# Normophonic Breathiness in Czech and Danish: Are Females Breathier Than Males?

**Article** · January 2020

DOI: 10.1016/j.jvoice.2019.10.019



**4 authors:**

Míša Hejná
Aarhus University


Pavel Šturm
Charles University in Prague


Lea Tyleckova
Charles University in Prague


Tomas Boril
Charles University in Prague

# Normophonic Breathiness in Czech and Danish: Are Females Breathier Than Males?

†Míša Hejná, *Pavel Šturm, *Lea Tylečková, and *Tomáš Bořil, †Aarhus C, Denmark, and *Praha 1, Czech Republic

**SUMMARY:** The present study compares the voice quality of female and male speech in two languages: Czech, a Slavic language, and Danish, a Germanic language. For both languages, the results based on a total of 120 vocally healthy speakers are in line with the claim that females are universally breathier than males. This was supported by the Cepstral Peak Prominence (CPP) and H1*-H2* measures, which are generally known as the most robust correlates of breathiness, and also by the H1*-A3* measure. However, the sex distinction was unsupported or even contradictory when using some other measures suggested to reflect breathiness, which provides an incentive to insist on employing a number of acoustic measures in future voice research. The perceptual component of the study nevertheless suggests that these contradictory findings are due to differences in perceived roughness rather than breathiness, and that CPP and H1*-H2* do reflect breathiness differences, and CPP in particular. We therefore conclude that it is indeed the case that female speakers are breathier than male speakers. Finally, in terms of the two robust measures (CPP and H1*-H2*), no language-specific differences in the magnitude of the effect of sex on breathiness were found.

**Key Words:** Breathiness−Phonation−Sex−Czech−Danish.

## INTRODUCTION

The current study focuses on a phonatory setting known as *breathiness* in a population of normophonic speakers. As sociophoneticians have pointed out, understanding language-specific characteristics of speech is important for speech pathologists as a baseline of what should be considered medically normal as opposed to abnormal speech patterns ([1]: p. 736). In nonpathological voices at least, breathy phonation can show variation that goes hand in hand with various socio-pragmatic functions. For instance, an increase in breathiness can cue the end of a turn,[2,3] identity-related aspects (eg, [4,5]), and interpersonal relationships and emotions (eg, [6,7,4]). Breathiness has also been identified as one of the phenomena sensitive to sex differences, with females being breathier than males (eg, [8]: pp. 229−230). At least two of these factors are potentially of interest to clinicians, and they are certainly of interest to sociolinguists: sex/gender and region. Presumably, these variables are the two demographic characteristics of speakers that clinicians deal with on a regular basis. It may be important to know what to expect in terms of breathiness levels of a female as opposed to a male speaker without a vocal pathology, and what to expect of speakers of different accents or languages. For instance, although the processes of TH-fronting (eg, /θ/ in <u>th</u>*ought* becomes /f/; /θɔːt/ → [fɔːt]) and R-labialisation (eg, *really* sounds more like *weally*; [ɹ] → [ʋ]) in British English have been approached as speech impediments ([9]: p. 120;[10]: p. 53;[11]: p. 492), sociolinguists have demonstrated their increasing presence in normal populations and identified them as sound changes in progress, with increasing numbers of cases found in incrementally younger generations (eg, [12−14]). To the linguist working within the fields of sociolinguistics and language variation and change, it is of interest to identify the constraints on linguistic variation and establish the nature of these constraints (eg, physiological as opposed to social, or a combination of the two). To the clinicians, understanding what the normal variation in speech is can inform their decision-making when working with their clients.

This study aims to describe variation in breathiness in the nonpathological speech of Czech and Danish speakers. These languages are targeted for two reasons, bearing the audiences of clinicians and linguists in mind. Firstly, to the best of our knowledge, values for the variation found in nonpathologically breathy voices have not been established for these two languages. The second motivation is driven by the limited evidence related to our understanding of breathiness variation in sociolinguistic research. We therefore identify such gaps first.

### Sex-related differences in breathiness

Nonpathological breathiness has been found to be higher in females than in males in a number of studies and languages: eg, Dutch,[15] British English (RP and a northern accent[16]), several varieties of American English,[17] French,[18] Japanese[19]: p. 47 compared with pp. 52, 58−60,62,121, Korean[20]: p. 811, and Spanish[21] (see also[22,23]). The pattern has been observed frequently enough to lead to speculations that females may be breathier than males due to physiological reasons[8]: pp. 229−230; [22]: p. 623. The physiological correlate is a more frequent presence of glottal gaps in female speech (eg,[24]: p. 2673; [25−27]), although to the best of our

Accepted for publication October 31, 2019.

Funding: The work was supported by the European Regional Development Fund-Project "Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World" (No. CZ.02.1.01/0.0/0.0/16_019/0000734). The third author was supported from the project "Language and its research tools," solved at Charles University from Specific university research in 2019.

From the †Department of English, Aarhus University, Aarhus C, Denmark; and the *Institute of Phonetics, Charles University, Praha 1, Czech Republic.

Address correspondence and reprint requests to Pavel Šturm, Institute of Phonetics, Charles University, nám. Jana Palacha 2, Praha 1, 116 38, Czech Republic. E-mail: pavel.sturm@ff.cuni.cz

Journal of Voice, Vol. ■■, No. ■■, pp. ■■−■■

0892-1997



https://doi.org/10.1016/j.jvoice.2019.10.019

knowledge it is not obvious whether these gaps are conditioned physiologically and/or socially, since it is not clear whether the higher presence of glottal gaps in healthy female population presents a learnt articulatory pattern rather than a result of morphological dimorphism. Helgason offers two explanations: women may on the one hand spread their vocal fold processes in order to "de-emphasize the fact that they radiate at higher frequencies than men do" (Titze, 1989 in[8]: p. 229) or, on the other hand, because of differences in vocal fold size ([8]: p. 229). Henton and Bladon ([16]: p. 226) speculate that there may be "a physiological basis for the association between breathiness and arousal. An accompaniment to the release of sex hormones by the hypothalamus is the release of other secretional lubrication to the body as a whole, and the larynx does not escape this effect. If the larynx receives extra lubrication, then this may inhibit the ability of the vocal folds to adduct fully, resulting in an inefficient phonation and producing breathy voice." Hejná[28] suggests that increased breathiness in females might be associated with the hormonal changes connected with the menstrual cycle, which have been shown to be correlated with physiological changes in the laryngeal structures.[29] Although these proposals have been put forward, it is not known what the physiological causes behind females producing more glottal gaps than males should be.

This study aims to contribute to our understanding of the sociolinguistic question of whether females tend to be universally breathier than males, similarly to what has been shown for $f_0$. Ideally, a larger-scale cross-linguistic comparison should be sought; however, analyzing two languages at a time is more manageable than including as many of the world's languages as possible. We focus on Czech and Danish because the sex-related information on breathiness is missing in the literature. Thus, one of the languages chosen for the analysis is a Slavic language (Czech), as this particular language group is understudied regarding normal phonatory properties of their speakers. Research looking into the question of sex-conditioned differences in breathiness exists for three Germanic languages: English, German, and Swedish.[30] As two of these are West Germanic languages, we selected Danish as our Germanic language of comparison to also explore whether the same sex-related difference in breathiness occurs in North Germanic more generally. This rationale is naturally of greater interest to linguists than to clinicians, who might not be concerned with different language backgrounds within their practice. However, clinicians might still be interested in a description of nonpathological voices: if females are generally breathier than males, the evaluation of voices in terms of a breathiness severity scale (from normal to severe pathology) will accommodate a different level of "normal" breathiness in females than in males, depending on the magnitude of such effects.

Based on the above findings, two hypotheses are specifically tested, pertaining to the research questions of whether female speakers are breathier than male speakers, and whether there is any language-specific variation in the levels of breathiness:

- H1: Female speakers will produce breathier phonation than male speakers. This will be tested primarily via an acoustic analysis employing a range of commonly-used measures of breathiness (see below).
- H2: The two languages will show different magnitudes of the sex-related differences in breathiness. This would suggest a sociocultural conditioning, whereas the absence of such an effect would be inconclusive about physiological vs social constraints.

Testing these hypotheses requires tackling another question: which measures predict sex differences in breathiness robustly (if any)? In order to address this question, we need to understand the nature of the measures available to identify variation in breathiness, and that of acoustic measures in particular. A range of measures have been proposed, targeting different aspects of voice quality. We predict that some of the commonly used measures will capture sex-related variation more strongly than others, and that the measures will show variation in how closely they correlate with one another. Moreover, it is also important to ask how these measures map onto a perceptual evaluation of breathiness for the reasons described in the next section, where we provide more detail on the issue.

## Physiological, acoustic, and perceptual correlates of breathiness

Breathiness can be described as a voice of soft quality, which reflects the holistic, perceptual nature of the phenomenon, with breathiness being the listener's indirect response to some acoustic (and articulatory) qualities of the voice. Some voices are perceived as breathier than others, but the specific acoustic correlates may differ (as discussed further below), and a variety of labels have been used in describing the different aspects of a breathy voice (eg, soft voice, breathy voice, whisper, whispery voice, and murmur; see[31]). It is especially important to differentiate between breathiness as defined here and roughness, since they might share a subset of similar acoustic correlates.[32,33] It is not surprising that establishing the most robust correlates of breathiness is not as straightforward as establishing correlates of some other types of variation found in speech. In articulatory terms, similarly to other phonation types, breathiness is a complex phenomenon. It involves a periodic vibration of the vocal folds and, at the same time, an abducted state of the glottis (eg,[31]). The resulting vibration is rather lax in comparison to other phonatory settings, as the vocal folds are not fully in contact ([34]: pp. 31 and 132; [35]: p. 418; [36]: p. 175), enabling more air to escape through the glottis (eg,[37]: p. 367; [38]: p. 385; [34]: p. 132; [35]: p. 418; [39]: pp. 87 and 447). As already mentioned, breathiness is also associated with the presence of glottal gaps, which again enable more air to pass through the glottis. Acoustically, breathiness manifests itself in multilateral ways. Firstly, the sound wave is associated with reduction in complexity, approaching a quasisinusoidal waveform shape.[40] In addition, the spectrographic information shows the loss of a clear formant structure, in particular above the second (F2) and the third (F3) formants ([39]: p. 448). Crucially, however, there is also an increased amount of glottal friction in these energy-attenuated frequencies (eg,[42]: p. 256; [41,43,44]; [45]: p. 4). It is not entirely clear which specific acoustic measures are the best to capture differences in breathiness (eg,[33]) and to what extent the most commonly used measures are correlated. When approaching phonatory variation, researchers typically use more than one measure (eg,[46,47] for clinical studies; and[48,49,50] for linguistic studies), and the measures employed differ researcher by researcher. The debates about which measures are the most appropriate are still ongoing (eg,[46,47]; [51]).

Furthermore, when considering which of the proposed measures to employ, it is useful to distinguish between clinical research or evaluation on the one hand, which seems to focus on the severity of vocal pathology, and linguistic investigation on the other. Techniques and standardized rating protocols used by voice clinicians are not necessarily appropriate for investigating and understanding vocal communication and specifically variation in voice quality within normophonic voices which is conditioned socially or which is due to sexual dimorphism. For instance, clinicians may rely on perceptual assessment procedures such as the Grade, Roughness, Breathiness, Asthenia and Strain scale (GRBAS[52]) and, more recently, on global tools such as the Acoustic Breathiness Index (ABI; eg,[53]). ABI has been shown to reliably predict pathological breathiness and offers the advantage of not being subject to the factors that human assessors may be, such as fatigue.[53] The reason why linguists may not turn to GRBAS (and similar assessment frameworks) is because they tend to be primarily interested in normophonic variation and because the magnitude of the differences of interest may be fairly small. However, considering what the results of tools such as ABI suggest, relying on one specific measure should be avoided. This is because different measures reflect different aspects of phonation and may also reflect different aspects of a single phonatory setting, such as breathiness. Including multiple measures should therefore be seen as beneficial, also for reasons of cross-disciplinary and better cross-linguistic comparisons. Reporting results relying on multiple individual measures may be even more useful if the researchers tap into the relationship between these measures. A number of studies look into the correlations between individual acoustic measures and the auditory percepts (eg,[46]), something which is not done in linguistics on a general basis. To bridge the gap between clinical phonetics and sociophonetics of phonation, we therefore use individual acoustic measures (as customary in linguistics) but combine these with intermeasure analyses as well as perceptual assessment (as customary in clinical work).

The acoustic measures used to quantify breathiness can be classified according to what aspects of the acoustic signal they capture: (1) periodicity of the signal/amount of noise in the signal, (2) spectral shape, and (3) short-term frequency/amplitude perturbation.[54] We provide a summary of the measures that are regularly used for the purposes of differentiating breathiness from other voice qualities or accounting for variation within breathiness levels by clinicians and linguists.

- *Cepstral Peak Prominence (CPP)* and *smoothed CPP* are considered to be rather robust acoustic measures of overall voice quality, especially of breathiness (eg,[55−57]), and of overall severity of dysphonia in both sustained vowel samples and continuous speech. Moreover, CPP is also the strongest contributor to predicting phonatory types in multidimensional measures, such as Acoustic Voice Quality Index[58] and Cepstral and Spectral Index of Dysphonia.[59] Additionally, CPP is included in ASHA's "Recommended Protocol for Instrumental Assessment of Voice" as the preferred acoustic measure to assess both the amount of noise, and the overall vocal quality.[60] The CPP measure represents the difference in amplitude level between the first "rahmonic" (anagram of harmonic and often associated with $f_0$) and the corresponding value on the linear regression line exactly below the peak relating "quefrency" (anagram of frequency) to cepstral magnitude.[33] The more periodic a voice signal is, the more it displays a well-defined harmonic structure. As a result, the cepstral peak will be more prominent and will produce higher CPP values.[55,58] CPP has been found to correlate substantially with perceptual evaluation of voice,[61] and its applications have been extended to the analysis of different phonation and dysphonia types.[62] For instance, Esposito[63] found that CPP can help distinguish breathy from modal and creaky phonations. On the other hand, it cannot help in discriminating between modal and creaky phonations. Cannito et al[64] reported that CPP, independently of other measures included in the study, was correlated with both perceived breathiness and roughness. Similar findings were reported by Barsties et al.[33]
- *Harmonics-to-noise ratio (HNR)* is a measure quantifying the amount of noise in a voice signal and is calculated as the ratio of the energy at harmonic frequencies relative to the amount of energy at inharmonic frequencies. The higher the HNR value, the more sonorant and harmonic the voice.[65] Some studies suggest that HNR is an acoustic correlate of breathiness,[56,57] while others provide contrastive findings.[66,67] The inconsistent findings about the role of HNR as an acoustic correlate of breathy voice could be explained by potentially erroneous HNR measurement when pitch ($f_0$) tracking is not sufficiently accurate.[55,61]
- Measures of the spectral shape/slope include parameters reflecting the harmonic source spectrum, comparing the differences in dB between the relative amplitudes of two harmonics, and parameters relating to the overall spectral shape/slope (spectral tilt) of the source spectrum, comparing the amount of energy in

different frequency bands and using an Long-Term Average Spectrum analytic method. The difference in relative amplitude between the first and second harmonics (*H1*-H2**) is associated with the relative length of the open phase of the glottal oscillation.[41,68] A slow or incomplete closing phase has been shown to increase the H1*-H2* ratio. Thus, in breathy voices, the amplitude of the first harmonic is relatively high compared to the following, relatively weaker harmonics.[61] Barsties v. Latoszek et al[33] found H1*-H2* to be one of the most promising acoustic measures to predict breathiness, and a number of studies also noted strong correlations with perceived breathiness.[43,55,56,61] In addition, H1*-H2* is considered a successful measure for distinguishing phonations in a variety of languages.[49,50,69] The findings reported by Hartl et al[66] are nevertheless somewhat contradictory (see also Simpson[23] for a critique of H1-H2 measures across sexes, drawing attention to a confound between breathiness and nasality).

- Similarly, *H2*-H4** has also been proved to be an important measure when distinguishing breathy from modal phonation (especially in combination with H1*-H2*) in languages with phonemic contrast in voice quality (eg, in White Hmong[49]). Increased values of H2*-H4* (as well as those of H1*-H2*) are likely to indicate perceived phonemic breathiness (ie, steep drop in harmonic energy in lower frequencies[70]).
- *H1*-A1** is a measure of the amplitude of the first harmonic relative to that of the first-formant prominence. Similarly, *H1*-A2** refers to the same measure, which nevertheless uses the second formant prominence, and *H1*-A3** the third. H1*-A1* and H1*-A3* can successfully distinguish between breathy and clear vowels in Khmer (along with H1*-H2*[69]). Garellek and Keating[50] also mention that, along with H1*-H2*, the H1*-A2* measure can best distinguish each phonation type in Jalapa Mazatec.
- Some studies investigating measures of the *overall spectral shape/slope* (*spectral tilt*), which compare energy at high and low frequency bands, have suggested that a higher spectral tilt (lower high frequency energy) is related to an increase in perceived breathiness.[43,71] The less abrupt opening and closing of the glottis during breathy phonation results in attenuation of high frequency harmonics. However, the measures of the overall spectral shape can be affected by the presence of aspiration noise at high frequencies caused by the increased amount of turbulent airflow during the glottal open phase.[41,66]
- *Jitter* and *shimmer* reflect short-term perturbation in fundamental frequency and amplitude, respectively. Jitter represents the amount of variability in the duration of successive pitch periods, while shimmer represents the amount of amplitude variability across successive glottal pulses.[65] Although these traditional perturbation measures are commonly used in voice research, especially in dysphonic speakers, their reliability might be somewhat limited, as they rely on the identification of cycles of vocal fold vibration, which could be problematic in severely dysphonic or aperiodic voice signals, or in continuous speech samples containing variations in pitch and loudness as well as rapid consonant-vowel and vowel-consonant transitions.[72,73]

To summarize, this paper focuses on two central hypotheses. The first hypothesis is that female speakers in Czech and Danish will be breathier than male speakers. The second hypothesis is that the two languages will show different magnitudes of the sex-related differences in breathiness (without predicting the direction of the effect). The findings might jointly suggest what mechanism is responsible for the sex-related effects on breathiness (ie, physiological or sociolinguistic/cultural). We primarily employ acoustic analyses based on CPP, H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*, and HNR to approach the hypotheses, but the interpretation of the measures is aided by perceptual correlates of breathiness as well, in order to bring a more holistic approach to the sociophonetics of breathiness.

## MATERIAL AND METHODS

### Speakers

The sample of participants comprised 60 native speakers of Czech (30 female, 30 male) and 60 native speakers of Danish (30 female, 30 male). All participants were aged between 18 and 45 years at the time of the data collection (see Figure 1), which took place in 2018. This age range was selected because we aimed to target menstruating females since it has been shown that the phonatory properties of menstruating female speakers can differ from those that have not commenced menstruation within their lifespan or that have started their menopause.[29,74,75] Menarche can be safely assumed to have begun for our female speakers by the age of 18. Regarding the

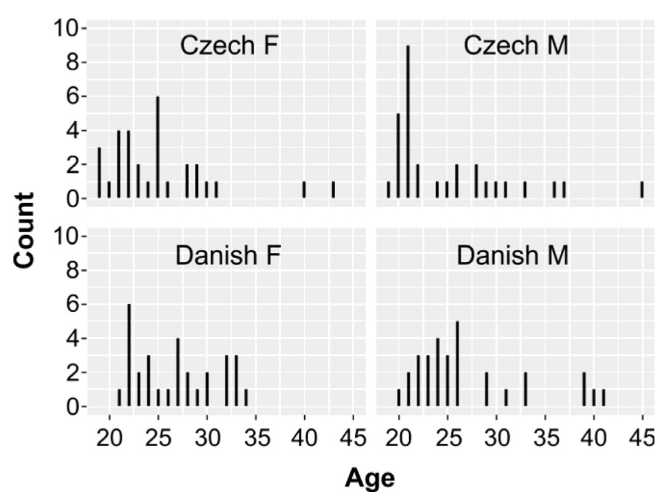


**FIGURE 1.** Age distribution with the number of speakers per specific age. Females (F) are shown on the left, males (M) on the right.

upper age limit, menopause rarely sets in prior to 40 years of age (eg,[76]). Moreover, female participants were also asked about their cycle and whether they used contraception at the time of the recording or before.

Seven of the Czech speakers were self-reported regular and/or chain smokers. Two of the Danish speakers reported to be regular smokers at the time of recording, one speaker reported quitting smoking 1 month prior to recording, and three speakers had quit at least 2.5 years before the data collection. Whether a participant was or was not a regular smoker at the time of recording was included as one of the control variables in the statistical analyses. The speakers' vocal health was assessed perceptually by the experimenters and confirmed by self-reporting. No subjects reported any diagnosed speech pathologies and no participants were vocally indisposed due to illness at the time of recording. Moreover, this was further confirmed in the perceptual evaluation by three experts (see below), who reported all speakers to fall within normal variation.

As regards the regional background of the participants, 35 Czech speakers grew up in Prague or Central Bohemia, 9 in North Bohemia, 5 in South Bohemia, 4 in East Bohemia, 4 in Moravia, and 3 in Silesia. 50 of the Danish speakers grew up in Jutland, 5 in Fyn, 1 in Lolland, 1 in Samsø, 1 in Zealand, and 2 in mixed regions.

## Material

Two sets of materials must be used if we aim to investigate two languages. From a clinician's perspective, it might be best to examine a sustained [a]-like vowel to ensure direct comparison between speakers, and to use a standardized protocol of voice evaluation. However, this is problematic for several reasons. First, a range of standardized protocols of voice evaluation is available and different protocols and software solutions are used in different countries (eg,[77,78]). In Denmark, the Max Manus software is used (Martin Wirenfeldt Nielsen, Danskpatologiselsab, personal communication 2019), although the Digital Voice Handicap Index software application has been designed for languages that also include Danish and Czech.[79] In the Czech Republic, the prevalent option among voice disorder clinicians is the GRBAS scale (Miroslava Hrbková, Phoniatrics Clinic in Prague, Mojmír Lejska and Radan Havlík, Centre for Speech, Voice and Hearing Impediments in Brno, personal communication 2019). Second and more importantly, our goal is to describe speech as linguistic behavior in a communicative setting. That precludes using sustained vowels, and favors (semi-)spontaneous recordings. However, a large number of factors, often unknown, come into play in terms of breathiness when using uncontrolled material. Therefore, controlled production of specific texts is needed. Reading a coherent passage is more natural than producing a list of unconnected sentences read in isolation. Moreover, having the same passage enables us to control for segmental and prosodic variation across the speakers of the given language as much as possible.

The inherent problem behind read passage materials nonetheless is that Czech and Danish have different phonological systems, so segmentally direct comparisons are therefore not possible. Voice quality studies examining English frequently use the Rainbow Passage,[80] which is balanced for use in the English language[81]: p. 1−2; [82]: p. 2) and is used by both clinicians[9,46] and linguists[83] for analyses of voice (see also references in Clopper and Pisoni[84]). Although its translation will not necessarily result in the same type of segmental balancing, and it could also be matched with other read passages segmentally more suitable for Czech and Danish, we translated the text into the two languages (see Appendices A and B). The motivation behind this choice is as follows. Formality and topics or specific lexical content are well-known to affect various properties of speech, including subsegments, segments, and suprasegments,[85,86,87,88,89], which is in our mind a bigger issue to tackle than a potential segmental imbalance. Our results will thus be more directly comparable to the English studies (frequently using this passage) and, more importantly, the two languages investigated here can be compared directly. If two different texts had been employed instead, they would have been unbalanced in the lexical and expressive meanings, which would have introduced a potential confound of language and text, known to affect phonetic aspects of language as well. Our primary concern was controlling for language-external variation maximally and language-internal variation as much as the typologically different phonological systems of the two languages allow. Finally, we partially controlled for the segmental imbalance by excluding high vowels (compare Simpson[90]) and by considering vowel height and vowel duration in the statistical analysis. This is further advantageous in capturing some of the variability due to vowel reduction occurring in connected speech (these processes mean that even corresponding phonemes in Czech and Danish would not in fact be realized identically). Also note that our analysis is concerned with vowels only, so many of the differences related to consonants are not relevant. Furthermore, coarticulatory processes of Danish are notoriously known for their typological markedness in linguistic phonology, making a direct *phonetic* comparison with other languages very problematic, if not impossible. The decisions we make here are in line with what has been acknowledged about the difficulties in cross-linguistic designs[1]: p. 721.

Crucially with regard to our analyses, the Czech material contained the phoneme /a(:)/, with 34 tokens per speaker in total. The Danish version of the text does not have a sufficient number of tokens of any nonhigh vowel phoneme, so the following nonhigh phonemes were chosen: /ɔ(:)/ (12 tokens), /ʌ/ (10 tokens), and /a(:)/ (7 tokens), yielding 29 tokens per speaker in total. Note that the short and long phonemes were not distinguished in the analyses in terms of their phonological length category but rather by phonetic duration. Appendices A and B enumerate the Czech and Danish target words. Following a reviewer's comment, we conducted a post hoc analysis comparing the [a]-like vowels, which correspond as directly as possible in the two languages (2 × 5 tokens). The mean values indicated similar trends in all the investigated parameters (except for Czech speakers with regard to H1*-A3*). Our overall results therefore agree with this small but balanced subset.

## Recording sessions

The recording sessions took place in a professional, sound-treated studio (either at the Institute of Phonetics at Charles University, Czech Republic, or at the Department of English at Aarhus University, Denmark) with practically no ambient background noise. The signal-to-noise ratio was well above the 15 dB value recommended for voice measurements.[91] For the Czech data, the mean SNR value was 42.6 dB (SD = 4.8 dB, min = 29.4 dB, max = 51.0 dB), and 35.1 dB for the Danish data (SD = 3.6 dB, min = 28.3 dB, max = 43.2 dB). A pair of comparable condenser microphones with a flat frequency response and a cardioid polar pattern was used for the entire data collection. Specifically, the Danish material was recorded on a Sony ECM-959A microphone into a Zoom H5 Handy Recorder, whereas the Czech material was obtained with an AKG C4500B-BC microphone connected directly to a computer's sound card. The recordings were saved as uncompressed mono .wav files (with 44.1 kHz sampling rate and 16-bit quantization). In both cases, the participants were seated 20−30 cm in front of the microphone and were instructed to move minimally during the recording session. Such a distance is recommended for recording human voice production accurately in terms of spectral properties as it avoids the proximity effect.[91,92] The recording conditions were thus kept as constant as possible across speakers and languages.

Given the lexical content of the text, we wanted to prevent any speaker-specific differences in the reading style, most importantly, a child-directed speech or "telling a story" speech style, which is associated with specific voice qualities.[93] The participants were therefore asked to imagine that they were weather experts speaking matter-of-factly on the radio to an adult audience. The instructions were the same for both the Czech and the Danish participants.

## Acoustic analysis

False starts and similar parts of the recordings were removed. After transliteration of Danish words into Czech orthography, the recordings of both languages were subjected to automatic forced alignment using the Prague Labeller algorithm.[94] The target vowel boundaries were afterward corrected manually based on the phonetically motivated recommendations for segmentation of the speech signal.[95]

The following restrictive criteria were set prior to analyses. First, as nonmodal phonation is known to cue the end of a turn or an utterance,[96,97,98] target segments in such contexts were not considered. For the sake of simplicity, utterance end was defined as the final syllable of a word which immediately preceded a full stop in the reading passage. Second, we omitted any portions of the target vowels which contained creaky phonation (this occasionally meant entire vowels). The motivation was that certain acoustic measures may not discriminate well between breathy and creaky phonation (eg, HNR[99]; CPP[48]), so excluding creaky intervals will disambiguate any potential measurement-related confounds. Importantly, creak was found only in 1.5% ($n = 31$) of the Czech data and 4% ($n = 57$) of the Danish data. Furthermore, this approach enabled us to avoid stød in the Danish data, at least when realized as creaky phonation (see[100] for more information on the phonetic realisation of stød). Finally, the minimal duration of a target interval was set to 40 milliseconds, ensuring a sufficient number of points in each token for parameter extraction. This means that vowels shorter than 40 milliseconds or vowels substantially shortened by creak were not considered. As mentioned above, there were approximately 30 tokens per speaker to be analyzed. Figure 2 shows an illustration of the target interval determination and parameter extraction.

Several measures related to breathiness, as well as other relevant variables, were extracted in VoiceSauce, a free stand-alone software[101,102] that is widely used in the study of phonation by linguists (eg,[103,104]). The following measures were extracted automatically with the software: $f_0$, F1 values, the spectral magnitudes of H1*-H2*[1], H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*, CPP, HNR, and the duration of the relevant vowel interval. For male speakers, $f_0$ range was set to the minimum of 60 Hz and the maximum of 400 Hz. For female speakers, the selected range was 90−500 Hz. $F_0$ was quantified with the default Straight algorithm at 1-millisecond intervals and the maximum duration selected in the settings was 10 seconds. F1 was established using the Snack algorithm, as this method is more optimal when the Straight algorithm is used to calculate $f_0$. (For more details on the Snack and Straight algorithms used for $f_0$ and formant values estimation, see[102,105,106]). When extracting F1, pre-emphasis was set to 0.96, window length to 25 milliseconds, and the frame shift to 1 milliseconds.

Regarding the VoiceSauce settings related to the remaining measures, CPP calculations are based on the algorithm described in Hillenbrand et al[55] and HNR measures extraction is derived by de Krom's algorithm.[107] In both parameters, a variable window length equal to five pitch periods was used for the calculations by default (for more detailed information, see[102,108]). However, unlike CPP, which covers the entire frequency range, HNR calculations are carried out in four different frequency ranges: 0−500 Hz, 0−1.5 kHz, 0−2.5 kHz, and 0−3.5 kHz. Only the last HNR measure (HNR35), with the largest frequency range, was selected to include calculations in higher frequency regions.[109] Finally, harmonic spectra magnitudes were computed pitch-synchronously using a default three-cycle window.

[1]The asterisk denotes harmonic amplitudes measures corrected for the effect of F1, F2, and F3 vocal tract resonances to enable researchers to make comparisons among different speakers and vowels.

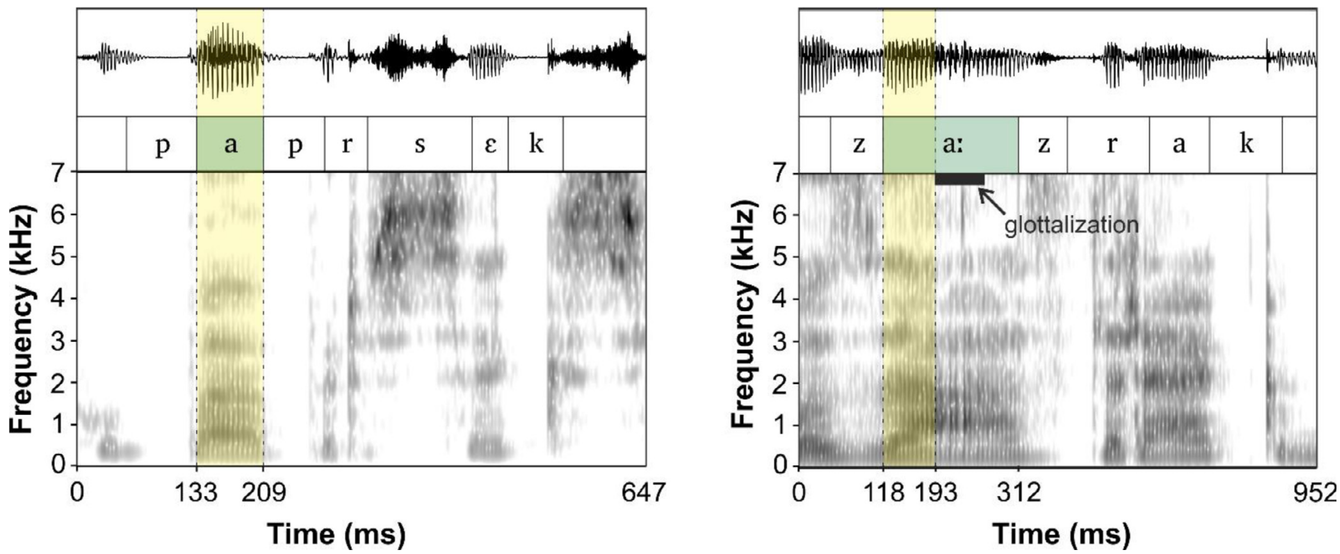


**FIGURE 2.** Illustration of segmentation and target interval selection. An unproblematic vowel on the left (target = vowel duration), a problematic vowel on the right (most of the vowel's duration is excluded due to creaky voice, its presence marked by the black rectangle). Parameters were extracted only from the target parts of the signal (yellow rectangles). Each interval was at least 40-millisecond long (in this example: 76 milliseconds and 75 milliseconds). (Color version available online.)

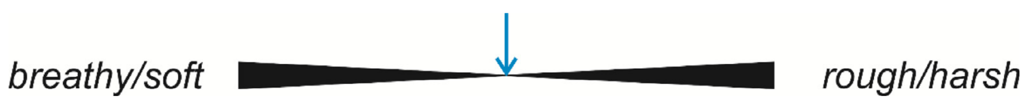


**FIGURE 3.** The rating scale used in the perceptual evaluation. The arrow can be moved anywhere from its initial position indicating modal phonation.

### Perceptual analysis

Since the acoustic measures we used may reflect different aspects of breathiness (or voice quality in general), perceptual evaluation of the voices was needed as a check to provide more conclusive interpretations. All the stimuli were evaluated by 3 speech therapists.[2] Preliminary auditory evaluation by the authors corroborated with a visual inspection of spectrograms did not identify any severe (pathological) breathiness in the sample (or any other voice disorder). Therefore, the evaluation scale was designed to describe nonpathological voices (unlike the GRBAS scale discussed earlier). Crucially, *breathiness*, a state of soft phonation, might be confounded with *roughness*, which is in the extreme (ie, pathological) case characterized as harsh, crackling, more aperiodic phonation with rapid amplitude/frequency fluctuations.[33] The experts were asked to evaluate how far a given stimulus deviates from modal phonation toward breathy/soft voice or toward rough/harsh voice. They indicated their choice on a continuous visual analogue scale ranging from 0 (modal) to the two extremes (see Figure 3). This procedure was adopted in order to make sure that the acoustic measures do not misidentify roughness for breathiness. Most speakers were expected to be modal or mildly deviating. Since the data did not include any portion with creaky phonation, this aspect of voice was not considered. The scale was explained in detail and it was ensured as much as possible that all raters interpreted the scale and task in a similar way.

In order for the results to be maximally relevant, the stimuli consisted of exactly the same material as the acoustic analyses. The target vowel intervals (see previous section) were cut out from each recording and concatenated with an overlap of 5 milliseconds between the neighboring intervals. The resulting sound was normalized to 70 dB RMS in Praat. The sound was duplicated to obtain sufficiently long stimulus durations (median = 4.2 seconds, min = 2.1 seconds, max = 5.7 seconds). The obvious advantage of this procedure is that all lexical meaning is lost, and the two languages can be evaluated more easily by the expert listeners, focusing the attention entirely on the voice of the speakers. A

[2]Rater 1 (female, 32 years, native speaker of Czech) graduated in Phonetics in Prague (MA) and also in Speech and Hearing Therapy in Brno (MA). She had worked as a clinical speech therapist for 4 years. Currently, she is employed at the Institute of Phonetics, Prague. Rater 2 (female, 33 years, native speaker of Czech) graduated in Speech and Hearing Therapy in Brno (MA). She had worked as a clinical speech therapist for 2 years. Currently, she is employed as a voice lecturer at the Theatre Faculty of the Academy of Performing Arts, Prague. Rater 3 (female, 40 years, native speaker of Slovenian, but L2 speaker of Czech) graduated in Speech and Language Therapy in Lublin (BA), Clinical Linguistics in Potsdam (MSc) and Speech Science in Edinburg (PhD). She had worked as a speech and language therapist for two years, followed by several years of academic research in the field. She is currently employed at the Institute of Phonetics, Prague.

practice session was necessary to become familiarized with the atypical aspect of the speech material and thus increase the ecological validity of the task (the listeners' internal evaluation criteria were expected to adjust during familiarization).

The experiment was administered individually in Praat via the Demo Window environment using headphones (Sennheiser HD599) in a sound-treated studio. The listeners were told that they were going to evaluate recordings of different speakers for the purposes of voice quality assessment. They were warned about the atypicality of the stimuli (eg, rapid alternations of vowels and frequent $f_0$ jumps between them). The arrow (slider) was initially positioned in the middle, indicating modal phonation. The raters were encouraged to move the arrow anywhere on the scale toward the extremes. They evaluated Czech voices first and Danish voices on the subsequent day. In each session, the listeners heard 60 speakers plus 10 repetitions as a check of intrarater consistency. The 60 target trials were randomized and presented in four blocks of 15 trials, separated by short breaks. Each trial was preceded by a short noise-like sweep to desensitize the listeners and reduce order effects. The raters could listen to the stimulus repeatedly by pressing a button. The 10 repeated trials were presented at the end of the session. The duration of a single session was approximately 30 minutes.

The perceptual ratings were then analyzed as such and also correlated with the acoustic measures of breathiness. However, it must be emphasized that the perceptual analysis is only viewed as complementary to make better sense of the primary experiment, allowing for a more conclusive interpretation of the acoustic analyses.

### Statistics

As the acoustic parameters were extracted in 1-millisecond steps, the values for each target token were first averaged, yielding one value per token. The statistical analyses were conducted in R[110] with RStudio.[111] Linear mixed effects (LME) regression models were constructed using the *lme4* package,[112] accounting for both fixed and random effects. The basic model was specified with the fixed effects of LANGUAGE (Czech × Danish), SEX (female × male), SMOKER (yes × no), AGE (in years), $F_1$ (in Hz), $F_0$ (in Hz), CREAK (yes × no), NASAL CONTEXT (yes × no), and DURATION (in milliseconds). The random effects structure included the random intercepts for SPEAKER and WORD, as well as the random slope for SEX. The statistical evaluation of the relevant predictors (SEX and LANGUAGE) was done by comparing the basic model with a reduced model lacking the fixed effect in question using likelihood ratio tests. All the other predictors, whether statistically significant or not, were always part of the model. Furthermore, the significance of the interaction between SEX and LANGUAGE was evaluated in the same way. The alpha level of 0.05 was used to establish the significance of any differences found; given the structure of our effects, alpha adjustment for post hoc comparison was not necessary. Summaries of the final model (with or without the interaction) are presented for each parameter in Appendix C. Effect plots were constructed using the *effects* package.[113]

## RESULTS

Mean between-group differences in raw values are shown in Table 1 for each of the seven breathiness measures. The

**TABLE 1.**
**Mean Values of Breathiness Measures Grouped by Sex and Language (F = female, M = male). Standard Deviation is Given in Brackets**

| | | Czech | | Danish | |
|---|---|---|---|---|---|
| Measure | Sex | Mean (SD) | Difference F-M | Mean (SD) | Difference F-M |
| CPP | F | 22.5 dB (2.4 dB) | −0.6 dB | 21.5 dB (2.3 dB) | −0.3 dB |
| | M | 23.1 dB (2.4 dB) | | 21.8 dB (2.5 dB) | |
| HNR | F | 34 dB (5.5 dB) | 9.5 dB | 41.6 dB (7.3 dB) | 9.9 dB |
| | M | 24.5 dB (6.1 dB) | | 31.7 dB (6.8 dB) | |
| H1*-H2* | F | 7.2 dB (4.2 dB) | 5.9 dB | 6.8 dB (4 dB) | 5.7 dB |
| | M | 1.3 dB (2.8 dB) | | 1.1 dB (4.2 dB) | |
| H2*-H4* | F | 1.1 dB (8.2 dB) | −9.6 dB | 1.6 dB (5.9 dB) | −4.7 dB |
| | M | 10.7 dB (5 dB) | | 6.3 dB (5.9 dB) | |
| H1*-A1* | F | 22.1 dB (7.7 dB) | −1 dB | 18.4 dB (5.2 dB) | −0.1 dB |
| | M | 23.1 dB (5.3 dB) | | 18.3 dB (6.6 dB) | |
| H1*-A2* | F | 17.8 dB (9 dB) | −1.2 dB | 18.7 dB (6.4 dB) | 0.5 dB |
| | M | 19 dB (6.5 dB) | | 18.2 dB (8.5 dB) | |
| H1*-A3* | F | 14.7 dB (9 dB) | 0.1 dB | 11.6 dB (8.3 dB) | 1.6 dB |
| | M | 14.6 dB (7.6 dB) | | 10 dB (11.4 dB) | |

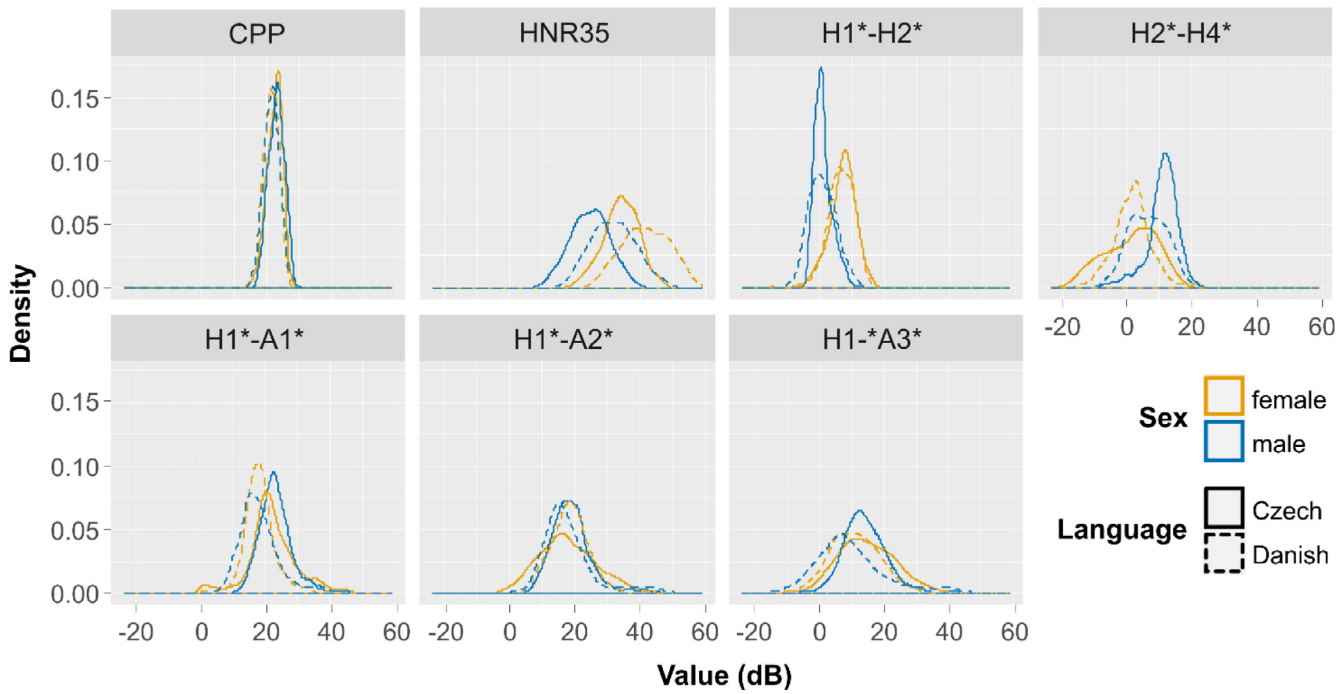


**FIGURE 4.** Density plots showing the distribution of the extracted values of breathiness measures grouped by sex (color) and language (line type).

distribution of these values is shown as density plots in Figure 4. Results of the statistical analysis are provided below for each measure separately, along with effect plots taking into account all the variables that have been included in the model (Figures 5–11). As a result, effect plots may differ from the raw values in Table 1 and Figure 4.

### Periodicity-related acoustic measures

Analyses of CPP (Figure 5) revealed that females are breathier in both languages, as they showed lower CPP values ($\chi^2$ (1) = 7.54, $P < 0.01$). Czech speakers exhibited lower degrees of breathiness than Danish speakers ($\chi^2$ (1) = 12.9, $P < 0.001$). However, no significant interactions between SEX and LANGUAGE were found ($\chi^2$ (1) = 0.9, $P = 0.35$).

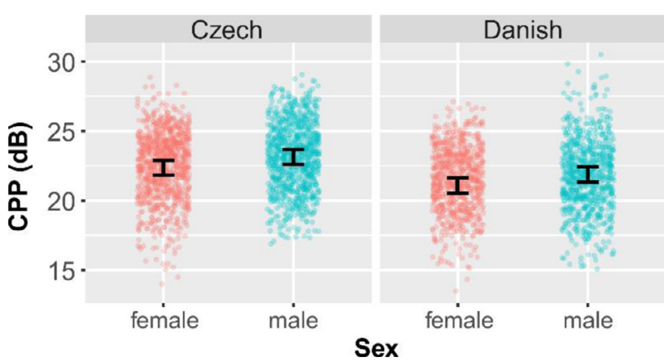


**FIGURE 5.** Effect plots of CPP by language and sex (mean and 95% CI intervals). The scatter plot shows the distribution of the raw data.

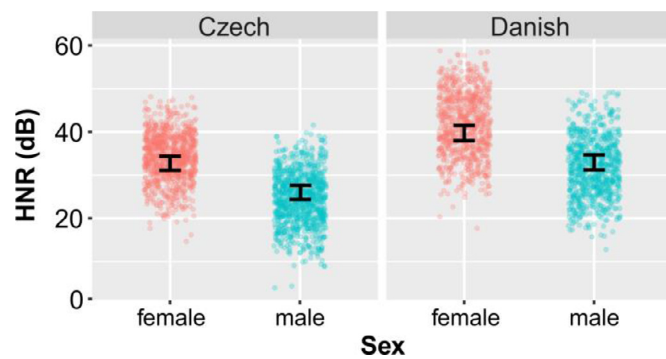


**FIGURE 6.** Effect plots of HNR by language and sex (mean and 95% CI intervals). The scatter plot shows the distribution of the raw data.

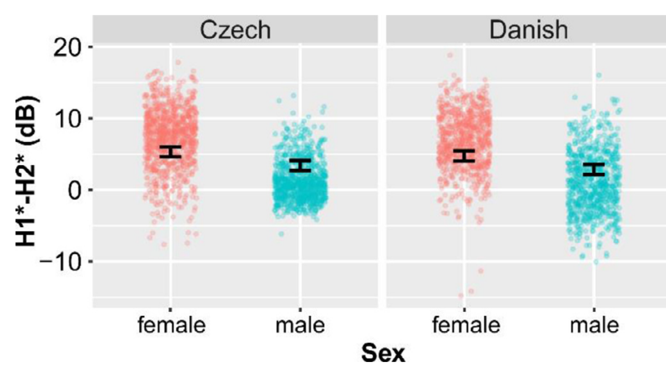


**FIGURE 7.** Effect plots of H1*-H2* by language and sex (mean and 95% CI intervals). The scatter plot shows the distribution of the raw data.

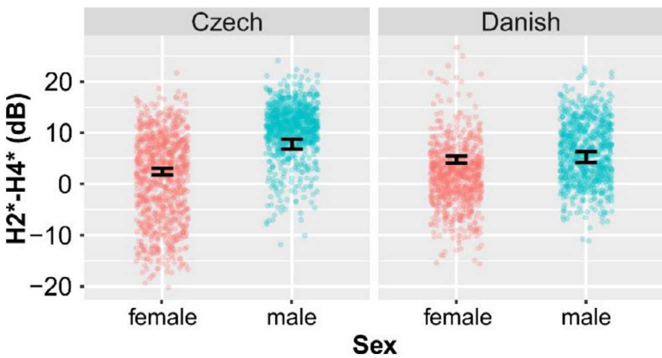


**FIGURE 8.** Effect plots of H2*-H4* by language and sex (mean and 95% CI intervals). The scatter plot shows the distribution of the raw data.

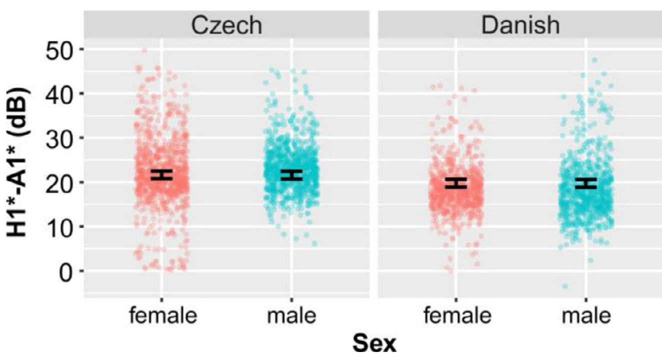


**FIGURE 9.** Effect plots of H1*-A1* by language and sex (mean and 95% CI intervals). The scatter plot shows the distribution of the raw data.

In the analysis of HNRs, males yielded lower HNR values than females in both languages ($\chi^2$ (1) = 83.9, $P$ < 0.001). This suggests that male voices have more aperiodic signal, which can be attributed to either breathiness or roughness (or indeed a combination of the two). In addition, Czechs proved to be breathier (or rougher) than Danes ($\chi^2$ (1) = 31.2, $P$ < 0.001). No significant interaction between SEX and LANGUAGE was found ($\chi^2$ (1) = 0.4, $P$ = 0.53), as seen in Figure 6.

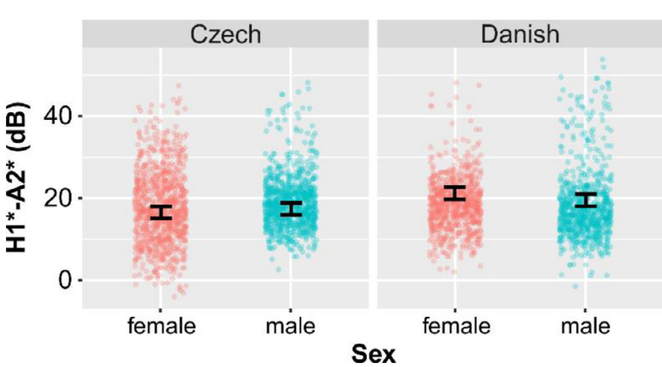


**FIGURE 10.** H1*-A2* by language and sex (mean and 95% CI intervals). The scatter plot shows the distribution of the raw data.

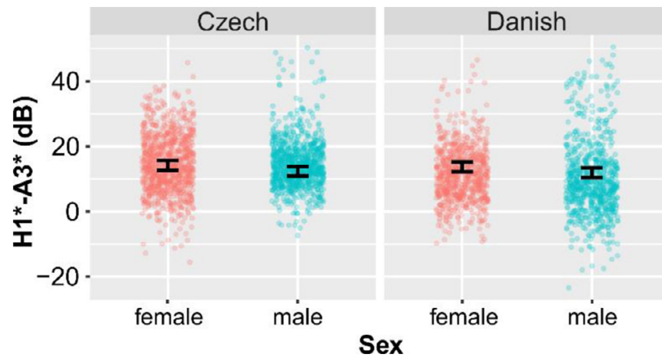


**FIGURE 11.** H1*-A3* by language and sex (mean and 95% CI intervals). The scatter plot shows the distribution of the raw data.

## Measures based on harmonic spectra magnitudes

Analyses of H1*-H2* show that females are breathier in both languages, as indicated by higher H1*-H2* values ($\chi^2$ (1) = 16.5, $P$ < 0.001). LANGUAGE was not a significant predictor ($\chi^2$ (1) = 2.0, $P$ = 0.16). Similarly to CPP, no significant interactions were found between the two factors ($\chi^2$ (1) = 0.1, $P$ = 0.8; see Figure 7).

H2*-H4* analyses indicate that males are breathier than females, showing higher H2*-H4* values ($\chi^2$ (1) = 18.2, $P$ < 0.001). In general, LANGUAGE did not have a uniform effect on the measure ($\chi^2$ (1) = 0.4, $P$ = 0.51). There was a significant interaction between SEX and LANGUAGE ($\chi^2$ (1) = 26.6, $P$ < 0.001); namely, the sex difference was apparent only in Czech (Figure 8).

The amplitude of the first harmonic relative to that of the first formant prominence, H1*-A1*, did not reveal a sex difference ($\chi^2$ (1) = 0.02, $P$ = 0.9). However, LANGUAGE affected this correlate of breathiness significantly ($\chi^2$ (1) = 13.2, $P$ < 0.001): Czech speakers seem to be breathier than Danish speakers, displaying higher values (Figure 9). The interaction between SEX and LANGUAGE did not reach significance ($\chi^2$ (1) = 3.1, $P$ = 0.08).

Analyses of H1*-A2* do not suggest any overall sex difference ($\chi^2$ (1) = 0.4, $P$ = 0.54), but the Danish speakers were generally breathier than the Czech speakers ($\chi^2$ (1) = 13.8, $P$ < 0.001). However, a significant interaction between SEX and LANGUAGE was found ($\chi^2$ (1) = 4.4, $P$ < 0.05). In particular, the effect plots in Figure 10 suggest that female speakers are breathier than male speakers in Danish, whereas the Czech speakers exhibit no or mild differences in the opposite direction.

Finally, with regard to H1*-A3*, female speakers displayed higher values (Figure 11), and are thus generally breathier than male speakers ($\chi^2$ (1) = 4.4, $P$ < 0.05). No effect of LANGUAGE is reported ($\chi^2$ (1) = 0.2, $P$ = 0.63), nor was there a significant interaction between the two factors ($\chi^2$ (1) = 2.4, $P$ = 0.12).

## Relationship between breathiness measures

Seeing that the individual measures do not show the same results and render an overall interpretation rather

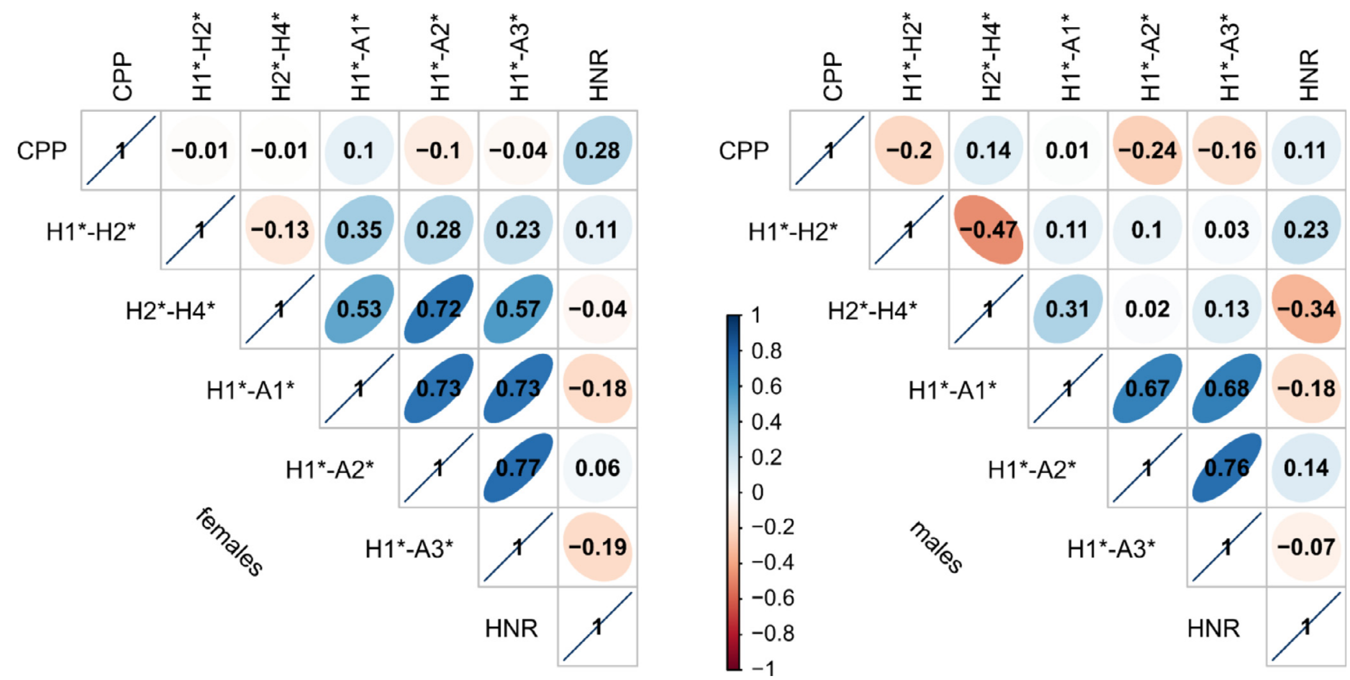


**FIGURE 12.** Correlations between breathiness measures (females on the left, males on the right).

challenging, we now turn to an analysis which aims to establish the relationship of these individual measures in our data. We first inspect the correlations between the individual measures for the two sexes and then the results of a Principal Component Analysis (PCA).

Generally, we find that most of the measures used here are uncorrelated or they correlate only moderately ($r$ < 0.40). The direction of the correlation is not consistent, even with measures that share the same monotonic relationship (eg, H1*-H2* and H2*-H4*). Figure 12 shows the results for females and males separately. For both groups, we observe strong positive correlations ($r$ = 0.67−0.77) between harmonic spectra magnitude measures that involve the amplitude of a formant. In contrast to males, females also exhibit moderate to strong positive correlations between these measures and H2*-H4*. In addition, the male data reveal a moderate negative correlation between H1*-H2* and H2*-H4*.

The correlations considered separately for the two languages did not show any particularly noticeable differences; again, the formant-related harmonic magnitude measures were positively correlated, and this correlation was strong or very strong ($r$ = 0.66−0.83).

The relationships between measures can further be explored in a PCA, an algorithm used for dimensionality reduction. Importantly, the original seven measures, correlated to some degree, are transformed so that the resulting seven components are mutually uncorrelated (orthogonal). The first component, PC1, captured 39% of the raw variability in the parameters, and was most strongly correlated with H1*-A1*, H1*-A2* and H1*-A3* ($r$ = −0.87 to −0.88). PC2 was most strongly correlated with H1*-H2* ($r$ = 0.87) and HNR ($r$ = 0.71). PC3 correlated with CPP ($r$ = 0.95). In addition, H2*-H4* correlated strongly with both PC1 ($r$ = −0.59) and PC2 ($r$ = −0.62). Taken together, the three PCs captured 80% of the raw variability in the seven measures (see Appendix D). The relevant conclusion about the seven measures employed here is that although the three formant-related harmonic amplitude measures seem to introduce a single source of variation to the data, CPP provides independent information, and so does to a large degree H1*-H2*. These results are consistent with the strong positive correlations found for the formant-related measures in the correlation analysis.

## Perceptual analysis

### *Rater consistency*

Because the data were assessed by three raters (all being speech therapists, see Material and methods), tests of both inter- and intrarater reliability were conducted. Intrarater reliability was checked by repeating 10 out of the 60 test items for each language and computing the difference between the second and the first occurrences. The three evaluators differed in their consistency: Rater 3 was more consistent (mean score difference = 0.09, SD = 0.11) than the other two raters (Rater 1: mean difference = 0.21, SD = 0.28; Rater 2: mean difference = 0.37, SD = 0.24). However, the better consistency of Rater 3 is due to a more frequent use of the modal category (middle of the scale) and a less frequent use of the extremes. Furthermore, the two other raters reported that it was sometimes difficult to decide whether a specific voice was extremely breathy or extremely rough, as it seemed to be a combination of both. Indeed, Rater 3 never switched the two sides of the scale in the repeated items, whereas Raters 1 and 2 both switched the direction 7 times (35 % of cases). When computed regardless of the direction, the mean differences were much lower for the two evaluators (0.14 for Rater 1, 0.06 for Rater 2).

Inter-rater reliability was estimated in terms of the correlation between the mean score for each item and the individual rater's evaluation. Two of the three raters showed a high degree of agreement. The correlation coefficients were $r = 0.85$ (CI = 0.79−0.89) for Rater 1, $r = 0.78$ (CI = 0.70−0.85) for Rater 2, and $r = 0.56$ (CI = 0.42−0.67) for Rater 3. Correlation plots (not included here) make it clear that Rater 3 used the modal option most frequently, resulting in more substantial disagreements with the mean score. The main difference between the three raters thus appears to be due to the extent to which the entire scale was employed, and how consistently they identified the deviation from modal phonation as breathiness vs roughness.

### *Speaker evaluation*

The perceptual evaluation by the three experts suggests that females are perceived as breathier than males and, at the same time, male voices are perceived as rougher than female voices (Figure 13). The 0.5 value indicates perceived modal phonation, the 0 value reflects breathy phonation, and roughness is indicated by the value of 1 on the rating scale. There is greater variability in the Danish speakers than in the Czech speakers. Specifically, some of the Danish male and female speakers are assessed as leaning toward the breathier end of the scale.

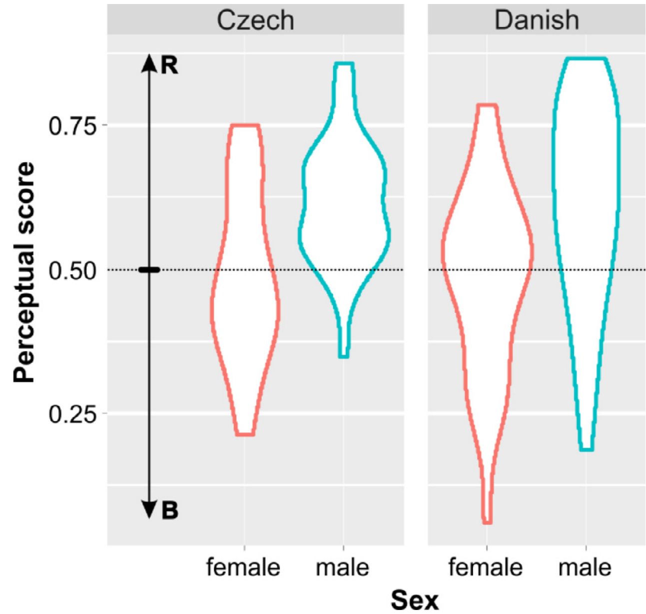


**FIGURE 13.** Perceptual assessment of the 60 speakers grouped by sex and language on a breathiness (B) to roughness (R) scale. Modal phonation corresponds to 0.5.

## Correlation between acoustic measures and perceptual evaluation

In order to interpret the individual analyses as a whole, we first performed an LME regression with perceptual score (the mean perceptual assessment of the three clinicians) as the dependent variable and the seven individual acoustic measures as predictors (with WORD as a random effect). Table 2 provides the estimates and, in the final column, the results of likelihood-ratio tests assessing the significance of a predictor's contribution to the perceptual score (by comparing the full model to a reduced model). The strongest predictors of the perceptual score are CPP and HNR, followed by H1*-H2*. The parameter estimates indicate that higher values of CPP were perceived as less breathy (or rougher), higher values of HNR as more breathy (less rough), and the same was true for H1*-H2*. The remaining four measures did not affect the model's goodness of fit substantially, and were therefore not statistically significant in the perceptual assessment. However, as the PCA analysis suggested, the formant-related measures are highly correlated—what this means is that omitting one of the three may not have an effect, but omitting all of them could. Indeed, when using only five measures (eg, with H1*-A1* in the full model to the exclusion of H1*-A2* and H1*-A3*), the contribution of the formant-related measure becomes statistically significant (and in the direction where higher parameter values are associated with less rough or breathier

**TABLE 2.**
**Estimates of Fixed Effects from an LME Model of Perceptual Score (0 = Breathiness, 0.5 = Modal Phonation, 1 = Roughness)**

| Fixed Effect | Estimate | SE | t value | LR test Result |
|---|---|---|---|---|
| Intercept | 0.344 | 2.82e−02 | 12.195 | |
| CPP | 2.14e−02 | 1.25e−03 | 17.153 | $\chi^2$ (1) = 249.2, $P < 0.001$ |
| HNR | −7.43e−03 | 4.48e−04 | −16.574 | $\chi^2$ (1) = 226.6, $P < 0.001$ |
| H1*-H2* | −3.65e−03 | 7.53e−04 | −4.846 | $\chi^2$ (1) = 22.4, $P < 0.001$ |
| H2*-H4* | 7.87e−04 | 4.73e−04 | 1.663 | $\chi^2$ (1) = 2.8, $P = 0.09$ |
| H1*-A1* | −8.81e−05 | 6.88e−04 | −0.128 | $\chi^2$ (1) = 0.01, $P = 0.90$ |
| H1*-A2* | −6.45e−04 | 6.61e−04 | −0.975 | $\chi^2$ (1) = 0.87, $P = 0.35$ |
| H1*-A3* | −4.19e−04 | 4.85e−04 | −0.865 | $\chi^2$ (1) = 0.68, $P = 0.41$ |

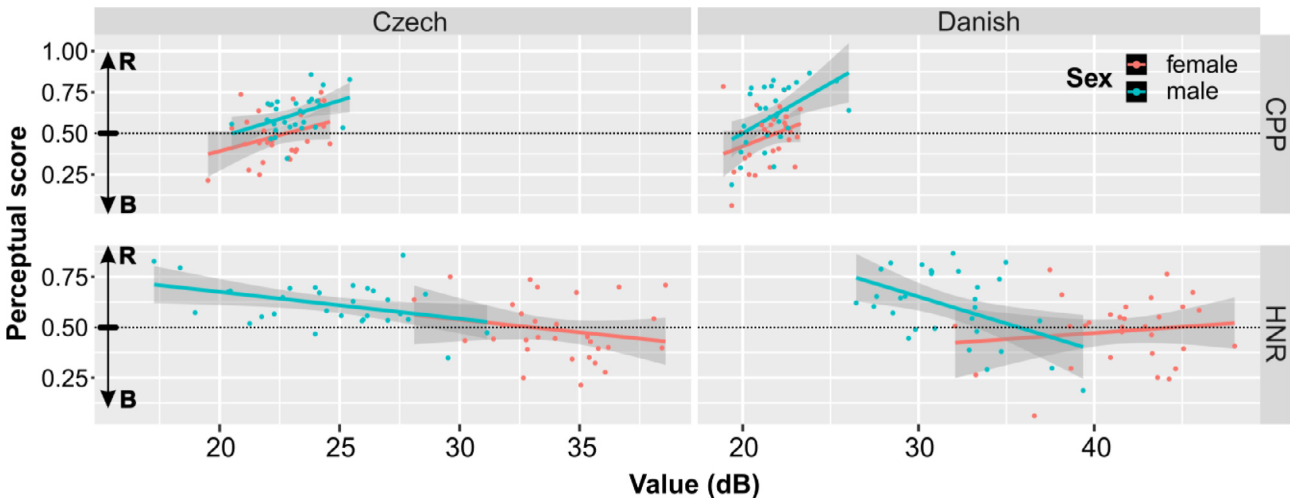


**FIGURE 14.** Scatterplots with trendlines (and 95% confidence bands). *X*-axis shows extracted mean values for CPP and HNR, *y*-axis displays the perceptual assessment of the 60 speakers.

evaluations). H2*-H4* thus remains the only predictor in the model without a significant effect on the evaluation.

Secondly, we inspected Pearson correlations between the perceptual assessment and the individual acoustic measures (averaged across words). Discussing each measure separately circumvents the problem of high covariance between some of the measures, and may shed light on the interpretation of the perceptual scale even further (eg, whether a decrease on the scale means higher breathiness or lower roughness).

In the following figures, the perceptual dimension of the scatterplots is always the same. Therefore, males are associated with values mostly above the line (indicating roughness), and females are dispersed more evenly around, above, and below the line (some voices being evaluated as rough, some as breathy, some approximately modal). The interesting information is how the acoustic values map onto these evaluations. A significant correlation means that the acoustic measure is relevant for predicting the perceptual score. In contrast, the lack of a correlation will indicate that the measure captures acoustic aspects different from those used by the evaluators as a basis for their decision.

In Figure 14, the acoustic values for the periodicity measures (CPP and HNR) are plotted against the perceptual evaluation score. Lower CPP values (ie, less periodic signal/less clearly defined harmonic structure) correlate with lower perceptual scores, ie, with more breathiness and less roughness. The correlation was significant in male speakers (Cz: $r = 0.43$, $P = 0.018$; Dn: $r = 0.51$, $P = 0.004$) but insignificant in females (Cz: $r = 0.32$, $P = 0.082$; Dn: $r = 0.31$, $P = 0.098$). The CPP measure seems to reflect perceived roughness as well as breathiness.

In contrast, lower HNR values (ie, more aperiodic signal/higher amount of noise) were associated with higher perceptual evaluation scores, that is, with more perceived roughness (lower panel of Figure 14). The HNR measure overall seems to reflect perceived roughness rather than breathiness in our speakers. Again, there was a strong correlation in the males speakers (Cz: $r = -0.41$, $P = 0.026$; Dn: $r = -0.47$, $P = 0.009$), indicating a link to perceived roughness, whereas there was no significant correlation in the female speakers (Cz: $r = -0.21$, $P = 0.260$; Dn: $r = 0.14$, $P = 0.476$).

Moving to the measures of the spectral shape/slope, firstly, we report that the H1*-H2* lower values (indicating less prominent/weaker first harmonic relative to the second one, thus suggesting less breathiness) mostly correlate with perceived roughness in both Czech and Danish male speakers (upper panel in Figure 15; Cz: $r = -0.45$, $P = 0.013$; Dn: $r = -0.49$, $P = 0.006$). In contrast, there was no significant correlation in the female speakers (only a tendency in Danish; Cz: $r = -0.004$, $P = 0.985$; Dn: $r = -0.22$, $P = 0.402$), suggesting that H1*-H2* seems to reflect perceived roughness rather than breathiness.

Secondly, no correlation between H2*-H4* values and perceptual assessment is apparent in females (Cz: $r = -0.08$, $P = 0.689$; Dn: $r = -0.01$, $P = 0.951$) and Czech males ($r = 0.09$, $P = 0.643$; see lower panel, Figure 15). In the Danish males, there is a slight tendency for higher values of the parameter (potentially suggesting more breathiness) to be evaluated as rougher ($r = 0.25$, $P = 0.179$). Thus, H2*-H4* seems to be associated with perceived roughness rather than breathiness (even more clearly than H1*-H2*, as males are to the right on the *x*-axis than females, not vice versa).

Finally, H1*-A1*, H1*-A2*, and H1*-A3* show no significant correlation with perceived breathiness in Czech speakers (Figure 16; for H1*-A3*: $P > 0.8$; otherwise $P = 0.2-0.4$). In Danish male speakers, however, stronger negative correlations between spectral shape/slope values and perceived breathiness can be observed ($r = -0.48$ to $-0.64$; $P < 0.01$). Even though Danish males are mostly perceived as rough rather than breathy, higher acoustic values are rated as more modal and, especially for the highest values, as breathier. There was no significant correlation in

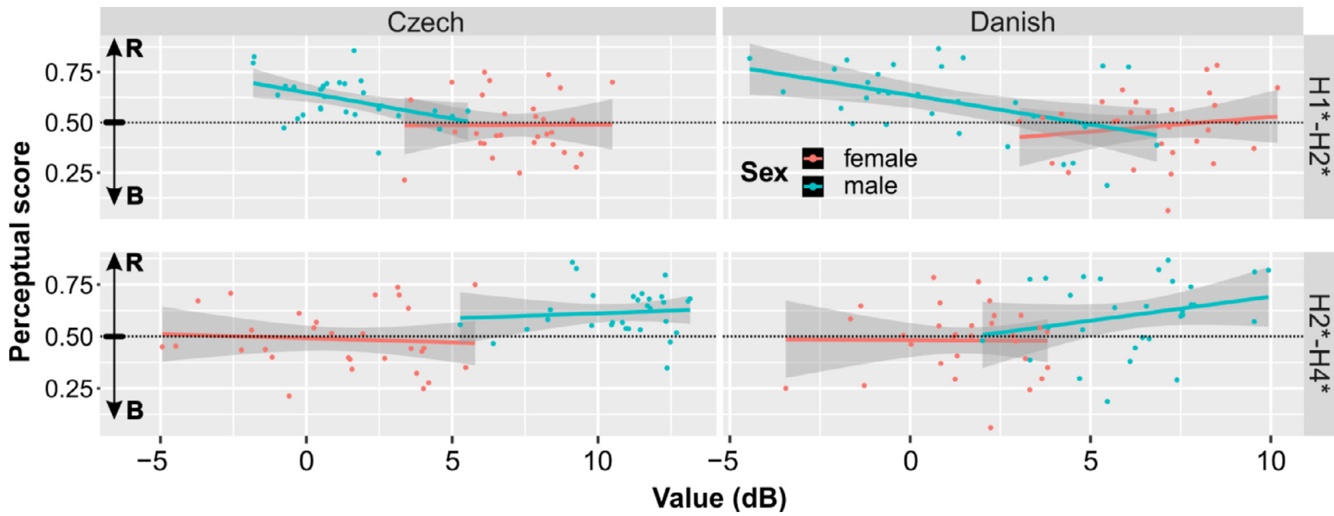


**FIGURE 15.** Scatterplots with trendlines (and 95% confidence bands). *X*-axis shows extracted mean values for H1*-H2* and H2*-H4*, *y*-axis displays the perceptual assessment of the 60 speakers.

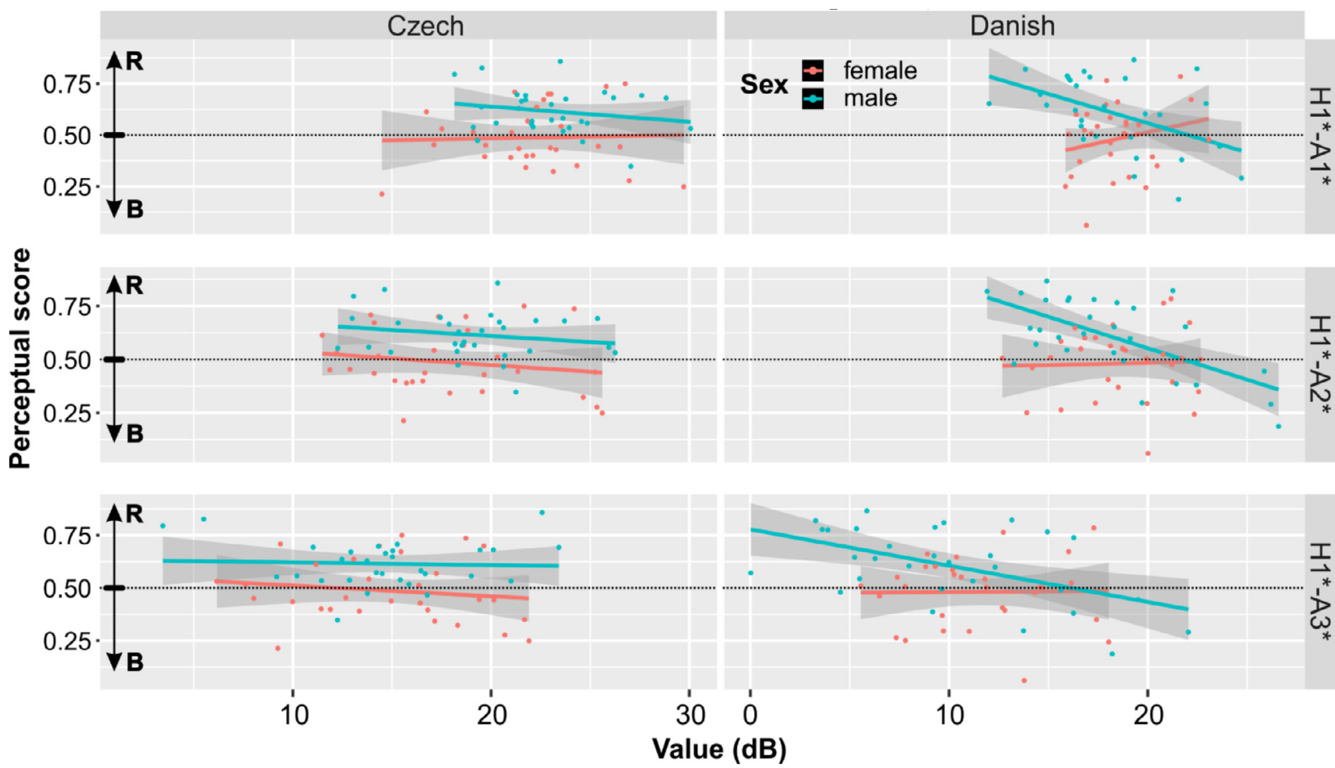


**FIGURE 16.** Scatterplots with trendlines (and 95% confidence bands). *X*-axis shows extracted mean values for H1*-A1* and H1*-A2* and H1*-A3*, *y*-axis displays the perceptual assessment of the 60 speakers.

Danish female speakers (for H1*-A1*: $r = 0.23$, $P = 0.213$; otherwise $P > 0.8$).

## DISCUSSION

The current paper focused on two primary, broader research questions and hypotheses. We report that our Hypothesis 1 is generally supported: female speakers are indeed breathier than male speakers in both Czech and Danish on the whole. Furthermore, some—although not many-language-specific differences in breathiness have been identified, which suggests that breathiness levels are indeed language specific to an extent in our dataset. However, the reported sex differences were generally consistent across the two languages (H2). In order to shed further light on these two hypotheses, we needed to address the issue of which measures predict sex differences in breathiness robustly (if any). The production and the perception components of our analyses consistently point toward CPP being the most reliable indicator of sex-related differences in breathiness (as well as roughness). We discuss the implications of the findings with respect to the hypotheses in more detail in what follows.

**Sex-related differences in breathiness**

From an acoustic point of view, the conclusion that female speakers in our dataset are generally breathier than male speakers is unambiguously suggested by the results based on CPP, H1*-H2* and H1*-A3*, and meshes well with the general claims about breathiness being a female characteristic.[16] Although this means only three measures out of seven, this is a crucial finding because CPP is considered a robust correlate of perceived breathiness,[55,61,56,57] and H1*-H2* is generally considered an acoustic measure well-suited to characterize differences along the glottal constriction continuum and is often associated with glottal open quotient[48,114]; it is also regarded as one of the most promising acoustic parameters to predict breathiness.[33] H1*-A3*, which reflects the source spectral tilt at higher formant frequencies,[115] also significantly yielded higher means in female speakers, suggesting a higher degree of breathiness in women. Moreover, perceptual evaluation by three experts (speech therapists) has shown that females in both languages are rated as breathier than males, who in contrast received higher roughness ratings. Nevertheless, the voices were still perceived within normophonic variation; no pathological cases were reported. Finally, when the seven acoustic parameters were taken as predictors of the perceptual score, CPP and H1*-H2* were among the best three predictors (along with HNR). H1*-A3* contributed to the model significantly only when considered alone (ie, without H1*-A1* and H1*-A2*), as the three measures appear to be highly correlated (this was also suggested by the correlation plots and by a PCA).

However, HNR and H2*-H4* point to the very opposite conclusion: male speakers' voices seem to be more aperiodic and breathier than female speakers'. In both Czech and Danish, our results show significant sex differences with higher HNR values in female speakers than in male speakers. This is in line with the results reported by Dehqan, Ansari, and Bakhtiar for Iranian speakers,[116] Ambreen et al for Pakistani speakers,[117] and Goy et al for English speakers.[118] Nevertheless, such findings are incongruous with the expectation that women should have lower HNR as they are assumed to be breathier than men, and their voice signals should thus exhibit higher amount of noise. However, HNR generally reflects aperiodicity rather than breathiness per se. Unlike CPP, HNR calculations might be affected by a number of factors, as explained by de Krom: "All kinds of signal properties may result in a noise-like appearance of the spectrum, such as a perturbation of the excitation signal (jitter and shimmer), rapid directional changes in fundamental frequency, formant transitions, and so forth." ([107]: p. 255). Thus, it can also reflect differences in perceived roughness. Indeed, the perceptual aspect of our analysis points out to HNR reflecting the fact that the male speakers were evaluated as rougher than the female speakers. Such an interpretation then renders the production results related to HNR not inconsistent with the claim that female speakers are generally reported to be breathier than male speakers. In our dataset, HNR reflects roughness rather than breathiness.

As for H2*-H4*, which is also commonly used to distinguish modal from breathy phonation,[49] our results do not suggest a higher degree of breathiness in female speakers, regardless of language. Acoustically, Czech males were breathier in terms of this measure than females, contrary to the perceptual evaluation (in which males were generally evaluated as rougher; furthermore, there was some degree of relationship between higher H2*-H4* values and increased perceived roughness in the Danish male speakers). The production-perception link shows clearly that H2*-H4* does not capture breathiness variation efficiently. Moreover, studies that have successfully used this measure[49] focused on phonemic contrasts rather than sociophonetic variation within phonemic categories. It might be the case that this measure is less sensitive to breathiness fluctuation than for instance H1*-H2*. Our finding corresponds to the results of Chen, Feng et al,[119] who also reported higher H2*-H4* mean values in male speakers, whereas Garellek, Samlan et al[68] reported lower H2*-H4* values in males. This discrepancy could be explained further by different methodological approaches: Chen, Feng et al's parameter extraction, similarly to ours, was based on read speech, whereas Garellek, Samlan et al's calculations used samples of sustained [a] which were inverse-filtered and copy-synthesized. It should also be borne in mind that H4 corresponds to 960 Hz in average female speakers and 480 Hz in males. Therefore, this measure may be more substantially affected by articulation (vowel reductions or vowel categories) than H1*-H2*, and may not necessarily be reliable for speech materials other than sustained low vowels. In any case, the H2*-H4* measure requires further exploration.

Finally, H1*-A1* did not show any significant sex difference in production. Similarly, H1*-A2* did not suggest a sex difference on the whole, but a closer inspection reveals that when the two languages are considered separately, females are breathier than males in Danish, thus following the generally reported trend for females to be breathier. As formants come into play with these measures, the same cautious note applies here as to H2*-H4*. Although the measurements are supposed to be corrected for the effects of F1, F2, and F3 vocal tract resonances on harmonic amplitudes, the VoiceSauce algorithm is automatic, and it is common knowledge that formant measurements frequently show errors and differ from those manually corrected. The extraction and correction of the investigated measures may thus not be reliable. In terms of perception, there seems to be no correlation between the three formant-related harmonic amplitude measures and perceived breathiness, except for the Danish male speakers. Moreover, the three parameters map onto perceptual scores almost identically (in contrast to the acoustic results, where only H1*-A3* showed a consistent difference). Taken together, we recommend that formant-related harmonic amplitude measures should be employed only in highly comparable conditions: identical (and unreduced) vowels, eg, sustained vowels used for voice disorder evaluations.

## Language-specific differences in breathiness

We were also interested in the question of language-specific details of breathiness. We postulated that if the two languages show different magnitudes of the sex-related differences in breathiness, this will suggest a sociocultural conditioning at play, whereas the absence of such an effect would be inconclusive about physiological vs social constraints (H2). We would like to argue that this hypothesis is not supported by the results reported here. At first blush, the fact that some measures point to differences between the Czech and Danish speakers do indeed suggest that we are dealing with language-specific differences in breathiness. A comparison of Czech and Danish in terms of CPP and H1*-A2* indicates that Czech speakers are less breathy than Danish speakers, whereas HNR and H1*-A1* suggest the opposite. More crucially, H2*-H4* and H1*-A2* showed a significant interaction between language and sex. With respect to H2*-H4*, the sex differences were more pronounced in Czech speakers than in Danish speakers (with female speakers being less breathy than male speakers), while in case of H1*-A2*, the sex differences displayed opposite tendencies across the two languages (Danish female speakers being breathier than male speakers as opposed to Czechs). However, as discussed in the previous section, those measures that most reliably reflect breathiness differences (CPP, H1*-H2*) do not in fact reveal language-specific patterns. There were no significant interactions of sex and language as reflected by these two measures, and only CPP showed a difference in the mean values between the languages. Moreover, H1*-A3* also behaved uniformly with regard to the two languages. This would therefore rather suggest that the sex differences reported for the Czech and Danish data in our study are not obviously due to sociocultural constraints.

In order to consider the issue further, apart from contrasting the Czech and Danish data, we also compared the obtained magnitudes of sex differences to studies examining other languages (Table 3). Focusing first on the measures which significantly indicated a greater degree of breathiness in women, the magnitude of the sex differences for CPP in our speakers is smaller compared to the (17-year-old) English speakers analyzed by Chen, Feng et al.[119] The differences found for H1*-H2* are somewhat greater than those for the speakers of Malayalam[115] but smaller than in English speakers reported by Hanson and Chuang[25] and Garellek, Samlan et al.[68] This seems to point out to breathiness being subject to specific cultural norms. Regarding H1*-A3*, our results suggest smaller sex differences than those available for Malayalam[115] and English,[25] which again points to culture-specific degree of variation in

**TABLE 3.**
**Comparison of Reported Effect Magnitudes Across Studies and Parameters (Only the Measures Yielding Significant Sex Differences in Our Study are Included)**

| Measure | Study | Language | Females | Males | F-M Difference |
|---|---|---|---|---|---|
| CPP | Our study | Czech | 22.5 | 23.1 | −0.6 dB |
| CPP | Our study | Danish | 21.5 | 21.8 | −0.3 dB |
| CPP | Chen et al (2010) | English | 23.2 | 24.7 | −1.5 dB |
| H1*-H2* | Our study | Czech | 7.2 | 1.3 | 5.9 dB |
| H1*-H2* | Our study | Danish | 6.8 | 1.1 | 5.7 dB |
| H1*-H2* | Narra et al (2015) | Malyalam | 11.5 | 7.2 | 4.3 dB |
| H1*-H2* | Hanson and Chuang (1999) | English | 3.1 | 0.0 | 3.1 dB |
| H1*-H2* | Garellek, Samlan et al (2013) | English | 8.9 | 6.1 | 2.8 dB |
| H1*A3* | Our study | Czech | 14.7 | 14.6 | 0.1 dB |
| H1*A3* | Our study | Danish | 11.6 | 10.0 | 1.6 dB |
| H1*A3* | Narra et al (2015) | Malyalam | 28.8 | 24.5 | 4.3 dB |
| H1*A3* | Hanson and Chuang (1999) | English | 23.4 | 13.8 | 9.6 dB |
| H2*-H4* | Our study | Czech | 1.1 | 10.7 | −9.6 dB |
| H2*-H4* | Our study | Danish | 1.6 | 6.3 | −4.7 dB |
| H2*-H4* | Chen et al (2010) | English | 2.4 | 6.2 | −3.8 dB |
| H2*-H4* | Garellek, Samlan et al (2013) | English | 11.6 | 8.9 | 2.6 dB |
| HNR | Our study | Czech | 34.0 | 24.5 | 9.5 dB |
| HNR | Our study | Danish | 41.6 | 31.7 | 9.9 dB |
| HNR | Chen et al (2010) | English | 32.7 | 25.3 | 7.4 dB |
| HNR | Dehqan et al (2008) | Persian | 18.8 | 18.4 | 0.4 dB |
| HNR | Ambreen et al (2017) | Urdu | 24.4 | 23.2 | 1.3 dB |

breathiness. Interestingly, HNR and H2*-H4* yield a greater degree of aperiodicity/breathiness in male speakers, and the magnitude of sex differences for HNR in our speakers is greater than in English speakers[119] and highly exceeds that found in Iranian speakers[116] and Pakistani speakers.[117] This might suggest either language specificity, or that other aspects of the voice are relevant for the given measures apart from breathiness. Sex differences for H2*-H4* in Czechs and Danes are greater than in English speakers.[119,68]

Although the sex differences reported in Table 3 might be affected by various factors (eg, speech material), the differences nevertheless suggest that breathiness is indeed language specific, and certainly not subject solely to physiological factors. While this is in no way a novel idea (eg,[4,120]), the quantitative analysis presented here, *when combined with those conducted by others*, provides some concrete, albeit not overwhelmingly obvious, support.

## Which measures best reflect breathiness?

As mentioned in the introduction, tackling the two hypotheses required an understanding of which measures predict sex differences in breathiness robustly (if any). Apart from the analyses targeting the individual measures employed here, we were also interested in whether the individual breathiness measures are correlated and how exactly and, if so, whether such correlations are affected by sex and/or language. Overall, females exhibited a larger number of strong correlations than males, and males yielded a moderate negative correlation between two of the measures: H1*-H2* and H2*-H4*. This can be explained by H1*-H2* reflecting, to some degree, perceived breathiness, and H2*-H4* being associated with roughness. The perceptual results also suggest that our male speakers seem to vary in how breathiness may be implemented in production exactly. On a more practical level, this trade-off should be borne in mind when interpreting results relying on single measures, and it may indeed also suggest that using a number of measures is a more prudent approach. The perceptual component of our analysis revealed that a difference in presumed breathiness quantified via an acoustic parameter, such as in CPP or H1*-H2*, does not uniformly correspond to a rise toward perceived breathiness, suggesting that different measures target different aspects of breathiness and possibly other voice qualities. Moreover, the perceptual evaluation presents a holistic approach in which some aspects of the voice may be stronger in the decision process than others. For instance, a decrease in CPP corresponded to an increase in breathiness for females and a decrease in roughness for males. In contrast, an increase in H1*-H2* was associated with a decrease in perceived roughness for males, but no obvious effect in perceived breathiness or roughness for females. The acoustic analyses revealed that the formant-related harmonic amplitude measures (H1*-A1*, etc.) are highly correlated, which was also the case in perception. They show very similar correlations with the perceptual score, and any of them is equally predictive in a multiple regression model (whereas the simultaneous presence of all three decreases their individual power). In the end, we conclude that CPP might be the single most relevant measure of breathiness (in line with[55−57]) since H1*-H2* (in perception) and especially HNR (in both perception and acoustics) seems to be closely connected to roughness.

## Limitations of the study

It might be argued that the dataset used in this study presents a fundamental confound: the two languages were studied through the lens of different vocalic segments and the recording equipment was not identical. In that case, any language-specific differences would be difficult to be labeled as such (for instance, some measures indicated language-specific differences in breathiness or roughness, on occasion interacting with the variable of sex as well). However, as pointed out in Methodology, the recording equipment and the recording conditions were comparable and do not present a confound in our sample. The instructions were identical, the microphones had a similar frequency response, and the SNR values were well above the recommended threshold.[91] Furthermore, as also argued in Methodology, the two languages are typologically not easily comparable from the point of view of the phonetic properties of their segmental inventories. Although we used a low central vowel /a/ for Czech, it was equally variable (due to reductions and contextual coarticulation) as the vowels from the Danish material, which underlyingly correspond to three different phonemes: /ɔ/, /ʌ/, and /a/. A narrower post hoc acoustic analysis of a more strictly comparable subset of the two vowel systems confirmed an absence of a confound. Moreover, vowel height and vowel length were included in the acoustic analysis as control variables (F1 and duration). Therefore, the issue should be minimalized. Finally, it could not directly affect the differences between male and female speakers, since both groups used the same speech material.

Another drawback that needs to be acknowledged is the fact that the perceptual component of our study included only three listeners (clinicians) and presented a fairly atypical and challenging task, albeit one that perfectly reflected the production data (ie, the exact same vowel tokens were used in the acoustic analysis and in the perceptual assessment). The challenge means that roughness ratings may have been due to aspects not to do with periodicity or amplitude perturbations, but rather due to the melodic variation (ie, frequent $f_0$ jumps between adjacent vowels). Roughness, which is connected to periodicity perturbations, might have thus been generally stronger in the evaluative process. This would explain why male voices were evaluated predominantly as rough and female voices as both rough and breathy. Moreover, as with many voice quality terms (see eg, Moisik et al[31]), roughness may have been interpreted to mean different things despite our best efforts to prevent this from happening. The assessors also reported that using a single scale for breathiness and roughness variation was challenging at times. We recommend that future research should use two separate scales to reflect these two phonatory dimensions, which may co-occur in a single token.

## CONCLUSIONS

When the most robust acoustic correlates of breathiness were used, it was found that female speakers are breathier than male speakers in both Czech and Danish. Although three formant-related harmonic amplitude measures seemed to be closely connected, the remaining acoustic measures were largely independent, which suggests that several measures should be used when examining breathiness, as each may describe different aspects of the phenomenon. The main finding corresponds to results reported for other languages and was also confirmed in a perceptual evaluation by three experts. Our study adds a novel angle on the analysis by looking into potential interactions between sex and language. Although we do find some, we conclude that these are not revealed by those measures that unambiguously and reliably correlate with perceived breathiness. However, a comparison of the effect magnitudes with studies of other languages suggests that the sex-related difference, reported in a range of languages, is to some degree conditioned socioculturally and not only physiologically. The implication for both researchers and clinicians is that, in the normophonic population of Czech and Danish speakers (as well as those of other languages), female speakers can be expected to show a higher degree of breathiness than male speakers, and that breathiness is more than a suitable variable of interest to target in sociolinguistic research. This may help clinicians and researchers in different countries to better understand that nonpathological vocal quality setting may vary according to the language and sex of the speaker.

### Acknowledgments

We are very grateful to Anna Jespersen for the translation of the Rainbow Passage into Danish. As always, thanks are due to our lovely participants, without whom this study could not have been carried out.

## APPENDICES

### Appendix A

Czech version of the Rainbow Passage. All target words (in bold) contain the low central vowel /a/ or /aː/ in the position indicated (underlined). Nouns and adjectives are in the nominative unless stated otherwise (gen. = genitive, acc. = accusative, loc. = locative, instr. = instrumental).

Když na dešťové **kapky** [/a/, *raindrops (acc.)*] dopadne ve vzduchu **paprsek** [/a/, *a ray*] světla, chovají se jako **hranol** [/a/, *a prism*] a vytvářejí duhu. Duha je rozdělení bílého světla do mnoha **nádherných** [/aː/, *beautiful (gen.)*] **barev** [/a/, *colours (gen.)*]. Ty **mají** [/a/, *have*] **tvar** [/a/, *the shape (acc.)*] dlouhého oblouku, vysoko na obloze, jehož konce **zdánlivě** [/aː/, *apparently*] končí za horizontem. Podle legendy se na jednom konci **nachází** [/a/, *there is*] kotlík se zlatem. Lidé ho hledají, ale nikdo ho ještě nenašel. Když se člověk **zadívá** [/a/, *looks (verb)*] do **dáli** [/aː/, *distance (gen.)*], jeho **přátelé** [/aː/, *friends*] říkají, že hledá kotlík **zlata** [/a/, *gold (gen.)*] na konci duhy. Během **staletí** [/a/, *centuries (gen.)*] lidé vysvětlovali duhu různě. Někteří ji přijali jako **zázrak** [/aː/, *a miracle (acc.)*] bez fyzikálního vysvětlení. Pro Hebrejce byla **znamením** [/a/, *a sign (instr.)*], že už nepřijde **žádná** [/aː/, *no (more)*] **další** [/a/, *more*] velká potopa. Řekové si představovali, že jde o **znamení** [/a/, *a sign (acc.)*] bohů předpovídající **válku** [/aː/, *a war (acc.)*] či silné deště. Norové považovali duhu za most, přes který bohové putují ze země do svého domu na nebesích. Jiní se **snažili** [/a/, *have tried*] vysvětlit fenomén duhy fyzikálně. Aristoteles mínil, že duhu vyvolává odraz slunečních **paprsků** [/a/, *rays (gen.)*] v dešti. Od té doby fyzici zjistili, že nejde o odraz, nýbrž o lom v dešťových **kapkách** [/a/, *raindrops (loc.)*], což **následně** [/aː/, *subsequently*] způsobuje duhu. Ohledně duhy se rozvinuly **také** [/a/, *also*] četné složité **nápady** [/aː/, *ideas*]. Rozdíly v duze **závisejí** [/aː/, *depend*] do velké míry na velikosti **kapek** [/a/, *raindrops (gen.)*] a šířka **barevných** [/a/, *colourful (gen.)*] pruhů **narůstá** [/a/, *increases (verb)*] spolu s tím. Skutečná **hlavní** [/a/, *primary*] duha, jíž pozorujeme, je údajně výsledkem **navršení** [/a/, *superimposition (gen.)*] několika duh. Pokud červená **barva** [/a/, *the colour*] druhého oblouku dopadá na zelenou prvního, vznikne ve výsledku oblouk s neobvykle širokým žlutým pruhem, jelikož červené a zelené světlo tvoří při smíchání žlutou. Toto je velmi běžný typ duhy, složený **hlavně** [/a/, *mainly*] z červené a žluté, s minimem zelené či modré.

### Appendix B

Danish version of the Rainbow Passage. Target words (in bold) are in the position indicated (underlined).

Når sollyset rammer **regndråber** [/ɔː/, *raindrops*] i **luften** [/ɔ/, *the air*], virker de som en prisme der **danner** [/a/, *forms*] (verb) en regnbue. En regnbue er en opbrydning af hvidt lys i mange **smukke** [/ɔ/, *beautiful*] farver. De **danner** [/a/, *forms*] (verb) en lang, **rund** [/ɔ/, *round*] hvælving, med **toppen** [/ʌ/, *the top*] af buen højt til vejrs og dens ender tilsyneladende udenfor **horisonten** [/ʌ/, *the horizon*]. Legender fortæller at der er en kogende **krukke** [/ɔ/, *cauldron*] guld i den ene ende. **Folk** [/ʌ/, *people*] kigger efter den men finder den aldrig. Når **folk** [/ʌ/, *people*] søger noget udenfor egen **formåen** [/ɔː/, *the capability*], siger man, at de kigger efter en **krukke** [/ɔ/, *cauldron*] guld for enden af regnbuen. Igennem tiderne har mennesker forsøgt at forklare regnbuen på forskellig vis. Nogen har accepteret den som et mirakel uden en fysisk forklaring. For hebræerne var den et tegn på at der ikke ville **komme** [/ʌ/, *to come*] flere syndfloder. Grækerne forestillede sig at den var et tegn fra guderne på krig eller voldsom regn. Nordboerne betragtede regnbuen som en bro som guderne passerede over fra jorden til deres hjem i himmelen. Andre har forsøgt at forklare fænomenet i fysiske termer. Aristoteles skrev at regnbuen skabtes gennem spejlinger af solens **stråler** [/ɔː/, *rays*] i regnen. Siden da har fysikere **fundet** [/ɔ/, *found*] ud af at det ikke er refleksion men refraktion i **regndråberne** [/ɔː/, *the raindrops*] som **skaber** [/aː/, *creates*] regnbuerne. Man har haft mange komplicerede ideer om regnbuen. Forskellene i regnbuen er stærkt afhængige af størrelsen på **dråberne** [/ɔː/, *the drops*], og bredden af det farvede **bånd** [/ʌ/, *band*] **vokser** [/ʌ/, *grow(s)*] når størrelsen på **dråberne** [/ɔː/, *the drops*] **vokser** [/ʌ/, *grow(s)*]. Den egentlige, primære regnbue, som man kan se, siges at være skabt af et antal buer lagt ovenpå **hinanden** [/a/, *each other*]. Hvis den røde bue overlapper med den første bues grønne **bånd** [/ʌ/, *band*] er resultatet et abnormt bredt gult **bånd** [/ʌ/, *band*], idet rødt og grønt lys **danner** [/a/, *forms*] (verb) gult lys når de **blandes** [/a/, *mix*] (verb). Det giver en meget almindelig type bue, som hovedsageligt består af rødt og gult, med lidt eller intet grønt eller blåt.

## Appendix C

Summaries of the optimal statistical models (i.e., with or without and interaction between SEX and LANGUAGE) for individual parameters. The baseline factor levels were Czech, female, nonsmoker, without creak, in non-nasal contexts. (Tables C.1, C.2, C.3, C.4, C.5, C.6, C.7)

**TABLE C.1.**
**Output of the Optimal Model for CPP**

| | Estimate | SE | *t* Value |
|---|---|---|---|
| (Intercept) | 21.228 | 0.736 | 28.827 |
| Language (Danish) | −1.259 | 0.342 | −3.688 |
| Sex (male) | 0.780 | 0.281 | 2.772 |
| Smoker (yes) | 0.012 | 0.435 | 0.028 |
| Age | −0.048 | 0.021 | −2.304 |
| $F_1$ | 0.001 | 0.0002 | 3.263 |
| $F_0$ | 0.003 | 0.002 | 1.770 |
| Creak (yes) | −0.566 | 0.235 | −2.409 |
| Nasal context (yes) | 0.527 | 0.259 | 2.036 |
| Duration | 0.015 | 0.002 | 7.647 |

**TABLE C.2.**
**Output of the Optimal Model for H1*-H2***

| | Estimate | SE | *t* Value |
|---|---|---|---|
| (Intercept) | −3.572 | 1.106 | −3.229 |
| Language (Danish) | −0.559 | 0.397 | −1.408 |
| Sex (male) | −1.922 | 0.460 | −4.179 |
| Smoker (yes) | −1.260 | 0.655 | −1.923 |
| Age | 0.085 | 0.031 | 2.732 |
| $F_1$ | 0.0001 | 0.0003 | 0.242 |
| $F_0$ | 0.045 | 0.003 | 14.965 |
| Creak (yes) | −0.124 | 0.388 | −0.319 |
| Nasal context (yes) | −0.378 | 0.211 | −1.793 |
| Duration | −0.005 | 0.003 | −1.793 |

**TABLE C.3.**
**Output of the Optimal Model for H2*-H4***

| | Estimate | SE | *t* Value |
|---|---|---|---|
| (Intercept) | 2.869 | 1.281 | 2.240 |
| Language (Danish) | 2.218 | 0.515 | 4.311 |
| Sex (male) | 5.020 | 0.700 | 7.169 |
| Smoker (yes) | 0.344 | 0.566 | 0.609 |
| Age | 0.005 | 0.027 | 0.165 |
| $F_1$ | 0.018 | 0.001 | 31.212 |
| $F_0$ | −0.061 | 0.004 | −14.251 |
| Creak (yes) | 1.225 | 0.640 | 1.916 |
| Nasal context (yes) | −0.053 | 0.279 | −0.190 |
| Duration | 0.001 | 0.004 | 0.177 |
| Lng (Dn): Sex (M) | −4.763 | 0.860 | −5.538 |

**TABLE C.4.**
**Output of the Optimal Model for H1*-A1***

| | Estimate | SE | *t* Value |
|---|---|---|---|
| (Intercept) | 7.606 | 1.364 | 5.58 |
| Language (Danish) | −1.913 | 0.511 | −3.74 |
| Sex (male) | −0.073 | 0.551 | −0.13 |
| Smoker (yes) | −1.194 | 0.817 | −1.46 |
| Age | 0.096 | 0.039 | 2.48 |
| $F_1$ | 0.027 | 0.0004 | 64.39 |
| $F_0$ | −0.018 | 0.0037 | −4.89 |
| Creak (yes) | 0.258 | 0.4627 | 0.56 |
| Nasal context (yes) | 1.374 | 0.2934 | 4.69 |
| Duration | −0.003 | 0.0033 | −1.01 |

**TABLE C.5.**
**Output of the Optimal Model for H1*-A2***

| | Estimate | SE | *t* Value |
|---|---|---|---|
| (Intercept) | −3.223 | 1.741 | −1.85 |
| Language (Danish) | 4.598 | 1.051 | 4.38 |
| Sex (male) | 0.791 | 0.871 | 0.91 |
| Smoker (yes) | −1.152 | 1.065 | −1.08 |
| Age | 0.2004 | 0.050 | 3.98 |
| $F_1$ | 0.037 | 0.0004 | 89.69 |
| $F_0$ | −0.025 | 0.004 | −6.50 |
| Creak yes | 0.035 | 0.465 | 0.07 |
| Nasal context (yes) | −0.476 | 0.681 | −0.70 |
| Duration | −0.0101 | 0.004 | −2.50 |
| Lng (Dn): Sex (M) | −2.407 | 1.136 | −2.12 |

**TABLE C.6.**
**Output of the Optimal Model for H1*-A3***

| | Estimate | SE | *t* Value |
|---|---|---|---|
| (Intercept) | −6.178 | 2.150 | −2.87 |
| Language (Danish) | −0.430 | 0.895 | −0.48 |
| Sex (male) | −1.788 | 0.850 | −2.11 |
| Smoker (yes) | −1.174 | 1.328 | −0.88 |
| Age | 0.237 | 0.063 | 3.76 |
| $F_1$ | 0.038 | 0.001 | 64.92 |
| $F_0$ | −0.029 | 0.005 | 5.29 |
| Creak (yes) | −0.007 | 0.658 | −0.01 |
| Nasal context (yes) | 0.501 | 0.571 | 0.88 |
| Duration | −0.021 | 0.005 | −3.99 |

**TABLE C.7.**
**Output of the Optimal Model for HNR**

| | Estimate | SE | *t* Value |
|---|---|---|---|
| (Intercept) | 22.657 | 1.830 | 12.381 |
| Language (Danish) | 6.949 | 1.141 | 6.089 |
| Sex (male) | −6.765 | 0.661 | −10.233 |
| moker (yes) | 0.367 | 1.046 | 0.350 |
| Age | 0.121 | 0.050 | 2.453 |
| $F_1$ | −0.002 | 0.001 | −3.348 |
| $F_0$ | 0.035 | 0.004 | 8.155 |
| Creak (yes) | −1.484 | 0.522 | −2.840 |
| Nasal context (yes) | −0.655 | 1.011 | −0.647 |
| Duration | 0.029 | 0.005 | 6.375 |

## Appendix D

**TABLE D.1.**
**Proportion of Variance Explained by the Principal Components in PCA and the Mapping of the First 3 Components onto the Acoustic Parameters**

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 |
|---|---|---|---|---|---|---|---|
| SD | 1.65 | 1.35 | 1.04 | 0.79 | 0.62 | 0.47 | 0.41 |
| Proportion of variance | 0.39 | 0.26 | 0.15 | 0.09 | 0.06 | 0.03 | 0.02 |
| Cumulative proportion | 0.39 | 0.65 | 0.80 | 0.89 | 0.95 | 0.98 | 1.00 |

| | H1*-H2* | H2*-H4* | H1*-A1* | H1*-A2* | H1*-A3* | CPP | HNR |
|---|---|---|---|---|---|---|---|
| PC1 | 0 | −0.59 | −0.88 | −0.88 | −0.87 | 0.09 | 0.24 |
| PC2 | 0.88 | −0.62 | 0.12 | 0.26 | 0.21 | −0.19 | 0.71 |
| PC3 | −0.02 | 0.12 | 0.16 | −0.01 | −0.03 | 0.95 | 0.36 |

## REFERENCES


1. Foulkes P, Scobbie J, Watt D. Sociophonetics. In: Hardcastle J, Laver J, Gibbon E, eds. *The Handbook of Phonetic Sciences*. 2nd ed. Blackwell Publishing; 2010:703–754.
2. Napoli J Di. The phonetics and phonology of glottalization in Italian. *Peter Lang*. Berlin; 2018.
3. Heldner M, Włodarczak M, Beňuš Š, et al. Voice quality as a turn-taking cue. In: *Proceedings of Interspeech 2019*. Graz; 20194165–4169.
4. Hall K. Lip service on the fantasy lines. In: Hall, Bucholtz, eds. *Gender Articulated: Language and the Socially Constructed Self*. 1995183–226. Routledge, NY.
5. Podesva R, Callier P. Voice quality and identity. *Ann Rev Appl Linguist*. 2015;35:173–194.
6. Liu X, Xu Y. What makes a female voice attractive? *17th ICPhS*. Hong Kong; 2011:1274–1277.
7. Gobl C, Ní Chasaide A. The role of voice quality in communicating emotion, mood and attitude. *Speech Commun*. 2003;40:189–212.
8. Helgason P. *Preaspiration in the Nordic Languages. Synchronic and Diachronic Aspects*. PhD thesis. Stockholm University; 2002.
9. Ball M, Lowry O, McInnis L. Distributional and stylistic variation in /r/-misarticulations: a case study. *Clin Linguist Phons*. 2009;20:119–124.
10. Gillon GT, Moriarty BC. Childhood apraxia of speech: children at risk for persistent reading and spelling disorder. *Semin Speech Lang*. 2007;28:48–57.
11. Phoon HS, Abdullah AC, Maclagan M. The effect of dialect on the phonological analysis of Chinese-influenced Malaysian English speaking children. *Int J Speech-Lang Pathol*. 2012;14:487–498.
12. Foulkes P, Docherty GJ. Variation in British English /r/. In: de Velde, van Hout, eds. *'r-atics: Sociolinguistic, Phonetic and Phonological Characteristics of /r/*. 4, Institut des Langues Vivantes et de Phonétique, Etudes & Travaux; 2001:27–43.
13. Foulkes P, Docherty GJ. Another chapter in the story of /r/: 'labiodental' variants in British English. *J Sociolinguist*. 2002;4:30–59.
14. Clark L, Trousdale G. Exploring the role of token frequency in phonological change: evidence from TH-fronting in east-central Scotland. *English Lang Linguist*. 2009;13:33–55.
15. van Borsel J, Janssens J, de Bodt M. Breathiness as a feminine voice characteristic: a perceptual approach. *J Voice*. 2009;23:291–294.
16. Henton CG, Bladon RAW. Breathiness in normal female speech: inefficiency versus desirability. *Lang Commun*. 1985;5:221–227.
17. Price PJ. Male and female voice source characteristics: inverse filtering results. *Speech Commun*. 1989;8:261–277.

18. Pépiot E. Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. *Speech Prosody*. Dublin; 2014305–309.
19. Teshigawara M. *Voices in Japanese Animation: A Phonetic Study of Vocal Stereotypes of Heroes and Villains in Japanese Culture*, PhD thesis. University of Victoria; 2003.
20. Winter, Grawunder S. The phonetic profile of Korean formal and informal speech registers. *J Phon*. 2012;40:808–815.
21. Mendoza E, Valencia N, Mufioz J, et al. Differences in voice quality between men and women: use of the Long-Term Average Spectrum (LTAS). *J Voice*. 1996;10:59–66.
22. Simpson A. Phonetic differences between male and female speech. *Lang Linguist Compass*. 2009;3:621–640. a.
23. Simpson A. The first and second harmonics should not be used to measure breathiness in male and female voices. *J Phon*. 2012;40:477–490.
24. Chen G, Kreiman J, Shue Y-L, et al. Acoustic correlates of glottal gaps. In: *Proceedings of Interspeech*. Florence; 20112673–2676.
25. Hanson HM, Chuang ES. Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *JASA*. 1999;106:1064–1077.
26. Rammage LA, Peppard RC, Bless DM. Aerodynamic, laryngoscopic, and perceptual-acoustic characteristics in dysphonic females with posterior glottal chinks: a retrospective study. *J Voice*. 1992;6:64–78.
27. Gorham-Rowan MM, Laures-Gore J. Acoustic-perceptual correlates of voice quality in elderly men and women. *J Commun Disord*. 2006;39:171–185.
28. Hejná M. A case study of menstrual cycle effects: global phonation or also local phonatory phenomena? *19th ICPhS*. Melbourne; 20192630–2634.
29. Abitbol J, Abitbol P, Abitbol B. Sex hormones and the female voice. *J Voice*. 1999;13:424–446.
30. Södersten M, Lindestad PÅ. Glottal closure and perceived breathiness during phonation in normally speaking subjects. *J Sp Hear Res*. 1990;33:601–611.
31. Moisik S, Hejná M, Esling J. Abducted vocal fold states and the epilarynx: a new taxonomy for distinguishing breathiness and whisperiness. *19th ICPhS*. Melbourne; 2019.
32. Dejonckere PH. Principal components in voice pathology. *Voice*. 1995;4:96–105.
33. Barsties v. Latoszek B, Maryn Y, Gerrits E, et al. Acoustic measurement of roughness and breathiness. *J Speech Lang Hear Res*. 2018;61:298–323.
34. Laver J. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press; 1980/2009.
35. Laver J. *Principles of Phonetics*. Cambridge: Cambridge University Press; 1994.
36. Zemlin WR. *Speech and Hearing Science: Anatomy and Physiology*. Boston: Allyn and Bacon; 1998.
37. Esling JH, Harris JG. States of the glottis: an articulatory phonetic model based on laryngoscopic observations. In: Hardcastle, Beck, eds. *A Figure of Speech: A Festschrift for John Laver*. NY / London: Routledge; 2005:347–383.
38. Gordon M, Ladefoged P. Phonation types: a cross-linguistic overview. *J Phon*. 2001;29:383–406.
39. Stevens KN. *Acoustic Phonetics*. Cambridge: The MIT Press; 1998.
40. Gordeeva O, Scobbie JM. Preaspiration as a correlate of word-final voice in Scottish English fricatives. In: Fuchs, Żygis, eds. *Turbulent Sounds: An Interdisciplinary Guide*. Berlin: Mouton de Gruyter; 2010:167–207.
41. Hanson HM. Glottal characteristics of female speaker: acoustic correlates. *JASA*. 1997;101:466. https://doi.org/10.1121/1.417991.
42. Gordeeva O, Scobbie JM. A phonetically versatile contrast: pulmonic and glottalic voicelessness in Scottish English obstruents and voice quality. *JIPA*. 2013;43:249–271.
43. Klatt D, Klatt L. Analysis, synthesis and perception of voice quality variations among female and male talkers. *JASA*. 1990;87:820–857.
44. Barsties v.Latoszek B, Lehnert B, Janotte B. Validation of the acoustic voice quality index version 03.01 and acoustic breathiness index in German. *J Voice*. 2018:1–9. [Articles in Press].
45. Lindqvist-Gauffin J. A descriptive model of laryngeal articulation in speech. *STL-QPSR*. 1972;12:1–14.
46. Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *J Voice*. 2003;18:299–304.
47. Carding PN, Steen IN, Webb A, et al. The reliability and sensitivity to change of acoustic measures of voice quality. *Clin Otolaryngol*. 2004;29:538–544.
48. Keating PA, Esposito C. Linguistic voice quality. *UCLA Working Papers in Phonetics*. 2007;105:85–91.
49. Garellek M, Keating P, Esposito CM, et al. Voice quality and tone identification in White Hmong. *J Acoust Soc Am*. 2013;133:1078–1089.
50. Garellek M, Keating P. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *JIPA*. 2011;41:185–205.
51. Maryn Y, de Bodt M, Roy N. The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders. *J Commun Disord*. 2010;43:161–174.
52. Hirano M. *Clinical Examination of Voice*. New York: Springer Verlag; 1981.
53. Barsties v. Latoszek B, Maryn Y, Gerrits E, et al. The Acoustic Breathiness Index (ABI): a multivariate acoustic model for breathiness. *J Voice*. 2017;31(4). 511.e11−511.e27.
54. Labuschagne IB, Ciocca V. The perception of breathiness: acoustic correlates and the influence of methodological factors. *Acoust Sci Tech*. 2016;37:191–201.
55. Hillenbrand J, Cleveland R, Erickson R. Acoustic correlates of breathy vocal quality. *J Sp Hear Res*. 1994;37:769–778.
56. Samlan RA, Story BH, Bunton K. Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling. *J Speech Lang Hear Res*. 2013;56:1209–1223.
57. Shrivastav R, Sapienza CM. Objective measures of breathy voice quality obtained using an auditory model. *JASA*. 2003;114:2217–2224.
58. Maryn Y, Weenink D. Objective dysphonia measures in the program Praat: smoothed cepstral peak prominence and acoustic voice quality index. *J Voice*. 2014;29:35–43.
59. Awan SN, Roy N, Dromey C. Estimating dysphonia severity in continuous speech: application of a multi-parameter spectral/cepstral model. *Clin Linguist Phon*. 2009;23:225–841.
60. Ferrer Riesgo CA, Nöth E. What makes the Cepstral Peak Prominence different to other acoustic correlates of vocal quality. *J Voice*. 2019:1–6. [Article in Press].
61. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Sp Hear Res*. 1996;39:311–321.
62. Phadke KV, Laukkanen A-M, Illomäki I, et al. Cepstral and perceptual investigations in female teachers with functionally healthy voice. *J Voice*. 2018:1–11. [Articles in Press].
63. Esposito CM. An acoustic and electroglottographic study of White Hmong tone and phonation. *J Phon*. 2012;40:466–476.
64. Cannito MP, Doiuchi M, Murry T, et al. Perceptual structure of adductor spasmodic dysphonia and its acoustic correlates. *J Voice*. 2012;26. 818.e5-818.e13.
65. Teixeira JP, Oliveira C, Lopes C. Vocal acoustic analysis - jitter, shimmer and HNR parameters. *Proc Technol*. 2013;9:1112–1122.
66. Hartl DM, Hans S, Vaissière J, et al. Objective acoustic and aerodynamic measures of breathiness in paralytic dysphonia. *Eur Arch Otorhinolaryngol*. 2003;260:175–182.
67. Kreiman J, Gerratt BR, Precoda K, et al. Individual differences in voice quality perception. *J Sp Hear Res*. 1992;35:512–520.
68. Garellek M, Samlan RA, Kreiman J, et al. Perceptual sensitivity to a model of the source spectrum. *Proc Meetings Acoust*. 2013;19:1–5.
69. Wayland R, Jongman A. Acoustic correlates of breathy and clear vowels: the case of Khmer. *J Phon*. 2003;31:181–201.

70. Kreiman J, Gerratt BR, Garellek M, et al. Toward a unified theory of voice production and perception. *Loquens*. 2014;1(1):e009.
71. Childers DG, Lee CK. Vocal quality factors: analysis, synthesis, and perception. *JASA*. 1991;90:2394–2410.
72. Awan SN, Solomon NP, Helou LB, et al. Spectral-cepstral estimation of dysphonia severity: external validation. *Ann Otol Rhinol Laryngol*. 2013;122:40–48.
73. Murphy PJ. Periodicity estimation in synthesized phonation signals using cepstral rahmonic peaks. *Speech Commun*. 2006;48:1704–1713.
74. Lindholm P, Vilkman E, Raudaskoski T, et al. The effect of postmenopause and postmenopausal HRT on measured voice values and vocal symptoms. *Maturitas*. 1997;28:47–53.
75. Raj A, Gupta B, Chowdhury A, et al. A study of voice changes in various phases of menstrual cycle and in postmenopausal women. *J Voice*. 2008;24:363–368.
76. Femke A, Bartelink M-LEL, Grobbee DE, et al. Postmenopausal status and early menopause as independent risk factors for cardiovascular disease: a meta-analysis. *Menopause*. 2006;13:265–279.
77. Jesus LMT, Barney A, Santos R, et al. Universidade de Aveiro's Voice Evaluation Protocol. In: *Proceedings of Interspeech*. Brighton; 2009971–974.
78. Friedrich G, Dejonckere PH. The voice evaluation protocol of the European Laryngological Society (ELS) – first results of a multicenter study. *Laryngo-Rhino-Otologie*. 2005;84:744–752.
79. Herbst CT, Oh J, Vydrová J, et al. Digital VHI—a freeware open-source software application to capture the Voice Handicap Index and other questionnaire data in various languages. *Logoped Phoniatr Vocol*. 2015;40:72–76.
80. Fairbanks G. *Voice and Articulation Drill Book*. NY: Harper & Row; 1960.
81. Jesus LMT, Valente ARS, Hall A. Is the Portuguese version of the passage 'The North Wind and the Sun' phonetically balanced? *JIPA*. 2015;45(1):1–11.
82. MacKanzie K, Millar A, Wilson JA, et al. Is voice therapy an effective treatment for dysphonia? A randomized controlled trial. *Br Med J*. 2001;232:1–6.
83. Levon E. Hearing "gay": prosody, interpretation, and the affective judgements of men's speech. *Am Speech*. 2006;81:56–78.
84. Clopper CG, Pisoni DB. The Nationwide Speech Project: a new corpus of American English dialects. *Speech Commun*. 2006;48:633–644.
85. Labov W. *Principles of Linguistic Change. Volume 1: Internal Factors*. Oxford: Basil Blackwell; 2001.
86. Labov W. *Principles of Linguistic Change. Volume II: Social Factors*. Oxford: Basil Blackwell; 2001.
87. Labov W. *Principles of Linguistic change. Volume III: Cognitive and Cultural Factors*. Oxford: Basil Blackwell; 2001.
88. Podesva R. Phonation type as a stylistic variable: the use of falsetto in constructing a persona. *J Sociolinguist*. 2007;11:478–504.
89. Schilling-Estes N. Self-conscious' speech: the performance register in Ocracoke English. *Lang Soc*. 1998;27:53–83.
90. Simpson A. Breathiness differences in male and female speech. Is H1-H2 an appropriate measure? *FONETIK 2009*. Stockholm; 2009. b.
91. Švec JG, Granqvist S. Guidelines for selecting microphones for human voice production research. *Am J Speech Lang Pathol*. 2010;19:356–368.
92. Schutte HK, Seidner W. Recommendation by the Union of European Phoniatricians (UEP): standardizing voice area measurement/phonetography. *Folia Phoniatr Logop*. 1983;35:286–288.
93. Uther M, Knoll MA, Birnham D. Do you speak E-NG-L-I-SH? A comparison of foreign- and infant-directed speech. *Speech Commun*. 2007;49(1):2–7.
94. Pollák P, Volín J, Skarnitzl R. HMM-based phonetic segmentation in Praat environment. In: *Proceedings of SPECOM 2007*. Moscow; 2007537–541.
95. Machač P, Skarnitzl R. *Principles of Phonetic Segmentation*. Prague: Epocha; 2009.
96. Ogden R. Turn transition, creak and glottal stop in Finnish talk-in-interaction. *JIPA*. 2001;31:139–152.
97. Ogden R. Non-modal voice quality and turn-taking in Finnish. In: Couper-Kuhlen, Ford, eds. *Sound Patterns in Interaction: Cross-Linguistic Studies from Conversation*. Amsterdam/ Philadelphia: John Benjamins; 2004:29–62.
98. Reed BS, Persson R. How speakers of different languages extend their turns: word linking and glottalization in French and German. *Res Lang Soc Interact*. 2016;49:128–147.
99. Keating P, Garellek M, Kreiman J. Acoustic properties of different kinds of creaky voice. *18th ICPhS*. Glasgow; 2015. ICPHS0821.
100. Fischer-Jørgensen E. Phonetic analysis of the stød in Standard Danish. *Phonetics*. 1989;46:1–59.
101. Y-L Shue, VoiceSauce: a program for voice analysis, version 1.36. http://www.phonetics.ucla.edu/voicesauce/(accessed 24 April 2019).
102. Shue Y-L, Keating P, Vicenik C, et al. VoiceSauce: a program for voice analysis. *17th ICPhS*. Hong Kong; 20111846–1849.
103. Harper Berkson K. Acoustic correlates of breathy sonorants in Marathi. *J Phon*. 2019;73:70–90.
104. Kelterer A, Schuppler B. Acoustic correlates of phonation type in Chichimec. In: *Proceedings of Interspeech 2019*. Graz, Austria; 20191981–1985. 15-19 September.
105. Kawahara H, Masuda-Katsuse I, de Cheveigné A. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction. *Speech Commun*. 1999;27:187–207.
106. K Sjölander, The Snack Sound Toolkit. http://www.speech.kth.se/snack/, 1997–2004.
107. de Krom G. A cepstrum-based technique for determining a harmonic-to-noise ratio in speech signals. *J Sp Hear Res*. 1993;36:254–266.
108. VoiceSauce Manual. http://www.phonetics.ucla.edu/voicesauce/documentation/parameters.html(accessed 15 May 2019).
109. Ferrand CT. Harmonics-to-noise ratio: an index of vocal ageing. *J Voice*. 2002;16:480–487.
110. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/(accessed 20 June 2019).
111. R Studio. Version 1.2.1335, 2009-2019. http://www.rstudio.com (accessed 20 June 2019).
112. D Bates, M Maechler, B Bolker, et al. lme4, (2019). https://cran.r-project.org/web/packages/lme4/lme4.pdf.
113. J Fox, S Weisberg, B Price, et al. R Core Team, Package 'effects'. https://cran.r-project.org/web/packages/effects/effects.pdf.
114. Chen G, Park SJ, Kreiman J, et al. Investigating the effect of F0 and vocal intensity on harmonic magnitudes: Data from high-speed laryngeal videoendoscopy. *15th Annual Conference of ISCA*. Singapore; 20141668–1672.
115. Narra M, Anu TD, Varghese SM, et al. Harmonic amplitude measures to note gender differences. *Adv Life Technol*. 2015;31:17–23.
116. Dehqan A, Ansari H, Bakhtiar M. Objective voice analysis of Iranian speakers with normal voices. *J Voice*. 2008;24:161–167.
117. Ambreen S, Bashir N, Tarar SA, et al. Acoustic analysis of normal voice patterns in Pakistani adults. *J Voice*. 2017;33. 124.e49–124.e58.
118. Goy H, Fernandes DN, Pichora-Fuller MK, et al. Normative voice data for younger and older adults. *J Voice*. 2013;27:545–555.
119. Chen G, Feng X, Shue Y-L, et al. On using voice source measures in automatic gender classification of children's speech. *11th Annual Conference of ISCA, Makuhari*. Japan; 2010673–676.
120. Yiu EM-L, Murdoch B, Hird K, et al. Cultural and language differences in voice quality perception: a preliminary investigation using synthesized signals. *Folia Phoniatr Logop*. 2008;60:107–119.