-conforming peers to identify as GLB or transgendered as adults, or to elect gender-reassignment surgery. Given this finding, this population provides a potential means for studying the development of the GLB speech style, at least for the subset of the adult GLB community who demonstrated gender nonconformity during childhood. Longitudinal studies of the development of gender identity and the gender-typicality of speech in children with and without GID have the potential to provide very powerful data regarding the origin of distinctively GLB-sounding speech styles.

Acknowledgments

This research was supported by a Grant-in-Aid of Research, Scholarship, and Artistry from the University of Minnesota Graduate School. The authors are extremely grateful to Sarah Anderson, Molly Babel, Frank Davis, and Micaela Zolltheis for their help in analyzing these data. We gratefully acknowledge Bert Remijsen for describing the procedure for calculating H1−H2, and for providing Praat scripts to implement it. We also thank Pauline Welby for help with the other Praat scripts used in this investigation, and Edward Carney both for help in calculating the nonlinear regressions described in Section 2.4.3, and for fitting the Bradley–Terry models described in Section 4.4.1.2. This research has been presented previously at the Fall 2003 biannual meeting of the Acoustical Society of America in Austin, TX, the Fall 2004 meeting of the American Speech–Language–Hearing Association, and in a number of invited colloquia. We gratefully acknowledge the comments provided by audiences at those locations. We are particularly grateful to comments and feedback from Molly Babel, Mary Beckman, Ann Bradlow, Sarah Jefferson, Henrietta Jonas-Cedergren, Miriam Krause, Norma Mendoza-Denton, Janet Pierrehumbert, and Pauline Welby.

References

ANSI. (1989). *Specifications for Audiometers*. Washington, DC: American National Standards Institute.

Bailey, J. M. (2003). *The boy who would be queen: The science of gender-bending and transsexualism*. Washington: Joseph Henry Press.

Boersma, P., & Weenink, D. (2003). *Praat v. 4.1.7 [computer software]*. Amsterdam: Institute of Phonetic Sciences.

Bradlow, A. (2002). Confluent talker- and listener-oriented forces in clear speech production. In C. Gussenhoven, T. Rietveld, & N. Warner (Eds.), *Papers in laboratory phonology VII* (pp. 241–273). Cambridge: Cambridge University Press.

Bradlow, A., Torretta, G., & Pisoni, D. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic–phonetic talker characteristics. *Speech Communication*, *20*, 255–272.

Busch, T., Hodgson, M., & Wakefield, C. (2003). Scale-model study of the effectiveness of highway noise barriers. *Journal of the Acoustical Society of America*, *114*, 1947–1954.

Carahaly, L. (2000). *Listener accuracy in identifying the sexual orientation of male and female speakers*. Unpublished M.A. thesis, Department of Speech and Hearing Science, The Ohio State University, Columbus, OH.

Cho, T., & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, *29*, 155–190.

Diehl, R. L., Lindblom, B., Hoemeke, K. A., & Fahey, R. P. (1996). On explaining certain male–female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, *24*, 187–208.

Eckert, P., & McConnell-Ginet, S. (1999). New generalizations and explanations in language and gender research. *Language in Society*, *18*, 185–201.

Fant, G. (1966). *A note on vocal tract size factors and non-uniform F-pattern scalings*. Speech Technology Laboratory: Quarterly Progress and Status Report, Vol. 4 (pp. 22–30).

Fant, G. (1975). *Non-uniform vowel normalization*. Speech Technology Laboratory: Quarterly Progress and Status Report, Vol. 2–3 (pp. 1–19).

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, *116*, 2365–2373.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America*, *106*, 1511–1522.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, P. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, *84*, 115–123.

Gaudio, R. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech*, *69*, 30–57.

Goldinger, S. (1997). Words and voices: Perception and production in an episodic lexicon. In K. Johnson, & J. Mullenix (Eds.), *Talker variability in speech processing*. San Diego, CA: Academic Press.

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

Gordon, E. M. (1997). Sex, speech, and stereotypes: Why women's speech is closer to the standard. *Language in Society*, *15*, 23–52.

Hanson, H. (1997). Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America*, *101*, 466–481.

Hazan, V., & Markham, D. (2004). Acoustic–phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America*, *116*, 3108–3118.

Henton, C. (1989). Fact and fiction in the description of male and female pitch. *Language and Communication*, *9*, 299–311.

Herman, R. (2000). Phonetic markers of global discourse structures in English. *Journal of Phonetics*, *28*, 466–493.

Hermes, D. J., & van Gestel, J. C. (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, *90*, 97–102.

Hillenbrand, J., Getty, L., Clark, M., & Wheeler, K. (1995). Acoustic characteristics of American english vowels. *Journal of the Acoustical Society of America*, *97*, 3099–3111.

IEEE (1969). *Standards Publication No. 297*. IEEE Recommended Practice for Speech Quality Measurements, au-17(3) (pp. 225–246).

Klatt, D. H., & Klatt, L. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*, 820–857.

Labov, W. (1994). *Principles of linguistic change. Internal factors*, Vol. I. Cambridge, MA: Blackwell.

Labov, W. (2001). *Principles of linguistic change. Social factors*, Vol. II. Cambridge, MA: Blackwell.

Labov, W., Ash, S., & Boberg, C. (in press). *Atlas of North American english*. Berlin: Mouton de Gruyter

Lass, N. J., Tecca, J. E., Mancuso, R. A., & Black, W. I. (1979). The effect of phonetic complexity on speaker race and sex identifications. *Journal of Phonetics*, *7*, 105–118.

Lass, N. J., Almerino, C. A., Jordan, L. F., & Walsh, J. M. (1980). The effect of filtered speech on speaker race and sex identifications. *Journal of Phonetics*, *8*, 101–112.

Laures, J., & Weismer, G. (1999). The effects of flattened fundamental frequency on intelligibility at the sentence level. *Journal of Speech, Language, Hearing Research*, *42*, 1148–1156.

Linacre, J. M. (1995). Paired comparisons with ties: Bradley–Terry and Rasch. *Rasch Measurement Transactions*, *9*, 425.

Linville, S. (1998). Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Pholia Phoniatrica et Logopaedica*, *50*, 35–48.

Moonwoman-Baird, B. (1997). Toward the study of lesbian speech. In A. Livia, & K. Hall (Eds.), *Queerly phrased* (pp. 202–213). New York: Oxford University Press.

Munson, B. (2004). Variability in /s/ production in children and adults: Evidence from dynamic measures of spectral mean. *Journal of Speech, Language, and Hearing Research*, *47*, 58–69.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, *47*, 1048–1058.

Murray, S.O. (1999). [Review of A. Livia and K. Hall (Eds.), *Queerly phrased*]. Language in Society, Vol. 28 (pp. 304–308).

Peterson, G., & Barney, H. (1952). Control methods used in the study of vowels. *Journal of the Acoustical Society of America*, *24*, 175–184.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, *28*, 96–103.

Pierrehumbert, J., Bent, T., Munson, B., Bradlow, A. R., & Bailey, M. (2004). The influence of sexual orientation on vowel production. *Journal of the Acoustical Society of America*, *116*(4), 1905–1908.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: MIT Press.

Pisoni, D., Nusbaum, H., Luce, P., & Slowiacek, L. (1985). Speech perception, word recognition, and the structure of the lexicon. *Speech Communication*, *4*, 75–95.

Rogers, H., & Smyth, R. (2001). Vowel reduction as a cue to distinguishing gay- and straight-sounding male speech. In *Proceedings of the annual meeting of the Canadian Linguistics Society*. Ottawa: University of Ottawa.

Rogers, H., & Smyth, R. (2003). Phonetic differences between gay- and straight-sounding male speakers of North American English. Paper presented at the *15th international congress of phonetic sciences*, Barcelona, Spain, August 3–9, 2003.

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

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburgh: Psychology Software Tools, Inc.

Sedaris, D. (2001). *Me talk pretty one day*. Boston: Back Bay Books.

Simpson, A. P. (2001). Dynamic consequences of differences in male and female vocal-tract dimensions. *Journal of the Acoustical Society of America*, *109*, 2153–2164.

Simpson, A. P. (2002). Gender-specific articulatory–acoustic relations in vowel sequences. *Journal of Phonetics*, *30*, 417–435.

Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society*, *32*, 329–350.

Smyth, R., & Rogers, H. (2002). Phonetics, gender, and sexual orientation. In *Proceedings of the Annual Meeting of the Canadian Linguistics Association* (pp. 299–301). Montreal: l'University du Quebec au Montreal.

Thompson, S., & Bellini, P. (1998). *Buddy Babylon: The autobiography of Buddy Cole*. New York: Dell.

van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38*, 253–265.

van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology*, *18*, 21–48.

van Dommelen, W. A. (1993). Speaker height and weight identification: A re-evaluation of some old data. *Journal of Phonetics*, *21*, 337–341.

van Dommelen, W. A., & Moxness, B. H. (1995). Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and Speech*, *38*, 267–287.

Waksler, S. (2001). Pitch range and women's sexual orientation. *Word*, *52*, 69–77.

Zucker, K. J., & Bradley, S. J. (2000). Gender identity disorder. In C. H. J. Zeanah (Ed.), *Handbook of infant mental health*, (2nd ed.) (pp. 412–424). New York: Guilford Press.

Zwicker, E., & Ternhardt, E. (1980). Analytical expression for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America*, *68*, 1523–1525.

Zwicky, A. (1997). Two lavender issues for linguists. In K. Hall, & A. Livia (Eds.), *Queerly phrased* (pp. 21–32). Oxford: Oxford University Press.