TOWARDS ACOUSTIC PROFILES OF PHONATORY QUALITIES

Ailbhe Ní Chasaide and Christer Gobl
Centre for Language and Communication Studies, Trinity College, Dublin, Ireland

ABSTRACT

Voluntary modulations in the mode of phonation constitute an important resource speakers use for the paralinguistic signalling of attitude and emotion. As a first step towards providing profiles of a range of phonatory qualities, this paper presents brief illustrative sketches of some acoustics characteristics associated with modal, tense, breathy/lax, whispery, and creaky voice, as described in [1]. The principal analytic technique on which these illustrations are based was interactive inverse filtering. Source characteristics were measured on the basis of a parametric model of the voice source (the LF model, [2]) fitted to the output of the inverse filter, and from spectral analyses.

INTRODUCTION

Individuals differ in terms of the habitual phonation quality they use, in a way that reflects not only the physical characteristics of their vocal apparatus, but also the linguistic and social group they belong to. As is outlined in the presentation by Laver (this session), speakers also make voluntary short term changes to their mode of phonation as a way of signalling their attitude, mood or emotion. The paralinguistic significance of some phonatory qualities may tend to be universal (e.g., whispery voice tends to give the impression of confidentiality), whereas in other cases, it may be culture or language specific (e.g., creaky voice is associated with bored resignation for some speakers of English).

Our understanding of this aspect of vocal communication is still quite limited. The relative paucity of systematic, quantitative work in this field does not reflect its importance but rather the lack of analytic tools and the many methodological difficulties presented by this kind of research. As a starting point, one needs to be able to make detailed and reliable analyses of the acoustic and physiological correlates of different voice qualities, laryngeal and supralaryngeal. Then there is the difficulty of eliciting appropriate samples of speech. On the one hand, emotionally coloured spontaneous speech would be highly desirable, but on the other hand, reliable analytic comparisons may depend on the speech material being very controlled (more on this in the Conclusions).

This paper presents illustrative sketches of a number of phonatory qualities, based on exploratory work carried out by the authors in recent years [3, 4]. The illustrations are of course tentative, being based on a detailed analysis of just a few utterances spoken with a few of the phonatory qualities that speakers are known to exploit for paralinguistic signalling. The qualities chosen were from the set described by Laver [1], whose descriptions serve as a starting point for our analyses.

DESCRIBING THE SOURCE

The main analysis technique involved inverse filtering of the speech pressure waveform. In order to obtain quantifiable results, a parametric model of differentiated glottal flow (the LF-model, [2]) was matched to the output of the inverse filter. Both the inverse filtering and the matching procedure were carried out for each glottal cycle, using specially designed interactive software allowing optimisation in both the time and frequency domains [5]. From the matched model a number of parameters were subsequently measured. The ones we focused on particularly were EE, RA, RK, RG, OQ and UP. These are explained briefly here, but for a more detailed description, see [7].

EE is the excitation strength and is measured as the negative amplitude of the differentiated flow at the moment of maximum discontinuity. It corresponds to the overall intensity of the signal, so that an increase in EE amplifies all frequency components.

RA is a measure of the return phase (dynamic leakage), which is the residual flow from excitation to complete closure. The acoustic consequence of the return phase is a steeper spectral slope. A large RA corresponds to greater attenuation of the higher frequencies.

RK is a measure of the skew of the glottal pulse: a larger value means a more symmetrical pulse shape. RG is a measure that relates to the duration of the opening branch of the glottal pulse. RK and RG together determine the open quotient, OQ, and they mainly affect the levels of the lower harmonics in the source spectrum. Note that in our definition of OQ, the open phase does not include the interval of the return phase.

UP is the peak glottal airflow, measured only for the oscillatory component of the glottal wave. In our data, UP was calculated indirectly from the other parameters, using a formula suggested by [6].

Aspiration noise is an important source parameter, particularly in breathy and whispery voice. We do not include it in our descriptions, simply because we had no reliable way of measuring it.

Spectral measurements from narrow band spectral sections were also carried out, both on the speech output signal (e.g., Figure 6) and on the source signal, output of the inverse filter (e.g., Figures 2, 3 and 4). Fuller details on these are provided below.

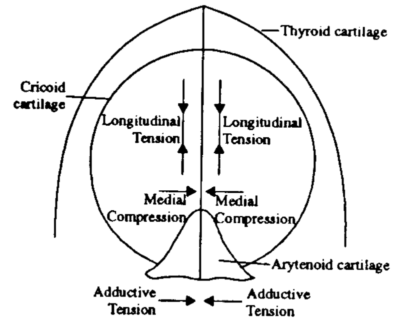

Figure 1. Three laryngeal parameters of muscular tension, after [1].

ACOUSTIC PROFILES

In this section we attempt to illustrate some of the acoustic characteristics of the following phonatory qualities: modal, tense, breathy/lax, whispery and creaky. Although the breathy and lax qualities were separately recorded and analysed, they are rather similar qualities and are treated under a single heading below. Following Laver's descriptions [1] the physiological correlates of these phonatory qualities are presented in terms of three hypothesised dimensions of muscular tension, schematically shown in Figure 1. *Adductive tension* results from contraction of the interarytenoid muscles and is the force which draws the arytenoids together so that the cartilaginous glottis is adducted. *Medial compression* results primarily from contraction of the lateral cricoarytenoid muscle (although the thyroarytenoid muscle can also contribute). It is defined as the force which causes approximation of the vocal processes of the arytenoids so that the ligamental glottis is closed. *Longitudinal tension* is the tension in the vocal folds which results from contraction of the vocalis and cricothyroid muscles, whose primary function is the control of pitch. (For additional descriptions of voice quality see [8, 9].)

The illustrations below of acoustic properties are based on analysis of two words spoken with the six qualities mentioned above. These words were extracted from recordings of materials read by a male phonetician, who was well practised in the Laver system. The recordings included the Rainbow passage and a number of nonsense words inserted into a carrier frame.

Figures 2, 3 and 4 below illustrate source characteristics in the first vowel of the nonsense word *babber*, which occurred in the frame *Say ---- again*. Figure 2 shows individual spectral sections of the voice source for four phonatory qualities at approximately the midpoint of the vowel. In Figure 3 are shown for these same qualities schematic source spectra, averaged for an interval corresponding approximately to four glottal cycles in the middle of the vowel. These were obtained in the following way. First of all, the spectrum was flattened by adding 6 dB per octave relative to L_0. The spectrum up to 4 kHz was then divided into four frequency bands of 1 kHz. Within each band the average amplitude of the harmonics was calculated and shown relative to L_0. By doing this we should get an idea of how the spectral slope for each of the four qualities deviates from the "ideal" slope of -6 dB per octave in the differentiated glottal flow (the horizontal zero line). In Figure 4 are shown for five qualities the relative levels of the first two harmonics in the source spectrum, measured at the approximate midpoint of the vowel. Figure 5 illustrates the pulse-by-pulse variation in some of the measured source parameters for the first six glottal cycles of the word *strikes*, taken from the Rainbow passage. Finally, Figure 6 illustrates the relative levels of F1 and H1 for approximately the same interval of this word.

Modal

Modal voice is the quality "which phonetic theory assumes takes place in ordinary voicing, when no specific feature is explicitly changed or added (p. 95)" [1]. For this quality, adductive tension, medial compression and longitudinal tension are thought to be moderate, and the ligamental and cartilaginous glottis are thought to vibrate as a single unit. The vocal fold vibration is further described as regularly periodic and efficient, with full glottal closure and thus, without audible glottal frication noise. Recent studies suggest, however, that incomplete glottal closure may be very common even in what is perceived as modal voice [10] and particularly in female speech.

In our analyses, this quality emerged as a relatively efficient mode of phonation, with a fairly strong excitation (EE) and fairly limited dynamic leakage (RA). For an illustration of some source parameter values in the word strikes, see Figure 5. The source spectrum for this quality in the vowel in *babber* exhibited a slope that is slightly greater than the "ideal" description of -12 dB per octave (or -6 dB in the differentiated flow: see Figure 3).

It is important to bear in mind that utterances spoken with modal (or indeed any) quality exhibit considerable dynamic variation as a function of the prosodic and segmental context [7, 11, 12].

Tense

At the laryngeal level, tense voice is thought to involve increased adductive tension and medial compression. The term "pressed phonation" is sometimes used for this quality. A higher degree of tension is likely to be found in the entire production system, and this will have consequences for the respiratory system (a raised subglottal pressure) as well as for the supralaryngeal articulation.

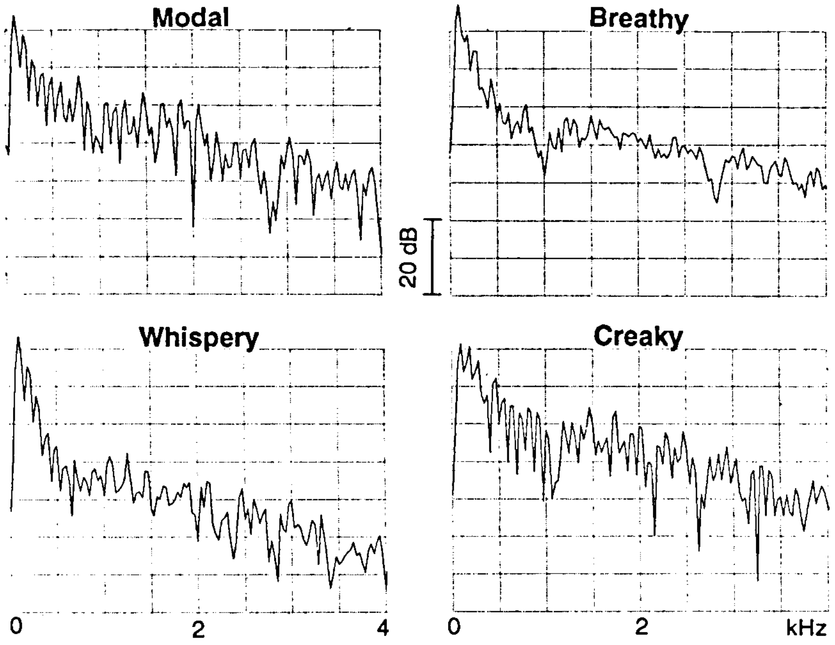

Figure 2. Spectral sections of the source signal at about the midpoint of [æ] *in babber, for four phonatory qualities.*

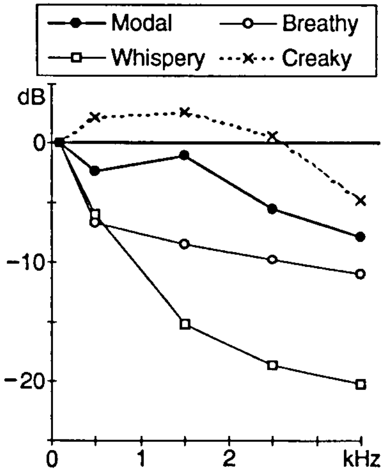

Figure 3. Schematic source spectra in [æ] *of babber, showing within four 1 kHz bands the average deviation from an "ideal" slope (see text).*

In our measures of tense voice (see, for example, source data in Figure 5) the glottal pulse exhibited a very low dynamic leakage (RA), showing a rather instantaneous closure of the vocal folds. The related frequency measure FA was generally higher than for all the other qualities measured in these contexts, and this would imply a flatter spectrum. Relative to modal phonation, the glottal pulse was rather skewed (low RK), the open quotient was lower (low OQ) and RG was higher. The effects of a high RG can be seen in the boosting of H2 relative to H1 for this quality (see illustration for *babber* in Figure 4). Overall, the higher frequencies in the spectrum are relatively dominant. One can get some impression of the relative balance of the lower and higher components of the spectrum from Figure 6 where, for *strikes*, the level of F1 (L_1) is shown relative to that of H1 (L_0). Note the very high L_1 of tense as compared to modal phonation. It is worth noting that for this particular utterance, the excitation (EE) was less strong for tense than for modal voice (see Figure 5). Thus, the greater intensity of the former in the speech output signal is determined in this instance by these other characteristics of the glottal pulse discussed above.

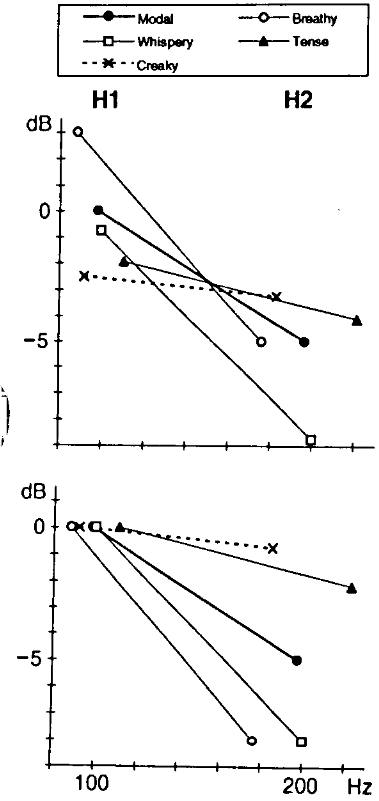

Figure 4. Relative amplitudes of H1 and H2 in source spectra of [æ] in babber, five qualities. Values shown in absolute terms (upper panel) and normalised to L_0 of modal voice (lower panel).

Breathy/Lax

Breathy voice is described as having minimal adductive tension, low longitudinal tension and weak medial compression, with the result that the vocal folds never come fully together and generate audible frication noise. At the laryngeal level, lax voice is described as being rather similar to breathy voice. It may differ in the extent to which laryngeal tensions are reduced: Laver suggests that lax voice may be slightly closer to modal than breathy voice. It is further postulated as having a lesser degree of tension in the entire speech apparatus.

As expected, the glottal pulse for both breathy and lax voice had rather high dynamic leakage (RA): see source values for lax voice in Figure 5. The related frequency measure FA is lower than for modal voice. The glottal pulse also has a high open quotient (OQ) with a long opening branch (low RG). These last characteristics contribute to the relative boosting of H1, a frequently observed spectral characteristic. See for example, the relatively strong H1 for breathy voice in the upper panel of Figure 4, and the rather low values for L_1 relative to L_0 for lax voice in Figure 6.

Whispery

At the laryngeal level whispery voice is thought to be characterised by low adductive tension and moderate longitudinal tension. The degree of medial compression may vary, and with it the size of the triangular opening of the cartilaginous glottis. With weak whisper medial compression is moderate and the opening may include part of the ligamental glottis. Whisper with a higher intensity is thought to have higher medial compression and a smaller opening of the cartilaginous glottis. It is suggested that laryngeal vibration is confined to the portion of the ligamental glottis which is adducted, and the whispery component to the opening between the arytenoids.

Whispery and breathy voice may form an auditory continuum. Although there may be no clear border line [1], they may nevertheless be auditorily distinguished by the relative dominance of the periodic and noise components: the noise component would be relatively greater for whispery voice, whereas the periodic component would be dominant in breathy voice.

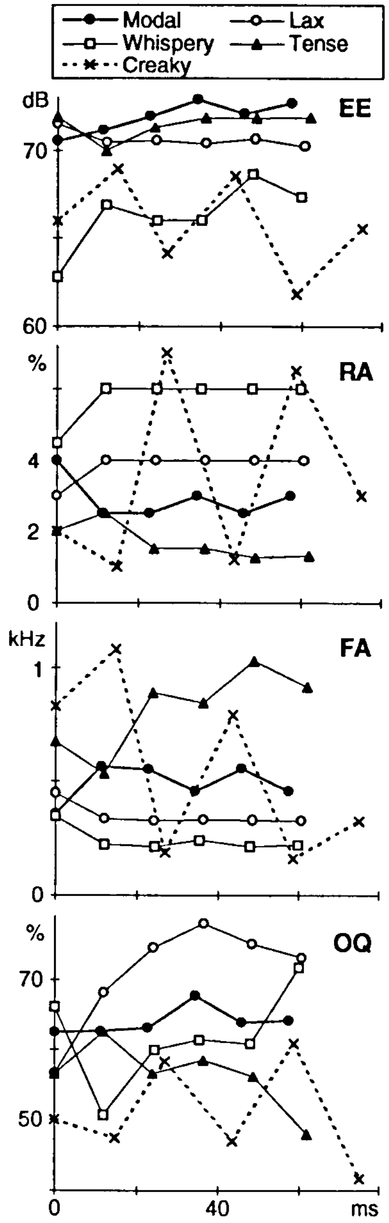

Figure 5. Source data for EE, RA, FA and OQ in the first six glottal cycles of strikes, for five phonatory qualities.

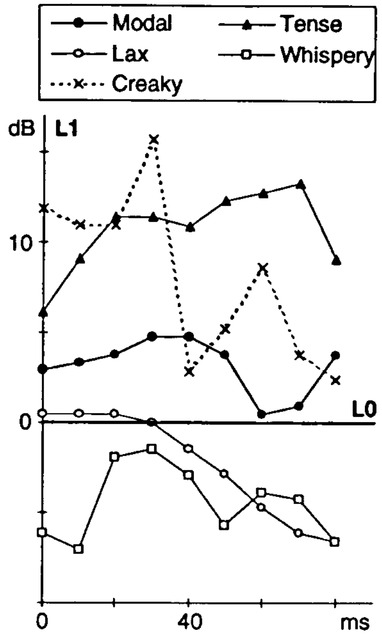

Figure 6. The levels of F1 (L_1) relative to H1 (L_0 = 0 dB) for the first 80 ms in strikes for five phonatory qualities.

In our measurements, the glottal pulse for whispery voice was found to differ from that of breathy voice in the considerably weaker excitation (EE), the higher dynamic leakage (RA), highest of all measured samples, and in the smaller open quotient (OQ). Calculated peak airflow values (UP) were also considerably lower. As a correlate of these last two factors, one would expect H1 to be less boosted than for breathy voice, and this can be seen, for example, in the upper panel of Figure 4, where the levels of H1 and H2 are shown in absolute terms for *babber*. As can be seen in the lower panel of this figure, however, the relative levels of H1 and H2 are very similar, presumably because of the steeper spectral slope of whispery voice.

This last can be seen in Figure 3, and is a clear consequence of the very high RA and low FA values.

Creaky

Creaky voice, being a mix of creak and voicing, is likely to involve the high adductive tension, high medial compression and low longitudinal tension, characteristic of creak. In the production of creak, the folds are thought to be relatively thick and compressed. The ventricular folds may also be somewhat adducted, so that their inferior surfaces come in contact with the superior surfaces of the true vocal folds, resulting in a rather thick vibrating structure. Because of the high adductive tension only the ligamental glottis is thought to vibrate. The f_0 and amplitude of consecutive glottal pulses are described to be irregular and the airflow rate has been observed to be very low [8].

In our recordings of creaky voice, the occurrence of creak was intermittent, in the sense of irregularity in successive glottal pulses. It did not occur in the word *babber* but did in the word *strikes*. In the latter, there is a clear alternation of two very different types of glottal pulse (see Figure 5). One is efficient, with a reasonably strong excitation (though not as strong as for modal), very low dynamic leakage (RA) and consequently a high FA. The other pulse is inefficient, with a very weak excitation, very high dynamic leakage (and a low FA). These two types of pulses should have very different source spectra. Both pulses show a relatively high degree of skew (low RK), a low open quotient and a high RG. The short open phase and the low calculated UP which are also found here are consistent with the low airflow rate reported for this quality and should affect the lower end of the spectrum by reducing its level. Note in Figure 6, that despite fluctuations, L_1 dominates L_0. Given the long closed phase, there is less damping and this could also contribute to the strong ringing of F1.

In the *babber* utterance successive glottal pulses were not irregular and resembled more the strong efficient type of pulse described above. Note the strong H2 and rather weak H1 for this quality in Figure 4 and the relative dominance of higher frequencies in Figure 3. In many ways, the spectral characteristics of this quality resembled those of tense voice.

CONCLUSIONS

We would emphasise that these illustrations are tentative and should not be taken as definitive accounts but rather as a first step toward more comprehensive profiles of different phonatory qualities. This is not only because they are based on a very limited number of utterances. Other factors need to be borne in mind when it comes to interpreting these kinds of source data.

First of all, cross-speaker variation can be quite large. For example, in the important source parameter RA, we have observed cross-speaker differences as great or greater than the differences shown here for a single speaker who has intentionally varied his voice quality. We feel therefore, that the relevant measures of phonatory quality may need to be expressed in terms of deviations from a given speaker's baseline values, rather than in absolute terms. A similar point has been made by other researchers, e.g., [13] as a result of trying to characterise linguistically distinctive phonatory qualities.

Another factor, alluded to briefly before, concerns the considerable dynamic variation that can occur within a single voice quality. These variations appear to be (at least sometimes) conditioned by the prosodic and segmental context and have been illustrated in earlier work [7, 11, 12, 14]. This consideration seriously constrains the kinds of

speech materials that can usefully be compared using these analysis techniques. Furthermore, profiles of individual phonatory qualities can not be static, but will need to take account of this dynamic modulation.

Clearly, there is much work to be done before one will have available comprehensive descriptions of the acoustic correlates of particular phonatory qualities. Yet even partial descriptions may provide useful reference material for looking at more spontaneously occurring speech, where the mode of phonation has been varied for paralinguistic purposes. Furthermore, such descriptions should eventually provide a basis for resynthesis, which should allow one to explore directly the paralinguistic colouring associated with individual voice qualities.

ACKNOWLEDGEMENTS

This work was supported by the Esprit- BRA Working Group, no. 6975, VOX. We are also grateful to Francis Nolan who was our informant.

REFERENCES

[1] Laver, J. (1980), *The phonetic description of voice quality*, Cambridge: Cambridge University Press.

[2] Fant, G., Liljencrants, J. and Q. Lin (1985), "A four-parameter model of glottal flow", *STL-QPSR*, Vol. 4/1985, pp. 1-13.

[3] Gobl, C. (1989), "A preliminary study of acoustic voice quality correlates", *STL-QPSR*, Vol. 4/1989, pp. 9-22.

[4] Gobl, C. and Ní Chasaide, A. (1992), "Acoustic characteristics of voice quality", *Speech Communication*, 11, pp. 481-490.

[5] Ní Chasaide, A., Gobl, C. and Monahan, P. (1992), "A technique for analysing voice quality in pathological and normal speech", *Journal of Clinical Speech & Language Studies*, Vol. 2, pp. 1-16.

[6] Fant, G and Lin, Q. (1988), "Frequency domain interpretation and derivation of glottal flow parameters", *STL-QPSR*, Vol. 2-3/1988, pp. 1-21.

[7] Ní Chasaide, A. and Gobl, C. (1993), "Contextual variation of the vowel voice source as a function of adjacent consonants", *Language and Speech*, 36, pp. 303-330.

[8] Catford, J.C. (1964), "Phonation types: the classification of some laryngeal components of speech production", in D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott, and J.L.M. Trim (eds.), *In Honour of Daniel Jones* (pp. 26-37), London: Longmans.

[9] Ladefoged, P. (1971), *Preliminaries to linguistic phonetics*, Chicago: The University of Chicago Press.

[10] Södersten, M. (1994), "Vocal fold closure during phonation", Ph.D. thesis, Studies in Logopedics and Phoniatrics No. 3, Huddinge University Hospital, Stockholm.

[11] Gobl, C. (1988), "Voice source dynamics in connected speech", *STL-QPSR*, Vol. 1/1988, pp. 123-159.

[12] Pierrehumbert J.B. (1989). A preliminary study of the consequences of intonation for the voice source. *STL-QPSR*, Vol. 4/1989, pp. 23-36.

[13] Traill, A. and Jackson, M. (1987), "Speaker variation and phonation types in Tsonga nasals", *UCLA Working Papers in Phonetics*, 67, pp. 1-28.

[14] Gobl, C., Ní Chasaide, A. and Monahan, P. (1995), "Intrinsic voice source characteristics of selected consonants", *Proc. of the XIIIth ICPhS*, Stockholm.