Review Article

Chloe Patman*, Paul Foulkes and Kirsty McDougall

Acoustic methods for analysing breathy and whispery voices: a systematic review

https://doi.org/10.1515/phon-2025-0007
Received January 24, 2025; accepted July 10, 2025; published online August 4, 2025

Abstract: Voice quality (VQ) is frequently analysed in phonetic and phonological research. Recently, there has been motivation to assess VQ more 'objectively' using acoustic analysis. Our systematic review revealed a notable research focus on pathological speakers, with 48 % of excluded studies being rejected for investigating pathological populations. Among studies involving non-pathological speakers, the analysed speech material is often restricted to sustained vowel productions. Therefore, the suitability of acoustic techniques when analysing more naturalistic speech and non-pathological speakers remains unclear. We present a systematic literature review of acoustic methods used to categorise breathy and whispery VQ in non-pathological speakers. The literature was surveyed using four databases (ProQuest, PubMed, SCOPUS and Web of Science) and ICPhS proceedings between 1999 and 2023. The selection criteria included peer-reviewed articles conducting an acoustic analysis of breathy and whispery voices for vocally healthy speakers. Initial searches yielded 754 papers. Once filtered, 21 papers remained. The results reveal some consistency in the main acoustic parameters for breathy VQ (higher spectral tilt and lower HNR/CPP). Whispery voice was only addressed in two studies, meaning no trends were observed. We conclude that there remain inconsistencies in methods and findings, and thus we cannot identify an agreed 'standard' approach generally applicable to non-pathological speakers.

Keywords: voice quality; acoustic analysis; breathy; whispery; systematic review

***Corresponding author: Chloe Patman**, Department of Theoretical and Applied Linguistics, University of Cambridge, Cambridge, CB3 9DA, UK; and Darwin College, Cambridge, CB3 9EU, UK, E-mail: cep72@cam.ac.uk. https://orcid.org/0009-0009-8790-2958
Paul Foulkes, Department of Language and Linguistic Science, University of York, York, YO10 5DD, UK, E-mail: paul.foulkes@york.ac.uk
Kirsty McDougall, Department of Theoretical and Applied Linguistics, University of Cambridge, Cambridge, CB3 9DA, UK, E-mail: kem37@cam.ac.uk

1 Introduction

1.1 Overview of vocal setting and voice quality analysis

This paper aims to clarify what is known empirically about the suitability of acoustic parameters to categorise breathy and whispery voices in non-pathological speakers. Our systematic review revealed a notable research focus on pathological speakers, with nearly 50 % of the studies we excluded being rejected on the basis that they investigated pathological populations. Among studies involving non-pathological speakers, the analysed speech material is most often restricted to sustained vowel productions. Furthermore, due to the subjective nature of many decisions involved in acoustic analysis, a wide variety of techniques and methods is present in the literature. Due to this variability, the present study conducted a systematic review to better understand the findings concerning acoustic voice quality analysis in non-pathological speakers. The specific applications of voice quality analysis relevant to the present study are discussed in Section 1.2.

Vocal settings (VSs), as defined by Laver (1980: 1), refer to the "overall auditory colouring of an individual speaker's voice". VSs encompass the general vocal configuration and characteristics of a speaker's voice, whereas voice qualities (VQs), describe more specific modifications within the larynx. Therefore, VSs capture the overall vocal profile, while VQs focus specifically on laryngeal adjustments (Laver 1980).

VSs are generally analysed using three broad families of techniques:

A. Perceptual analysis (i.e., how VSs are interpreted by listeners).
B. Analysis of vocal production (e.g., laryngoscopy, which involves the instrumental examination of articulatory configuration).
C. Analysis of acoustic output (e.g., spectral outcomes).

Given their multidimensional nature, VSs are often analysed perceptually. Various protocols have been developed for the perceptual assessment of VSs, many of which originate in speech and language pathology. In the UK, the most widely used system is probably the Vocal Profile Analysis (VPA) scheme (Laver 1980). The VPA allows an analyst to categorise the presence and degree of several different VSs on a gradient scale. Various versions of the VPA can be found. The majority refer to between 30 and 40 vocal settings, primarily defined by their main articulatory regions, along with dimensions to capture overall muscular tension. The VPA includes several laryngeal settings, and assessments are made perceptually, with each setting judged in relation to a 'neutral' or 'modal' voice. Deviations from these baseline settings are marked on a scale of 1–3 for non-pathological speech (e.g., 1 = slight, 2 = marked, 3 = extreme) and 4 to 6 for speech classed as pathological.

Perceptual assessments, including VPA analysis, are not without limitations. First, they are of course subjective. They rely on listener classification, which can lead to variability in assessments across different raters and even within the same rater (Beck 2007). For instance, a rater's judgement of a particular voice may be influenced by other voices evaluated during the same session (Klug 2023). Second, the assessment can be dependent on both the speaker and the analyst. Some speakers may be easier to evaluate if they display unusual but clear and consistent settings. Additionally, analysts often report varying strengths and weaknesses when assessing different settings (see e.g., San Segundo et al. 2019: 369). This variation means that, although inter-rater agreement is achievable, it typically requires calibration and training to ensure consistency (San Segundo et al. 2019). Other limitations of a VPA analysis include the challenge of mapping articulatory labels to perceptual assessments; for example, what does the difference between a lowered tongue body rated at grade 3 vs 4 sound like? (Kreiman and Sidtis 2011). Finally, there is the issue of interdependence among the 30 to 40 settings: how easy is it to perceptually distinguish these settings?

As a result of these limitations, there has been growing motivation to assess VSs more objectively, particularly through acoustic analysis. While we refer to this approach as being "more objective" we acknowledge that various subjective decisions are nevertheless involved in acoustic analyses (see e.g., Klug and Niermann 2023). There are several methods for acoustically capturing breathy and whispery VQs, with some of the more frequently used measures including cepstral peak prominence (CPP), harmonics-to-noise ratio (HNR), and spectral tilt. CPP is a measure of the amount of respiratory noise in the signal relative to the overall amplitude. Higher values indicate a more prominent peak relative to the noise in the signal (Hillenbrand and Houde 1996). HNR measures the ratio of periodic (harmonic) to aperiodic (noise) sound energy, with a higher HNR indicating a clearer, more periodic signal with less noise (Kreiman et al. 2014). Finally, spectral tilt quantifies how energy in the signal decreases as frequency increases, illustrating the distribution of energy across the frequency range (Chai and Garellek 2022). This review will examine the suitability of these parameters in categorising long-term VQs (defined as semi-permanent vocal characteristics (Laver 1980)) in non-pathological speakers with breathy and whispery VQs.

1.2 Applications of vocal setting and voice quality analysis

The VPA has been adapted for use in both sociophonetics (Esling 1978; Stuart-Smith 1999) and forensic speech science (San Segundo et al. 2019), with variations in VSs arising due to both biological and social factors. Biological factors encompass the

anatomy and physiology of a speaker's vocal tract, including variations in larynx and vocal fold sizes (Zhang 2021). Social factors include learned behaviours, such as the use of creak to mark the end of a turn in Finnish (Ogden 2001). Both biological and social factors affect speech acoustics, leading to variation in VSs between and within speakers.

Dallaston and Docherty (2020) provided a systematic review on the prevalence of creak in varieties of English, aiming to better understand two things: (1) the social and regional distribution of speakers using creak, and (2) the methods used across studies to measure the prevalence of creak. The main aim of their review was to gain an objective, quantified understanding of the prevalence of creaky voice across different varieties of English. They sampled the literature for research on native speakers of English who were vocally healthy and were not told by the researchers to manipulate their VQ. Quantitative analysis of whole stretches of continuous speech were included. Results illustrated consistency in the speakers investigated, namely young, female college students. Variability was found, however, in the formulae used to calculate creaky voice, highlighting the substantial diversity in the methods used to analyse this particular VQ. The authors concluded that although the body of research on creaky voice is limited, focusing predominantly on young American women, it is heterogenous in its methods. This work highlighted substantial diversity in the methods used to analyse creaky voice, motivating us to ask whether current knowledge regarding breathy and whispery voice qualities was in a similar position.

Prior research on non-contrastive uses of VQ has often omitted female speakers from analysis altogether. For example, Chan (2023) had access to a database of 552 Australian English speakers – 332 females and 231 males. However, he only analysed a subset of 75 male speakers, excluding all the available data from the female speakers. In other work, analysts extract acoustic measures from both male and female speakers but do not analyse them as separate groups, despite established differences in vocal tract physiology and phonation patterns (Simpson 2009). Pommée and Morsomme (2022), for example, used a mixed sample of 14 female and 10 male speakers, reporting results for the combined sample without considering sex-specific results. Breathy and whispery VQ are particularly relevant for female speakers (Klatt and Klatt 1990), a demographic which has been under-researched in phonetics more generally but particularly in forensic speech science. For example, of the three forensically relevant papers in this review, none analysed female speakers (Chan 2023; Klug et al. 2019; Xu et al. 2023). Overall, all 21 papers reviewed included male speakers, whereas only 16 included female speakers.

The present paper therefore uses the systematic review approach not only to better understand the suitability of acoustic parameters for non-pathological speakers but also in a range of other contexts including (i) different speakers (e.g., male vs female), (ii) different speech materials (e.g., isolated vowels vs continuous speech), and (iii) different analysis settings (e.g., window shift/length and f0/formant ranges).

1.3 Research questions

To the best of our knowledge, there has been no thorough review of the suitability of acoustic parameters in categorising breathy and whispery voices. The focus of this paper is therefore to provide an empirical overview of the performance of acoustic parameters in categorising breathy and whispery VQs for non-pathological speakers across varied speaking contexts, different speaker groups and individual speakers.

The following questions are posed for research on the acoustic parameters of breathy and whispery voice for vocally healthy speakers:

1. What demographic profiles of speakers are investigated?
2. What acoustic parameters have been measured?
3. What speech materials have been analysed?
4. What analysis settings have been used?
5. What do these studies collectively reveal about the suitability of acoustic parameters for breathy and whispery voices?

The remainder of this paper is structured as follows. Section 1.4 offers a brief overview of the VQs relevant to this systematic review, focusing initially on modal voice (commonly used as a baseline), followed by breathy and whispery voice. Section 2 provides information about the systematic review method, outlining the criteria used for the present study. The results are presented in Section 3, addressing the five research questions outlined above. Finally, the discussion in Section 4 summarises the current state of knowledge and concludes with considerations for future research.

1.4 Overview of laryngeal voice qualities

Modal phonation, as a concept, is fundamental to our understanding of other VQs (modifications in the larynx). It serves as a baseline from which deviations to other VQs such as breathy and whispery voices are described. However, Laver (1980) describes modal phonation as an abstract or imaginary concept, stating that it is acoustically defined by the complete adduction of the glottis, producing sounds with minimal friction noise. In practice, achieving this idealised form is nearly impossible, with all speakers naturally varying in their baseline productions. Therefore, in this paper we use the term “default” to refer to the typical or normal VQ of a given speaker.

Breathy phonation involves both the abduction and adduction of the glottis. However, adductive tension is minimal, resulting in weak medial compression and longitudinal tension (Laver 1980). Additionally, the posterior end of the glottis

remains open (Esling and Harris 2005), causing substantial turbulence as air escapes through this gap (Laver 1980). Acoustically, this escape of airflow often results in a strong first harmonic (Klatt and Klatt 1990), a steeper spectral slope, and an overall reduction in acoustic intensity (Gordon and Ladefoged 2001).

The primary difference between breathy and whispery phonation is the degree of glottal opening, which is narrower for whispery voices, resulting in less linear airflow and thus more turbulence (Esling et al. 2019). While the state of the glottis is a key factor in distinguishing these voices, another critical distinction involves the state of the epilarynx (the upper two-thirds of the larynx located just above the vocal folds; Moisik et al. 2019). Whispery phonation involves constriction of the epilarynx, whereas breathy voice is characterised by the absence of such supraglottic constriction, typically accompanied by larynx lowering. The additional laryngeal constriction in whispery phonation produces the auditory effects of whispering, leading to a more turbulent airflow and consequently more noise (Moisik et al. 2019). Although the escaping air in whispery phonation generates turbulent and non-harmonic components in the signal (Ishi et al. 2010), it lacks phonation, producing an aperiodic signal. The perceptual transition between breathy and whispery phonation is best understood on a continuum, as both share characteristics of turbulent aspiration noise, making them difficult to distinguish. San Segundo et al. (2019) explain that the overlap in perceptual effects results in a lack of consensus on how to differentiate the two.

Since breathy and whispery phonation are often perceptually intertwined and are especially relevant to female speakers, we conducted a systematic review of the literature on these two VQs.

2 Materials and methods

When surveying the literature on a given topic, researchers can choose between a narrative approach and a systematic review method. In linguistic and phonetic research, the narrative approach is more frequently implemented. The systematic review is a relatively recent addition to literature reviews, evidenced by a Google Ngram search, which shows a rapid increase in the usage of the term since the 2000s.

The goal of a narrative approach review is to cover as much of the relevant literature as possible. Alternatively, a systematic review involves following a clear structured method to survey the literature and filter relevant papers. Advantages of the systematic review include both its comprehensiveness and replicability as it minimises the likelihood of snowball sampling. Systematic reviews are also said to be especially valuable for topics where methodological approaches are diverse (e.g., linguistic and phonetic analysis). For instance, Dallaston and Docherty (2020) used

the systematic review method to gain a clearer understanding of both the range of speakers and the analytical methods used in studies of creaky voice. Their review aimed to (1) dispel myths about creaky voice by distinguishing anecdotal claims about VQ from empirical research, and (2) clarify the range of methodological approaches used to investigate creaky voice. While they found young American English speakers were frequently studied in relation to creaky voice, the authors also observed substantial variability in the methods used to measure it. The success of their review in highlighting such methodological diversity provides a clear example of how systematic reviews can be insightful for topics where the consensus is less established. Following Dallaston and Docherty (2020), the present study uses the systematic review approach (Pickering and Byrne 2013) to review the literature on the suitability of acoustic parameters to categorise breathy and whispery voices in non-pathological speakers, across a range of speakers, speech materials and analysis settings.

As with any literature review, one limitation of the systematic review approach is that publications released after the initial search date are excluded. Additionally, as noted by Pickering and Byrne (2013), some useful literature may be excluded if it does not match the key terms used in the search. While this issue can also occur in narrative reviews, it is arguably less likely in systematic reviews, which employ a more structured and comprehensive approach to scanning the relevant literature.

In conducting a systematic review, researchers must establish clear criteria for accepting or rejecting papers. Terminological criteria, as noted by Pickering and Byrne (2013), require the researcher to select key terms that will capture a comprehensive range of relevant literature while minimising the number of unrelated results. Sampling criteria refine the search by specifying the sample of interest to reduce the number of relevant papers. Analysis criteria define the type of analysis required for inclusion. Finally, publication criteria outline the publication standards that papers must meet to be accepted. Below we outline the specific inclusion criteria applied in the present study.

2.1 Criteria for inclusion

2.1.1 Terminological criteria

Table 1 outlines the terminological criteria used in the present study. Variations of the term ‘voice quality’ were included in the title search. Abstract key terms, on the other hand, focused on more specific keywords, such as variation of ‘breathy’, ‘whispery’, or ‘nonmodal voice’.

Table 1: The search terms used to locate the relevant literature across the online databases.

Database	Search strategy
ProQuest (search limited to peer reviewed articles) **Key:** ti = title Noft = anywhere except the full text	ti((“voice quality” OR “voice qualities” OR “phonation type” OR “phonation types” OR “acoustic voice” OR “acoustic vocal quality” OR “acoustic characteristics of voice”)) AND noft((“nonmodal voice” OR “non modal voice” OR “non-modal phonation” OR “non modal phonation” OR “breathy” OR “breathy voice” OR “breathy quality” OR “breathy phonation” OR “breathiness” OR “whispery” OR “whispery voice” OR “whispery quality” OR “whispery phonation”))
PubMed **Key:** Title = title Title/Abstract = title, collection title, abstract, other abstract and keywords	“voice quality” [Title] OR “voice qualities”[Title] OR “phonation type”[Title] OR “phonation types” [Title] OR “acoustic voice” [Title] OR “acoustic vocal quality” [Title] OR “acoustic characteristics of voice” [Title] AND “nonmodal voice”[Title/Abstract] OR “non modal voice”[Title/Abstract] OR “nonmodal phonation” [Title/Abstract] OR “non modal phonation” [Title/Abstract] OR “breathy” [Title/Abstract] OR “breathy voice” [Title/Abstract] OR “breathy quality” [Title/Abstract] OR “breathy phonation” [Title/Abstract] OR “breathiness” [Title/Abstract] OR “whispery”[Title/Abstract] OR “whispery voice”[Title/Abstract] OR “whispery quality”[Title/Abstract] OR “whispery phonation”[Title/Abstract]
SCOPUS **Key:** ARTICLE TITLE = titleTITLE-ABS-KEY = Document title, abstract, keywords	ARTICLE TITLE (“voice quality” OR “voice qualities” OR “phonation type” OR “phonation types” OR “acoustic voice” OR “acoustic vocal quality” OR “acoustic characteristics of voice”) AND TITLE-ABS-KEY(“nonmodal voice” or “non modal phonation” or “non modal voice” or “breathy” or “breathy quality” or “breathy phonation” or “breathiness” or “whispery” or “whispery voice” or “whispery quality” or “whispery phonation”)
Web of Science **Key:** TI = title TS = topic (title, abstract, author keywords and keywords plus)	TI = (“voice quality” OR “voice qualities” OR “phonation type” OR “phonation types” OR “acoustic voice” OR “acoustic vocal quality” OR “acoustic characteristics of voice”) AND TS = (“nonmodal voice” or “non modal phonation” or “non modal voice” or “breathy” or “breathy quality” or “breathy phonation” or “breathiness” OR “Whispery” OR “Whispery voice” OR “Whispery quality” OR “Whispery phonation”)

2.1.2 Sampling criteria

The sampling criteria are applied to refine the search and reduce the number of relevant papers. The following criteria were established:

1. Only peer-reviewed work written in English was included, to ensure that the authors could accurately interpret the results.
2. In terms of the analysed speech, any language was accepted. The only exception to this was for languages in which VQ is used to make phonemic contrast distinctions. This decision was made since we were interested in the use of breathy and whispery voice as a long-term semi-permanent vocal characteristic (Laver 1980), rather than for short-term segmental contrasts. Having said this, we acknowledge that future work could explore the acoustic similarities and differences between phonologically contrastive and non-contrastive VQ.
3. Only real speech data were included, with our interests lying in naturalistic and continuous speech. Synthesised or artificially modified speech was therefore excluded.
4. Paralinguistic uses of VQ were excluded, such as VQ adjustments to mark sarcasm, or associated with post-traumatic stress disorder and depression. Again, this was because our research focus is on long-term VQs as opposed to context-dependent realisations of VQ.
5. Finally, only vocally healthy participants were included. Papers where participants had poor vocal health or brain conditions were excluded. Additionally, studies including both vocally healthy and unhealthy subjects were excluded as the analysis in these papers compared these two groups, using healthy subjects as the control condition. We decided to focus on vocally healthy participants to better understand the suitability of these techniques when applied to non-pathological voices.

2.1.3 Analysis criteria

The aim of this review was to better understand the suitability of acoustic parameters for categorising non-pathological breathy and whispery voices in naturalistic, continuous speech. Therefore, the main analysis criteria were that the speech must have undergone some type of acoustic analysis. Papers combining acoustic analysis with other methods (e.g., perceptual assessments) were accepted. Additionally, clinical measures, such as the "acoustic breathiness index" (ABI) and "acoustic voice quality index" (AVQI) were included as they encompass various types of acoustic parameters. Data from any vocally healthy speaker, regardless of speaking style, were included.

2.1.4 Publication criteria

For the present review, only original research articles published in peer-reviewed academic journals or conference proceedings were included. This criterion ensured that the methodological standards employed in the papers selected had been subject to academic scrutiny. Other systematic reviews and meta-analyses were rejected as these reviews were based on a different set of research questions and aims.

2.2 Search strategy

Four online databases were used to search for relevant papers. These were ProQuest, PubMed, SCOPUS, and Web of Science. Following Dallaston and Docherty (2020), these databases were chosen for their coverage of relevant fields, including linguistics, phonetics, and speech pathology. While the focus of the present review was on vocally healthy subjects, the inclusion of speech pathology databases was particularly beneficial, as a substantial amount of research on vocally healthy subjects is published within this field. Moreover, these databases also include relevant peer-reviewed conference papers, such as Interspeech Proceedings (https://www.isca-archive.org/). Table 1 outlines the full search strategy employed across the four databases. Additionally, manual searches, using the same key terms in Table 1, were conducted to include papers from the International Congress of Phonetic Sciences (ICPhS), a conference not covered by the online databases but reporting a range of VQ research. This review encompassed papers from 1999 to 2023, with the conference held every four years. Data extraction was completed by the first author.

3 Results

We initially present the papers which satisfied the criteria outlined above (3.1). Next, we address each of the research questions in detail: Section 3.2 explores the demographic profile of the speakers studied, Section 3.3 reviews the parameters measured, Section 3.4 outlines the speech materials analysed, Section 3.5 the analysis settings used, and Section 3.6 summarises what the studies collectively reveal about the suitability of acoustic parameters for breathy voices. Finally, Section 3.7 briefly discusses the concept of within speaker variability and Section 3.8 provides an overview of the results for the limited number of papers investigating whispery voice.

3.1 Included papers

We begin by examining the papers included in the online databases, where our key term searches produced a total of 754 records: 122 from ProQuest, 256 from SCOPUS, 219 from Web of Science and 157 from PubMed. After combining the results from the four databases, duplicate papers (n = 440) were removed automatically using an R script (R Core Team 2021). This left a total of 314 papers for further screening.

The titles and abstracts of the 314 papers remaining were reviewed against the criteria outlined in Section 2. When reviewing the titles and abstracts it was evident that 259 papers did not meet the required criteria. For example, some papers investigated vocally unhealthy speakers (e.g., 'Voice Quality of female teachers with vocal fatigue', Kovacic and Emica 2013) whereas others used synthesised speech stimuli (e.g., 'Cultural and language differences in voice quality perception: a preliminary investigation using synthesised signals', Yiu et al. 2008).

At this point in the review process, we erred on the side of caution, only rejecting papers where it was immediately evident that the sampling criteria were not met. Examples of papers that were accepted at this stage include the following: 'An exploration of voice quality in mothers speaking Canadian English to infants' (Cheng et al. 2023), and 'Modulation spectral features for objective voice quality assessment: the breathiness case' (Markaki and Stylianou 2009).

The remaining 55 papers were subject to a more comprehensive review, during which the methods and results section were closely evaluated to determine if they met the sampling and analysis criteria. For example, the paper titled 'An exploration of voice quality in mothers speaking Canadian English to infants' (Cheng et al. 2023) was rejected at this stage as the speech sample focused on paralinguistic uses of VQ, e.g., child-directed speech. This process identified that there were 16 papers from the online databases which fully met the criteria outlined in Section 2.

After reviewing the results from the online databases, we conducted manual searches of the ICPhS proceedings between 1999 and 2023. From this, five additional relevant papers were identified, bringing the total number of papers to 21. These 21 papers are listed in Table 2, in reverse chronological order. The paper numbers indicated in the first column are subsequently used to refer to the specific studies throughout the remainder of this paper.

3.2 What is the demographic profile of speakers investigated?

Table 3 summarises the demographic samples of speakers across the 21 studies. While most studies analysed English speech, some did not specify the language. A trend towards particular accents was also observed with Australian, British and

Table 2: The 21 studies that met the criteria, organised by year of publication. For the complete list of authors, please refer to the full references.

Paper number	Author (year)	Title	Source
1	Duarte-Borquez et al. (2024)	Utterance final voice quality in American English and Mexican Spanish bilinguals	Database
2	Chan (2023)	Evidential value of voice quality acoustics in forensic voice comparison	Database
3	Xu et al. (2023)	Contributions of acoustic measures to the classification of laryngeal voice quality in continuous English speech	ICPhS
4	Gierlich et al. (2023)	Test-Retest reliability of the acoustic voice quality index and the acoustic breathiness index	Database
5	Schultz et al. (2023)	A cross-sectional study of perceptual and acoustic voice characteristics in healthy aging	Database
6	Pommée et al. (2022)	Voice quality in telephone interviews: A preliminary acoustic investigation	Database
7	Klug et al. (2019)	Analysing breathy voice in forensic speaker comparison: Using acoustics to confirm perception	ICPhS
8	Park et al. (2019)	Categorization in the perception of breathy voice quality and its relation to voice production in healthy speakers	Database
9	Kadiri et al. (2019)	Melfrequency cepstral coefficients of voice source waveforms for classification of phonation types in speech	Database
10	Feng et al. (2019)	Identification of voice quality variation using ivectors	Database
11	Yokonishi et al. (2016)	Relationship of various open quotients with acoustic property phonation types fundamental frequency and intensity	Database
12	Borsky et al. (2017)	Modal and nonmodal voice quality classification using acoustic and electroglottographic features	Database
13	Szakay and Torgersen (2015)	An acoustic analysis of voice quality in London English: The effect of gender, ethnicity and f0	ICPhS
14	Kreiman et al. (2012)	Variability in the relationships among voice quality, harmonic amplitude, open quotient and glottal area waveform shape in sustained phonation	Database
15	Shue et al. (2010)	On the interdependencies between voice quality glottal gaps and voice-source related acoustic measures	Database
16	Lehto et al. (2007)	Comparison of two inverse filtering methods in parameterization of the glottal closing phase characteristics in different phonation types	Database
17	Gorham-Rowan et al. (2006)	Acoustic-perceptual correlates of voice quality in elderly men and women	Database
18	Zetterholm (1999)	Auditory and acoustic analysis of voice quality variations in normal voices	ICPhS
19	Trittin et al. (1995)	Voice quality analysis of male and female Spanish speakers	Database
20	Gobl et al. (1992)	Acoustic characteristics of voice quality	Database

Table 2: (continued)

Paper number	Author (year)	Title	Source
21	Pittam et al. (1987)	Predicting impressions of speakers from voice quality acoustic and perceptual measures	Database

American English being the most frequently specified. Across the majority of studies, a somewhat broader range of languages (e.g., German, French and Spanish) were used, with most including speech from adult females. The overall sample sizes varied markedly across the papers, with some conducting small-scale investigations using three or four speakers, while others analysed up to 82 speakers. The mean sample size across all studies was 29. We then focused on the demographic profiles of the sampled speakers in the forensically motivated studies. As seen in Table 3, three papers (14 %) were forensically motivated, all of which sampled adult male English speakers.

3.3 What acoustic parameters have been measured?

Table 4 summarises the acoustic parameters used across studies.

As expected, Table 4 illustrates substantial variability in acoustic parameters across studies, with 17 different parameters analysed across the 21 papers. The most frequently analysed were spectral tilt (8 papers), harmonic to noise ratio (HNR, 7 papers), cepstral peak prominence (CPP, 6 papers) and f0 (6 papers). It is worth noting, however, that some studies analysed several different parameters, acknowledging that no single parameter is perfect or easy to interpret. Additionally, with these parameters being easy to extract automatically, it is possible to generate a range of parameters quickly. As a result, some researchers either focus on a few of the extracted parameters or reduce the dimensionality of a large set via statistical techniques such as principal components analysis. For instance, while f0 measurements were extracted in several papers, few analysed f0 explicitly in relation to breathy voice. Therefore, the following section focuses on the three most frequently analysed parameters: spectral tilt, HNR and CPP. It is also worth noting that we chose not to focus on MFCCs (the only other parameter measured more than twice) as they are mainly intended to analyse supralaryngeal vocal tract resonances instead of laryngeal characteristics (Hughes et al. 2023: 2; Jurafsky and Martin 2008: 18). MFCCs represent how the power spectrum of a signal changes across different frequencies over time (Hughes et al. 2023).

Table 3: A summary of the sample size, speaker sex, language and dialect explored, and whether the paper was forensically motivated or not.

Author	Sample size	Sex distribution	Language (Accent)	Forensic
Duarte-Borquez et al. (2024)	21	M = 4, F = 17	Bilingual: English and Spanish (US; San Diego-Tijuana border region)	No
Chan (2023)	75	M = 75, F = 0	English (Australian)	Yes
Xu et al. (2023)	4	M = 4, F = 0	English (unspecified)	Yes
Gierlich et al. (2023)	39	M = 7, F = 32	German (unspecified)	No
Schultz et al. (2023)	150	M = 68, F = 82	Unspecified (unspecified)	No
Pommée et al. (2022)	24	M = 10, F = 14	French (unspecified)	No
Klug et al. (2019)	22	M = 22, F = 0	English (British)	Yes
Park et al. (2019)	20	M = 0, F = 20	English (unspecified)	No
Kadiri et al. (2019)	11	M = 5, F = 6	Finnish (unspecified)	No
Feng et al. (2019)	6	M = 2, F = 4	NA (unspecified)	No
Yokonishi et al. (2016)	6	M = 6, F = 0	NA (unspecified)	No
Borsky et al. (2017)	28	M = 7, F = 21	NA (unspecified)	No
Szakay and Torgersen (2015)	41	M = 25, F = 16	English (London)	No
Kreiman et al. (2012)	6	M = 3, F = 3	NA (unspecified)	No
Shue et al. (2010)	6	M = 3, F = 3	NA (unspecified)	No
Lehto et al. (2007)	13	M = 7, F = 6	NA (unspecified)	No
Gorham-Rowan et al. (2006)	112	M = 56, F = 56	NA (unspecified)	No
Zetterholm (1999)	20	M = 12, F = 8	Swedish (North central dialect)	No
Trittin et al. (1995)	10	M = 5, F = 5	Spanish (Castilian)	No
Gobl et al. (1992)	1	M = 1, F = 0	English (unspecified)	No
Pittam et al. (1987)	12	M = 6, F = 6	English (Australian)	No

3.4 What speech materials have been analysed?

We also aimed to better understand the suitability of acoustic parameters across different speech samples. For instance, we aimed to find out if acoustic parameters work equally well when tested on continuous or spontaneous speech samples as opposed to sustained vowel production. Table 5 summarises the VQ investigated in each paper, the speech style(s) collected and analysed, and the software/settings used.

Table 4: The range and frequency of acoustic parameters measured across papers. The number references in the third column indicate the relevant paper – see Table 1.

Acoustic parameter	Number of papers	Paper number
Spectral tilt	8	1, 2, 3, 7, 11, 13, 17
Harmonic to noise ratio (HNR)	7	1, 2, 3, 7, 8, 15, 17
Cepstral peak prominence (CPP)	6	1, 2, 3, 7, 8, 15
f0	6	1, 3, 5, 17, 19, 20
Mel Frequency cepstral coefficients (MFCCs)	3	9, 10, 12
Acoustic breathiness index	2	4, 6
Acoustic voice quality index	2	4, 6
Amplitude	2	3, 17
Formants	2	17, 20
Jitter	1	5
Shimmer	1	5
High-low spectral ratio	1	8
Inverse filtering	1	16
Long term average spectrum	1	21
Open quotient	1	15
COVAREP features	1	12
NA	1	18

Of the 21 papers reviewed, all analysed breathy voices, while only two (papers #20 and #21) addressed whispery voices. A range of speech styles were present in the recordings analysed. The most frequently recorded speech style was sustained vowel production, found in 11 papers, followed closely by read speech in nine papers. Spontaneous speech was less common, appearing in only three papers. Finally, one paper (#1, Duarte-Borquez et al. 2024) used a picture naming task and another (#20, Gobl et al. 1992) included nonsense words. Despite the variability in recorded speech styles, over half of the papers (13) analysed vowel productions. Some studies, however, analysed a broader range of speech material, including all voiced speech or sonorants (vowels, nasals, liquids, and glides).

When reviewing the papers in detail, it became evident that there was substantial variability in how the voices classified as breathy were selected for analysis. Four papers (#2, 4, 13, 19) relied solely on acoustic analysis to categorise the VQ, meaning a sample of voices were selected and then classified retrospectively via acoustic parameters. In the remaining studies exemplar or representative voices were selected and classified prior to acoustic analysis, e.g., via a perceptual approach (#5, 6, 7, 8, 15) or controlled recordings (#3, 9, 10, 11, 14, 15, 16, 18, 20). In perceptual studies, assessments were made by either lay listeners, expert practitioners, or speech and language therapists. In production studies, speakers – often professional

Table 5: A summary of the voice quality of interest, the speech material collected and analysed, and the software/setting used to conduct the analysis.

Paper number	Voice quality	Speech recorded	Speech analysed/ measured	Software	Software settings
1	Breathy	Read speech and picture naming task	Segmented intervals (details unspecified)	VoiceSauce	STRAIGHT algorithm
2	Breathy	Spontaneous speech (conversational police interview)	Vowels	VoiceSauce	20ms window length, 10ms window shift
3	Breathy	Read speech	Voiced portions (f0>0)	VoiceSauce	25ms frame, 10ms frame shift
4	Breathy	Read speech and vowel production	Vowels	voxPLOT	Unspecified
5	Breathy	Reading task and vowel production	Vowels	Multidimensional voice program	Unspecified
6	Breathy	Read sentences and vowel production	Unspecified	Unspecified	Unspecified
7	Breathy	Spontaneous speech	Sonorants	VoiceSauce	25 ms window length, 1ms window shift
8	Breathy	Vowel production	Vowels	Praat	Unspecified
9	Breathy	Sustained vowel production	Vowels	Unspecified	MFCCs extracted: Zero-time windowing cepstral coefficients 25-ms Hamming windowed frames with a 5-ms shift
10	Breathy	Spontaneous speech and read sentences	Spontaneous speech and read sentences	An i-vector based algorithm	MFCCs extracted with a frame length of 25ms and a sliding length of 10ms.
11	Breathy	Sustained vowel production	Vowels	Unspecified	Unspecified
12	Breathy	Sustained vowel production	Vowels	Support vector machines, random forests, deep neural networks and Gaussian mixture model classifiers	MFCC feature extraction: 25 ms window and 10 ms overlap. A Hamming window was applied. The number of filters in the mel-filter bank was set to 22, in the frequency range from 50 Hz to 4,000 Hz, and 13 MFCCs were computed.

Table 5: (continued)

Paper number	Voice quality	Speech recorded	Speech analysed/ measured	Software	Software settings
13	Breathy	Unspecified	Vowels (midpoint)	Unspecified	Unspecified
14	Breathy	Vowel production	Vowels	VoiceSauce	STRAIGHT algorithm
15	Breathy	Sustained vowels	Vowels	VoiceSauce	Resolution of 1ms
16	Breathy	Vowels	Vowels	Unspecified	Custom made DECAP program
17	Breathy	Vowel production	Vowels	Multidimensional voice program	Unspecified
18	Breathy	Read continuous speech	Vowels	Unspecified	Unspecified
19	Breathy	Natural sentences	Unspecified	PCVox	Unspecified
20	Breathy and whispery	Read passage and nonsense words integrated into a sentence	/straiks/	Inverse filtering	Unspecified
21	Breathy and whispery	Reading passage	Unspecified	Hewlett and Packard 2582A digital spectrum analyser	Hanning window followed by an FFT

phoneticians – were instructed to produce speech using specific phonation types (e.g., breathy or modal voice) and acoustic parameters were then compared between different VQs. Two studies (#12, 21) used both production and perception for speech classification. For example, the speakers were asked to produce speech in a breathy voice, and perceptual methods were then used to confirm that the production was indeed breathy. Finally, one study (#1) used the visual inspection of the waveform and spectrogram to categorise the VQ. Here modal voice was identified by the absence of aspiration noise or voicing irregularity, while breathy voice was recognised by the presence of aspiration.

3.5 What analysis settings have been used?

The majority of papers utilised automatic segmentation tools (e.g., Montreal Forced Aligner, McAuliffe et al. 2017) to extract vowels for their analysis. Only a few studies

(#2, 7, 15) opted for manual segmentation. A range of software packages were employed to extract acoustic parameters. VoiceSauce was the most frequently used, in six of the 21 papers. Other software packages included voxPLOT, Multidimensional Voice Program PCVox, and the Hewlett and Packard 2582 Digital Spectrum Analyser. Finally, there was considerable variability across the papers in both the amount of detail provided about the software settings and the settings themselves. For example, some papers included information on window length and shift, while others detailed the custom programs or algorithms employed.

3.6 What do these studies collectively reveal about the suitability of acoustic parameters for characterising breathy voices?

This section focuses exclusively on the results for breathy voice, with findings for whispery voice being summarised in section 3.8. It is important to note that while different studies may ostensibly analyse the same acoustic parameters, the methodological diversity across studies makes direct comparisons challenging. We provide descriptive comparisons below, but these should be interpreted with caution, bearing in mind the methodological sources of variation contributing to the results. That said, a general trend across studies is that breathy voice demonstrates higher spectral tilt and lower CPP and HNR. Table 6 summarises the results for the papers that analysed breathy voice samples using HNR, CPP and spectral tilt.

Starting with HNR (see the orange section of Table 6), four different versions of this parameter were analysed across the papers. Some papers conducted HNR analysis in restricted frequency ranges (e.g., HNR05: 0–5000 Hz, HNR15: 0–1500Hz, HNR25: 0–2500 Hz, HNR35: 0–3500 Hz), while some averaged the HNR across the frequency range. Four papers (#1, 3, 8, 15) found that breathy voice had a significantly lower HNR than non-breathy voice. One further study (#7, Klug et al. 2019) also reported that breathy voice samples had a lower HNR than non-breathy samples, but this result was not statistically significant. On the other hand, one study (#2, Chan 2023) reported the opposite effect. However, this paper did not conduct any auditory assessments of the samples or statistical checks to ensure that breathy voice samples did indeed have a significantly lower HNR. Instead, the authors took an assumption from prior literature and used this to categorise the VQ (Chan 2023, p. 348).

Table 6: A summary of results for the three most frequently analysed parameters of breathy voice: HNR, CPP and spectral tilt. Top two rows: paper number and statistical analysis. Subsequent rows: individual acoustic parameters. Statistically significant differences are marked with asterisks. Overall differences in the relevant parameter values between breathy and non-breathy are indicated by < or >. Results that contrast with the general trend in a given row are marked in bold.

Paper number	**1**	**2**	**3**	**7**	**8**	**11**	**13**	**14**	**15**	**17**
Statistical analysis	Logistic mixed-effect models		PCA analysis	Mixed effects linear regression	Spearman correlation	*t*-tests + Bonferroni correction		Two-way ANOVA	ANOVA comparisons	Pearson correlation
Mean HNR		**Breathy > modal**	**Breathy < modal and creak		**Breathy < non-breathy					
HNR05	**Breathy < modal	**Breathy > modal**	**Breathy < modal and creak	Breathy < non-breathy					**Breathy < normal	
HNR15		**Breathy > modal**	**Breathy < modal and creak	Breathy < non-breathy					**Breathy < normal	
HNR25		**Breathy > modal**		Breathy < non-breathy					**Breathy < normal	
HNR35		**Breathy > modal**								
CPP	**Breathy < modal	Breathy < modal		**Breathy < non-breathy	**Breathy < non-breathy				**Breathy < normal	
H1*−A1*			**Breathy > modal and creak	**Breathy > non-breathy						
H1*−A2*										
H1*−A3*			Breathy > modal and creak							
H1−A1		Breathy > modal							Breathy > normal	**Breathy > non-breathy
H1−A2		Breathy > modal							**Breathy > normal	
H1−A3		Breathy > modal							**Breathy > normal	
H1−H2		Breathy > modal	**Breathy > modal and creak				**Breathy > non-breathy		**Breathy > normal	
H1*−H2*	**Breathy-creaky < modal**		Breathy > modal and creak	**Breathy > non-breathy		**Breathy > modal		**Breathy > non-breathy/ modal		
H2−H4		Breathy > modal							Breathy > normal	
H2*−H4*				Breathy > non-breathy						

Moving onto CPP (see the blue section in Table 6), the five papers that analysed this parameter all found that breathy voice samples had a lower CPP than non-breathy samples. This finding was statistically significant in four out of the five papers. By contrast, in paper #3 (Xu et al. 2023), a PCA analysis revealed that CPP did not contribute to the principal components that accounted for the variability between breathy, 'modal' and creaky voice.

Finally, the most frequently analysed parameter was spectral tilt (see the green section in Table 6). As with HNR, there were several different metrics applied to quantify this parameter, with ten different combinations of harmonic properties analysed (H_n = frequency of the nth harmonic, A_n = amplitude of the harmonic nearest the nth formant, * = corrected to compensate for bias and errors). Across the eight papers that analysed spectral tilt, seven found that breathy voice samples had a higher spectral tilt than non-breathy samples, with six papers reporting this effect to be significant. The spectral tilt parameters that were significant in more than one paper include: H1*–H2*, H1–H2 and H1*–A1*. One study (#1, Duarte-Borquez 2024), reported the opposite effect for spectral tilt. Specifically, in H1*–H2* measures breathy-creaky voice samples exhibited significantly lower spectral tilt values compared to modal samples. However, it is worth noting that this effect was observed in breathy-creaky voice samples, rather than in purely breathy phonation. Moreover, the authors are transparent in acknowledging that H1*–H2* does not consistently reflect glottal spreading, thereby recognising the inherent variability associated with this acoustic measure.

Overall, these results highlight the following trends: breathy voice is typically associated with increased spectral tilt but a decrease in HNR and CPP. Table 6 shows a relatively consistent pattern across findings, despite considerable methodological variation, such as in the comparison of breathy voice with "modal", "non-breathy" and "normal" voice types. This review also highlights the current lack of research on the suitability of these acoustic parameters in various applied contexts, such as in forensic speech samples involving female speakers.

3.7 Within-speaker variability

Across the 21 papers reviewed, very few make reference to within-speaker variability. One exception to this was paper 14 (Kreiman et al. 2012). This paper found that when accounting for within-speaker variability, a range of strategies were used to produce breathy voice, for instance, manipulating the glottal gap, changing the open quotient, varying the f0, and altering the skewness of glottal pulses. The authors therefore suggested that speakers might not be adjusting one spectral feature in isolation but are instead coordinating changes across several aspects of the glottal

source to influence the overall harmonic pattern. They found that the open quotient, asymmetry coefficient, and/or fundamental frequency (f0) accounted for a substantial proportion of the variance in H1*–H2* across utterances, ranging from 57 % to as much as 93 %. F0, however, accounted for most of the variance.

3.8 Acoustic analysis of whispery voice

Only two of the 21 papers that were reviewed addressed whispery voice (#20, 21). Paper 21 (Pittam et al. 1987) investigated six male and six female Australian-born English speakers who were asked to read a passage in the following voice types: breathy, creaky, nasal, tense and whispery. A long-term average spectrum analysis was conducted on the voice samples, producing 256 amplitude points across the frequency range. For statistical analysis, the data were condensed into a normalised set of eight values, reflecting the successive differences in mean amplitude across 1.5 Bark intervals from 0 to 2152 Hz. The results indicated that whispery voice was strongly linked to changes in energy across the entire frequency range above 4.5 Bark. The other paper (#20) (Gobl et al. 1992) investigated a single male native speaker of English, a phonetician, producing a range of voice qualities in read speech, including whispery voice. By analysing the data using inverse filtering it was found that whispery voice shows the most extreme return phase values, which measures the residual airflow from the point of excitation to complete closure. This affects the steepness of the source spectrum, with a larger return phase leading to greater attenuation of higher frequencies. With so few findings available, it is difficult to draw generalisations from the small amount of work that has investigated whispery voice. Additionally, it is worth noting that neither of these papers addresses the parameters most frequently analysed for breathy voice (HNR, CPP and spectral tilt).

4 Discussion

The primary aim of this review was to assess the suitability of acoustic parameters in categorising breathy and whispery voices among non-pathological speakers. A secondary aim was to examine the consistency of results for these parameters across different speakers, speech materials and analysis settings. The remainder of the paper will explore implications for future research based on observations from the current literature.

4.1 Considerations for further research

4.1.1 Speech classification

A notable source of variability across the 21 papers was how the voices classified as breathy were selected for analysis. Some papers relied solely on acoustic analysis to categorise the VQ, meaning a sample of voices were selected and then classified retrospectively via acoustic parameters. In the remaining studies exemplar or representative voices were selected and classified prior to acoustic analysis, e.g., via a perceptual approach or controlled recordings. Although this review generally found that breathy voices have a higher spectral tilt and lower CPP/HNR, this was not always the case, with one or two papers finding the opposite effect. Given this variability, using retrospective classification via the acoustics may be problematic, as we cannot be entirely confident that the parameters will categorise the samples accurately. Furthermore, there are no thresholds within these acoustic parameters to differentiate between breathy and non-breathy voice reliably. Finally, with there being several different parameters that contribute to the categorisation of VQ, it is potentially problematic to rely on just one or two parameters as a way of categorising a speech sample.

The methodological diversity presented across the papers ultimately raises a question around whether the voices labelled as breathy or whispery in one study are comparable to those in another study. In turn this makes it challenging to understand how well some parameters perform when categorising breathy or whispery voice. Also there might be different types of breathy or whispery voice, just as we have seen with creaky voice in studies by Keating et al. (2015) and Klug (2023). Finally, by relying on the acoustics to categorise the VQ, an auditory assessment of the speech material is not necessarily conducted to check the voice quality is as expected. However, a human analyst's evaluation of the raw data is arguably important if theoretical claims are to be made based on those measurements (Foulkes et al. 2018, p. 6).

Classification via production is advantageous in some ways, as it produces highly controlled speech, eliminating reliance on subjective judgements. However, it is also likely to produce unnatural and extreme realisations. It is also difficult for a speaker to maintain an entire read passage in a breathy or whispery VQ. Finally, the unnatural and extreme realisations can make it challenging to understand how well these parameters apply to more subtle or intermittent realisations of VQ. Classification via perception, on the other hand, is advantageous because an auditory assessment of the speech is conducted. However, a limitation is that low inter-rater reliability is often observed for perceptual assessments (San Segundo et al. 2019). For instance, what one rater perceives as moderately breathy might only be perceived as

mildly breathy by another, with both biological and social factors influencing not only speech production but also perception.

4.1.2 Speaker and speech sample

In general, the studies examined a relatively diverse range of participants including females and speakers of languages other than English. However, it was often the case that not enough data were captured to investigate between-speaker variability across the groups. Finally, the speech styles investigated across the papers were also somewhat limited, with most studies measuring sustained vowel production and only a handful looking at continuous or spontaneous speaking styles.

Based on the demographic profile of the speakers and the speech samples investigated, a gap in the literature has been identified, particularly, the lack of forensically motivated work for female speakers and languages other than English. Given the clear biological differences between men and women in the size of the larynx and vocal tract, it is likely that acoustic parameters will vary across speakers of different sexes. For example, female speakers are often said to exhibit a posterior glottal gap during phonation (Linville 2002), where separation of the arytenoids leads to increased aspiration noise and perceived breathiness. With both biological (sex) and social factors (e.g., gender, accent, language, speech style, ethnicity) affecting speech acoustics, it is important to test how acoustic parameters of these VQs perform with different speakers and speech samples.

4.1.3 Acoustic parameters and extraction methods

The findings of this review reveal that CPP, HNR and spectral tilt are the most widespread and reliable parameters for categorising breathy voice quality. Since only two studies investigated whispery voice, it is difficult to summarise the most reliable parameters for this VQ.

Starting with audio processing, the present review highlighted that both manual and automatic approaches have been used to segment the files for analysis. While automatic approaches are time-efficient, manual checks for a portion of the data allow the researcher to assess the accuracy and margin of error for the automated methods (Foulkes et al. 2018). Fully manual segmentation therefore remains the most precise method. Moving onto the software used for analysis, the review illustrated that although VoiceSauce is the most frequently used software for the extraction of acoustic parameters in VQ studies more broadly, it was used in less than 50 % of the reviewed studies. Variability was also observed in how or whether the analysis settings were documented. For replicability purposes it would be helpful if future

work were to specify the settings used to extract the acoustic parameters, e.g., the window length and shift.

Finally, while most studies have understandably focused on the analysis of vowels, others have expanded their analysis to include sonorants or all voiced segments. Paper 7 (Klug et al. 2019) explained that the authors' rationale for extending the measurements beyond vowels was primarily to overcome the challenges of limited data in forensic casework. Other papers, such as paper 3 (Xu et al. 2023), explained that they extracted voiced speech to assess whether one can apply parameters on connected speech without the need to separate vowels from other voiced sections.

4.1.4 Summary and implications

To summarise, our review found that research on vocally healthy speakers presents breathy voice as *problematically heterogeneous* in terms of methodology. As a result, it remains difficult to identify a universal acoustic measure that reliably represents breathy voice. Similar to Keating et al. (2015) and Dallaston and Docherty (2020), who argue that no single parameter can characterize creaky voice due to its varied manifestations, the same appears to apply to breathy voice, where different vocal qualities may interact. For instance, Duarte-Bórquez et al. (2024) examine both breathy-creaky and purely breathy voices.

Across the 21 studies reviewed we observed considerable variability in how voices were classified as breathy. Consequently, it is unclear whether the starting point – for example, the type of breathy phonation being analysed – is consistent across studies. This issue is further exacerbated by the limited use of auditory assessments to confirm that the samples indeed reflected a breathy voice or if there were elements of breathy and creaky voice combined, for example. Given this lack of standardisation, it is perhaps unsurprising that findings on the acoustics of breathy voice vary. Inconsistencies are likely further amplified by variation in the acoustic analysis process itself, with differences in settings such as window length and shift contributing additional variability to the results.

In light of these findings, we suggest some considerations for future research. First, given the absence of a reliable, universal acoustic marker for breathy voice, we recommend that researchers avoid relying solely on acoustic analysis when determining VQ. For instance, it would not be advisable to classify a sample as breathy based only on a low CPP value. Second, while perceptual assessment can be subjective and time-consuming, it remains a valuable tool – particularly while the field continues to investigate the suitability of various acoustic parameters. We therefore recommend that researchers incorporate some auditory checks, at least on a subset of their data. A simplified version of the VPA scheme (e.g., San Segundo et al. 2019)

may provide a practical and consistent framework for such assessments. Finally, we encourage researchers to remain aware of the complexity inherent in VQ categorisation. Labels such as "breathy voice" may lack the nuance needed to account for the varied manifestations of breathiness observed in non-pathological speech with comparable issues discussed in relation to creaky voice (e.g., Keating et al. 2015; Klug et al. 2024). Finally, for the sake of replicability and transparency, it would be useful if future studies report the specific settings used to extract acoustic parameters, including details such as window length and shift. This would enable subsequent researchers to evaluate how settings may influence the results. Together, these steps will support the development of more consistent methodologies for the acoustic analysis of VQ and help move the field towards more reliable and interpretable findings.

5 Conclusions

In reviewing the suitability of acoustic parameters in categorising breathy and whispery voices in non-pathological speakers and more naturalistic/continuous speech, the present study identified the key acoustic parameters of breathy voice as being a higher spectral tilt, lower CPP and lower HNR. However, like Dallaston and Docherty's (2020: 15–16) finding for creaky voice, the present review highlights that in research on vocally healthy speakers, breathy and whispery voice are also "problematically heterogenous" with respect to its method. Therefore, at present it is not fully clear how the findings on breathy and whispery voice generalise to different speaker demographics and speech materials, especially considering the variability in how voices classified as breathy are selected for analysis.

Directions for future work include the following: (1) continue to investigate the capabilities of acoustic parameters by testing a wider range of speaker samples, including female speakers and languages beyond English, (2) use non-contemporaneous recordings to allow for an investigation of within-speaker variability, (3) analyse whispery voice samples, and finally, (4) consider the method used to categorise speech prior to analysis. While the suitability of acoustic parameters is still under review, classification prior to the acoustic analysis via perception or production would help to ensure that the voices labelled as breathy and whispery are as comparable as possible across studies.

Acknowledgements: The first author is supported by a Harding Distinguished Postgraduate Scholarship at the University of Cambridge. Thanks to Katharina Klug and Chenzi Xu for valuable conversations surrounding the acoustic analysis of VQ. Additionally, thanks are given to the University of Cambridge Phonetics Laboratory

for hosting Paul Foulkes' workshop on Vocal Profile Analysis in 2023, which provided motivation for this review.

Author contributions: Chloe Patman. Conceptualization, Data Curation, Formal Analysis, Methodology, Resources, Visualization, Writing – Original Draft Preparation, Review & Editing. **Paul Foulkes.** Conceptualization, Methodology, Writing – Review & Editing. **Kirsty McDougall.** Conceptualization, Methodology, Writing – Review & Editing.

Competing interests: The authors have no conflicts of interest.

Research funding: The first author is supported by a Harding Distinguished Postgraduate Scholarship at the University of Cambridge.

Data availability: Search outputs from databases are available on request.

References

Beck, J. 2007. *Vocal profile analysis scheme: A user's manual*. Edinburgh: Queen Margaret University College, Speech Science Research Centre.

Borsky, M., D. D. Mehta, J. H. Van Stan & J. Gudnason. 2017. Modal and non-modal voice quality classification using acoustic and electroglottographic features. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25(12). 2281–2291.

Chai, Y. & M. Garellek. 2022. On H1–H2 as an acoustic measure of linguistic phonation type. *Journal of the Acoustical Society of America* 152(3). 1856–1870.

Chan, R. K. W. 2023. Evidential value of voice quality acoustics in forensic voice comparison. *Forensic Science International* 348. 111725.

Cheng, A., E. McClay & H. H. Yeung. 2023. An exploration of voice quality in mothers speaking Canadian English to infants. *Language Learning and Development*. 1–18. https://doi.org/10.1080/15475441.2023.2256708.

Dallaston, K. & G. Docherty. 2020. The quantitative prevalence of creaky voice (vocal fry) in varieties of English: A systematic review of the literature. *PLoS One* 15(3). 1–18.

Duarte-Borquez, C., M. Van Doren & M. Garellek. 2024. Utterance-final voice quality in American English and Mexican Spanish bilinguals. *Languages* 9(3). 70.

Esling, J. 1978. The identification of features of voice quality in social groups. *Journal of the International Phonetic Association* 8(2). 18–23.

Esling, J. H. & J. G. Harris. 2005. States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In W. J. Hardcastle & J. M. Beck (eds.), *A figure of speech*, 347–383. London: Routledge.

Esling, J. H., S. R. Moisik, A. Benner & L. Crevier-Buchman. 2019. *Voice quality: The laryngeal articulator model*. Cambridge: Cambridge University Press.

Feng, C., E. van Leer & D. V. Anderson. 2019. Identification of voice quality variation using I-Vectors. In *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 105–109. NY, USA: New Paltz.

Foulkes, P., G. Docherty, S. Shattuck-Hufnagel & V. Hughes. 2018. Three steps forward for predictability: Consideration of methodological robustness, indexical and prosodic factors, and replication in the laboratory. *Linguistics Vanguard* 4(2). 1–9.

Gierlich, J. & V. Latoszek Barsties. 2023. Test-retest reliability of the acoustic voice quality index and the acoustic breathiness index. *Journal of Voice*. [Advance online publication]. https://doi.org/10.1016/j.jvoice.2023.08.016.

Gobl, C. & A. Ní Chasaide. 1992. Acoustic characteristics of voice quality. *Speech Communication* 11(4). 481–490.

Gordon, M. & P. Ladefoged. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29. 383–406.

Gorham-Rowan, M. M. & J. Laures-Gore. 2006. Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders* 39(3). 171–184.

Hillenbrand, J. & R. A. Houde. 1996. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research* 39(2). 311–321.

Hughes, V., A. Cardoso, P. Foulkes, P. French, A. Gully & P. Harrison. 2023. Speaker-specificity in speech production: The contribution of source and filter. *Journal of Phonetics* 97. 101224.

Ishi, C. T., Ishiguro, H. & Hagita, N. 2010. Analysis of the roles and the dynamics of breathy and whispery voice qualities in dialogue speech. *EURASIP Journal on Audio, Speech, and Music Processing*. 1–12, https://doi.org/10.1186/1687-4722-2010-528193.

Jurafsky, D. & J. Martin. 2008. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River: Pearson Prentice Hall.

Kadiri, S. R. & P. Alku. 2019. Mel-frequency cepstral coefficients of voice source waveforms for classification of phonation types in speech. In *Proceedings of Interspeech*, 2508–2512.

Keating, P., M. Garellek & J. Kreiman. 2015. Acoustic properties of different kinds of creaky voice. In *Proceedings of the 18th International Congress of Phonetic Sciences.* The Scottish Consortium for ICPhS 2015 (Ed.), Glasgow. Available at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0821.pdf.

Klatt, D. H. & L. C. Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87(2). 820–857.

Klug, K. 2023. *Assessing a speaker's voice quality for forensic purposes: Using the example of creaky voice and breathy voice*. University of York PhD dissertation. Available from: https://etheses.whiterose.ac.uk/34778/.

Klug, K., C. Kirchhübel, P. Foulkes & P. French. 2019. Analysing breathy voice in forensic speaker comparison: Using acoustics to confirm perception. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 795–799. Australasian Speech Science and Technology Association Inc., and International Phonetic Association.

Klug, K. & M. Niermann. 2023. Assessing the suitability of f0 estimators with respect to recording condition and voice quality. In K. Klug, *Assessing a speaker's voice quality for forensic purposes: Using the example of creaky voice and breathy voice*, 58–77. York, UK: University of York PhD dissertation. Available from: https://etheses.whiterose.ac.uk/34778/.

Klug, K., C. Kirchhübel, P. Foulkes, A. Braun & P. French. 2024. Assessing creaky voice quality for forensic purposes. In *Proceedings of the 2nd International Conference on the Foundations of Speech. Proceedings of the 2023 Aarhus International Conference on Voice Studies*, 16–26. Sciendo.

Kovacic, G. & F. Emica. 2013. Voice quality of female teachers with vocal fatigue. *Hrvatska Revija Za Rehabilitacijska Istrazivanja* 49. 92–107.

Kreiman, J. & D. Sidtis. 2011. *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Malden, MA, USA: Wiley-Blackwell.

Kreiman, J., Y.-L. Shue, G. Chen, M. Iseli, B. R. Gerratt, J. Neubauer & Abeer Alwan. 2012. Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation. *Journal of the Acoustical Society of America* 132(4). 2625–2632.

Kreiman, J., B. R. Gerratt, M. Garellek, R. Samlan & Z. Zhang. 2014. Toward a unified theory of voice production and perception. *Loquens* 1(1). 1–19.

Laver, J. 1980. *The phonetic description of voice quality*. Cambridge, UK: Cambridge University Press.

Lehto, L., M. Airas, E. Björkner, J. Sundberg & P. Alku. 2007. Comparison of two inverse filtering methods in parameterization of the glottal closing phase characteristics in different phonation types. *Journal of Voice* 21(2). 138–150.

Linville, S. E. 2002. Source characteristics of aged voice assessed from long-term average spectra. *Journal of Voice* 16(4). 472–479.

Markaki, M. & Y. Stylianou. 2009. Modulation spectral features for objective voice quality assessment: The breathiness case. In *Proceedings of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*. 1–4. Florence, Italy: Department of Information Engineering, University of Florence.

McAuliffe, M., M. Socolof, S. Mihuc, M. Wagner & M. Sonderegger. 2017. Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. *Proceedings of Interspeech* 2017. 498–502.

Moisik, S. R., M. Hejná & J. H. Esling. 2019. Abducted vocal fold states and the epilarynx: A new taxonomy for distinguishing breathiness and whisperiness. In Skarnitzl & J. Volín (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 220–224. Canberra. Available at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_269.pdf.

Ogden, R. 2001. Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association* 31(1). 139–152.

Park, Y., J. S. Perkell, M. L. Matthies & C. E. Stepp. 2019. Categorization in the perception of breathy voice quality and its relation to voice production in healthy speakers. *Journal of Speech, Language, and Hearing Research* 62(10). 3655–3666.

Pickering, C. & J. Byrne. 2013. The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *Higher Education Research and Development* 33. 534–548.

Pittam, J. & C. Gallois. 1987. Predicting impressions of speakers from voice quality: Acoustic and non-acoustic correlates. *Journal of Language and Social Psychology* 6(3–4). 231–242.

Pommée, T. & D. Morsomme. 2022. Voice quality in telephone interviews: A preliminary acoustic investigation. *Journal of Voice* 563.e1–563.e20.

R Core Team. 2021. *R: A language and environment for statistical computing [Computer Software]*. Vienna: R Foundation for Statistical Computing Available from .

San Segundo, E., P. Foulkes, P. French, P. Harrison, V. Hughes & C. Kavanagh. 2019. The use of the Vocal Profile Analysis for speaker characterization: Methodological proposals. *Journal of the International Phonetic Association* 49(3). 353–380.

Schultz, B. G., S. Rojas, M. St John, E. Kefalianos & A. P. Vogel. 2023. A cross-sectional study of perceptual and acoustic voice characteristics in healthy aging. *Journal of Voice* 37(6). 969.e23–969.e41.

Shue, Y. L., G. Chen & A. Alwan. 2010. On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic parameters. In *Proceedings of Interspeech*, 34–37. Chiba, Japan.

Simpson, A. P. 2009. Phonetic differences between male and female speech. *Language and Linguistics Compass* 3(2). 621–640.

Stuart-Smith, J. 1999. Glasgow: Accent and voice quality. In P. Foulkes & G. J. Docherty (eds.), *Urban voices: Accent studies in the British Isles*, 201–222. Leeds, UK: Arnold.

Szakay, A. & E. Torgersen. 2015. An acoustic analysis of voice quality in London English: The effect of gender, ethnicity, and f0. In *Proceedings of the 18th International Congress of Phonetic Sciences.* The Scottish Consortium for ICPhS 2015 (Ed.), Glasgow. Available at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0996.pdf.

Trittin, P. J. & A. de Santos y Lleó. 1995. Voice quality analysis of male and female Spanish speakers. *Speech Communication* 16(4). 359–368.

Xu, C., P. Foulkes, P. Harrison, V. Hughes & J. H. Wormald. 2023. Contributions of acoustic measures to the classification of laryngeal voice quality in continuous English speech. In R. Skarnitzl & J. Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*, 1806–1810. Canberra. Available at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2023/full_papers/772.pdf.

Yiu, E. M., B. Murdoch, K. Hird, P. Lau & E. M. Ho. 2008. Cultural and language differences in voice quality perception: A preliminary investigation using synthesised signals. *Folia Phoniatrica et Logopaedica* 60(3). 107–119.

Yokonishi, H., H. Imagawa, K. Sakakibara, A. Yamauchi, T. Nito, T. Yamasoba & N. Tayama. 2016. Relationship of various open quotients with acoustic property, phonation types, fundamental frequency, and intensity. *Journal of Voice* 30(2). 145–157.

Zetterholm, E. 1999. Auditory and acoustic analysis of voice quality variations in normal voices. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (eds.), *Proceedings of the 4th International Congress of Phonetic Sciences*, 973–976. San Francisco. Available at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_0973.pdf.

Zhang, Z. 2021. Contribution of laryngeal size to differences between male and female voice production. *Journal of the Acoustical Society of America* 150(6). 4511.