DISCRIMINATION OF ASPIRATION NOISE IN BREATHY VOWELS

Ilse Labuschagne *1 and Valter Ciocca †1
[1]School of Audiology and Speech Sciences, The University of British Columbia.

1 Introduction

The presence of aspiration noise (AH) in vowel utterances is one of the key acoustic features that characterizes breathy speech [1, 2]. AH difference limens (DLs) for tokens of the vowel /a/ were measured by Shrivastav and Sapienza [3], and by Kreiman and Gerratt [2], using same-different (SD) and matching tasks, respectively. AH DLs were found to be between 11 and 21 dB in these studies. Shrivastav and Sapienza [3] noticed that DLs depended on the amount of AH present in the standard vowel stimulus: DLs decreased as the amount of AH of the standard stimulus increased. Listeners also exhibited different AH DLs for different stimuli. Since these studies co-varied the amount of AH with changes in other glottal parameters, (e.g., open quotient, spectral tilt, flutter), it is difficult to evaluate the extent to which these changes contributed to the observed AH DLs. The present study investigated DLs for AH by keeping other glottal parameters constant.

There is evidence that changes in the amount of aspiration noise may be more difficult to detect in stimuli with larger amounts of high frequency harmonic energy [3, 4]. Therefore, listeners' sensitivity to the amount of AH in vowels could be affected by spectral differences. In order to test this hypothesis, the present study compared AH DLs for two vowels (/æ/ and /i/).

2 Method

2.1 Stimuli

The implementation of a parallel Klatt synthesizer [5] in the Praat software [6] was used to synthesize six-formant utterances of /æ/ and /i/. These stimuli were modelled after the vowels produced by three male speakers (S08, S30, S44) selected from the Hillenbrand vowel database [7]. The database contains recordings in an h-vowel-d format. The vowel portions of these six utterances (two vowels by three speakers) were synthesized at a sampling rate of 44.1 kHz. This sampling rate ensured that high frequency noise components were present in the stimuli, unlike previous studies which limited the frequency range of their stimuli to 5.5 kHz. Vowels were synthesized to resemble natural utterances by using formant and fundamental frequency estimates at eight time points within each vowel. The amplitude of voicing was set to reproduce the intensity contour of the original excised vowels, and ranged from 67 to 73 dB (average = 70 dB). Glottal waveform parameters for all stimuli included an open quotient of 0.4; all other glottal parameters were set to default values. AH was set to 35 dB for the standard stimuli, with two test conditions differing in AH from the standard stimuli by 2 dB or 4 dB (AH = 37 or 39 dB). Vowel duration ranged from 276 to 302 ms (average = 290 ms).

2.2 Participants

Eight naïve listeners between the ages of 21 and 27 years were screened to ensure that their hearing thresholds were within normal limits (15 dB HL at octave frequencies from 250 to 4000 Hz). Listeners were remunerated for their participation.

2.3 Procedure

First, listeners completed a short training session in order to familiarize themselves with the type of stimuli used in the study. In a subsequent screening session, listeners completed a two-alternative forced choice (2AFC) task in which 16 stimulus pairs with large AH differences (AH = 0 dB for the standard stimulus; AH = 33 dB or 38 dB for the comparison stimuli) were presented. Stimuli were presented binaurally through headphones at a level of 70 dBA in a soundproof booth. Within each trial, listeners heard a sequence of two vowels with different levels of AH. Listeners were asked to select which of the two stimuli sounded "breathier". Feedback was provided after each response. A minimum test score of 12 out of 16 correct identifications was required for a listener to be included in the main experiment. All participants reached this criterion.

The experiment consisted of two sessions, blocked by vowel quality (first /æ/, and then /i/ stimuli). The procedure was similar to that of the familiarization task, except that the two intervals consisted of a standard stimulus with a 35-dB AH level, and one of two comparison stimuli (ΔAH either +2 dB or +4 dB relative to the standard). These differences were chosen on the basis of the results of a pilot study. During each of the two sessions, listeners were presented with 120 trial pairs (3 speakers × 2 AH levels × 2 presentation orders × 10 repetitions). The experiment, including the familiarization and screening sessions, lasted about 1 hour.

3 Results

3.1 Perceptual data

Responses were converted into d' scores [8]. As expected, sensitivity was consistently higher for the 4-dB than for the 2-dB ΔAH stimuli. Figure 1 shows the differences in d' score averages across listeners for the two AH levels and the two vowels.

A three-way ANOVA with AH level, vowel quality and speaker as repeated factors was performed on the d' scores. The main effect of AH level showed a significant difference ($F(1, 86) = 20.8$, $p < 0.01$), with an average d' of 0.56 for the 2-dB and 1.28 for the 4-dB conditions. Main effects of

* ilse.labuschagne@alumni.ubc.ca
† vciocca@mail.ubc.ca

speaker and vowel quality, and interaction effects did not reach significance.

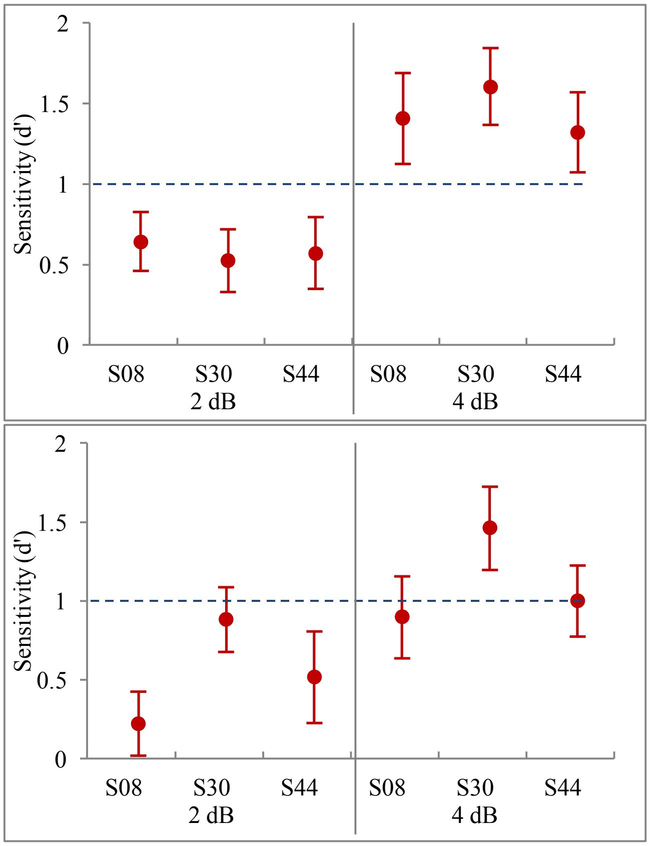

Figure 1: Averages with standard error bars of the sensitivity scores across listeners for the three /æ/ and /i/ vowels -panels (a) and (b), respectively). The empirical threshold level (d' = 1) is indicated with a dotted line.

Although the main effect of vowel quality was not significant, some of the evidence supports the idea that listeners may be more sensitive to AH differences in /æ/ vowels compared to /i/ vowels. For instance, in the 4-dB condition only 4 of the 24 d' scores calculated for the /æ/ vowel (3 vowels × 8 listeners) were smaller than 1, while for the /i/ vowels more than half (14 out of 24) of the d' scores were smaller than 1. While the average d' values for the two vowels in the 2-dB condition were similar (0.58 and 0.54 for /æ/ and /i/, respectively), the difference in the 4-dB condition was larger (1.44 and 1.12 for /æ/ and /i/, respectively). A one-way ANOVA showed that the difference in d' scores between vowels for the 4-dB condition was not significant ($F(1, 46) = 3.02$, $p = 0.089$).

4 Discussion and conclusions

The present results indicate that the AH DLs for the vowels /æ/ and /i/ are between 2 and 4 dB. These values are substantially smaller than those reported in previous studies [2, 3]. Possible reasons for this difference include the differences in the tasks used to measure thresholds, differences in vowel quality, and different definitions of AH DLs. For example, it has been found that an SD task such as the one used by Shrivastav and Sapienza [3] results in poorer performance than a 2AFC task [8, p. 228]. Furthermore, Shrivastav and Sapienza defined the AH DL as the level at which listeners were able to correctly identify a difference in AH in 70.7% of the stimuli. This corresponds to d' = 1 for yes-no tasks, but to d' = 2 in SD tasks. In the current experiment, ΔAH would have to increase above 4 dB to reach a d' value of 2. It is also possible that the higher sampling rate of the stimuli used in the present study, and the omission of low-pass filters in the form of the spectral tilt synthesis parameters, resulted in higher noise energy at higher frequencies; higher noise energy may have resulted in improved discrimination. Differences in vowel quality between the present studies and previous studies may also have contributed to the observed differences in AH DLs. Future studies using the same procedure as the present study with /a/ stimuli will shed light on the contribution of this factor to AH DLs.

The difference in AH DLs between /æ/ and /i/ did not reach statistical significance, probably due to the small sample size (n = 8). It is also possible that the order of presentation of the vowel sessions (/æ/ was always presented first) produced a learning effect that inflated d' scores for the /i/ vowels. Differences in d' scores observed for the two vowels may be due to higher amounts of high frequency formant energy in /i/ compared to /æ/ vowels [3, 4]. Systematic changes to higher formant amplitudes and/or bandwidths, and acoustic analysis measurements may provide insight into the relationship between AH DLs and the acoustic properties of breathy vowels differing in vowel quality.

Acknowledgments

This research was supported by a start-up fund provided to Valter Ciocca by the Faculty of Medicine, University of British Columbia.

References

[1] D. H. Klatt and L. C. Klatt. Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.*, 87(2), 820–857, 1990.

[2] J. Kreiman, and B. R. Gerratt. Perception of aperiodicity in pathological voice. *J. Acoust. Soc. Am.*, 117(4):2201–2211, 2005.

[3] R. Shrivastav, and C. M. Sapienza, Some difference limens for the perception of breathiness. *J. Acoust. Soc. Am,* 120(1):416-423, 2006.

[4] B. R. Gerrat and J. Kreiman. Measuring vocal quality with speech synthesis. *J. Acoust. Soc. Am.*, 110(5):2560-2566, 2001.

[5] D. H. Klatt. Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.*, 67(3):971–995, 1980.

[6] Boersma, P., and Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 5.4.19, retrieved 22 September 2015 from http://www.praat.org/, 2015.

[7] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.*, 97(5):3099-3111, 1995.

[8] N. A. Macmillan, and C. D. Creelman. Detection Theory: A user's guide (2nd ed.). Mahwah, New Jersey: Lawrence Earlbaum Associates, 2005.