Detection of high-frequency energy level changes in speech and singing

Brian B. Monson[a)]
National Center for Voice and Speech, University of Utah, 136 South Main Street, Suite 320, Salt Lake City, Utah 84101

Andrew J. Lotto and Brad H. Story
Department of Speech, Language, and Hearing Sciences, University of Arizona, P.O. Box 210071, Tucson, Arizona 85721

(Received 26 April 2013; revised 6 September 2013; accepted 25 October 2013)

Previous work has shown that human listeners are sensitive to level differences in high-frequency energy (HFE) in isolated vowel sounds produced by male singers. Results indicated that sensitivity to HFE level changes increased with overall HFE level, suggesting that listeners would be more "tuned" to HFE in vocal production exhibiting higher levels of HFE. It follows that sensitivity to HFE level changes should be higher (1) for female vocal production than for male vocal production and (2) for singing than for speech. To test this hypothesis, difference limens for HFE level changes in male and female speech and singing were obtained. Listeners showed significantly greater ability to detect level changes in singing vs speech but not in female vs male speech. Mean differences limen scores for speech and singing were about 5 dB in the 8-kHz octave (5.6–11.3 kHz) but 8–10 dB in the 16-kHz octave (11.3–22 kHz). These scores are lower (better) than those previously reported for isolated vowels and some musical instruments. © *2014 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4829525]

PACS number(s): 43.71.Es, 43.71.An, 43.71.Bp, 43.75.Rs [MAH] Pages: 400–406

I. INTRODUCTION

Some researchers have recently been motivated to examine the high-frequency portion of the speech spectrum. The draw has been due to sparse but increasingly conclusive evidence that high-frequency energy (HFE, defined here as the energy in the 8- and 16-kHz octave bands or 5.7–22 kHz) plays a larger perceptual role in speech perception than previously assumed. The percepts impacted by HFE that have been reported thus far are speech and voice quality (Olson, 1947; Moore and Tan, 2003; Moore *et al.*, 2010; Monson *et al.*, 2011), speech source localization (Best *et al.*, 2005), speech intelligibility (Lippmann, 1996; Stelmachowicz *et al.*, 2001; Apoux and Bacon, 2004; Moore *et al.*, 2010), and child word-learning (Stelmachowicz *et al.*, 2007; Pittman, 2008). Some studies indicate that HFE could also play a role in the perception of voice disorders (de Krom, 1995; Hartl *et al.*, 2003) and in talker identification (Hayakawa and Itakura, 1995; White, 2001; Liss *et al.*, 2010). These findings bear on communication devices and augmentative hearing devices that are just recently being manufactured to represent this frequency range (Moore, 2012; Pulakka *et al.*, 2012). For example, it has been proposed that the widely used classical telephone bandwidth (300–3400 Hz) be replaced with an adaptive multi-rate wideband (50–7000 Hz) standard (3GPP TS 26.190, 2005). Two new wider bandwidths (super wideband, 50–14 000 Hz, and full band, 20–20 000 Hz) are also undergoing standardization procedures, reflecting the desire for improved representation of the high frequencies in telecommunication (Geiser, 2012).

Given these results and potential applications, characterization of the impact of HFE on typical speech communication will be useful. Toward this goal, Monson *et al.* (2012) reported significant HFE level differences ranging from 2 to 5 dB between vocal production modes (speech vs singing), 3 to 8 dB between production levels (soft, normal, loud), 3 to 20 dB between phonemes (voiceless fricatives), and 2 to 5 dB between genders. These results lead to a natural follow-up: Is the auditory system sensitive enough to detect these level differences in HFE?

A previous study by Monson *et al.* (2011) showed that human listeners could detect small changes in HFE level in isolated vowels produced by two male singers of differing vocal timbre. That study consisted of listeners participating in a perceptual discrimination task using 500-ms excerpts of sustained /a/ vowels produced by the singers as stimuli with the 8- and 16-kHz center-frequency octave bands individually and incrementally attenuated. Difference limens (DLs) were obtained for the individual octaves and voices at two production levels (loud/soft). The major findings of the study were: (1) All listeners showed at least some ability to detect HFE level changes in sustained vowel sounds (i.e., without high-frequency noise generated by consonants); (2) listeners exhibited greater ability to detect changes in the 8-kHz octave than in the 16-kHz octave, although minimum DL scores (indicating greatest sensitivity) were comparable for

[a)]Author to whom correspondence should be addressed. Current address: Department of Newborn Medicine, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, CWN418, Boston, MA 02115. Electronic mail: brian.monson@duke-nus.edu.sg

both octaves; (3) median and minimum DL scores were negatively correlated with the sound pressure level (SPL) of the HFE octave band (for the 8-kHz octave); (4) there were large individual differences between normal-hearing listeners in their ability to detect HFE changes, which were not predicted by age or pure-tone thresholds; and (5) listeners' DL scores were comparable to—and in some cases better than—scores reported in previous studies using white noise (Viemeister, 1983; Moore *et al*., 1989) and musical instruments (Gunawan and Sen, 2008), suggesting that humans are equally sensitive to HFE in voice as they are to HFE from other sound sources despite the low relative level of HFE in voice.

Because Monson *et al*. (2011) used brief isolated voice samples from men, the generalization of this study to male and female running speech and singing is uncertain. First, the major contributors to HFE in running speech are consonants (Monson *et al*., 2011), and particularly voiceless fricatives (Jongman *et al*., 2000; Maniwa *et al*., 2009; Monson *et al*., 2012). Second, it has been suggested that HFE is of greater perceptual consequence for female talkers (Stelmachowicz *et al*., 2001), who tend to have higher HFE levels, at least in the 16-kHz octave (Monson *et al*., 2012). Thus some studies that have reported strong perceptual effects of HFE used only female speech (Lippmann, 1996; Stelmachowicz *et al*., 2007; Pittman, 2008). Finally, in the past, HFE has largely been associated with sound quality (e.g., Moore and Tan, 2003; Monson *et al*., 2011), suggesting the primary perceptual effect of HFE changes is a change in sound quality. Because sound quality is of primary concern in singing voice production, further examination of the role of HFE in singing voice (i.e., beyond isolated vowels) is warranted.

If higher HFE levels are correlated with listener sensitivity as suggested by Monson *et al*. (2011), one would predict that listeners will show greater sensitivity to HFE level changes in running speech vs isolated vowels and in female speech/singing vs male speech/singing. Furthermore, because higher HFE levels are also found in normal singing than in normal speech (Monson *et al*., 2012), and because HFE has been implicated in sound quality, it is hypothesized that HFE level differences are more detectable in singing voice production than in speech (where sound quality is presumably of lesser importance). To further explore human sensitivity to HFE changes in normal speech and singing, listeners participated in a perceptual discrimination task using speech and singing stimuli with the 8- and 16-kHz center-frequency octave bands individually and incrementally attenuated.

II. METHODS

A. Stimuli

Stimuli were drawn from a database of high-fidelity anechoic recordings of speech and singing. Details of the recordings can be found in Monson *et al*. (2012). Briefly, recordings consisted of spoken and sung versions of 20 six-syllable low-predictability phonetically representative phrases with alternating syllabic strength taken from Spitzer *et al*. (2007). Phrases were spoken and sung by 15 subjects (8 female), who were native speakers of American English with no reported history of a speech or voice disorder. All subjects also had at least 2 yr of post-high school private singing voice training. Sung versions were sung on a descending five-note scale.

One test phrase from the database was selected for the creation of the stimuli used in this experiment. Test phrase selection was based on the criteria that the phrase contain one voiceless fricative and be useful for comparison to the previous study. The phrase "amend the slower page" was selected because it contained the fricative /s/ and the sung vowel /a/, which was the vowel used for the stimuli in Monson *et al*. (2011).

To make the experiment generalizable, one female subject and one male subject were selected who exhibited average characteristics of HFE for speaking and singing of the test phrase. Because one of the main goals of the experiment was to compare listeners' ability to detect HFE changes in speech vs singing, the "average" subject was taken as a subject who exhibited the mean difference in HFE level between speech and singing. The normal speech and normal singing versions of the phrase "amend the slower page" for each subject were adjusted in overall level to the mean overall level at 1 m for normal speech (62 dB SPL) and singing (74 dB SPL), respectively. Octave band analysis was then performed (shown in Table I), and the mean level difference between speech and singing was calculated for each HFE octave band (8 and 16 kHz) for each gender. As a measure of deviation, a "sum of squared deviation" from the mean was calculated across HFE octave bands for each subject (compared against the mean for his/her gender) according to

$$(X_{8\text{kHz}} - \mu_{8\text{kHz}})^2 + (X_{16\text{kHz}} - \mu_{16\text{kHz}})^2 , \quad (1)$$

TABLE I. HFE octave levels at 1 m (re 20 μPa) for the spoken and sung versions of "amend the slower page" from each subject. Levels are relative to an overall signal level of 62 dB for speech and 74 dB for singing. Mean HFE octave levels for each gender are also shown.

Subject	Speech 8 kHz	Speech 16 kHz	Singing 8 kHz	Singing 16 kHz
f1	48.9	39.0	53.1	46.1
f2	42.2	36.5	49.3	39.5
f3	48.8	40.5	54.4	50.9
f4	45.6	39.7	47.2	41.5
f5	45.2	37.3	47.2	40.4
f6	43.6	39.7	50.0	41.6
f7	55.8	45.3	55.5	52.7
f8	46.5	36.5	50.0	39.4
Female mean	**47.1**	**39.3**	**50.8**	**44.0**
m1	45.1	28.3	53.8	38.3
m2	48.0	39.1	47.4	35.1
m3	40.8	35.0	48.8	37.8
m4	48.4	38.2	53.5	42.1
m5	43.5	32.7	49.6	36.5
m6	43.2	32.7	50.3	36.7
m7	43.9	33.8	48.5	42.9
Male mean	**44.7**	**34.2**	**50.3**	**38.5**

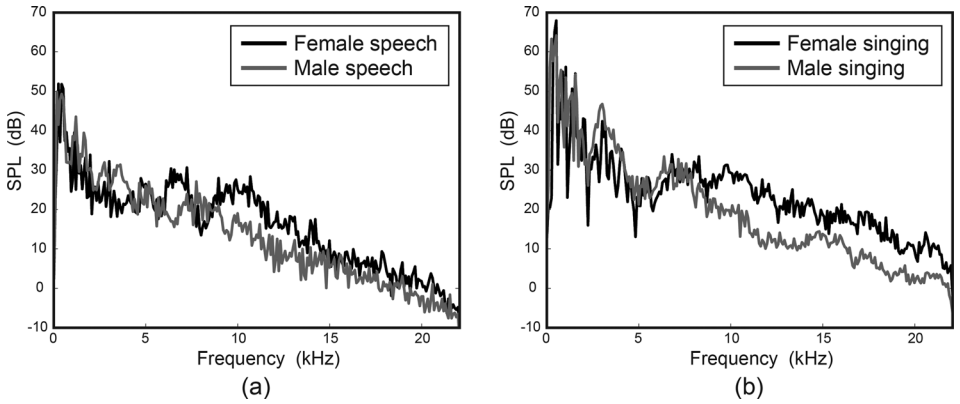

FIG. 1. Long-term average spectra (LTAS) for the female and male (a) spoken and (b) sung versions of "amend the slower page."

where X is the level difference (in dB) between speech and singing for the HFE band of interest for that subject and μ is the mean level difference (in dB) between speech and singing in that band for the subject's gender. The subject in each gender who exhibited the minimum sum was selected as the average subject. The minimum deviation from the mean for the female subjects was subject f1 and for the male subjects was subject m6.

The phrases for both subjects were sung in the key of F (five-note descending scale: C-B♭-A-G-F), one octave apart, resulting in fundamental frequencies of 523-466-440-392-349 Hz for the female and 262-233-220-196-175 Hz for the male. The time length varied for each stimulus. The lengths of the speech stimuli were 1.47 s (female) and 1.48 s (male), while the lengths of the singing stimuli were 4.41 s (female) and 5.45 s (male). Figure 1 shows the long-term average spectrum (LTAS) of each of the four stimuli, normalized to the overall mean levels at 1 m for normal speech (62 dB SPL) and singing (74 dB SPL). As expected, the female HFE octave band levels are higher than the male levels in every case, the smallest difference of about 3 dB found in the 8-kHz octave for singing (see Table I). Figure 1(a) also reveals speech spectral shape differences between genders in the 8-kHz octave: Male speech HFE exhibits a single peak at approximately 7.5 kHz, while female speech HFE shows a spectral dip at 7.5 kHz, between two spectral peaks at approximately 6.5 and 10 kHz.

The selected stimuli were played back using the experimental setup (see Sec. II C) without the listener. They were re-recorded for analysis with a Larson Davis 2541 type 1 precision microphone at 1 m. They were adjusted in playback level to the average levels at 1 m as determined from the original recordings.

B. Stimulus modification

Each stimulus was modified using a previously reported method (Monson *et al.*, 2011). Each stimulus was passed through a digital Parks-McClellan equiripple finite impulse response (FIR) bandstop filter to remove the octave band centered at 8 kHz (5657–11 314 Hz). Each stimulus was also passed separately through a bandpass filter to extract the 8-kHz octave. The output signal from the bandpass filter was then attenuated in 10-, 3-, or 1-dB steps to the desired level and summed with the output signal from the bandstop filter. This same procedure was also used for the 16-kHz octave (11 314–22 050 Hz), using a low-pass filter in place of the bandstop filter. Figure 2 illustrates this filtering process with attenuation of the separate octave bands in 10-dB steps for each of the stimuli. To eliminate the possible influence of audible artifacts of the filtering process, the "original" signal used (solid line) was regenerated by filtering the signal as described earlier and summing the extracted octave with 0 dB attenuation.

C. Listening conditions

The experiment took place in a standardized double-walled sound booth. The control room desktop computer running the experiment was installed with a Lynx L22 sound card connected directly to a Mackie HR624 high resolution studio monitor loudspeaker located in the booth. The frequency response of the sound card and loudspeaker had a good response (±5 dB) out to 20 kHz. Listeners sat in a desk directly in front of the loudspeaker with the ear located a distance of 1 m from the loudspeaker. Listeners were asked to avoid large deviations from their sitting positions, and

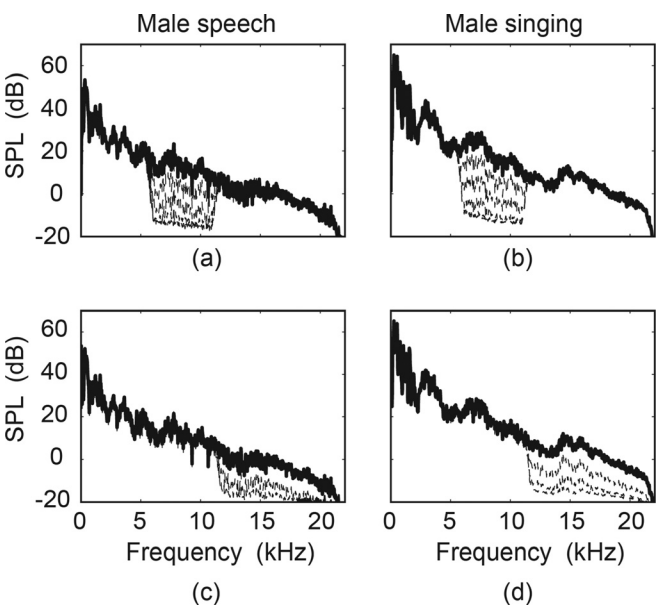

FIG. 2. LTAS of the octave band filtering process for (a) male speech 8-kHz octave, (b) male singing 8-kHz octave, (c) male speech 16-kHz octave, and (d) male singing 16-kHz octave.

specifically not to lean forward toward the loudspeaker, but were not physically constrained.

D. Participants

Twenty-three listeners participated in the experiment (13 female). The experiment was approved by the Institutional Review Boards at the University of Arizona and Brigham Young University. Participant age ranged from 19 to 32 yr with a mean age of 23 yr. Monaural audiometric thresholds were measured for all octave frequencies from 250 Hz to 16 kHz with a GSI 61 clinical audiometer with high-frequency capability. Telephonics TDH-50P (294D200-2) headphones were used for regular audiometric frequencies (250 Hz to 8 kHz), and Sennheiser HDA 200 headphones were used for high-frequency audiometry (8 and 16 kHz). This resulted in two thresholds obtained at 8 kHz for each ear. All subjects had thresholds better than or equal to 15 dB hearing level (HL) in at least one ear at all frequencies up to 8 kHz. Four subjects had thresholds worse than 15 dB HL in both ears at 16 kHz.

E. Procedure

An adaptive three-alternative forced-choice oddity task algorithm was used, implemented with the AFC software package developed by Stephan Ewert at the University of Oldenburg, Germany. For each trial, the listener was presented with three signals (separated by 500 ms of silence): Two identical signals consisting of an original version of one of the four stimuli, and one odd signal consisting of this same stimulus with one of the two HFE octave bands attenuated in level. The order in which the odd signal was presented was randomized for each trial. Instructions to the listener were given by the investigator prior to beginning the experiment and thereafter on the computer screen located next to the loudspeaker. Following the presentation of the signals, the listener was instructed to choose the odd signal with the prompt "Which of the voice samples sounded different?" Using a computer keyboard, s/he input a numerical response (1, 2, or 3). As is standard practice for level discrimination studies, the listener was given feedback ("correct" or "incorrect") on the response before moving on automatically to the next trial.

The experiment tested eight different stimulus conditions (2 production modes $\times$ 2 genders $\times$ 2 HFE octaves). The listening session consisted of a training block followed by two experimental blocks. The training block consisted of eight separate training runs, each using only one of the eight stimulus conditions. Each of the two experimental blocks, however, used all four stimuli (2 production modes $\times$ 2 genders) for one of the two HFE octaves. Block presentation order for the experimental blocks was matched across listeners. Stimulus presentation within each experimental block was interleaved and randomized across gender and production mode.

The first trial presented for each stimulus always had the greatest attenuation and the attenuation was incrementally decreased thereafter. For each run in the training block, a one-up one-down rule was used, and the attenuation step size changed from 10 to 3 dB on the first upper reversal. The training block run continued until two more reversals occurred, and these two reversals were averaged to obtain a preliminary DL estimate. For the experimental blocks, a one-up two-down rule was used and the attenuation step size changed from 3 to 1 dB on the first upper reversal. Following the first upper reversal, four more reversals were obtained and were averaged to obtain a DL estimate. The DL scores were reported as attenuation in decibels. From examination of the HFE levels and the booth noise floor, a practical value of 35 dB was selected as the maximum tolerable DL score for the 8-kHz octave and 30 dB for the 16-kHz octave. Any DL score greater than these values was interpreted as an inability to detect the HFE band.

III. RESULTS

All listeners showed ability to detect HFE level changes made to the stimuli. Figure 3 shows the percentage of listeners (on the ordinate) able to detect the HFE attenuations (on the abscissa) for each stimulus. All listeners detected the complete attenuation (DL scores < 35 dB) of the 8-kHz octave for all four stimuli. Approximately 65% of listeners detected the complete attenuation (DL scores < 30 dB) of the 16-kHz octave for each of the four stimuli.

Figure 4 shows boxplots of the DL scores obtained for each octave in each stimulus, revealing outlying scores in the 8-kHz octave, which were scored by three subjects. One subject had DL scores greater than two standard deviations away from the mean in three of the four stimuli. Two subjects accounted for the other five outlying scores (though these scores were within two standard deviations).

For the 16-kHz octave, seven subjects had all four scores worse than 30 dB, showing inability to perform the task. One subject had only one score better than 30 dB, while one subject had only one score worse than 30 dB, totaling

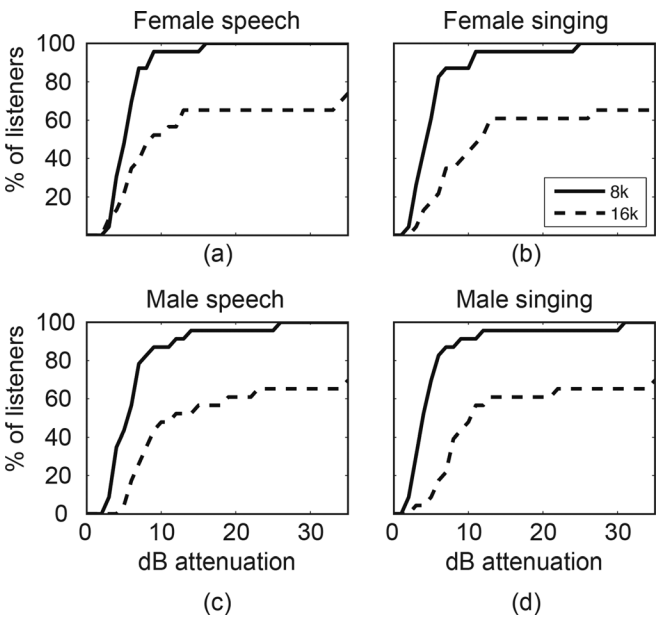

FIG. 3. Percentage of listeners (ordinate) able to detect the attenuations (abscissa) of the 8-kHz (solid line) and 16-kHz (dotted line) octave bands in (a) female speech, (b) female singing, (c) male speech, and (d) male singing.

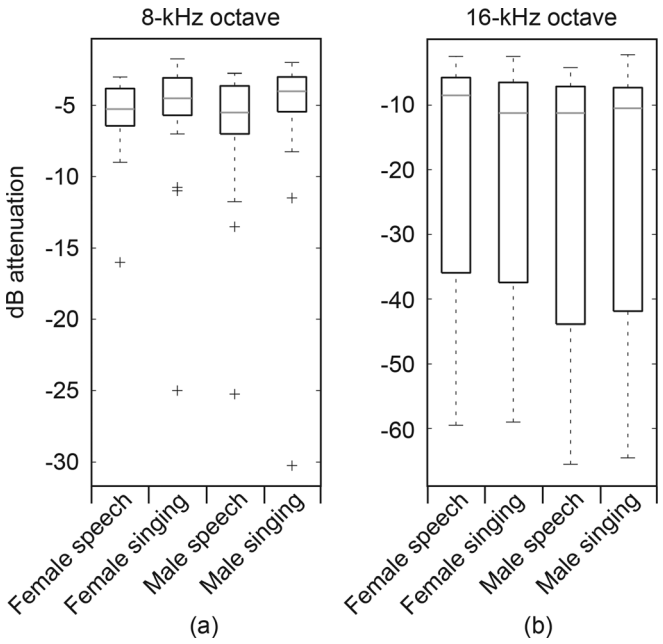

FIG. 4. Boxplots of DL score distributions for the (a) 8-kHz and (b) 16-kHz octaves for each stimulus.

nine subjects showing inability on the 16-kHz octave task for at least one stimulus. Given these initial results, it was determined to (1) perform separate statistical analyses for each octave, (2) exclude one subject scoring three outlying (>2 standard deviations) scores from the 8-kHz octave statistical analysis when calculating and comparing means, and (3) exclude the eight subjects scoring worse than 30 dB on at least three stimuli in the 16-kHz task from the 16-kHz octave statistical analysis when calculating and comparing means. These subjects were still included in other statistical analyses (e.g., calculation of medians, regression analyses, etc.). Table II gives the resulting mean, median, and minimum DL scores (and standard deviations) for each stimulus and HFE octave band.

A. Effects of production mode

It was hypothesized that human listeners are more able to detect HFE alterations made in singing than in normal speech. A repeated-measures analysis of variance (ANOVA) showed significant differences between DL scores for singing vs speech in the 8-kHz octave [$F(1,21) = 18.262$, $p < 0.001$, $\eta^2 = 0.465$] but not in the 16-kHz octave [$F(1,14) = 11.926$, $p = 0.411$, $\eta^2 = 0.049$]. These results suggest that listeners have greater sensitivity to the 8-kHz octave HFE in singing than in speech but not so for the 16-kHz octave HFE. It is of interest that the lowest mean and median DL score in the 16-kHz octave was for female speech despite the 16-kHz octave level being 7 dB higher in female singing. The mean, median, and minimum scores for male singing and speech, on the other hand, appeared to follow the trend seen in the 8-kHz octave scores.

TABLE II. Mean, median, and minimum DL scores and standard deviations for each HFE octave in each stimulus condition.

		Stimulus			
Octave	Score	Female speech	Female singing	Male speech	Male singing
8 kHz	Mean	5.6	4.8	5.8	4.5
	Median	5.1	4.6	5.5	4
	Minimum	3	1.8	2.8	2
	Std. Dev.	2.8	2.4	2.8	2.2
16 kHz	Mean	6.8	9.5	9.5	8.6
	Median	9.4	11.8	12.9	10.6
	Minimum	2.5	2.5	4.3	2.3
	Std. Dev.	3.1	8.3	5.2	4.5

B. Effects of gender

It was hypothesized that human listeners are more able to detect HFE alterations made in female speech/singing than in male speech/singing. A repeated-measures ANOVA showed no significant differences between DL scores for female vs male speech/singing in the 8-kHz octave [$F(1,21) = 0.027$, $p = 0.872$, $\eta^2 = 0.001$] nor in the 16-kHz octave [$F(1,14) = 1.875$, $p = 0.118$]. There was no significant interaction between gender and mode for either the 8-kHz octave [$F(1,21) = 0.568$, $p = 0.459$, $\eta^2 = 0.026$] or the 16-kHz octave [$F(1,14) = 3.574$, $p = 0.08$, $\eta^2 = 0.203$]. These results give little evidence to support our hypothesis. In fact, both median and mean scores were lower for male singing than for female singing (for both octaves) despite the male HFE level being lower in level. Differences in the spectral envelope [see Fig. 1(a)] did not seem to affect performance either. These results suggest that detection of HFE changes is of equal importance in male speech/singing as it is in female speech/singing despite female HFE levels being higher.

IV. DISCUSSION

In general, listeners were more sensitive to HFE changes in normal singing and running speech than previously reported for brief samples of isolated voice (i.e., 500-ms excerpts of sustained /a/ vowels produced by two male singers). All listeners successfully performed the 8-kHz octave task for every stimulus used. This was true for only one of the isolated voice stimuli, which was presented at a level of 75 dB SPL in the Monson *et al.* (2011) study. Listeners were also much more successful on the 16-kHz task with speech and singing. As reported previously, a total of 60% of listeners successfully performed the 16-kHz task across the isolated voice stimuli but no more than 50% were successful for any given isolated voice stimulus. Conversely, 65% of listeners were successful on every stimulus here. Furthermore, statistical measures (medians, means, etc.) were better in every case for speech and singing than for isolated voice stimuli, except for one isolated voice stimulus that showed somewhat comparable values (see Monson *et al.*, 2011).

These results indicate that the percentage of the population able to detect the presence of the 16-kHz octave is greater than the estimate provided by Monson *et al.* (2011) or what might be predicted from a previous report that decreasing the upper cutoff frequency from 16 854 to

10 869 Hz in band-limited speech shows no significant increase in perceived naturalness (Moore and Tan, 2003). While Monson *et al.* (2011) previously argued that applications such as voice synthesis and telecommunications that have the objective of providing a natural sounding speech signal should focus particularly on the 8-kHz octave, this large percentage of successful listeners suggests the 16-kHz octave is also important.

It is possible that stimulus duration had an effect on ability to detect HFE level changes, accounting for the poorer DL scores in speech than in singing. It would follow that mean DL scores should be anti-correlated with signal duration. This relationship was observed in the 8-kHz octave for the four stimuli used here ($r = -0.99$, $p < 0.05$) but was not observed in the 16-kHz octave ($r = 0.31$, $p = 0.69$), making it difficult to speculate on the effect of stimulus duration. Because sung phrases are generally time-extended versions of spoken text, however, controlling for duration of phrases is not possible without compromising phonetic content. Here, rather than truncate the longer tokens, the sung recordings were used in their entirety to retain the phonetic content of their speech counterparts.

It was difficult to find any predictors of individual listener ability to detect HFE changes. The average 8-kHz DL score was calculated for each listener and was used as the dependent variable in a step-wise linear regression analysis with age, years of musical training, and the minimum pure-tone thresholds (of the two ears) at each octave as possible predictors. None of these variables significantly predicted performance ($\alpha = 0.05$). This is perhaps due to the lack of variance in the 8-kHz octave performance. For the 16-kHz octave, however, a logistic regression analysis was used with these same variables to classify listeners into either the successful group (DLs < 30 dB) or unsuccessful group (DLs > 30 dB) with no variables being significantly predictive. On the other hand, the average 16-kHz DL scores calculated for listeners in the successful group ($n = 15$) were strongly correlated with the minimum pure-tone threshold measured at 16 kHz ($r = 0.826$, $p < 0.001$).

Comparison of the DLs obtained in speech/singing stimuli vs non-speech stimuli from previous studies supports the idea that listeners are equally or more sensitive to HFE in speech and singing than in white noise and musical instruments. Gunawan and Sen (2008), using 1.5-s samples of musical instruments presented at 65 dB SPL, reported mean 8-kHz DLs of approximately 3, 14, and 17 dB for clarinet, trumpet, and viola, respectively. The means achieved here for the samples of speech (also approximately 1.5 s in duration) were just over 5.5 dB, scoring much better than the trumpet and viola and rivaling the clarinet scores. Means for singing were even lower (about 4.5 dB), although again the stimulus times were longer, which could have affected performance. The mean 8-kHz DL score for white noise stimuli presented to three listeners by Moore *et al.* (1989) was slightly less than 4 dB, although these listeners were given at least 6 h of practice before participating in the experiment. Again, mean DLs achieved here rival this score, and listeners were trained here for less than 30 min. Interestingly, those three listeners' DLs for a lower frequency white noise band (500-Hz bandwidth centered at 1 kHz) were approximately 1.5, 2, and 3 dB—the latter two scores being achieved here 20 times across stimuli and in both high-frequency octaves.

While not tested here, one might question whether listeners would show improved DL scores for attenuation of the entire HFE band (i.e., both HFE octaves, 5.7–22 kHz). This possibility follows from the Moore *et al.* (1989) study wherein listeners showed increased ability to detect level changes of spectral notches in broadband noise stimuli as the notch bandwidth increased (with the same center frequency). However, this effect was not as apparent in the Gunawan and Sen (2008) study, which tested detectable band notches for musical instruments. DL scores for wider-bandwidth notches were usually comparable to, but not necessarily better than, narrowband notch DL scores for the musical instruments. In a few cases, narrow band notches were more readily detectable than wide band notches. Because these cases were rare, however, we speculate that attenuation of the full HFE band would result in DL scores at least comparable to, and perhaps better than, DL scores achieved for the 8-kHz octave alone.

Listeners were asked to report on their experience after the experiment by describing how the odd sample differed from the other two and how they performed the task. The reports given use a variety of qualitative descriptors including "muffled," "bright/dark," "not as full," "not as clear," "tinny," "less bright," "resonance," "muddled," "not as pure," "raspy," "tunnel sound," "less brilliant," "difference in clarity," "muted," "not crisp," "damped," and "dull." A few others described more quantitative differences in loudness, volume, and pitch, and still others used physiological descriptors of "more nasal," "more closed," "like a lisp," "further back," and "covered." Some listeners reported one of the genders being easier than the other (two for female, two for male), and some reported likewise for one of the production modes (three for speech, two for singing). DL scores generally corroborated these perceptions for these individuals, except in the perception of speech being easier, where all three individuals scored better DLs for singing than for speech.

The qualitative descriptors used by the listeners are consistent with the idea that HFE plays a role in the perception of sound quality. These descriptors might be of particular interest for singers and teachers of singing whose objective it is to attain a vocal sound of a certain quality. The results here indicate that—because listeners are sensitive to HFE changes—the quality of a vocal performance will be affected by the amount of HFE present in a singing voice. The level of HFE could potentially be controlled by the singer but might also be manipulated electronically during sound equalization and mixing procedures (i.e., by manipulating the amount of "treble"). This speaks to the importance of adequate high-frequency hearing for sound engineers' and singing teachers' ability to determine the quality of a singing performance.

Most listeners reported to which specific phonemes and/or syllables they would attend to perform the task (some specified more than one phoneme). Included were the phonemes /s/ (reported by ten subjects), /Σ/ (three), /ə/ (three), /d/ (two), /m/ (one), and /ð/ (one); the syllables "mend" (one)

and "slow" (one); the glide /w/ in the word "slower" (one); and the "sound of the consonants" (one). Interestingly, only about half of the anecdotal reports specified voiceless fricatives, which are assumed to be the main contributor to HFE perception due to their HFE profile. Many of the responses suggest that HFE is of perceptual significance in vowels, nasals, and glides. They also indicate HFE could play a role in the perception of certain speech disorders, vowel placement, loudness, and pitch. It will be instructive in follow-up experiments to assess these aspects of listeners' perceptions more quantitatively.

V. CONCLUSIONS

The results here indicate that normal-hearing listeners can detect HFE level differences of 4–6 dB. Because level differences this large are found between genders, between voiceless fricatives, and between production modes (speech vs singing) (Monson *et al.*, 2012), listeners may be able distinguish these separate classes solely on the basis of HFE level differences. Thus HFE potentially provides perceptual information useful for classification. About 1/3 of listeners could not detect the 16-kHz octave band, but this inability was not predicted by audiometric thresholds nor by age (though listeners here were all between 19 and 32 yr old). In fact, it was difficult to find any significantly predictive measure for ability to detect HFE level differences.

We find little evidence to support the idea that HFE is of greater perceptual consequence for female speech production than for male speech production despite the increased levels of HFE in female speech and other differences in the spectral envelope. It is possible that gender effects on perception of HFE are task specific since Stelmachowicz *et al.* (2001) reported an effect for intelligibility, but their study examined the perception of only one phoneme (/s/) using one female talker and one male talker. Here an attempt to circumvent the problem of generalizability was made by choosing average female and male talkers. HFE does appear to be of greater perceptual consequence in singing than in speech as might be expected due to its contribution to qualitative percepts (Olson, 1947; Moore and Tan, 2003). These results are meaningful for communication and hearing devices that attempt to represent this frequency range. They are also useful for sound reinforcement, voice synthesis, and speech perception experiments where faithful (or poor) reproduction of HFE will have an effect on listeners' percepts.

ACKNOWLEDGMENTS

This work was funded by NIH Grant Nos. F31DC010533, R01DC8612, and R01DC6859. The authors thank Richard Harris, Kent Gee, Eric Hunter, and the Brigham Young University Acoustics Research Group for use of facilities and equipment.

3GPP TS 26.190 (**2005**). *Adaptive Multi-Rate-Wideband (AMR-WB) Speech Codec, Transcoding Functions*, 3rd Generation Partnership Project, Valbonne, France, version 6.1.1.

Apoux, F., and Bacon, S. P. (**2004**). "Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise," J. Acoust. Soc. Am. **116**, 1671–1680.

Best, V., Carlile, S., Jin, C., and van Schaik, A. (**2005**). "The role of high frequencies in speech localization," J. Acoust. Soc. Am. **118**, 353–363.

de Krom, G. (**1995**). "Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments," J. Speech Hear. Res. **38**, 794–811.

Geiser, B. (**2012**). "Paths toward HD-voice communication," in *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on (VDE)*, pp. 1–4.

Gunawan, D., and Sen, D. (**2008**). "Spectral envelope sensitivity of musical instrument sounds," J. Acoust. Soc. Am. **123**, 500–506.

Hartl, D. M., Hans, S., Vaissiere, J., and Brasnu, D. F. (**2003**). "Objective acoustic and aerodynamic measures of breathiness in paralytic dysphonia," Eur. Arch. Otorhinolaryngol. **260**, 175–182.

Hayakawa, S., and Itakura, F. (**1995**). "The influence of noise on the speaker recognition performance using the higher frequency band," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 321–324.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Lippmann, R. P. (**1996**). "Accurate consonant perception without mid-frequency speech energy," IEEE Trans. Speech Audio Process. **4**, 66–69.

Liss, J. M., LeGendre, S., and Lotto, A. J. (**2010**). "Discriminating dysarthria type from envelope modulation spectra," J. Speech Lang. Hear. Res. **53**, 1246–1255.

Maniwa, K., Jongman, A., and Wade, T. (**2009**). "Acoustic characteristics of clearly spoken English fricatives," J. Acoust. Soc. Am. **125**, 3962–3973.

Monson, B. B., Lotto, A. J., and Story, B. H. (**2012**). "Analysis of high-frequency energy in long-term average spectra (LTAS) of singing, speech, and voiceless fricatives," J. Acoust. Soc. Am. **132**, 1754–1764.

Monson, B. B., Lotto, A. J., and Ternstrom, S. (**2011**). "Detection of high-frequency energy changes in sustained vowels produced by singers," J. Acoust. Soc. Am. **129**, 2263–2268.

Moore, B. C. (**2012**). "Effects of bandwidth, compression speed, and gain at high frequencies on preferences for amplified music," Trends Amplif. **16**, 159–172.

Moore, B. C. J., Fullgrabe, C., and Stone, M. A. (**2010**). "Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task," J. Acoust. Soc. Am. **128**, 360–371.

Moore, B. C. J., Oldfield, S. R., and Dooley, G. J. (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. **85**, 820–836.

Moore, B. C. J., and Tan, C. T. (**2003**). "Perceived naturalness of spectrally distorted speech and music," J. Acoust. Soc. Am. **114**, 408–419.

Olson, H. F. (**1947**). "Frequency range preference for speech and music," J. Acoust. Soc. Am. **19**, 549–555.

Pittman, A. L. (**2008**). "Short-term word-learning rate in children with normal hearing and children with hearing loss in limited and extended high-frequency bandwidths," J. Speech Lang. Hear. Res. **51**, 785–797.

Pulakka, H., Laaksonen, L., Yrttiaho, S., Myllyla, V., and Alku, P. (**2012**). "Conversational quality evaluation of artificial bandwidth extension of telephone speech," J. Acoust. Soc. Am. **132**, 848–861.

Spitzer, S. M., Liss, J. M., and Mattys, S. L. (**2007**). "Acoustic cues to lexical segmentation: A study of resynthesized speech," J. Acoust. Soc. Am. **122**, 3678–3687.

Stelmachowicz, P. G., Lewis, D. E., Choi, S., and Hoover, B. (**2007**). "Effect of stimulus bandwidth on auditory skills in normal-hearing and hearing-impaired children," Ear Hear. **28**, 483–494.

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., and Lewis, D. E. (**2001**). "Effect of stimulus bandwidth on the perception of vertical bar s vertical bar in normal- and hearing-impaired children and adults," J. Acoust. Soc. Am. **110**, 2183–2190.

Viemeister, N. F. (**1983**). "Auditory intensity discrimination at high-frequencies in the presence of noise," Science **221**, 1206–1208.

White, P. (**2001**). "Long-term average spectrum (LTAS) analysis of sex- and gender-related differences in children's voices," Logoped. Phoniatr. Vocol. **26**, 97–101.