Precision Voice Analysis in Speaking, Singing and Pathology

Adrian Fourcin† **and Gea DeJong**‡
† University College London, UK
‡ City University, UK
E-mail: a.fourcin@ucl.ac.uk, g.dejong@city.ac.uk

ABSTRACT

Temporal periodicity is the main physical correlate used experimentally in relating perceived pitch to vocal fold frequency. The accuracy of most measurement methods does not, however, match the pitch difference limens of normal hearing.

The experimental results presented here are concerned with the observations, and resulting understanding, that can be derived when routine analyses correspond to a 0.1% level of frequency difference detection at 1kHz.

Three experimental situations are briefly discussed. The first two concern the relation between sustained vowel production and the modal structures found in the analysis of representative samples of voice in connected speech for two groups of subjects. Two of the main factors that contribute to our ability to control voice pitch come from auditory processing and proprioceptive laryngeal feedback. The analysis of representative samples of fluent speech makes it possible to define the main modal values of vocal fold vibration available to a speaker. In order to maintain pitch stability in a sustained sound we have found that the normal speaker chooses a dominant normal modal value of vocal fold vibration – defined from within the characteristics of connected speech. The normally hearing speaker but with a pathological voice adopts a related strategy. However, for such speakers the choice of dominant mode may lead to stably controlled voice pitch but at a quite abnormal vibratory frequency. These results have practical implications in the clinical management of voice pathology since it is not present practice to relate estimates of degree of pathology derived from sustained vowels to measurement of connected speech. The third set of related experiments concern the links between singing and auditory monitoring. There is an inverse relation between our ability to detect pitch change and the duration of the sound. This leads to an approximately tenfold difference between detectable pitch changes in sung vowels and the intonational changes of conversational speech. Crossplot [period by period] analyses of voice frequency irregularity in speech and singing lead, in consequence, to large predictable differences between these modalities of production.

1. INTRODUCTION

Our ability to perceive changes in the pitch of sounds is to an important extent a function of the nature of the sounds themselves. Table 1 gives voice frequency range examples.

pure tone [1] [2]	**vowel**		**connected speech** [6]
	sustained [3] [4]	fall or rise [5]	
0·4% to 0·7%	0·15% to 0·3%	8%	6% to 20%

Table 1 Pitch difference limens – static & dynamic

The levels of pitch discrimination for steady sounds imply an **average** ability to detect temporal differences of about 4 μs and for some individuals, 1μs. The implication for voice pitch range perception is that the voiced sounds of running speech are auditorily processed quite differently from those of sustained vowels and, in consequence, to a degree in singing. Since the ability to productively control the sounds of speech is dependent on and mediated by the speaker's auditory processing we can expect that the measurement of the detailed course of pitch variation in speaking and singing will lead to the discovery of corresponding differences of vocal fold frequency variation. Special techniques are needed for the accuracy required.

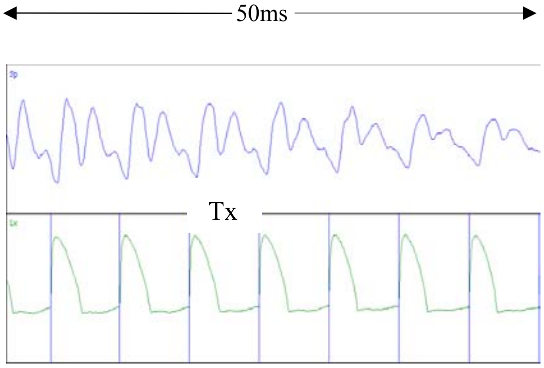

figure 1 The acoustic signal, Sp, above is synchronous with the laryngograph/egg signal, Lx, below; automatically generated markers indicate the closed phase onsets used for measurement with very high accuracy.

ISBN 1-876346-48-5

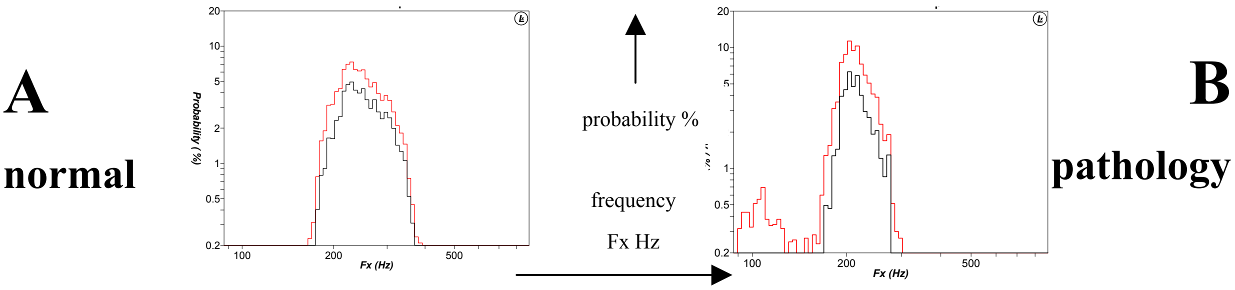

Figure 2 DFx1&2 – The graph sets of figure 2 show probability of occurrence of a vocal fold frequency plotted against the instantaneous frequency, Fx. The inner, joint probability, digram distribution indicates the pitch bearing components of the speaker's voice. The outer distribution, however, shows the frequency range of all the vocal fold periods in the whole of the voiced sample. For the normal voice, A, there is little difference between the two distributions since the normal voice has very little of the irregularity that tends to make successive vocal fold contacts differ in timing. In the distribution for the abnormal voice there is a marked difference between the two distributions arising from the presence of voice irregularity.

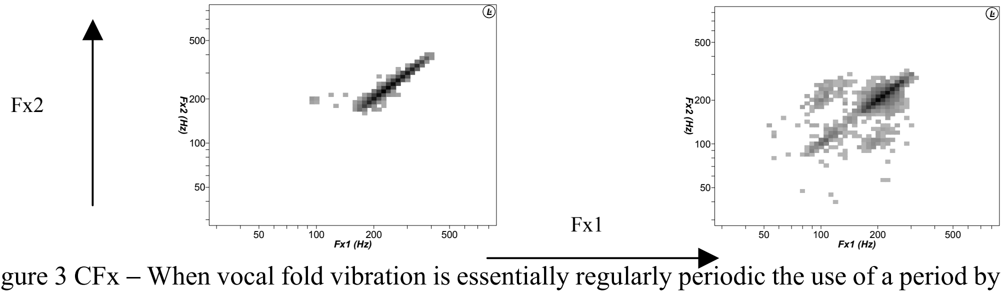

Figure 3 CFx – When vocal fold vibration is essentially regularly periodic the use of a period by period crossplot, as shown here for A, gives a clearly defined diagonal line – since successive periods (Fx1 – Fx2) have almost the same values, apart from the variations arising from the intonational frequency related changes of connected speech. For the pathological voice, B, however, the shape of the crossplot is not so simply defined because successive vocal fold periods are very often markedly different and are not totally under the speaker's cognitive control.

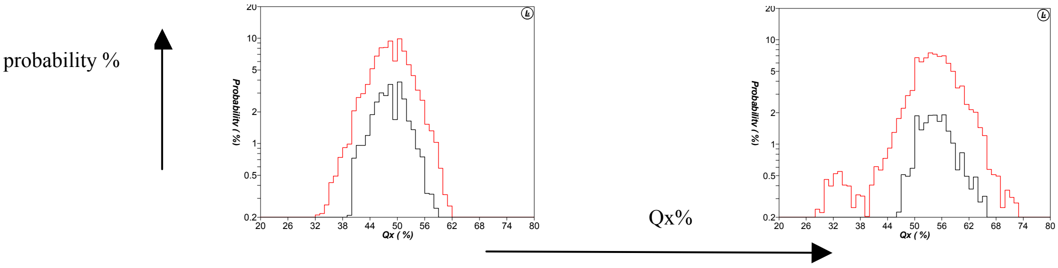

Figure 4 DQx 1&2 – distributions of first and second order "closed phase" (Qx%) as a function of vocal fold frequency, Fx. Voice quality is a complex attribute of voice but one important aspect comes from the regularity and duration of the closed phase from vocal fold cycle to cycle. First and second order plots can often give important information in regard to the physical nature of a pathological voice and here it is evident that speaker B has poor closed phase coherence.

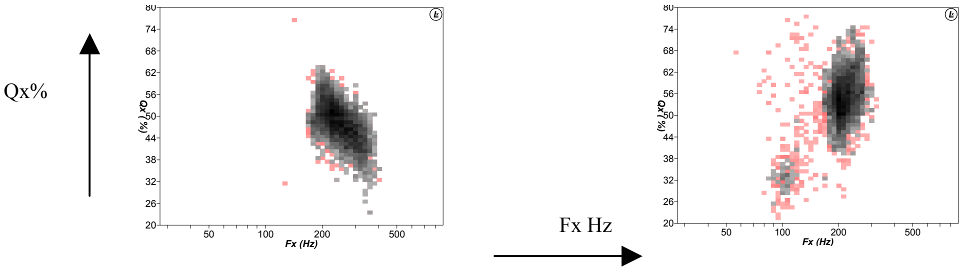

Figure 5 QxFx – "Closed phase" ratio Qx as a function of vocal fold frequency. In normal speech the variation of the closed phase at the level of discourse is a rule governed phenomenon whose rules differ from one person to another. Speaker A is a young woman and shows an effect of increasing breathiness with larynx frequency. The pathological voice, B, is substantially deviant and gives a range of Qx [the closed phase measure based on trans-glottal conductance] which is never found in the normal voice and shows the irregularity as a function of frequency which can be clearly heard in her speaking voice.

2. EXPERIMENTAL PROCEDURE

Subjects

Subjects for this study were five female speakers, aged 32-39 (mean age 32.0y) and two male speakers, aged 28-45 years (mean age 36.5 years), who were demonstrably healthy and exhibited no speech or hearing disorders. All speakers were native speakers of British English.

Task and materials

Each session consisted of four different types of phonation tasks. All subjects was asked to 1) sustain the vowel [A] for a minimum duration of 15 seconds at a comfortable pitch level at the beginning and end and in the middle of each session, 2) read verse 1, 3 and 4 of the British National Anthem, 'God save the King/Queen', 3) sing the same verses of the Anthem, and 4) read the tongue twister 'Peter Piper picked a peck of pickled peppers' three times.

In order to familiarize the subject with the material, each subject started with a warm–up session, practicing each task. All tasks were repeated again in the second session with their speech and vocal fold vibration patterns being recorded using the SpeechStudio Laryngograph® System developed by Laryngograph, Ltd. London. During the second session, the intensity levels exhibited by the subject performing each subtask were measured and recorded using a calibrated sound level meter (RadioShack Sound Level Meter, 33-2055). In session three, the subject was asked to perform the same phonation tasks again while being binaurally masked by a pink masking noise (e.g. spectral frequency of 1/f). The masking signal was produced using the speech analysis software Cool Edit Pro, version 1.2. To ensure an optimal masking level that was not painful or damaging to the listener's hearing, the subject was asked to adjust the intensity level him/herself by first increasing the level until it was painful (but below the maximum intensity level allowed for binaural masking) and subsequently decreasing it again by about 5 dB. In order to combat the Lombard Effect (Lane et al 1961, Lane and Tranel, 1971, Siegel and Pick 1974) and to control for intensity, each subject was asked to maintain the intensity level exhibited and registered during the second session by checking the sound level meter. The Lombard Effect is the term used for the phenomenon where speakers, when subjected to a masking sound, automatically increase their speaking intensity in order to restore auditory feedback. The final session consisted of two subtasks only: 1) sustaining the vowel [A] for a minimum duration of 15 seconds at a comfortable pitch level and 2) reading verses 1,3 and 4 of the British National Anthem. In this session the subjects were asked to reduce their speaking intensity to such a level, that their speech was masked completely. Masking a speaker's speech entirely is practically impossible without subjecting the speaker to an intensity level that is between uncomfortable and painful. However, the aim of the final session was to achieve complete masking by adjusting the intensity level of the speech output. In other words, complete masking was attempted by adjusting the intensity levels of the (masked) signal of interest instead of increasing the intensity levels of the masking signal.

3. RESULTS

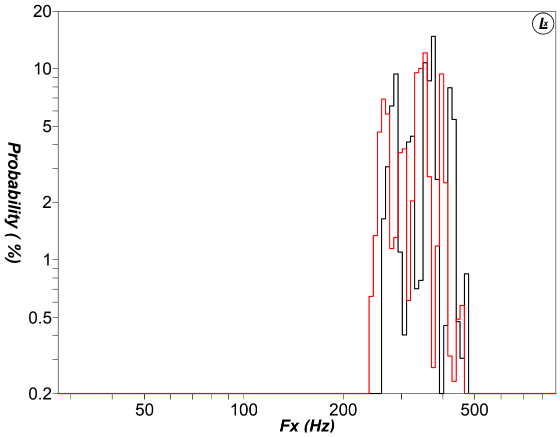

Figure 6 Second Order "pitch based" vocal fold frequency distributions – black, unmasked singing; red, masked singing – showing vocal fold (vf) frequency congruence; for woman subject CB

A variety of larynx excitation parameters were calculated and studied and some exploratory observations were made. Firstly, the Frequency Distribution graphs (DFx2) showed that for most speakers [figure 6 is an example], the distribution remained fairly similar for both unmasked and masked singing conditions. Masking appeared to affect neither the range nor the distribution of vf frequencies.

Figure 7 Vocal fold frequency crossplots for reading [left] and for singing [with slight vibrato, right]; giving a qualitative indication of period to period vocal fold vibrational irregularity.

Second, when looking at a period-by-period cross plot, it was noted that CFx irregularity for singing tended tobe less than the irregularity for the reading condition and that the difference could be quite significant. This trend was found for both male and female speakers. For a number of speakers, however, CFx irregularity increased for the singing condition compared with the reading condition. After close examination of the original singing data, it was found that the use of (unintentional) vibrato caused the increase. Even relatively small samples of vibrato could have an effect.

ISBN 1-876346-48-5 © 2003 UAB

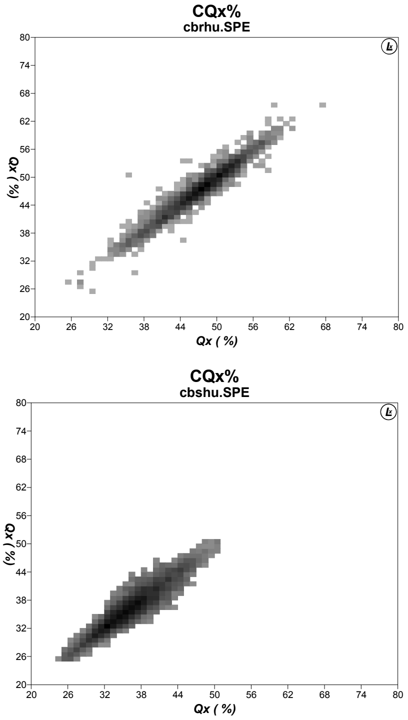

Figure 8 The upper figure shows the closed phase crossplot CQx for CB reading and the lower plot for singing the same text – CFx width] closure variability.

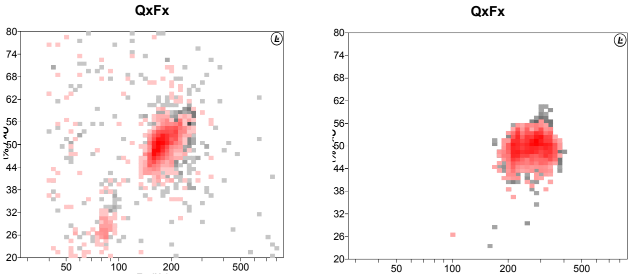

Figure 9 QxFx gives an overview of the relation between closure and vf frequency control for subject AW – left reading and right singing; black unmasked, right masked – Qx% ordinate; Fx Hz abscissa

Thirdly, when studying the closed phase ratio in relation to Fundamental Frequency (i.e. QxFx), it was noted that, as one may expect, this parameter was much better controlled for the singing condition compared with the reading condition. A particular example is shown above in figure 9. The one subject, GW, who failed to show this pattern, was also suspected of diplophonia. Although, she did control for pitch accurately, GW exhibited a rather weak singing voice and patterns of diplophonia.

When observing timing and pitch data, it was found that those subjects who exhibited a correct timing and pitch relationship during the unmasked singing condition, in other words, those who maintained the correct tune and rhythm, also controlled those parameters during the masked condition. Their singing skills did not seem to be affected by masking.

Further developments:

Further experiments will be carried out adjusting experimental procedures utilised in the pilot study. In order to obtain irregularity measures that are a good reflection of voice quality, it seems crucial to control for vibrato. In addition, it was noticed that the singing material, ‘God save the Queen’, may have too many (emotional) associations attached to it that influence the manner of singing. Using more anodyne material may give more opportunity to standardise and control the manner of reading and singing. Another parameter related to this was timing. This study did not include any real or professional singers. The control and use of their voice is the object of current work.

4. CONCLUSIONS

Although the main inferences to be drawn from the difference limen implications for the pitch perceptual ability of speakers and singers to control voice frequency appear to be correct, the detailed mechanisms of pitch mediation are of much greater complexity. Effects associated with vibrato and the need expressively to control voice quality via closed phase variation introduce experimental variables that require the detailed extension of our work.

REFERENCES

[1] Nelson, D.A., Stanton, M. E., and Freyman R. L., 1983, “A general equation describing frequency discrimination as function of frequency and sensation level” *J.A.S.A. 73, 2117-2123*

[2] Vance, T.F.,1914, see page 340 in E.G. Boring, *“Sensation and Perception in the History of Experimental Psychology”*, Appleton-Century-Crofts Inc. New York 1942

[3] Wier, C.C., Jesteadt, W., Green, D.M., 1977, “Frequency discrimination as a function of frequency and sensation level”, *J.A.S.A. 61, 178-184*

[4] Moore, B.C.J. Glasberg, B.R., Peters, R.W., 1985, “Relative dominance of individual partials in determining the pitch of complex tones” *J.A.S.A. 75, 550-561*

[5] Klatt, D. H. 1973, “Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception”, *J.A.S.A. 53, 8-16*

[6] t’Haart, J.,1981,”Differential sensitivity to pitch distance, particularly in speech” *J.A.S.A. 65, 811-821*

[7] H.L., Lane, A.C. Catania,, and S.S. Stevens, “Voice Level: Autophonic Scale, Perceived Loudness, and Effects of Side Tone,” in *J. Acoust. Soc. Am.*, 33: 160-167, 1961.

[8] H.L. Lane and B. Tranel, “The Lombard Sign and the Role of Hearing in Speech”, *in J. Speech Hear. Res*, 14: 677-709, 1971.

[9] G.M. Siegel and H.L. Pick, “ Auditory Feedback in The Regulation of Voice”, *J.Acoust. Soc. Am*, 56: 1618-1624, 1974.