Breathiness differences in male and female speech. Is H1-H2 an appropriate measure?

Adrian P. Simpson
Institute of German Linguistics, University of Jena, Jena, Germany

Abstract

A well-established difference between male and female voices, at least in an Anglo-Saxon context, is the greater degree of breathy voice used by women. The acoustic measure that has most commonly been used to validate this difference are the relative strengths of the first and second harmonics, H1-H2. This paper suggests that sex-specific differences in harmonic spacing combined with the high likelihood of nasality being present in the vocalic portions make the use of H1-H2 an unreliable measure in establishing sex-specific differences in breathiness.

Introduction

One aspect of male and female speech that has attracted a good deal of interest are differences in voice quality, in particular, breathy voice. Sex-specific differences in breathy voice have been examined from different perspectives. Henton and Bladon (1985) examine behavioural differences, whereas in the model proposed by Titze (1989), differences in vocal fold dimensions predict a constant dc flow during female voicing. In an attempt to improve the quality of female speech synthesis, Klatt and Klatt (1990) use a variety of methods to analyse the amount of aspiration noise in the male and female source.

A variety of methods have been proposed to measure breathiness:

- Relative lowering of fundamental frequency (Pandit, 1957).
- Presence of noise in the upper spectrum (Pandit, 1957; Ladefoged and Antoñanzas Barroso, 1985; Klatt and Klatt, 1990).
- Presence of tracheal poles/zeroes (Klatt and Klatt, 1990).
- Relationship between the strength of the first harmonic H1 and amplitude of the first formant A1 (Fischer-Jørgensen, 1967; Ladefoged, 1983).
- Relationship between the strength of the first and second harmonic, H1-H2 (Fischer-Jørgensen, 1967; Henton and Bladon, 1985; Huffman, 1987; Ladefoged and Antoñanzas Barroso, 1985; Klatt and Klatt, 1990).

It is the last of these measures that has most commonly been applied to measuring sex-specific voice quality differences.

In this paper I set out to show that relating the strength of the first harmonic to other spectral measures as a way of comparing breathiness between male and female speakers is unreliable. The line of argumentation is as follows. The frequency of the first nasal formant (F_{N1}) can be estimated to in the region of 200–350 Hz for both male and female speakers (Stevens et al., 1985; Maeda, 1993). At a typical male fundamental frequency of 120 Hz this will be expressed in an enhancement of the second and third harmonics. By contrast, at a typical female fundamental frequency of over 200 Hz it may well be the first harmonic that is more affected by F_{N1}. Comparison of H1 and H2 as a measure of breathiness has to be carried out on opener vowel qualities in order to minimise the effect of the first oral resonance, F1. Lowering of the soft palate is known to increase with the degree of vowel openness. Although the ratio of the opening into the oral cavity and that into the nasal port is crucial for the perception of nasality (Laver, 1980), acoustic correlates of nasality are present when the velopharyngeal port is open. It cannot be excluded, then, that any attempt to compare the male and female correlates of breathiness in terms of the first harmonic might be confounded by the sex-specific effects of F_{N1} on the first two harmonics, in particular a relative strengthening of the first female and the second male harmonic. Establishing that female voices are breathier than male voices using the relative intensities of the first two harmonics might then be a self-fulfilling prophecy.

Data

The data used in this study are drawn from two sources. The first data set was collected as part of a study comparing nasometry and spectrography in a clinical setting (Benkenstein, 2007). Seven male and fifteen female speakers were recorded producing word lists, short texts and the map task using the Kay Elemetrics Nasometer 6200. This method uses two microphones separated by an attenuating plate (20 dB separation) that capture the acoustic output of the nose and the mouth. A nasalance measure is calculated from the relative intensity of the two signals following bandpass filtering of both at ca. 500 Hz. The present study is interested only in a small selection of repeated disyllabic oral and nasal words from this corpus so only the combined and unfiltered oral and nasal signals will be used. The second data set used is the publicly available *Kiel Corpus of Read Speech* (IPDS, 1994), which contains data from 50 German speakers (25 female and 25 male) reading collections of sentences and short texts. Spectral analysis of consonant-vowel-consonant sequences analogous to those from the first dataset was carried out to ensure that the signals from the first dataset had not been adversely affected by the relatively complex recording setup involved with the nasometer together with the subsequent addition of the oral and nasal signals.

Sex-specific harmonic expression of nasality

The rest of this paper will concern itself with demonstrating sex-specific differences in the harmonic expression of nasality. In particular, I will show how F_{N1} is responsible for a greater enhancement of the second male and the first (i.e. fundamental) female harmonic. Further, I will show that F_{N1} is expressed in precisely those contexts where one would want to measure H1-H2. In order to show how systematic the patterns across different speakers are, individual DFT spectra from the same eight (four male, four female) speakers will be used. It is important to emphasise that there is nothing special about this subset of speakers – any of the 22 speakers could have been used to illustrate the same patterns.

We begin with spectral differences found within nasals, i.e. uncontroversial cases of nasality. Figure 1 contains female and male spectra taken at the centre of the alveolar nasal in the word *mahne* (“warn”). From the strength of the lower harmonics, F_{N1} for both the male and female speakers is around 200–300 Hz, which is commensurate with sweep-tone measurements of nasals taken from Båvegård et al. (1993). Spectrally, this is expressed as a strengthening primarily of the first female and the second male harmonic.

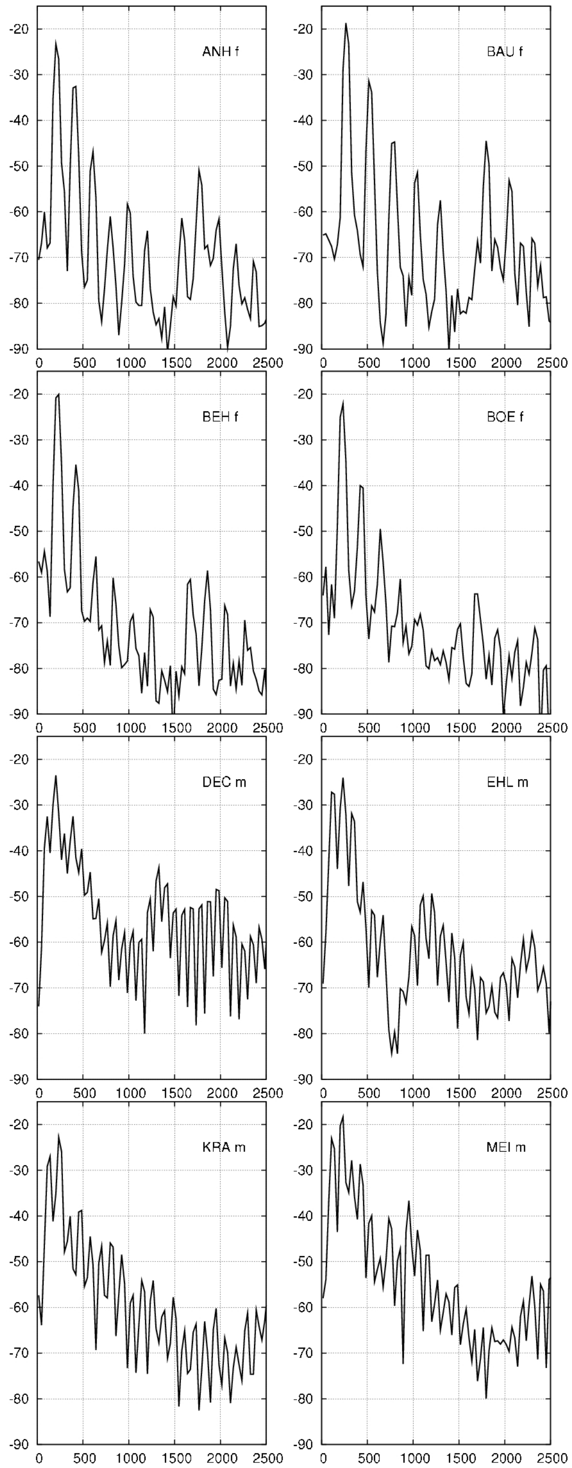

Figure 1: DFT spectra calculated at midpoint of [n] in *mahne* for four female (top) and four male speakers.

It is reasonable to assume that the velum will be lowered throughout the production of the word *mahne.* Figure 2 shows spectra taken from the same word tokens as shown in Figure 1, this time calculated at the midpoint of the long open vowel in the first syllable. While there is a good deal of interindividual variation in the position and amplitude of F1 and F2, due to qualitative differences as well as to the amount of vowel nasalisation present, there is clear spectral evidence of F_{N1}, again to be found in the increased intensity of the second male and the first female harmonic.

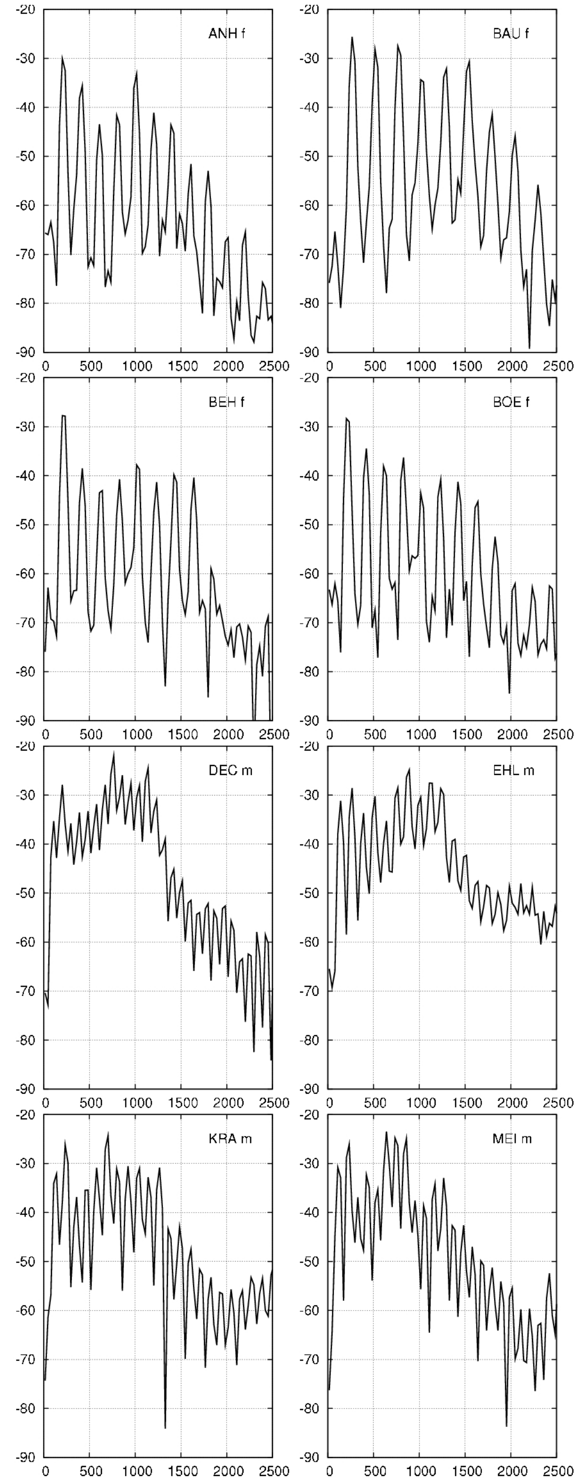

Figure 2: DFT spectra calculated at midpoint of the open vowel in the first syllable of *mahne* (same speakers).

Neither nasals nor vocalic portions in a phonologically nasal environment would be chosen as suitable contexts for measuring H1-H2. However, they do give us a clear indication of the spectral correlates of both vocalic and consonantal nasality, and in particular we were able to establish systematic sex-specific harmonic differences.

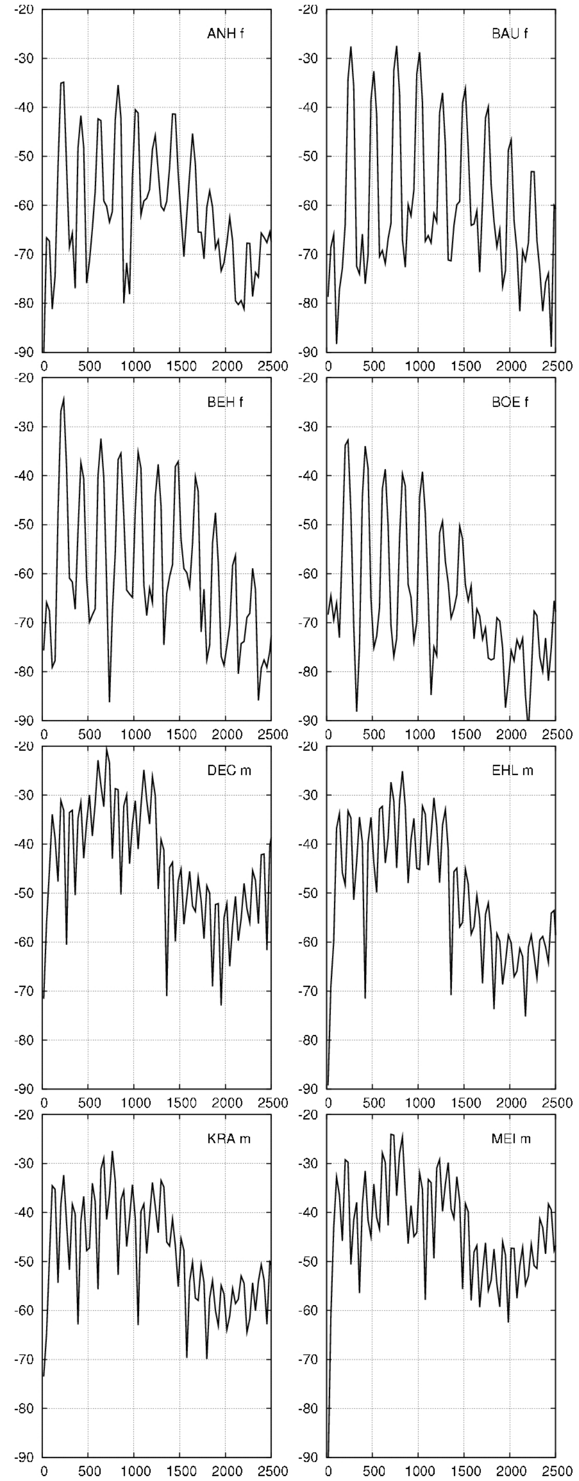

Figure 3: DFT spectra calculated at midpoint of the open vowel in the first syllable of *Pate* (same speakers).

Let us now turn to the type of context where one would want to measure H1-H2. The first syllable of the word *Pate* ("godfather") contains a vowel in a phonologically oral context with an open quality which maximises F1 and hence minimises its influence on the lower harmonics. Figure 3 shows DFT spectra calculated at the midpoint of the long open vowel in tokens of the word *Pate* ("godfather") for the same set of speakers. In contrast to the categorically identical tokens in the nasalised environment in Figure 2, it is somewhat easier to estimate F1 and F2. However, the most striking similarity with the spectra in Figure 2 is evidence of a resonance in the region 200–300 Hz, suggesting that here too, nasality is present, and as before is marked primarily by a strengthened female fundamental and a prominent second male harmonic.

Discussion

Increased spectral tilt is a reliable acoustic indication of breathy voice. However, as I have attempted to show in this paper, using the strength of the first harmonic, without taking into consideration the possibility that nasality may also be acoustically present, makes it an inappropriate point of reference when studying sex-specific differences. Indeed, data such as those shown in Figure 3 could have been used to show that in German, too, female speakers are more breathy than males. The female spectral tilt measured using H1-H2 is significantly steeper than that of the males.

So, what of other studies? It is hard to make direct claims about other studies in which individual spectra are not available. However, it is perhaps significant that Klatt and Klatt (1990) first added 10 dB to each H1 value before calculating the difference to H2, ensuring that the H1-H2 difference is almost always positive (Klatt and Klatt, 1990: 829; see also e.g. Trittin and de Santos y Lleó, 1995 for Spanish). The male average calculated at the midpoint of the vowels in reiterant [ʔɑ] and [hɑ] syllables is 6.2 dB. This is not only significantly less than the female average of 11.9 dB, but also indicates that the male H2 in the original spectra is consistently stronger, once the 10 dB are subtracted again.

I have not set out to show in this paper that female speakers are less breathy than male speakers. I am also not claiming that H1-H2 is an unreliable measure of intraindividual differences in breathiness, as it has been used in a linguistic context (Bickley, 1982). However, it seems that the method has been transferred from the study of intraindividual voice quality differences to an interindividual context without considering the implications of other acoustic factors that confound its validity.

References

Benkenstein R. (2007) Vergleich objektiver Verfahren zur Untersuchung der Nasalität im Deutschen. Peter Lang, Frankfurt.

Båvegård M., Fant G., Gauffin J. and Liljencrants J. (1993) Vocal tract sweeptone data and model simulations of vowels, laterals and nasals. STL-QPSR 34, 43– 76.

Fischer-Jørgensen E. (1967) Phonetic analysis of breathy (murmured) vowels in Gujarati. Indian Linguistics, 71–139.

Henton C. G. and Bladon R. A. W. (1985) Breathiness in normal female speech: Inefficiency versus desirability. Language and Communication 5, 221–227.

Huffman M. K. (1987) Measures of phonation type in Hmong. J. Acoust. Soc. Amer. 81, 495–504.

IPDS (1994) The Kiel Corpus of Read Speech. Vol. 1, CD-ROM#1. Institut für Phonetik und digitale Sprachverarbeitung, Kiel.

Klatt D. H. and Klatt L. C. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. J. Acoust. Soc. Amer. 87, 820–857.

Ladefoged P. (1983) The linguistic use of different phonation types. In: Bless D. and Abbs J. (eds) Vocal Fold Physiology: Contemporary Research and Clinical Issues, 351–360. San Diego: College Hill.

Ladefoged P. and Antoñanzas Barroso N. (1985) Computer measures of breathy phonation. Working Papers in Phonetics, UCLA 61, 79–86.

Laver J. (1980) The phonetic description of voice quality. Cambridge: Cambridge University Press.

Maeda S. (1993) Acoustics of vowel nasalization and articulatory shifts in French nasalized vowels. In: Huffman M. K. and Krakow R. A. (eds) Nasals, nasalization, and the velum, 147–167. San Diego: Academic Press.

Pandit P. B. (1957) Nasalization, aspiration and murmur in Gujarati. Indian Linguistics 17, 165–172.

Stevens K. N., Fant G. and Hawkins S. (1987) Some acoustical and perceptual correlates of nasal vowels. In Channon R. and Shockey L. (eds) In Honor of Ilse Lehiste: Ilse Lehiste Püdendusteos, 241–254. Dordrecht: Foris.

Titze I. R. (1989) Physiologic and acoustic differences between male and female voices. J. Acoust. Soc. Amer. 85, 1699–1707.

Trittin P. J. and de Santos y Lleó A. (1995) Voice quality analysis of male and female Spanish speakers. Speech Communication 16, 359–368.