Principal components of glottal waveforms: towards parameterisation and manipulation of laryngeal voice-quality

Parham Mokhtari[1], *Hartmut R. Pfitzinger*[1,2], *Carlos Toshinori Ishi*[1]

[1] JST/CREST at ATR Human Information Science Laboratories, Kyoto, Japan
[2] Department of Phonetics and Speech Communication, University of Munich, Germany
`parham@atr.co.jp` `hpt@phonetik.uni-muenchen.de` `carlos@atr.co.jp`

Abstract

We propose a more holistic approach to modelling of the glottal volume-velocity waveform and its derivative. Principal components analysis (PCA) is applied to single-cycle glottal waveforms measured in Laver's (1980) recorded speech data representing several phonation-types. The first four principal components lend themselves to qualitative interpretations, and account for over 90% of the total variance. Decision trees generated using those components yield around 64% accuracy in classifying the 13 voice qualities analysed. This approach is capable of modelling a wide variety of voice qualities, and should therefore be useful in speech synthesis.

1. Introduction

Voice quality in a broad sense refers to suprasegmentally persisting attributes of both the laryngeal (or phonatory) and the supralaryngeal (or vocal-tract) components of articulatory *settings* adopted during speech production (Laver, 1980). In this paper we consider only those settings which are attributed to the larynx, or the voice source. Of particular concern to us in the context of the Japan Science and Technology (JST) project on expressive speech processing, are the numerous types of phonation which speakers regularly use in daily conversation to convey paralinguistic information, over and beyond the phonetically-induced differences in voice quality. However, as attested by the sheer diversity and recurrent modifications of existing methods of voice quality analysis (cf. Buder, 2000), we are still far from a complete understanding of the acoustic correlates of phonation types, let alone their paralinguistic functions.

Whether the acoustic correlates of phonation types are studied in the time-domain (e.g., the glottal-flow waveform and its derivative) or in the frequency-domain (e.g., the magnitude-spectra of glottal-flow waveforms), *parameterisation* has been of fundamental importance both in revealing the underlying physiologic-acoustic relationships and in yielding models which are of practical use. For example, a number of parametric time-domain models of glottal airflow have been proposed (e.g., Rosenberg, 1971; Holmes, 1973; Fant et al., 1985; Klatt & Klatt, 1990), which have also found extensive use in both speech analysis and synthesis. Their differences notwithstanding, these models parameterise essentially what have come to be regarded as salient features of the glottal volume-velocity waveform or its first derivative (for a survey, cf. Cummings & Clements, 1995). Such features include the peak-to-peak amplitude of the glottal waveform, the duration of the glottal cycle (pitch period), the relative duration of the open phase (open quotient) and of the closing phase (closing quotient), the maximum slope of the closing phase, and the asymmetry of the main glottal pulse. Numerous studies have already reported on the relationships among these parameters, their relations with frequency-domain parameters such as spectral slopes and the relative amplitudes of harmonics, and their correlations with various types of phonation.

An alternative to imposing such pre-defined (albeit salient) parameters which control mathematically tractable functions, is to allow the underlying basis-functions to be determined empirically by statistical analyses of measured data. Such an approach was taken, for example in Harshman et al.'s (1977) factor analysis of tongue-shapes, which provided new perspectives on vowel articulation complementary to previous parametric models of the vocal-tract area-function (e.g., Stevens & House, 1955). Analogously but in a different context, the series of studies by Pols and his colleagues (e.g., Klein et al., 1970) showed that a factor analysis of vowel spectra yields basis-functions which can be related both to a perceptual vowel space and to the classical formant parameters.

By extension, it would seem interesting to ask whether similar statistical methods could not be used in search of an underlying structure to a space spanned by a range of laryngeal voice qualities – i.e., in search of basic dimensions of variability which underlie observed patterns of glottal airflow (and therefore indirectly, of vocal-fold vibration) across a range of phonation types. As described in the following sections, we propose to apply Principal Components Analysis (PCA) to glottal-flow waveforms estimated by inverse-filtering of the acoustic speech signal, specifically using data selected as representative of a broad range of laryngeal voice qualities.

2. Control Data and Analysis Methods

2.1. Laver's (1980) recording of voice quality

While we are aiming ultimately to use the methods developed here to analyse the large amounts of spontaneous speech being recorded as part of the JST project, it is important to begin our analyses with some control data. In particular, we turn to one of the leading authorities on voice quality – John Laver – whose recordings were made available as a demonstration tape accompanying his book (Laver, 1980). In the part of that recording relevant to our current study, Laver produces 17 repetitions of the phrase "*Learning to speak well is an important and fruitful task*", each time adopting a distinct laryngeal voice quality across the entire phrase. Figure 1 summarises the hierarchy of phonation types described in Laver's (1980) descriptive framework, and highlights the 15 categories for which there exists a corresponding utterance in the acoustic recording. The remaining two of the 17 categories included on the demonstration tape involve the overall muscular-tension settings of *tense voice* and *lax voice*; in so

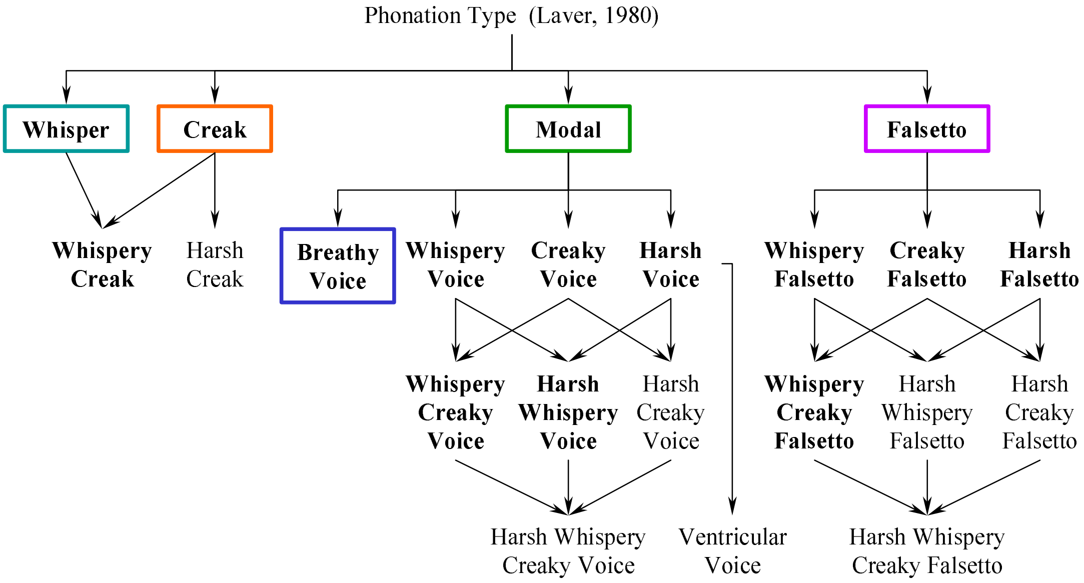

Figure 1. Summary of Laver's (1980) descriptive framework for laryngeal voice quality. The 15 voice qualities for which there exists a sample recording on Laver's demonstration tape, are highlighted in bold font.

far as these settings also influence the laryngeal part of the vocal system, we include them as part of our current analysis.

2.2. Acoustic measurement of single glottal cycles

Starting with a version of the recordings digitized at a sampling rate of 16kHz, the 17 utterances relevant to our study were manually delimited and stored as separate wave-files. Each utterance was then subjected to the automatic methods of acoustic analysis described in Mokhtari & Campbell (2003), in order to estimate representative glottal-flow waveforms within acoustic centres of reliability.

Briefly, our unsupervised algorithms first locate centres of measurement reliability in the speech stream, by a combination of acoustic-prosodic and cepstral analyses – (i) the contour of sonorant energy is used to find quasi-syllable boundaries and, in combination with a voicing decision, to demarcate syllabic nuclei; (ii) linear transformations of the linear-prediction (LP) cepstrum are used to obtain initial estimates of the first four formant frequencies and bandwidths (Broad & Clermont, 1989); (iii) the cepstra are also used to compute a contour of local spectral change, which is combined with a contour of distance between the original and the formant-generated spectra to obtain a simultaneous measure of formant reliability and spectral continuity; (iv) all the significant local minima along that composite contour which also lie within syllabic nuclei, are then retained as reliable centres. At each reliable centre, which is by definition five analysis frames (72msec) in duration, (i) the initial formant estimates are optimised by analysis-by-synthesis, then used to generate a time-varying inverse-filter; (ii) the original speech signal is first high-pass filtered to attenuate low-frequency rumble, then low-pass filtered to attenuate spectral information above the fourth formant, then inverse-filtered to remove the effect of the first four resonances of the vocal-tract; (iii) the resulting signal is simply integrated to obtain an estimate of the glottal flow; and (iv) as the absolute scale of the waveform amplitude remains unknown, the waveform is normalised by subtracting its mean value computed across the entire 72msec segment. Finally, a single cycle of the glottal waveform is automatically selected near the middle of each reliable centre, starting from a minimum in the waveform and extending by one period (as determined by the reciprocal of the fundamental frequency of voicing F0).

Despite the explicit attempt to locate reliable centres, there are admittedly various potential sources of error in the above procedures that inevitably contribute to imperfections in the estimated glottal cycles. As it is precisely the shape of the glottal-flow within a single period that is of interest in this study, and as our subsequent statistical analyses may be sensitive to errors which excessively alter the shape of the estimated waveform, the results were manually checked in order to retain only those measurements for which (i) the duration of the glottal cycle was not in error (e.g., owing to F0 measurement errors); and (ii) the effects of the vocal-tract resonances were reasonably removed by inverse-filtering (as judged visually by checking for the absence of formant-ripples in the glottal waveform, by a reasonable match of the optimised formants superimposed on the spectrogram, and by the absence of formant peaks in the short-time spectrum of the glottal waveform).

2.3. Principal components analysis (PCA) of glottal cycles

With the aim of revealing the underlying modes of variability, the manually-checked ensemble of single-cycle waveforms was subjected to principal components analysis. One possible approach would be to resample each waveform at a fixed number of equal-length intervals along the time-axis, thereby normalising differences in period length (or F0) and retaining variations in only the amplitude dimension. However, a more general approach is to acknowledge co-variations along both the amplitude-axis and the time-axis, thereby allowing the PCA greater flexibility to model variability across waveforms

Voice Quality	Number of measurements	Phonetic distribution of measurements within the phrase / ləːnɪŋtəspikwɛl ɪzənɪmpɔːʔəntənfruːtfʊltɑːsk /
M – *modal*	9	… **əː ɪ** **i wɛ** **ɔː** **uː ʊ** ……
F – *falsetto*	11	… **ɪ** **i ɛ ɪ** **ɔː** **uː ʊ ɑː** …
C – *creak*	3	… **əː** **ɪ ɪ** … … … …
WV – *whispery voice*	6	… **ɪ** **ə ɪ** **ɔː** **uː** … …
WF – *whispery falsetto*	6	… … **ɛ** **ɔː** **uː ʊ ɑː** …
CV – *creaky voice*	6	… **i** **ɪ ɪ** **ɔː** **uː ʊ** …
CF – *creaky falsetto*	2	… **ɪ** …
WCV – *whispery creaky voice*	3	… **ɪ** **ɔː** …
BV – *breathy voice*	7	… **əː ɪ** **i ɛ** **ɔː** **ʊ ɑː** …
HV – *harsh voice*	3	… **ɪ** **ə** **uː** …
HWV – *harsh whispery voice*	6	… **əː ɪ** **i** **ɪ** **uː** …
TV – *tense voice*	9	… **əː ɪ** **i ɛ** **ɪ** **ɔː** **uː ʊ ɑː** …
LV – *lax voice*	6	… **əː** **i** **ɔː** **uː** **ɑː** …

Table 1. Total number and phonetic distribution of measurements in each of the 13 laryngeal voice qualities of Laver (1980) for which the glottal-flow waveform could be estimated reliably. The approximate phonetic identity of acoustic measurements are indicated by the relevant symbol (in bold font) corresponding to a simplified phonetic transcription of the utterance.

simultaneously in both dimensions. This latter approach was adopted, by resampling each waveform at a fixed number of intervals equal in length along the path of the waveform itself, and retaining the values of both amplitude and time at each sample. As a compromise between preserving details of each waveform and keeping the number of parameters as low as possible, each period was resampled at 30 equal-length intervals, yielding 31 time-amplitude pairs (which, according to the sampling theorem, implies retaining approximately the first 15 harmonics of each glottal spectrum). Furthermore, as the amplitude offset of the glottal-flow estimated by inverse-filtering is unknown and the arbitrary amplitude differences between waveforms might contribute an undesirable artefact in the PCA, the first-difference of each resampled waveform was computed in order to obtain the *glottal-flow derivative* sampled at 30 coordinate pairs, for a total of 60 parameters on which the PCA would operate. Moreover, as the unrelated dimensions of time and relative-amplitude would bias the PCA to favour the dimension with larger variability, prior to PCA every amplitude-difference and time value was standardised by subtracting the overall mean and dividing by the overall standard-deviation of values in that dimension.

3. Results

3.1. Prototype glottal-flow waveforms

Owing to either an absence of, or errors in, measurements of F0 and the formants, no reliable glottal period could be obtained for the voice qualities *whisper* (which is unvoiced, by definition), *whispery creak*, *whispery creaky falsetto*, and *harsh falsetto*. As listed in Table 1, a total of 77 reliable measurements could therefore be made, covering only 13 of the 17 voice qualities analysed. Table 1 also shows that while the number and phonetic identity of measurements are not identical across the 13 voice qualities, nor is there an imbalance sufficient to raise doubts regarding the statistical analyses to follow (e.g., the voice qualities which are represented by only two or three examples may be assimilated into physiologically and acoustically neighbouring categories).

In order to gain a visual impression of the raw data prior to PCA, one representative glottal-flow waveform was computed for each of the 13 voice qualities, simply by averaging (both amplitude and time parameters) across the number of reliable measurements. The prototype waveforms thus obtained, are arranged along standardised coordinates in Figure 2. The three waveforms along the top are representative of the glottal airflow during three kinds of phonation in the *falsetto* register, displaying characteristically short periods and, for this speaker, relatively low peak-to-peak amplitudes; interestingly, the two waveforms contributing to our prototype for *creaky falsetto* are perhaps best categorised in the *falsetto* group, as the *creak* component appeared to be absent in the segment measured. The three waveforms along the bottom on the right represent three kinds of *creaky* phonation; while *creak* by itself appears to have a single, flat-topped pulse, *creaky voice* and especially *whispery creaky voice* appear to be mildly diplophonous. However, the waveform with the most accentuated diplophony

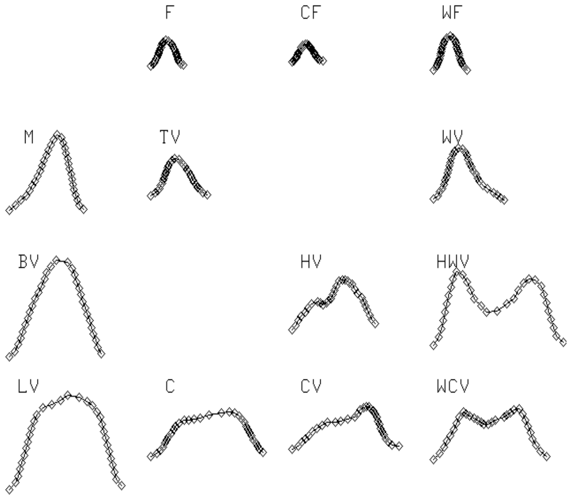

Figure 2. Prototype (mean) glottal-flow waveforms obtained for each of Laver's (1980) 13 voice qualities.

is that of *harsh whispery voice*; together with the prototype for *harsh voice*, these waveforms accord with evidence that a harsh voice-quality is often perceived with sufficient degrees of frequency jitter and amplitude shimmer (Laver, 1980), the modulations being caused physiologically by quasi-independent vibrations of supra-glottal structures such as the ventricular folds (e.g., Sakakibara et al., 2003). While all four of the waveforms in the rightmost column share a whispery quality, the prototypical *whispery voice* appears as almost a mirror image of *modal* phonation, with a more rapid opening and a longer closing phase. By contrast, the prototype for *tense voice* appears to be more symmetric and, similar to *falsetto*, of a shorter period and smaller amplitude. The prototypical *breathy voice* and *lax voice* (both produced with low muscular tension and inefficient glottal airflow) appear smoother, more sinusoidal, and (for this speaker) of greater overall amplitude and longer fundamental period.

3.2. Principal components of waveform variation

As described in the previous section, PCA was performed on the ensemble of 77 glottal-flow *derivative* waveforms to avoid artefacts that might arise from having an arbitrary amplitude scale. However, to ease visual comparison with the prototype waveforms of Fig. 2, the results of the PCA were reverted to an amplitude scale by first-order integration, arbitrarily setting the first sample to zero amplitude. Fig. 3 shows the overall mean glottal-flow waveform thus obtained (the solid line in each panel), together with both positive and negative perturbations by one standard-deviation in each of the first four PCs. The cumulative percentages of the total variance explained by these top four PCs are 57.6%, 80.8%, 88.2%, and 92.1%; hence, PCA of data originally represented in a 60-dimensional space has yielded orthogonal basis-functions, four of which suffice to account for over 90% of the variance.

The first principal component, shown in the top panel of Fig. 3, explains 57.6% of the original variance and appears to account for mainly the variations in the duration of the waveform, i.e., changes in the fundamental period of the glottal-cycle. This is not surprising, given that the data include such a wide range of fundamental periods spanning three registers of phonation from *falsetto* (with the shortest period), through *modal*, to *creak* register; however, the longest periods are attributed to *harsh whispery voice*, owing to its characteristic diplophony. Furthermore, together with changes in the fundamental period, the first PC also appears to account for concomitant variations in the shape of the waveform: the shorter periods are more symmetric and round-peaked, while the longer periods are broader and flat-topped; i.e., the first PC does not account for changes in the slopes of the opening and closing phases.

The second principal component, shown in the second panel of Fig. 3, explains a further 23.2% of the original variance and appears to account mainly for variations in the shape of the open phase of the glottal-flow waveform. In particular, the central portion of the waveform is either of high amplitude with a single peak slightly skewed to the right, or of low amplitude to account for the depression between two pulses in the diplophonic types of phonation. By contrast with the first PC and consistent with the orthogonal nature of the principal components, the second PC does not appear to model any significant variation in the fundamental period.

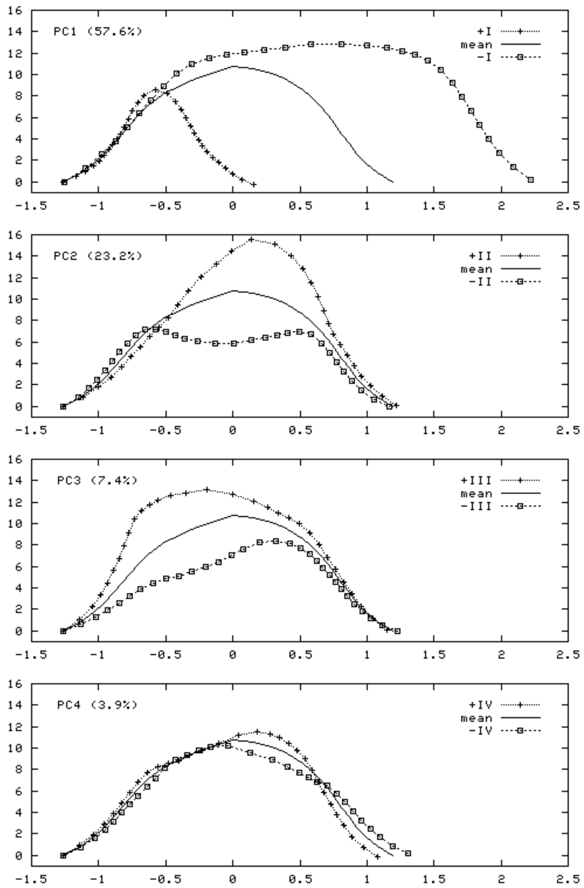

Figure 3. Mean and ±1σ in each of the first four principal components of glottal-flow, by PCA of the 77 flow-derivative waveforms measured in Laver's (1980) data. The abscissae are standardised, while the ordinates show arbitrary units of flow amplitude.

The third principal component shown in Fig. 3, also does not account for any variations in the period length; it explains a further 7.4% of the original variance and appears to account for variations in mainly the slope of the opening phase and the shape of the peak of the glottal pulse. At one extreme, there is a more rapid opening-rise followed by a relatively flat top; while at the other extreme a more gentle opening-rise is followed by a second, also gentle rise to the peak, the overall waveform resembling the prototypes for *creaky voice* and *harsh voice*.

The fourth principal component and the final one that we shall consider here, explains only a further 3.9% of the original variance and appears to account for mainly a combination of the skew of the pulse and the speed of the closing phase. At one extreme, the glottal-flow waveform is relatively symmetric with a more gradual closing-phase; while at the other extreme the pulse is slightly skewed to the right and is followed by a more rapid closing-phase. While the fifth and higher PCs account for much finer details of the waveform shape, they each explain no more than 2% of the original variance and will therefore not be considered here.

3.3. Voice quality classification by principal components

Interpretation of the relevance of the principal components to the voice qualities analysed is better visualised by projecting all 77 data-points onto two-dimensional planes defined by pairs of PCs. Fig. 4 shows that the data projected onto the PC1-PC2 plane do appear to cluster reasonably according to the intended voice qualities. As already noted in the previous section, PC1 models mainly the duration of the fundamental period, with the three types of *falsetto* clustered tightly at the positive extreme and the diplophonic *harsh whispery voice* at the negative extreme. Also as noted previously, PC2 models mainly a high-amplitude single-peaked pulse characteristic of *lax voice* and *breathy voice* at the positive extreme, versus a lower-amplitude double-peaked pulse characteristic of *harsh whispery voice* at the negative extreme. In addition to confirming these observations, Fig. 4 shows that *whispery falsetto* is somewhat separated from the other two types of *falsetto* by more positive PC2 values; *tense voice* and *whispery voice* form a closely overlapping cluster; *modal*, *breathy voice* and *lax voice* form distinct neighbouring clusters; *creak* and *creaky voice* also overlap together with one sample of *whispery creaky voice*; and *harsh whispery voice* is clearly distinguished by negative values of both PC1 and PC2.

By contrast with the PC1-PC2 plane, the data shown in Fig. 5 for the PC3-PC4 plane are not as clearly clustered according to the predefined voice qualities, and there appear a profusion of data points near the origin. Nevertheless, PC3 does seem to distinguish between *harsh whispery voice*, *breathy voice* and *lax voice* at positive values, and *creak* and *creaky voice* at negative values; PC4 could be useful in distinguishing *harsh voice* at positive values (rightward skew and a more abrupt closure), from *tense voice* and *whispery voice* whose projected data have mostly negative values (more symmetric pulse and gradual closure).

To test the efficacy of these top four PCs in classifying the 13 voice qualities, decision trees (Quinlan, 1993) were generated using the projection of the samples into the 4-dimensional space. A jack-knife method was adopted, where decision trees were built using subsets of 75 samples at a time and tested using the remaining 2 samples not included in the training (except for one case of 74 training samples and 3 test samples), thus making a total of 38 separate experiments. The overall classification accuracy obtained in this way, was 49/76 or 64%, which is much higher than the chance accuracy (around 8%) for 13 categories. It is interesting to note that *breathy voice* was the only category to yield a perfect classification score. On the other hand, the confusion matrix in Table 2 shows that the misclassifications are mostly intuitive: e.g., there were some confusions among the three *falsetto* categories, with both samples of CF misclassified as F (as noted earlier, the *creaky* element of CF is missing in our measured samples); there were confusions among WV and TV (which do have similarities in their prototype waveforms as shown earlier in Fig. 2); and two samples of WCV were respectively misclassified as C and LV (both having broad, flat-topped waveforms) while the third was misclassified as HWV (similarly diplophonous).

	CF	F	WF	TV	WV	M	BV	LV	HWV	HV	WCV	C	CV
CF	0	2											
F	1	**6**	4										
WF		3	**3**										
TV			1	**6**	2								
WV				2	**4**								
M					1	**7**	1						
BV							**7**						
LV								**5**	1				
HWV								1	**5**				
HV					1	1	1			0			
WCV								1	1		0	1	
C										1		**2**	
CV										1	1		**4**

Table 2. Confusion matrix showing classification of the 77 glottal-flow derivative waveforms into the 13 voice qualities, by decision trees constructed using PC1 through PC4.

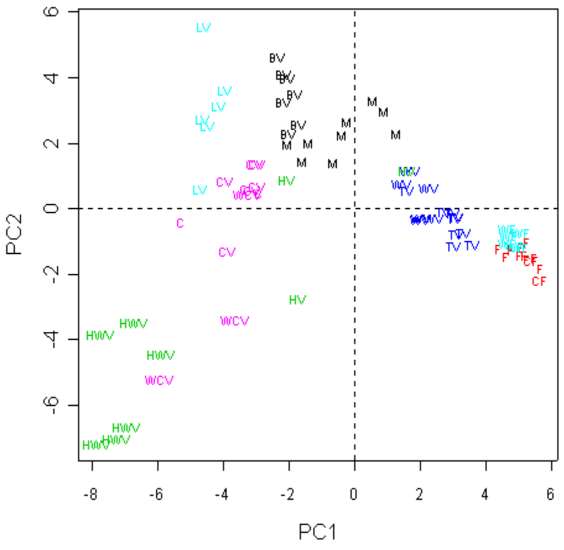

Figure 4. All 77 data-points projected onto the PC1-PC2 plane and labelled by Laver's (1980) intended voice quality.

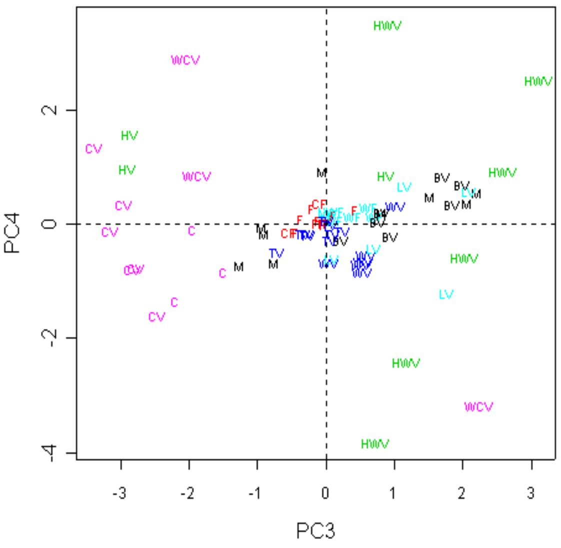

Figure 5. All 77 data-points projected onto the PC3-PC4 plane and labelled by Laver's (1980) intended voice quality.

4. Conclusions and Future Work

We have proposed a new, more holistic approach to modelling of the glottal volume-velocity waveform (or its derivative). By contrast with classical models based on pre-defined parameters which control salient features of standard stylizations of the glottal waveform or its derivative, we have attempted to derive the underlying basis-functions of a single glottal cycle, by statistical analysis of real data extracted from the acoustic speech signal of one male speaker demonstrating a variety of laryngeal voice qualities within a physiologically and auditorily-motivated descriptive framework. Within these constraints, four principal components were found which together account for over 90% of the variance in the data, and which have the following qualitative interpretations: PC1 – duration of the fundamental period, contrasting *falsetto* and *creak* registers; PC2 – convexity/concavity of the open-phase, contrasting a single-peaked glottal pulse and a diplophonic waveform; PC3 – speed of the opening-phase and consequent broadness and energy of the pulse peak; and PC4 – forward-skew of the pulse and speed of the closing-phase. Combinations of these principal components were used to construct decision trees which yielded an accuracy of around 64% in distinguishing among the 13 voice qualities.

Potential improvements to our methods include an increase in the resolution in representing each glottal cycle. The present choice of around 30 samples was driven by a compromise to retain sufficient resolution (equivalent to retaining about 15 spectral harmonics), while at the same time avoiding excessive ill-conditioning of the matrix input to the PCA by keeping the dimensionality of the parameter-space (60) lower than the total number of measurements (77). Use of more robust methods to make reliable measurements in a larger amount of data will obviate the necessity of compromising the time (and spectral) resolution.

Indeed, in ongoing work we are extending our methods to deal with much larger amounts of more spontaneous speech. Pending further research, the results reported in this paper can at best be regarded as speaker-specific and constrained within the context of prototypical phonation-types produced to demonstrate a range of voice qualities. For example, the high separability of groups of voice qualities along the first principal dimension alone (as evidenced visually in Fig. 4) may stem from Laver's use of fairly narrow and distinct ranges of intonation (or F0) across each entire utterance. Spontaneous, conversational speech data of the type we are collecting in the framework of the JST project, may well reveal a less distinct clustering along PC1, but also new and perhaps unforseen characteristics of the second and higher principal components.

Work in progress is also addressing questions regarding the acoustic and auditory-perceptual correlates of the principal components. In particular, we are exploring the relations of the PCs with acoustic-prosodic and spectral parameters measured in the same data, and their potential relations with more traditional parameters used to model the glottal-flow waveform and its derivative. Our PCA modelling approach is also being adapted in the framework of analysis-modification-resynthesis of speech, firstly in order to better understand the perceptual consequences of individually varying the principal components of the glottal excitation, but also with the practical aim of adding the flexibility of voice-quality control in speech-to-speech and text-to-speech synthesis. Our holistic approach is particularly suited to this task, as the data-driven modelling by principal components is by definition capable of adapting to a wide variety of phonation-types or laryngeal voice qualities.

5. Acknowledgements

This work is supported by the Japan Science and Technology (JST) Corporation under CREST Project 131. We are grateful to Mihoko Teshigawara (University of Victoria, Canada) for making available to us Laver's (1980) acoustic recordings; to Ken-Ichi Sakakibara (NTT, Japan) and Nick Campbell (ATR, Japan) for their invaluable discussions and advice; and to all members of the JST-CREST project at ATR for their support and encouragement.

6. References

[1] Broad, D. J. and Clermont, F., "Formant estimation by linear transformation of the LPC cepstrum", *J. Acoust. Soc. Am.* 86(5): 2013-2017, 1989.

[2] Buder, E. H., "Acoustic analysis of voice quality: a tabulation of algorithms 1902-1990", in Kent, R. D. and Ball, M. J. (eds.), *Voice Quality Measurement*, 119-244, Singular Publishing Group, San Diego, 2000.

[3] Cummings, K. E. and Clements, M. A., "Glottal models for digital speech processing: a historical survey and new results", *Digital Signal Processing* 5: 21-42, 1995.

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

[5] Harshman, R., Ladefoged, P. and Goldstein, L., "Factor analysis of tongue shapes", *J. Acoust. Soc. Am.* 62: 693-707, 1977.

[6] Holmes, J., "The influence of glottal waveform on the naturalness of speech from a parallel formant synthesizer", *IEEE Trans. Aud. Electroacoust.* AU21: 298-305, 1973.

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

[8] Klein, W., Plomp, R. and Pols, L. C. W., "Vowel spectra, vowel spaces, and vowel identification", *J. Acoust. Soc. Am.* 48(4): 999-1009, 1970.

[9] Laver, J., *The phonetic description of voice quality*, Cambridge University Press, Cambridge, 1980.

[10] Mokhtari, P. and Campbell, N., "Automatic measurement of pressed/breathy phonation at acoustic centres of reliability in continuous speech", *IEICE Trans. on Information and Systems*, E-86-D (3): 574-582, 2003.

[11] Quinlan, J. R., *C4.5: Programs for machine learning*, Morgan Kauffman, 1993. (C4.5 program available for download at www.cse.unsw.edu.au/~quinlan).

[12] Rosenberg, A. E., "Effect of glottal pulse shape on the quality of natural vowels", *J. Acoust. Soc. Am.* 49(2): 583-590, 1971.

[13] Sakakibara, K.-I., Imagawa, H., Niimi, S. and Osaka, N., "The laryngeal flow model for pressed-type singing voices", *Proc. Stockholm Music Acoust. Conf.*, 2003.

[14] Stevens, K. N. and House, A. S., "Development of a quantitative description of vowel articulation", *J. Acoust. Soc. Am.* 27: 484-493, 1955.