

Detection of high-frequency energy changes in sustained vowels produced by singers

Brian B. Monson^{a)} and Andrew J. Lotto

Department of Speech, Language, and Hearing Sciences, University of Arizona, P.O. Box 210071, Tucson, Arizona 85721

Sten Ternström

Department of Speech, Music, and Hearing, Kungliga Tekniska Högskolan, Lindstedtsvägen 24, SE-100 44 Stockholm, Sweden

(Received 5 August 2010; revised 17 January 2011; accepted 19 January 2011)

The human voice spectrum above 5 kHz receives little attention. However, there are reasons to believe that this high-frequency energy (HFE) may play a role in perceived quality of voice in singing and speech. To fulfill this role, differences in HFE must first be detectable. To determine human ability to detect differences in HFE, the levels of the 8- and 16-kHz center-frequency octave bands were individually attenuated in sustained vowel sounds produced by singers and presented to listeners. Relatively small changes in HFE were in fact detectable, suggesting that this frequency range potentially contributes to the perception of especially the singing voice. Detection ability was greater in the 8-kHz octave than in the 16-kHz octave and varied with band energy level.

© 2011 Acoustical Society of America. [DOI: 10.1121/1.3557033]

PACS number(s): 43.75.Rs, 43.75.Cd, 43.71.Bp, 43.70.Gr [NHF]

Pages: 2263–2268

I. INTRODUCTION

The human voice produces acoustic energy above 5 kHz, as both harmonics and noise. This high-frequency energy (HFE) has traditionally received little attention in speech and voice research. Reasons for this may include: (1) presence of HFE is not necessary for highly intelligible speech communication and reproduction, (2) the HFE level is much lower than the spectral energy level below 5 kHz, (3) early signal processing technology did not allow for accurate measurement or analysis of HFE, and (4) as frequency increases, voice acoustics become increasingly difficult to analyze and model (the planar wave propagation assumption in vocal tract modeling becomes invalid, scattering behavior increases in complexity, etc.). The result is a lack of detailed characterization of the content of HFE from human voice and its perceptual effects.

Despite the lack of scientific scrutiny, HFE in voice does receive attention in some practical applications. Recording studio and live sound engineers regularly manipulate the “treble” frequency range for speaking and singing voices, using equalization techniques. The sensitivity of the engineer, or an audience of listeners, to changes in HFE is unknown. Additionally, while standard telephony has been restricted to the bandwidth below 4 kHz, so-called “wideband” telephony is now being integrated into modern telephony domains. This technology extends the bandwidth of transmission to 7 kHz. The perceptual effect of this improvement has not been studied in detail. Claims have also been made that the bandwidth of augmentative hearing devices should be extended to include HFE (e.g. Ref. 1). The

usefulness of such bandwidth extension, and what frequency range is actually necessary, is still heavily debated.

The limited number of studies on HFE information in speech and voice have examined its contribution to overall acoustic energy in normal speech and voice;^{2–8} its contribution to overall acoustic energy in voice disorders;^{9–11} its effect on localization ability;¹² and its effect on speech intelligibility.^{13–16}

A few studies that have examined speech and voice quality have suggested that HFE plays a role in the perceived quality of voice. As early as 1957, Olson¹⁷ reported that listeners preferred live full bandwidth speech (and music) to that acoustically low-pass filtered at 5 kHz. He claimed that “a limited frequency range impairs the quality and artistic value” of speeches and songs. Qian and Kabal¹⁸ observed that “voice quality is perceived as being much worse” for standard telephony narrowband speech (300–3400 Hz) than for wideband speech (50–7000 Hz). In a listening experiment they found that listeners preferred wideband speech quality to that of narrowband speech, though no attempt was made to isolate the separate effects of low-frequency vs high-frequency extension. Moore and Tan,¹⁹ however, found that the percept of “naturalness” was affected specifically by HFE. Part of their study examined naturalness scores for bandpass filtered speech and music. Changing the upper cut-off frequency from approximately 11 kHz to 7 kHz (with a lower cutoff of 55 Hz) markedly decreased perceived naturalness scores for the speech signal.

The aforementioned studies have established that HFE in speech is audible. Understandably, any perceptual work done related to HFE in these studies has focused on running speech or isolated consonants.^{1,13–16,18,19} It is generally understood that consonants (fricatives, in particular) will exhibit higher levels of HFE than vowels. To illustrate this,

^{a)}Author to whom correspondence should be addressed. Electronic mail: bbmonson@email.arizona.edu

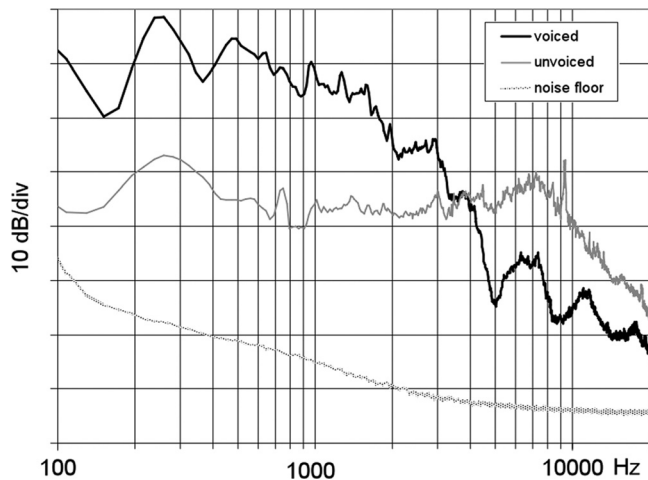


FIG. 1. Long time average spectrum of 60 s of running speech, illustrating the separate contributions of voiced and unvoiced speech segments. The speaker was a female reading out loud from a novel. (The sharp peak at just over 9 kHz in the unvoiced LTAS is due to a single whistling “s” consonant.)

Fig. 1. shows an example of the separate contributions of voiced and voiceless speech to overall acoustic energy for running speech. The question remains then as to whether HFE might contribute to percepts of isolated vowel sounds. This question is particularly pertinent to singing, in which the majority of the acoustic signal consists of sustained vowel sounds. The potential importance of HFE for singing is compounded when considering that previous studies have indicated that HFE contributes to percepts of quality in speech. As qualitative percepts are especially significant for singing, studying HFE in singing voice seems a logical step.

Previous studies have made no attempt to quantify listener sensitivity to HFE in that they have only presented speech signals with either the normal level of HFE or complete attenuation of HFE bands. One purpose of the current experiment was to explore listener sensitivity to changes of the HFE in voice signals by attenuating normal HFE levels in small increments. In other words, previous studies have taken an “all-or-nothing” approach to determining human ability to hear HFE in speech. This study instead examines human sensitivity in this frequency range by incrementally changing the “landscape” of HFE. While such perceptual experiments have been performed using non-speech stimuli (such as broadband noise^{20,21} and musical instrument recordings²²), it is not certain that results from such studies can predict sensitivity to HFE in human voice stimuli due to the differing spectral levels of HFE in voice. In performing experiments using human voice stimuli we hoped to begin to determine the extent to which HFE in speech and voice provides useful perceptual information. In particular, before embarking on the difficult task of including HFE in voice synthesis and vocal tract modeling, it would be helpful to know if listeners are highly sensitive to subtle changes in HFE.

It is unknown how HFE production or detection might be affected by overall sound pressure level or differing voice type. It is also not well known how dependent on frequency the detection of HFE may be (though results from Moore and Tan¹⁹ indicated that filtering out energy above 11 kHz

did not significantly change the perception of naturalness for speech). In the current study, vocalizations produced at different sound levels were used, and separate frequency bands were attenuated in isolation. Two male singers of differing voice quality and timbre were used. The specific questions addressed by this research were: (1) Are listeners able to detect small level changes in HFE of isolated vowel sounds? (2) Are listeners able to detect these changes equally across frequency bands? (3) Are listeners able to detect these changes equally in the same voice at different sound pressure levels? (4) Are listeners able to detect these changes equally for different voices/voice types?

II. METHOD

To determine the audibility of changes in HFE, the energy levels of the 8- and 16-kHz center-frequency octave bands were individually attenuated in sustained vowel sounds produced by singers.

A. Stimuli

Recordings were taken from a database of high-fidelity (16-bits, 44.1 kHz) recordings of singers in sound-treated rooms, with a microphone distance of 30 cm directly in front of the singer’s mouth. Singers were asked to phonate without vibrato on the vowel /a/ at various intensity levels. The subjects chosen for this study were two male tenors. Two 500-ms excerpts of each subject were selected based on overall sound pressure level (SPL) to represent “loud” and “soft” phonation. The fundamental frequency of phonation was approximately 170 Hz.

Each stimulus was passed through a digital Parks–McClellan equiripple FIR bandstop filter to remove the octave band centered at 8-kHz (5657–11 314 Hz). Each stimulus was also passed separately through a digital Parks–McClellan equiripple FIR 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 used for the 16-kHz octave (11 314–22 050 Hz), using a low-pass filter in place of the bandstop filter, and a high-pass filter in place of the bandpass filter. Figure 2 illustrates this filtering process with attenuation of the separate octave bands in 10-dB steps for one of the stimuli. To eliminate the possible influence of audible artefacts of the filtering process, the “original” signal was regenerated by filtering the signal as described earlier and summing the extracted octave with 0-dB attenuation.

The generated stimuli were played back using the experimental set-up (see Sec. II B) without the listener. They were re-recorded for analysis with a Brüel and Kjær Type 4003 microphone at 1 m from the loudspeaker. The spectra of the four stimuli (loud and soft × two singers) re-recorded with the B&K microphone in this arrangement are shown in Fig. 3. The stimuli were ordered according to SPL, and each was slightly amplified or attenuated in level so that overall SPLs at 1 m were 80, 75, 70, and 65 dB for the stimuli as labeled. These values approximate the actual levels of the stimuli

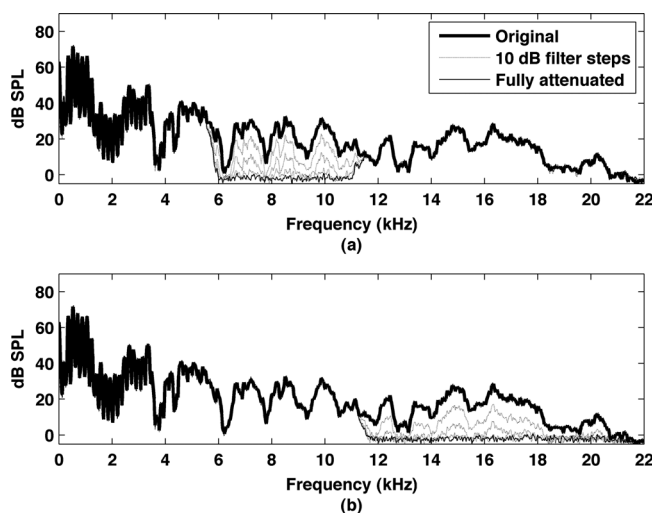


FIG. 2. Spectra of a sustained /a/ after attenuation of (a) the 8-kHz octave and (b) the 16-kHz octave.

when produced at 1 m based on the initial levels recorded at 30 cm.

Table I gives the overall SPLs for the stimuli. The absolute and relative levels of the HFE octave bands are also reported. For each voice, HFE was seen to decrease with decrease in overall intensity. It can be seen, however, that voice 2 “loud,” while having a lower overall SPL than voice 1 “loud” (75 dB compared to 80 dB), exhibited higher absolute and relative levels of HFE for both octave bands. The same phenomenon was seen when comparing voice 2 “soft” to voice 1 “soft.” Subjectively, voice 1 would be described as having a “darker” tone quality, while voice 2 would be described as having a “brighter” tone quality. It is likely that HFE level contributes to these perceptual voice qualities. This idea is strengthened by the fact that relative HFE levels appeared to be voice specific and relatively stable across phonation level. Voice 2 was approximately 10 dB greater in relative level for the 8-kHz octave than voice 1, and more

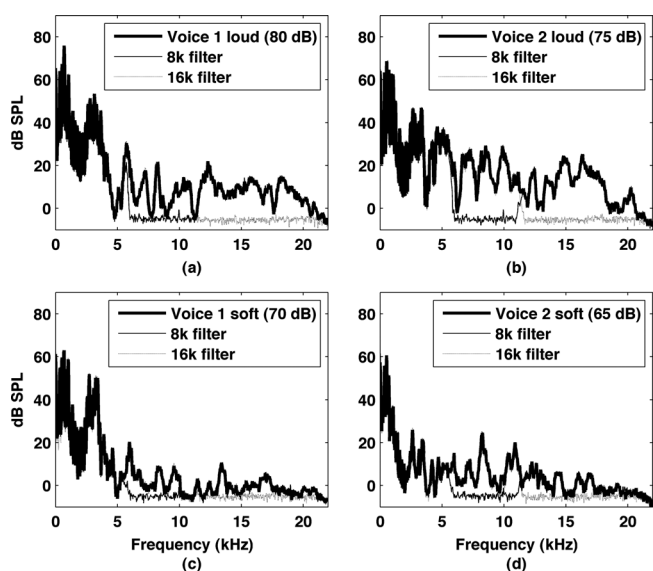


FIG. 3. Spectra of two voices on the sustained vowel /a/ during loud (a and b) and soft (c and d) phonation.

TABLE I. Overall levels at 1 m (re 20 μ Pa) of a sustained vowel /a/ produced by two male singers and two phonation levels. Absolute levels of the 8- and 16-kHz octave bands as well as levels relative to overall stimulus level are also given.

Stimulus	Overall level (dB)	8-kHz Octave level		16-kHz Octave level	
		Abs (dB)	Rel (dB)	Abs (dB)	Rel
Voice 1 loud	80	40.3	-39.7	37.1	-42.9
Voice 2 loud	75	45.1	-29.9	41.3	-33.7
Voice 1 soft	70	30.9	-39.1	27	-43
Voice 2 soft	65	36.2	-28.8	28.4	-36.6

than 6 dB greater in the 16-kHz octave. It is interesting that this held true for both the loud and soft conditions. It is also interesting that, while overall HFE level increased with sound level, relative HFE level for each voice changed very little with change of sound level. These patterns suggest that HFE can provide information on stable individual differences in voice quality. In all cases, the removal of both HFE octave bands had very little effect on the overall SPL of the stimulus (<0.01 dB).

B. Procedure

Thirty listeners participated in the experiment (14 female). Age ranged from 20 to 61 years, with a mean age of 29.5 years. Binaural audiometric thresholds were measured for octave frequencies from 250 Hz to 8 kHz. All subjects had thresholds better than or equal to 15 dB HL for all frequencies tested.

The experiment took place in a sound-treated room. Stimuli were played over an Audio Pro Type A4-14 Mk II (Audio Pro, Sweden) full-range powered loudspeaker with good high-frequency response (± 4 dB from 6 to 20 kHz). Stimuli were subjected to 10-ms raised-cosine fade-in and fade-out functions during playback. Listeners sat directly in front of the loudspeaker at a distance of 1 m.

An adaptive three-alternative forced-choice oddity task was used. For each trial the listener was presented with three stimuli (separated by 500 ms of silence): two of the original signal and one test signal with an attenuated HFE level. The order in which the test signal was presented was randomized for each trial. Instructions to the listener were given on a computer screen located next to the loudspeaker. The listeners were to choose the odd signal and were given feedback on their response before moving on to the next trial. The experimental session consisted of a training block followed by two experimental blocks. Each of the two experimental blocks tested one of the two separate dB HFE octaves. Block presentation order was matched across listeners. Stimulus presentation within each block was interleaved and randomized across voice number and phonation level.

The first trial presented for each stimulus always had the greatest attenuation and the attenuation was incrementally decreased. For the training block, a 1-up 1-down rule was used and the attenuation step size was set at 10 dB; the training block continued until two reversals occurred. For the experimental blocks, a 1-up 2-down rule was used and the attenuation step size changed from 3 to 1 dB on the first

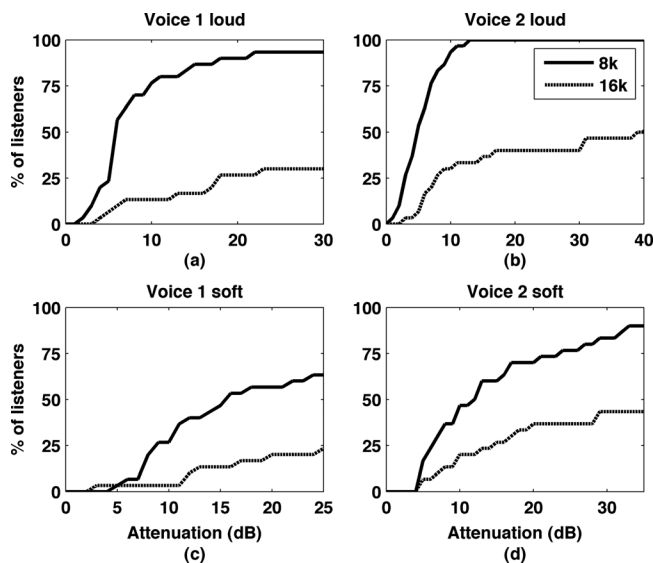


FIG. 4. Percentage of listeners able to detect the attenuations of the 8-kHz (solid line) and 16-kHz (dotted line) octave bands in the four stimuli. The maximum attenuation shown is based on maximum attenuation before reaching the noise floor, which varied for each stimulus.

upper reversal. The block then continued until three reversals occurred. The levels at these three reversals were averaged to obtain a difference limen (DL) estimate. (This procedure was similar to that used by Gunawan and Sen.²²)

III. RESULTS AND DISCUSSION

All listeners showed at least some ability to detect HFE level changes. Figure 4 shows the percentage of listeners (on the ordinate) able to detect the HFE attenuations (on the abscissa) for each stimulus. Table II gives the median and minimum DL values for each stimulus and HFE octave band. The DLs are reported here as octave band attenuation in dB. Results are reported comparing octave band, subject voice, and sound level.

A. Octave band

Overall, listeners showed much greater ability to detect HFE level changes in the 8-kHz octave band than in the 16-kHz octave band. All listeners were able to detect changes in level of the 8-kHz octave band for at least one of the stimuli. For three of the four stimuli used, 90% or more of listeners could detect full attenuation of the 8-kHz band (voice 1 soft

TABLE II. Median DL scores for the 8-kHz octave (with standard deviations) and minimum DL scores for both octave bands in each stimulus. Median DL scores for the 16-kHz octave are not included because no more than 50% of listeners could detect attenuation of the 16-kHz octave band for any stimulus.

Stimulus	8-kHz			16-kHz
	Med (dB)	SD (dB)	Min (dB)	Min (dB)
Voice 1 loud	6	8.5	1.3	4
Voice 2 loud	4.8	2.9	1	3
Voice 1 soft	16	13.9	5	2.7
Voice 2 soft	11.8	11.9	4.3	4.3

being the exception). Conversely, no more than half of the listeners could detect full attenuation of the 16-kHz octave for any given stimulus. However, 60% of listeners showed at least some ability to detect changes in the 16-kHz octave band level.

The 8-kHz octave appears at first to be of greater perceptual significance than the 16-kHz octave. However, it should be noted that minimum DL scores for the 16-kHz octave were, at most, only 2.7 dB greater than their 8-kHz octave counterparts. The range of DL scores was quite large, and certain listeners were able to detect changes in the 16-kHz octave much better than others. In some cases, listeners did better detecting changes in the 16-kHz octave than in the 8-kHz octave. It is difficult to account for these individual differences in perception. Follow-up regression analyses predicting HFE DLs by age and pure-tone thresholds provided little indication of the variables underlying this listener variability. The average DLs for each listener collapsed across the four 8-kHz stimuli were not correlated with age (Pearson's $r = -0.065$), pure-tone threshold at 8 kHz ($r = 0.014$), nor the average of all collected pure-tone averages ($r = -0.080$). Because DL values could not be obtained for many of the listeners in the 16-kHz conditions, a linear regression analysis was not appropriate. Instead, listeners were separated into those for whom a DL could be obtained on two or more stimuli ($n = 14$) and those for whom a DL was not obtained on at least three of the four stimuli ($n = 16$). A logistic regression analysis was computed to predict membership in these two groupings. Again, age ($\beta = -0.141$) and 8000-Hz threshold ($\beta = -0.105$) were not significantly predictive. On the other hand, the average pure-tone threshold did moderately predict group membership ($\beta = -0.395, p = 0.01$).

B. Voice

The ability to detect changes in HFE level varied with voice type. Specifically, while all listeners could detect a 13-dB attenuation of the 8-kHz octave for voice 2 loud, only 80% of listeners detected this same change in voice 1 loud, despite voice 1 being 5 dB greater in overall SPL. A similar trend was seen for the soft condition. It was also found that listeners detected the 16-kHz octave band level changes more readily for voice 2 than for voice 1. This implies that listeners' ability to detect HFE is dependent upon the timbre of the voice, which will be specific to the individual singer/talker.

These results might not be terribly surprising, given that voice 2 exhibited both higher absolute and relative HFE levels for both soft and loud conditions (see Table I). It is not unlikely that these stable differences are important for perception of singer/talker-specific voice quality. From a vocal production standpoint, however, it is intriguing to speculate on the cause of such a difference in HFE level. Is this difference strictly attributable to biological differences in the vocal tract or vocal fold vibratory characteristics over which an individual has no control? Or, more interestingly, could this difference be attributed to learned behaviors of vocal tract modification or vocal fold phonation? If the latter, then it could be significant for singers or talkers who desire to

change the quality of their voice by controlling the shape of the HFE spectrum. Is it possible to attain a certain vocal tract configuration or a vibration of the vocal folds such that the shape of the HFE spectrum optimizes the aesthetic of the voice? And, if so, what is that optimal spectral shape? Would it differ for different vocalization styles and genres? These questions are the subject of future research.

C. Sound pressure level

For both voices in this experiment, an increase in phonation SPL gave an increase in ability to detect HFE level changes. Approximately 93% of listeners had DLs of 25-dB attenuation or less for the 8-kHz octave for voice 1 loud, compared to only 63% for voice 1 soft. For voice 2, 100% of listeners had DLs of 13 dB or less in the loud condition, while only 60% did in the soft condition, and only 90% detected complete attenuation. A similar trend was seen for the 16-kHz octave, though much less pronounced.

To explore why this is the case, a comparison can be made between DL scores and HFE level (see Tables I and II). Median and minimum DL scores for the 8-kHz octave HFE were all highly correlated with absolute HFE level and peak amplitude. Pearson correlation coefficients between absolute HFE levels and DL scores were $r = -0.97$ and $r = -0.93$ for the median and minimum scores, respectively. Correlation coefficients between peak amplitudes and DL scores were $r = -0.98$ and $r = -0.94$ for the median and minimum scores, respectively. Absolute level and peak amplitudes also correlated well (r values ≥ 0.89) with the percentage of listeners able to detect full attenuation of the 8-kHz octave level.

Conversely, correlations between relative levels and DL scores were less than 0.25. Inspecting the 8-kHz octave levels given in Table I reveals that decreases in phonation intensity had little effect on the relative HFE level, with voice 1 around -40 dB and voice 2 close to -30 dB. Yet detection performance decreased on the task for the softer phonation intensity of each voice. With relative levels remaining fairly consistent, this suggests that the detection of HFE changes is not influenced greatly by the masking of the HFE by the lower frequencies.

These results give two possible predictors of listeners' ability to detect HFE changes in the 8-kHz octave: (1) total HFE energy level and (2) peak amplitude. (These predictors would account for the singer-specific DL score differences discussed in Sec. III B.) Presumably, then, any change of phonation that increases the HFE level would increase a listener's ability to detect it. Such a level change may occur by changing from low- to high-intensity phonation (as in singing), which was observed for both recorded voices in this experiment. Level changes, and peak amplitude changes in particular, may occur by changing vocal tract shape for differing vowels or styles of vocalization, changing phonation type, or introducing consonants (as in running speech). It should be recognized that the SPLs of the stimulus sounds in this experiment were chosen so as to be similar to those of the corresponding live voice, at the same distance as the loudspeaker. In typical music applications such as headphone listening or concerts, voices will often be played at a higher

level, and thus the HFE content may be perceived differently. Furthermore, while head position was not controlled in this study, it is likely that simply turning one's ear toward the voice source will affect the audibility of the HFE.

D. Discussion

It is instructive to compare these results to those obtained from previous experiments using non-speech stimuli. Gunawan and Sen²² reported DLs for five human subjects detecting HFE spectral level changes in clarinet, trumpet, and viola spectra. Stimuli were 1.5 s in duration and were presented monaurally over headphones at a level of approximately 65 dB. Mean DLs reported for attenuating a band centered at 8 kHz with comparable bandwidth to that used in this study were approximately 3, 14, and 17 dB for clarinet, trumpet, and viola, respectively. While direct comparison is difficult due to differences in listening environment and stimulus duration, the mean 8-kHz octave DL for the 65-dB stimulus used in this study (voice 2 soft) was 13.1 dB (for listeners from whom a DL could be obtained, $n = 27$). This suggests that listeners may exhibit greater sensitivity to voice HFE changes than HFE changes in some musical instruments, at least for the 8-kHz octave. On the other hand, whereas Gunawan and Sen give no report of any issues in obtaining DLs for the 16-kHz octave from the five listeners in their study, 16-kHz DLs for voice 2 soft could only be obtained from less than half of the listeners here. It is possible that age may have had an effect as all of their listeners were between the ages of 20 and 26 years. Even so, means reported by Gunawan and Sen for the 16-kHz octave were nearly 25 dB for all three musical instruments, while the mean DL for successful listeners here ($n = 14$) was 13.9 dB.

Moore *et al.*²¹ reported thresholds obtained from three subjects, detecting the change in level of spectral notches centered around 8 kHz in a broadband white noise spectrum. Stimuli were presented monaurally over an insert earphone. The thresholds they report for a notch with comparable bandwidth to that used for the 8-kHz octave in this study were less than 5 dB for all three listeners, suggesting greater sensitivity to white noise HFE changes than voice HFE changes. For every voice stimulus in this study, however, at least one subject achieved a DL of 5 dB or less, and for one stimulus (voice 2 loud), over 50% of listeners achieved DLs of 5 dB or less. It is also noted that the three listeners in the study of Moore *et al.* were given at least 6 h of practice before beginning the data collection process.

As a final note, in an attempt to ascertain what percepts may be affected by HFE changes, participants were asked to report what was different about the "odd" signal and what they listened for to perform the task. Several qualitative descriptors were used to describe the differences in sound, such as "shrillness," "roughness," "thrill," "sharpness/softness," "lighter/darker," "scratchy," "free or clear," "natural," "muted," and "flat." One trained singer described a difference in the "fullness of sound" of the voice, while another described it as a difference in the "ring" of the voice. Given the results from earlier studies relating HFE to sound quality (e.g. Ref. 19), these comments were not too

surprising. What was more intriguing, however, were comments made by some participants (usually with musical training) that were not necessarily related to qualitative aspects of the voice. For example, some described the difference as a pitch change, including one trained singer. A few trained singers also commented that the vowel sounded different, describing it as “more open or closed.” One trained musician and singer described it as a “vowel change,” a difference in “roundness,” and said sometimes it sounded more or less “nasally.” While anecdotal, these descriptions indicate that HFE may affect percepts other than those related strictly to quality and naturalness (e.g., intelligibility of vowels or perception of sung pitch). They further suggest possible methods of changing the HFE spectral contour by vocal tract modification (i.e., opening or closing vowels).

IV. CONCLUSIONS

The current experiment was an exploratory effort to determine human ability to detect changes in HFE in human voice, specifically in sustained vowel sounds (i.e., without fricative consonants and with little aspiration noise). All of the subjects could detect level changes made to HFE, and many could detect relatively small changes in HFE, in samples of isolated sustained vowel sounds. Thus, there is perceptually relevant information in HFE that may be important for determining the aesthetic of singing voice (see also Refs. 17 and 19). Additionally, we have found large individual differences between normal-hearing listeners in their ability to discriminate HFE differences (see standard deviations reported in Table II). It follows that, if HFE is indeed part of sound and voice quality, then one would predict large variability in listener appreciation of these qualities even among normal listeners.

The present results justify future efforts to more closely examine HFE in general, and in applications such as voice synthesis and vocal tract modeling, where little attention has been given to a large portion of this frequency range. Special effort ought to be given to the 8-kHz center-frequency octave band where listeners appear to be most sensitive to HFE changes, though it should be noted that more than half of the subjects could detect changes in the 16-kHz center-frequency octave band as well. The results reported here also confirm the need for better characterization of perceptual information found within the voice spectrum above 5 kHz. This line of research would potentially lead to improved analysis and possible training techniques for voice quality, significant to singers and other professional voice users.

ACKNOWLEDGMENTS

This work was carried out at KTH in Stockholm, with support from the Swedish Research Council, Contract No. 2007-4460, and approval from the Regional Ethical Review Board. This research was funded in part by an initiative for a

Center for Science, Medicine, and the Performing Arts, University of Arizona, directed by the late Tom Hixon. We are grateful to the participating listeners. We are also grateful to Dr. Brian C. J. Moore and the other reviewers for their helpful suggestions.

- ¹P. G. Stelmachowicz, A. L. Pittman, B. M. Hoover, and D. E. Lewis, “Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults,” *J. Acoust. Soc. Am.* **110**, 2183–2190 (2001).
- ²B. C. J. Moore, M. A. Stone, C. Fullgrabe, B. R. Glasberg, and S. Puria, “Spectro-temporal characteristics of speech at high frequencies, and the potential for restoration of audibility to people with mild-to-moderate hearing loss,” *Ear Hear.* **29**, 907–922 (2008).
- ³D. Byrne, H. Dillon, K. Tran, S. Arlinger, K. Wilbraham, R. Cox, B. Hagerman, R. Hetu, J. Kei, C. Lui, J. Kiessling, M. N. Kotby, N. H. A. Nasser, W. A. H. E. Kholly, Y. Nakanishi, H. Oyer, R. Powell, D. Stephens, R. Meredith, T. Sirimanna, G. Tavartkiladze, G. I. Frolenkov, S. Westerman, and C. Ludvigsen, “An international comparison of long-term average speech spectra,” *J. Acoust. Soc. Am.* **96**, 2108–2120 (1994).
- ⁴H. K. Dunn and S. D. White, “Statistical measurements on conversational speech,” *J. Acoust. Soc. Am.* **11**, 278–283 (1940).
- ⁵C. H. Shadle and C. Scully, “An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences,” *J. Phonetics* **23**, 53–66 (1995).
- ⁶A. Jongman, R. Wayland, and S. Wong, “Acoustic characteristics of English fricatives,” *J. Acoust. Soc. Am.* **108**, 1252–1263 (2000).
- ⁷K. Shoji, E. Regenhagen, J. D. Yu, and S. M. Blaugrund, “High-frequency components of normal voice,” *J. Voice* **5**, 29–35 (1991).
- ⁸S. Ternström, “Hi-fi voice: Observations on the distribution of energy in the singing voice spectrum above 5 kHz,” in *Proceedings of Acoustics 2008*, Paris, France (July 2008), pp. 3171–3176.
- ⁹K. Shoji, E. Regenhagen, J. D. Yu, and S. M. Blaugrund, “High-frequency power ratio of breathy voice,” *Laryngoscope* **102**, 267–271 (1992).
- ¹⁰N. V. Naranjo, E. M. Lara, I. M. Rodriguez, and G. C. Garcia, “High-frequency components of normal and dysphonic voices,” *J. Voice* **8**, 157–162 (1994).
- ¹¹D. M. Hartl, S. Hans, J. Vaissiere, and D. F. Brasnu, “Objective acoustic and aerodynamic measures of breathiness in paralytic dysphonia,” *Eur. Arch. Otorhinolaryngol.* **260**, 175–182 (2003).
- ¹²V. Best, S. Carlile, C. Jin, and A. van Schaik, “The role of high-frequencies in speech localization,” *J. Acoust. Soc. Am.* **118**, 353–363 (2005).
- ¹³A. L. Pittman, “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 (2008).
- ¹⁴F. Apoux and S. P. Bacon, “Relative importance of temporal information in various frequency regions for consonant identification in quiet and noise,” *J. Acoust. Soc. Amer.* **116**, 1671–1680 (2004).
- ¹⁵R. P. Lippmann, “Accurate consonant perception without mid-frequency speech energy,” *IEEE Trans. Speech Audio Process* **4**, 66–69 (1996).
- ¹⁶B. C. J. Moore, C. Fullgrabe, and M. A. Stone, “Effect of spatial separation, extended bandwidth, and compression speech on intelligibility in a competing-speech task,” *J. Acoust. Soc. Am.* **128**, 360–371 (2010).
- ¹⁷H. F. Olson, *Elements of Acoustical Engineering* (Van Nostrand, New York, 1957), pp. 587–603.
- ¹⁸Y. Qian and P. Kabal, “Combining equalization and estimation for bandwidth extension of narrowband speech,” in *IEEE Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1 (Montreal, Quebec, 2004), pp. 713–716.
- ¹⁹B. C. J. Moore and C. T. Tan, “Perceived naturalness of spectrally distorted speech and music,” *J. Acoust. Soc. Am.* **114**, 408–419 (2003).
- ²⁰N. F. Viemeister, “Auditory intensity discrimination at high frequencies in the presence of noise,” *Science* **221**, 1206–1208 (1983).
- ²¹B. C. J. Moore, S. R. Oldfield, and G. J. Dooley, “Detection and discrimination of spectral peaks and notches at 1 and 8 kHz,” *J. Acoust. Soc. Am.* **85**, 820–836 (1989).
- ²²D. Gunawan and D. Sen, “Spectral envelope sensitivity of musical instrument sounds,” *J. Acoust. Soc. Am.* **123**, 500–506 (2008).