Perceptual sensitivity to a model of the source spectrum

Marc Garellek, Robin A. Samlan, and Jody KreimanBruce R. Gerratt

Citation: Proc. Mtgs. Acoust. **19**, 060157 (2013); doi: 10.1121/1.4799531
View online: http://dx.doi.org/10.1121/1.4799531
View Table of Contents: http://asa.scitation.org/toc/pma/19/1
Published by the Acoustical Society of America

Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal

Montreal, Canada

2 - 7 June 2013

Speech Communication

Session 4aSCb: Voice and F0 Across Tasks (Poster Session)

4aSCb4. Perceptual sensitivity to a model of the source spectrum

Marc Garellek, Robin A. Samlan, Jody Kreiman* and Bruce R. Gerratt

***Corresponding author's address: Division of Head and Neck Surgery, UCLA School of Medicine, 31-24 Rehab Center, Los Angeles, CA 90095-1794, jkreiman@ucla.edu**

A psychoacoustic model of the source spectrum has been proposed in which four spectral slope parameters describe perception of overall voice quality: H1-H2 (the difference in amplitude between the first and second harmonics), H2-H4, H4-2000 Hz (i.e., the harmonic nearest 2000 Hz), and 2000-5000 Hz. The goals of this study are to evaluate perceptual sensitivity in the mid-to-high frequency range of the model and determine how sensitivity to one parameter varies as a function of another. To determine listener sensitivity to slope changes for each parameter, just-noticeable differences were obtained for series of stimuli based on synthetic copies of one male and one female voice. Twenty listeners completed an adaptive up-down paradigm. To provide a baseline of listener sensitivity to each spectral slope parameter, the synthetic voices were manipulated so that spectral slope varied by 0.5 dB increments for each parameter while other parameters remained constant. We then assessed how listener sensitivity to a given harmonic slope parameter changes when the others covary. These results will help assess the validity of the model and determine what sources of cross-voice variability in spectral configuration are perceptible.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4799531]
Received 17 Jan 2013; published 2 Jun 2013

INTRODUCTION

Valid and reliable characterization of voice quality has been a longstanding challenge. One approach to this problem has been to establish links between acoustic aspects of the source signal and voice quality. A psychoacoustic model of the voice source spectrum has been proposed in which unique acoustic parameters combine to capture the overall voice quality (Kreiman & Gerratt, 2010; Kreiman et al., 2011; Kreiman & Gerratt, 2012, Garellek et al. 2013). This model has four spectral slope parameters: H1-H2 (the amplitude difference between the first and second harmonics), H2-H4 (the amplitude difference between the second and fourth harmonics), H4-2kHz (the amplitude difference between the fourth harmonic and the harmonic nearest 2000 Hz), and 2kHz-5kHz (the amplitude difference between the harmonic nearest 2000 Hz and that nearest 5000 Hz). A schematic of the source spectrum model is shown in Figure 1.

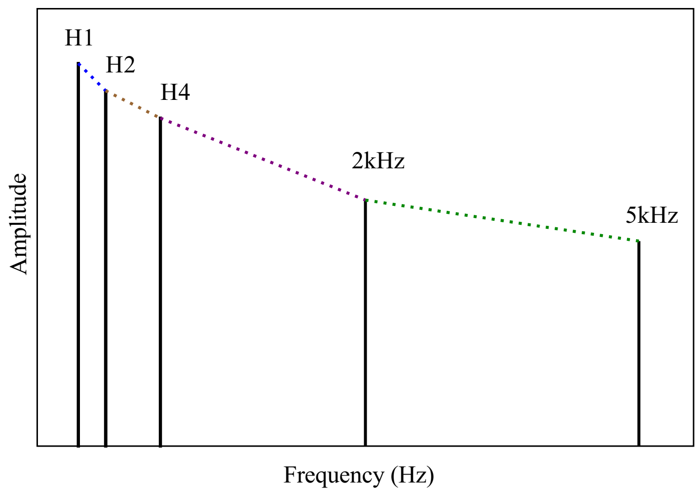

FIGURE 1. Schematic of the proposed source spectrum model.

As an important step in establishing the validity of the model, it must be shown that listeners perceive changes in each of these parameters. Previous work has demonstrated that listeners are sensitive to individual components of the model, particularly H1-H2 (Kreiman & Gerratt 2010) and to higher frequencies (Kreiman & Gerratt 2012). Garellek et al. (2013) also found that changes to both H1-H2 and H2-H4 (but not to H4-2kHz or 2kHz-5kHz) were important for the perception of contrastive breathiness in White Hmong. The ultimate goal of this study is to determine listener sensitivity to all four of the proposed component slopes in the model and to specify interactions among these parameters in determining quality. But in order to test listener sensitivity to these different parameters, we must first know how they vary, and how the slope of one parameter might influence that of another. Thus, the goal of this paper is to determine the variability in H1-H2, H2-H4, H4-2kHz, and 2kHz-5kHz in a large sample of voices, and the extent to which the value of a given parameter can be predicted based on the others. The paper presented at Montréal will include data on perceptual sensitivity and interactions among these parameters.

METHOD

Voices

In this study, we analyzed the source spectral configuration of 144 normal and disordered voices (79 female, 65 male). These voices were 1s-long samples of /a/, which were recorded at a sampling rate of 20 kHz using a Brüel & Kjær ½ in. microphone. The voices were inverse-filtered and copy-synthesized to obtain a synthetic copy of the original voice that was a close perceptual match to the original. For more details of the copy-synthesis process, see Javkin et al. (1987) and Kreiman et al. (2010). In brief, the synthesizer’s sampling rate was fixed at 10 kHz. Parameters describing the harmonic part of the voice source were estimated by inverse filtering a representative cycle of phonation for each voice using the method described by Javkin et al. (1987). The spectral characteristics of the inharmonic part of the source (the noise excitation) were estimated using cepstral-domain analysis similar to that described by de Krom (1993). Spectrally-shaped noise was synthesized by passing white noise through a 100-tap finite impulse response filter fitted to that noise spectrum. To model the fundamental frequency (F0) and amplitude contours, F0 was tracked pulse by pulse on the time domain waveform. Formant frequencies and bandwidths were estimated using autocorrelation linear predictive coding analysis with a window of 25.6 ms. The complete synthesized source was then filtered through the vocal tract model, and all parameters were adjusted until the synthetic copy formed an acceptable match to the original natural voice sample.

Once the voice was copy-synthesized, we obtained the values of each of the spectral slope parameters. We also collected the NHR and F0 of each voice, because these parameters are expected to influence the perception of spectral slope.

RESULTS

The mean value of each spectral slope parameter is provided in Table 1. Female voices generally have steeper spectral slopes below 2000 Hz, but male voices have on average larger values of H4-2kHz.

TABLE 1. Mean values, in dB, for H1-H2, H2-H4, H4-2 kHz, and 2kHz-5kHz for the male and female voices. Standard deviations are provided parenthetically.

Voice	H1-H2	H2-H4	H4-2kHz	2kHz-5kHz
Female	8.93 (4.55)	11.57 (4.99)	18.08 (6.66)	16.20 (9.20)
Male	6.13 (4.11)	8.93 (3.74)	24.58 (6.58)	15.49 (8.23)

Next, we ran linear regression models to investigate how the configuration of one parameter may vary as a function of the other spectral parameters, as well as F0 and noise. For H1-H2, the results of the regression model indicate that only F0 and H2-H4 are significant predictors [$F(6,138) = 6.54$, $p < 0.0001$]. As expected, higher values of F0 are associated with higher values of H1-H2 ($p < 0.001$). On the other hand, higher values of H2-H4 are associated with a decrease in H1-H2 ($p < 0.01$). As shown in Figure 2 (left panel), this effect is driven by the male (Spearman's $r = -0.30$) rather than the female voices (Spearman's $r = -0.03$).

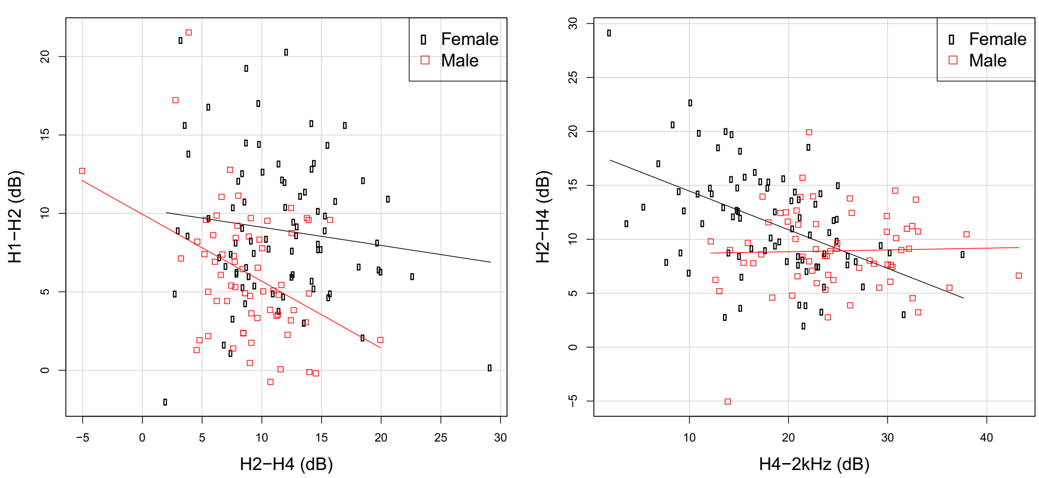

FIGURE 2. (Left) H1-H2 plotted as a function of H2-H4; (right) H2-H4 plotted as a function of H4-2kHz. Red squares are male voices, and black circles are female voices.

For H2-H4, the results of the regression model indicate that F0, H1-H2, and H4-2kHz are significant predictors [$F(6,138) = 10.2$, $p < 0.0001$]. As with H1-H2, higher values of F0 are also associated with higher values of H2-H4 ($p < 0.0001$). Lower values of H1-H2 result in higher values of H2-H4 ($p < 0.01$). This is expected given the results shown in the left panel of Figure 2. Finally, higher values of H4-2kHz are associated with a decrease in H2-H4 ($p < 0.05$). However, as shown in the right panel of Figure 2, the significant effect of H4-2kHz on H2-H4 is driven primary by female voices (cf. Spearman's r = -0.45 for female voices and -0.05 for male voices).

The linear regression model with H4-2kHz as the dependent variable [F (5,138) = 15.58, $p < 0.0001$] reveals that lower values of both H2-H4 and 2kHz-5kHz result in an increase in H4-2kHz ($p < 0.05$). Furthermore, higher F0 is associated with a decrease in the measure ($p < 0.0001$).

Lastly, the results of the regression model predicting 2kHz-5kHz as a function of the other parameters, F0, and noise reveal a significant effect of H4-2kHz and a (marginally) significant effect of noise on the final spectral parameter [F (5,138) = 2.29, $p < 0.05$]. As shown in Figure 3, an increase in 2kHz-5kHz, is associated with a decrease in H4-2kHz ($p < 0.05$) and an increase in noise ($p = 0.05$).

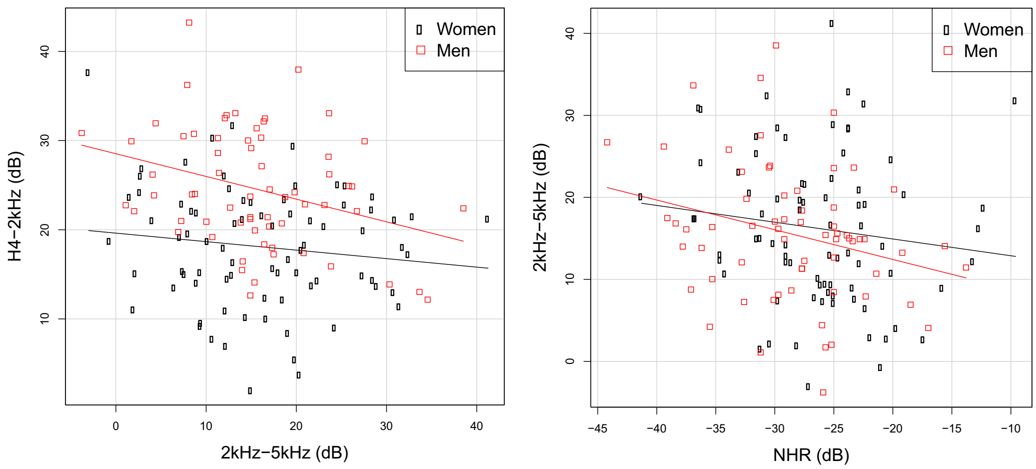

FIGURE 3. (Left) H4-2kHz plotted as a function of 2kHz-5kHz; (right) 2kHz-5kHz plotted as a function of NHR. Higher (i.e., less negative) values of NHR indicate an increase in noise. Red squares are male voices, and black circles are female voices.

DISCUSSION

The results from 144 voices indicate that the slope of one spectral parameter is to some extent dependent on the slope of the adjacent parameter. Thus, H1-H2 covaries with H2-H4 (especially for male voices), whereas the latter covaries with both the former and H4-2kHz (especially for female voices). Interestingly, the relationship is usually negative, such that an increase in slope for one parameter is associated with a decrease in slope for its adjacent parameter(s). Thus, a steep spectral slope is associated with a flatter spectral roll-off in the adjacent parameters, but a sharply falling (i.e., steep) spectral slope between H1 and H2 is not necessarily associated with an overall flatter spectrum in the higher frequencies. Therefore, if H2-H4 is flat, then both H1-H2 and H4-2kHz are likely to be steep. But it is less likely to have two adjacent parameters that are relatively flat, or two that are relatively sharply sloping (see Figure 4).

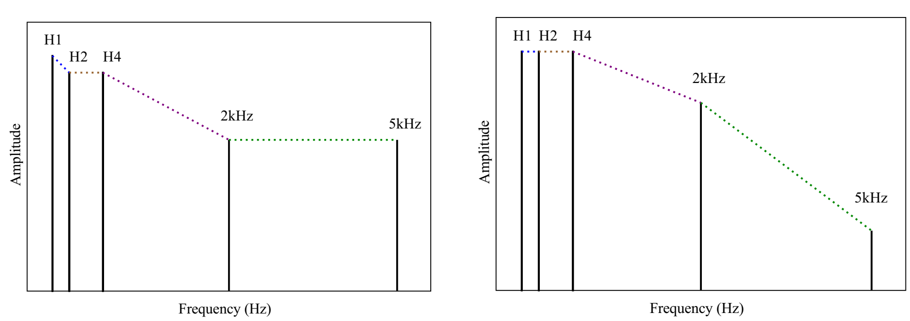

FIGURE 4. (Left) Schematic of a spectrum that is likely to occur; a relatively flat parameter (H2-H4) is flanked by steeper adjacent parameters (H1-H2 and H4-2kHz). (Right) Schematic of a spectrum that is not likely to occur; a relatively flat parameter (H2-H4) is adjacent to another relatively flat parameter (H1-H2), and there are two adjacent steep parameters (H4-2kHz and 2kHz-5kHz).

The spectral slope parameters are also influenced by F0, such that higher F0 values usually involve a sharper spectral slope up to H4, but a flatter slope from H4 to 2kHz. Finally, the noise level is found to influence only the highest spectral parameter; when the noise level is high, 2kHz-5kHz is flatter. It is possible that the noise is interacting with higher-frequency harmonic amplitudes, which suggests that more work is needed to assess the relative contributions of the harmonic vs. inharmonic components in the highest frequencies.

These results suggest that listener sensitivity to a particular parameter should vary with respect to its adjacent parameters. For example, all else being equal, sensitivity to H1-H2 should vary with H2-H4, whereas sensitivity to H2-H4 should vary with both H1-H2 and H4-2kHz. We also expect that F0 should affect sensitivity to the first three parameters, whereas noise should be especially important in perceiving 2kHz-5kHz.

CONCLUSION

In this study, we investigated the source spectral configurations of 144 voices, whose sources were obtained by inverse filtering and copy synthesis. The variability of four spectral parameters and their interactions were investigated. H1-H2 and H2-H4 vary within a smaller range than the higher-frequency parameters H4-2kHz and 2kHz-5kHz. The slope of each parameter is negatively correlated with its adjacent parameters. H1-H2, H2-H4, and H4-2kHz vary with respect to F0, whereas 2kHz-5kHz varies more with the noise level. Future work will determine whether listeners are sensitive to these variations in spectral configuration.

REFERENCES

Garellek, M., Esposito, C. M., Keating, P., and Kreiman, J. **(2013)**. “Voice quality and tone identification in White Hmong,” J. Acoust. Soc. Am. **133**, doi: 10.1121/1.4773259.

Javkin, H. R., Antoñanzas-Barroso, N., and Maddieson, I. **(1987)**. “Integrated software for analysis and synthesis of voice quality,” Behavior Research Methods **42**,1030-1041.

Kreiman, J., and Gerratt, B. R. **(2010)**. “Perceptual sensitivity to the first harmonic amplitude in the voice source,” J. Acoust. Soc. Am. **128**, 2085-2089.

Kreiman, J., Garellek, M., and Esposito, C. M. **(2011)**. “Perceptual importance of the voice source spectrum from H2 to 2kHz,” J. Acoust. Soc. Am. **130**, 2570.

Kreiman, J., and Gerratt, B. R. **(2012)**. “Perceptual interaction of the harmonic source and noise in voice,” J. Acoust. Soc. Am. **131**, 492-500.

de Krom, G. **(1993)**. “A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals,” J. Speech Hear. Res. **36**, 254-266.