CLINICAL VOICE MEASUREMENT USING EGG/LX SIGNALS 4TH INTERNATIONAL MAVEBA WORKSHOP

Adrian Fourcin[1 2]
[1]Department of Phonetics & Linguistics, University College London, UK, fourcin@btinternet.com
[2]Laryngograph Ltd. www.laryngograph.com

Abstract

Nearly fifty years ago, Philippe Fabre[1] initiated a method for the non-invasive electrical measurement of vocal fold vibration that is now known generically as "electro-glottography" — egg. The name has arisen from the initial misinterpretation of the waveform, but the technique itself has now come into widespread daily use in the voice clinic, although for vocal fold contact rather than glottal opening measurements. The present extremely brief discussion is concerned with three particular aspects of the ways in which the approach can be usefully linked to basic aspects of voice perception and production — the psychophonetic use of the data in the measurement of sustained vowels and connected speech production; the use of these criteria in "pitch" based quantitative assessments; and the application of the technique in vocal fold closed phase duration appraisal. Particular attention is given to pathological voice analysis.

I. Introduction

The use of non-invasive electrical sensing, during fluent speech production, of vocal fold contact has especial research and clinical advantage in the definition of:

- contact closure epoch
- instantaneous period & intra period irregularity
- peak acoustic excitation
- closure duration value and variability
- precision stroboscopic trigger instants.

The approach also makes it feasible to link objective measurement to pitch perceptual processing.

II. Methodology

Using pitch perception to guide voice measurement

For most practical purposes the really important aspects of voice are those that can be heard, and the dominant dimension in hearing voice is pitch. This simple concept leads to the possibility of using some simple quantitative criteria to detect and quantify the differences between "good" and "bad" voices.
Classically, pure tones provide a basic reference for both the definition and perceptual investigation of pitch. Subjective psychophysical data have been stably established over many years. Maximum discriminability is reached between 1 kHz, C6, near the top of the soprano register, and 2kHz with an average best just noticeable difference, jnd, of about 0.7% at 200Hz and 0.4% or 4 Hz in the region of 1kHz with individual jnd sensitivities going down to 0.1% [2]. Auditory pitch detection for the frequency ranges of the speaking and singing voice appear to employ mechanisms which operate on the basis of temporal processing [2]. This level of pitch discrimination implies an average ability to detect temporal differences between successive periods of about 4 μs, and for some individuals, 1μs. This temporal signal processing ability for pitch perception is paralleled in auditory lateralisation where interaural time differences of about 2μs to 10μs are detectable.

For steady complex tones and vowels in the fundamental frequency range of conversational speech, the pitch discrimination jnds are even smaller than those obtained with pure tones. Wier and Moore [2], within the range 200 to 600 Hz, reported jnd values from about 0·15% to 0·3%. For vowel-like sounds with simple changing fundamental frequency contours, however, the ability to perceive differences in fundamental frequency is drastically reduced and the jnd may be 8% at about 100 Hz . This increase in, and magnitude of, jnd has also been found for whole word utterances with simple intonation contours, the jnd here never being less than 6%. When more complex contours are used, the differences needed to achieve reliable detection may be as great as 20% . The subjective results for these stimulus types are not as well established as for sustained sounds and there is a dependence on the duration of the tone. There is, however, a good working consensus between a large number of reported observations [t'Hart, 2]. These established observations give clear implications in respect of the accuracy criteria which should be aimed at for the analysis of the separate categories of sustained sounds and connected speech.

A basic set of tools for accurate voice pitch measurement

Tool 1

Most methods of voice pitch analysis depend on the use of the acoustic signal of speech sampled at low rates which do not correspond to the requirements imposed by the pitch dL performance of the ear. The essential need which has to be met is best defined by an example taken from the singing voice. At a voice frequency of 1000Hz in the soprano range the human ear can detect changes of

*Models and analysis of vocal emissions for biomedical applications. 4th international workshop. October 29-31, 2005 – Firenze, Italy.
Edited by C. Manfredi. ISBN 88-8453-320-1 (online)*

around 0.1% [2]. In order to do as well as this, but for only a single cycle, it is necessary to use a sampling frequency of 1MHz. This is what is done for the following measurements.

Tool 2

A second basic problem associated with conventional approaches to voice analysis comes from the inadequacy of pitch extraction algorithms based on the acoustic signal. A more reliable technique is to use the electrolaryngographic [Lx, positive peak corresponding to maximumclosure] output from the speaker's voice activity. This gives the basis for an accurate determination of each individual pitch period, Tx, that can be sampled at 1MHz to support measurements which, although very highly detailed for many purposes, are linked to the best that the ear can do and that provide for considerable flexibility in the choice of bin widths in analysis and graphical displays.

It is, of course, quite easy to deal with sustained sounds but the method must also be reasonably robust when applied to the rapidly changing waveform of running speech. An output for a practical system is shown below.

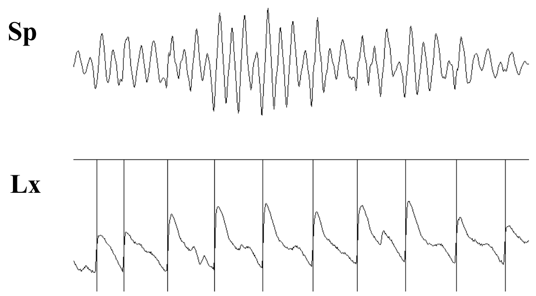

Figure 1 acoustic,Sp, and electrolaryngograph ,Lx, waveforms for a sample of pathological speech with automatically detected closure markers (55ms duration)

The figure shows the process of marker generation used for the definition of Tx for a pathological voice sample; the excerpt is from a sample of fluent speech. The speech acoustic waveform, Sp above, illustrates the difficulty of cycle definition if only acoustic data is available.

The use of period by period sampling gives a very clear view of the difficulties that may be encountered by a speaker with a voice disability and shows the remarkable precision of normal voice pitch control. This type of precision data analysis is also basic to the provision of measurements both for sustained sounds and for the analysis of running speech.

The current clinical techniques for the quantification of voice abnormality depend to an appreciable extent on the use of sustained sounds and the standard protocol uses the steady state in the centre of the sample. Period by period analysis gives a clear indication of the onset and offset transients which this approach misses. The pathological speaker in general has difficulty in producing smooth voice onset and offset. This is clearly seen initially, where diplophonic breaks in the voice precede more steady production, and in the voice breaks at the end. As must be expected, perception leads production but it is striking and commonly observed, that the pathological voice does not have a jitter commensurate with the disability. This small difference results partly from the choice of the centre interval of a sound sustained at a comfortable pitch – and partly from the speaker's auditory monitoring ability for sustained sounds, and phonatory choice of a dominant mode.

III RESULTS & DISCUSSION

Connected speech and sustained vowels

For the majority of the population, speech communication is at the heart of our daily lives. Clinical voice measurement, however, is mostly directed towards the appraisal of the ability to produce a sustained vowel. Since there are quite substantial perceptual differences between our ability to hear pitch regularity in sustained vowel sounds as opposed to fluent speech, it would be of interest to make at least an initial appraisal of the ways in which perception and production may interact in the voice pitch structures of the two types of phonatory activity. There may additionally be an advantage in comparing pitch regularity inspired analyses based on the two types of spoken material simply with a view to contributing to filling the gap between clinical indices of severity of dysphonia based on vowel measurement and those using a perceptual evaluation of continuous speech. Most important of all, however, is both to make use of pitch criteria and to take account of the nature of pitched sounds. Regular repetition of an acoustic event and perceived pitch go hand in hand.

Analyses of ordinary running speech

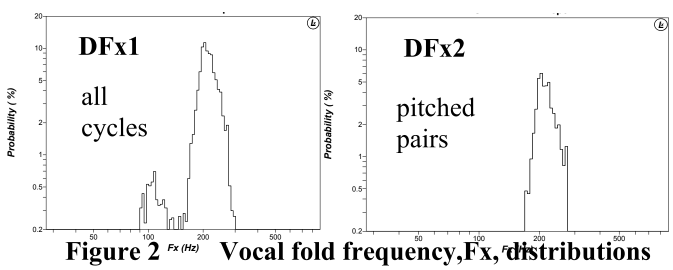

Figure 2 Vocal fold frequency,Fx, distributions for a 2m sample of pathological connected speech — speaker B

The two distributions in Fig 2 are very dissimilar. DFx1, on the left, shows the distribution of Fx values for every vocal fold period in the whole 2m. sample. DFx2 shows only those Fx values for which two successive periods have been essentially the same. Two modes of vocal fold vibration are shown. The main at about 200 Hz is well defined. At about an octave below, the lower mode is more diffuse and is evidently associated with considerable period to period irregularity since the values of DFx1 and DFx2 are so different. Different pathologies give rise to different types of modal structural differences but for most cases the presence of voice pathology will be associated with marked discrepancies in magnitude and shape between these two forms of representation.

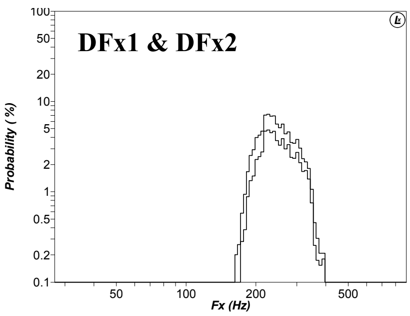

Figure 3 Overlaid Fx distributions for a normal sample (speaker A) AND for its "pitched" components

The use of accurate period by period information makes it easy to plot the occasions when two successive pitch periods have essentially the same value. For the normal voice this happens very often indeed. The pathological voice, however, is very easily identified by the ear as having period to period irregularity. This important feature is shown in the inner of the two distributions above in Figure 3. The two distributions together give an immediate insight into important aspects of voice quality – pitch height and range, modal structure, and regularity. These first and second order distributions are especially useful in general in pathological voice analysis.

Jitter and irregularity in connected speech

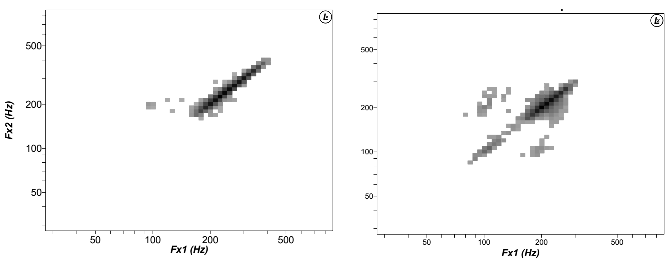

Figure 4 Vocal fold period crossplots, CFx speaker A on the left, B on the right

Jitter and intonation

The procedure basic to the ordinary application of the jitter criterion is applied only to sustained sounds and requires that the voiced sound being measured is held at as constant a pitch as possible by the speaker. The essential concept, however, is directed at obtaining a quantitative assessment of pitch variability. The idea is just as applicable to ordinary connected speech so as to get an appraisal of the irregularity which may be inherent to the social use of a pathological voice.

An obvious first approach to the measurement of pitch irregularity in a sample of running speech is to determine the standard deviation of the spread of cycle to cycle differences in regard to periods or frequencies. A difficulty with this approach is that it will necessarily include ordinary intonational variations as part of the estimate of irregularity. The problem is perhaps best illustrated with reference to actual data. When vocal fold vibration is essentially regularly periodic the use of a period by period crossplot, as in Figure 4 A, gives a clearly defined diagonal line – since successive periods have almost the same values, apart from the variations arising from the intonational frequency related changes of connected speech. For the pathological voice, 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. This method of plotting the range of variability in period to period coherence is effectively similar to the application of the jitter criterion, used for sustained sounds, to the whole of a connected speech sample.

The interpretation of jitter in running speech, however, is not at all the same as that for sustained sounds. First, the pitch dLs are quite different in the two cases. The bin sizes needed for the adequate representation of significant changes in the present data involves 6% steps. The 0·1% resolution required for the analysis of sustained sounds is not appropriate. Second, the presence of intonational changes makes it necessary to ignore variations which are part of the normal patterning of vocal fold frequency change in running speech. Figure 4 A shows that there is indeed a centre continuous core of variation for the whole of the vocal fold frequency range and this is found for all normal speakers.

If the pitch difference limen value of 6% is applied to this data then it becomes possible to apply a theoretically founded criterion which makes it feasible in practice to separate the variability arising from intonation from that due to other causes. It is then only necessary to determine all the pitch deviations which are more than 6% away from the centre line in the graph showing Fx1 against Fx2 – where Fx1 is the frequency value of the first vocal fold cycle in any pair of cycles in the whole utterance and Fx2 is the frequency value of the immediately following cycle of the pair. Fx is used to denote the frequency value of a single vocal fold cycle, the period of this cycle being measured from point of closure to point of closure.

Normal and pathological voice examples

The comparison of Figure 4A with 4B shows how the relatively small jitter differences sustained sounds from these speakers, of .3% & .8%, is related to quite marked structural changes in their samples of connected speech. In these particular instances, irregularity is 3·2% for the normal speaker and 14·7% for pathological speaker B. Both values were measured in the way described above as a percentage of the number of vocal fold periods, outside the centre core of intonation-dependent pitch change, relative to the total number of vocal fold periods in the whole spoken sample.

Loudness and Quality

Connected speech phonetogram

The standard phonetogram was designed to provide an overview of the dynamic range of a singer's voice and was based on the separate production of sustained sounds. The same principle can be applied to the analysis of the speaking voice to give first and second order "Dynamic Phonetogram" derived from the amplitude –frequency analyses of a complete sample of connected speech (also called Speech Range Profile).

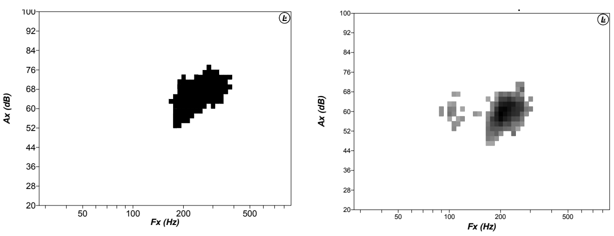

Figure 5 Second Order Dynamic Phonetograms derived from 2m. samples of connected speech: normal speaker A, left; abnormal voice speaker, right

In both Figure 5A and B, only the second order distributions are shown. This has little effect on the presentation for speaker A; it does have a profound influence on the form and range of the data presentation for speaker B since the presence of a bimodal peak in loudness is very evident in the first order distribution but not in the "pitch" related second order plot.

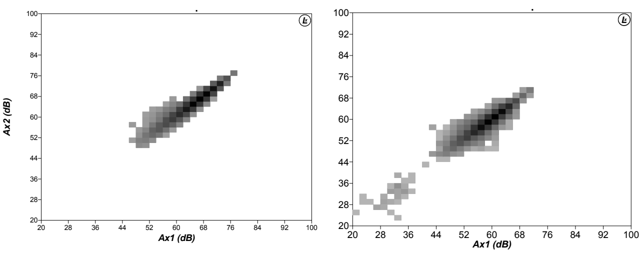

Figure 6 Period by period amplitude crossplots – CAx

A factor contributing to our perception of hoarseness comes from the irregularity of successive amplitude peaks in the cycle to cycle excitation of the vocal tract. This is especially evident in connected speech and speaker A on the left, Fig. 6, has a smaller spread in these analyses than B. Using a similar measure of irregularity to that employed for CFx gives values respectively of 3·3% and 6·5%.

IV IN CONCLUSION

Voice quality, "closed" phase and pitch

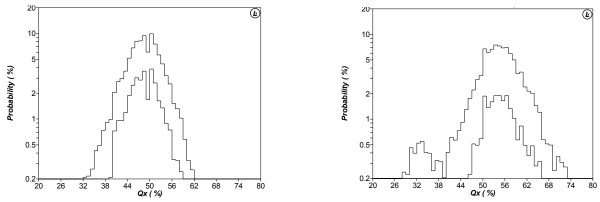

Figure 7 DQx 1&2 – distributions of first and second order "closed phase" as a function of vocal fold frequency, Fx

Voice quality is a complex attribute of voice but one important additional 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, in Fig 7 it is evident that speaker B has poor closed phase coherence, and range.

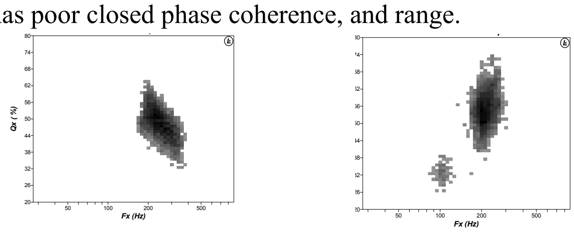

Figure 8 "Closed phase" ratio Qx as a function of vocal fold frequency — A left, B right

The pathological voice, B Fig 8, 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 relates to the irregularity as a function of pitch which can be clearly heard in her speaking voice. More generally, the Lx waveform can provide a sensitive basis for analysis that can be used effectively from the operatic voice to more extreme conditions [3].

I would like to acknowledge the great help received from Julian McGlashan FRCS and Dr Evelyn Abberton.

REFERENCES

[1] P.Fabre,"Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonationBull. Acad. Méd, 1957, pp.66-70

[2]please see "Meauring Voice in the Clinic", www.laryngograph.com for reference list

[3] A.Fourcin, "Precision Stroboscopy, Voice Quality and Electrolarlyngography", in Ch 13 Voice Quality Measurement ed RD Kent & MJ Ball, 2000