

# What Voice Clinicians Need from Bioengineers

Mette F. Pedersen

The Medical Centre, Oestergade 18, Dk-1100 Copenhagen Denmark

*Purpose:* Discuss issues regarding pragmatic interdisciplinary cooperation that will benefit patients, clinician and academics carrying for the human voice

*Method:* Retrospective meta review of current literature and needs.

## Background

There are several newer services for the diagnostic approach of voice disorders; nanotechnologists, biotechnologists and bioengineers. Nano- and biotechnologists are trained from the biological approach and bioengineers from the technological aspect. This means that discussing pharmacological aspects of diagnosis and treatment is a point for the nano- and biotechnologists and mechanical instruments for the bioengineer, working in the biological field. There is great need for bioengineers, e.g. in radiotherapy and cardiology. The problem is that their evidence and transparency of formulas in software till now have not followed the rules of biological documentation when it comes to voice research as accepted in nanobiology and biotechnology (1)

As earlier concluded, the biotechnology industry will have to use the same methods as in pharmacology to document their technology (2). The exact formulas e.g. of the evidence based cestrum measure as used in software from factories like Wolf Ltd. has been difficult to get and so has the possibility to compare to other formulas, in order to make prospective randomized studies, and to evaluate them with statistical relevant power calculation between groups.

The clinical need for voice therapists in the future seems to be clinical evidence to document biotechnological measures and transparency of hardware and software.

The following examples are given.

## High speed films

The speed of the closing/opening of the vocal folds during phonation at fundamental frequency approximates 110 times per second in males and 220 times in females, and the velocities approach one meter per second (3). In order to visualize this speed to the human eye, the speed must be slowed down significantly (4). One technology, that is especially helpful for this purpose, is high speed imaging. The principle behind the high speed technology is that the camera is able to photograph the vocal-fold vibrations with speeds exceeding the frequency of the vibration, where after the images are presented to the viewer at slower rates.

The first high speed film machine, that was produced in the 1930s, resulted in the conduction of many studies examining vocal-fold vibrations, and provided crucial understanding of the laryngeal physiology (5, 6). The high speed cameras have been developed throughout the years, and today, the cameras are able to record up to 4000 (and higher) frames per second.

In the field of laryngology, high speed imaging can greatly aid in evolving the diagnostic approach and investigations of different aspects of voice problems, especially in the investigation of disorders where the less accurate video-stroboscopy measures cannot be used. The high speed technology has recently been used to gain insights into voice production of patients who have been through endoscopic phonosurgical treatment of early glottic cancer (7). High speed imaging can furthermore aid in developing methods for restoration of damaged biomechanical properties of the mucous membrane lining the vocal folds. In neurological voice disturbances, focus has been on dystonia where spasms are clearly seen. It is therefore of great importance that the biotechnology of the high speed cameras are further developed based on quantitative evidence in the voice clinics.

#### **Fundamental frequency and intensity variation on sustained tones and during reading of a standard text, and cepstral peak prominence**

The fundamental frequency is defined as the lowest frequency produced by any instrument (including the vocal folds). The average adult male has a fundamental frequency of 85-180 Hz and the average female has a fundamental frequency of 165-255 Hz. Control of the fundamental frequency has an important role in vocal communication. By regulating the fundamental frequency, one is able to express information about for example speech inflections (8).

The intensity variation (in decibels) measures how much one is able to control the voice, while producing a tone (typically (/ah/)) or while reading a standard text. A large variation in the intensity is widely seen during puberty, especially in male teenagers where the phenomenon of their voices 'breaking' can be observed. (9)

'Cepstrum' refers to the phenomenon, that the logarithm of a power spectrum of a signal with echo has an additional periodic component due to the echo. In other words, the 'cepstrum' refers to the spectrum analysis of a spectrum analysis.

The cepstral peak prominence measurements are useful parameters in voice research, and have shown to "correlate well with perceptions of breathiness" and should predict roughness of voice (10). A meta-analysis from 2009 evaluated different acoustical analyses with listeners test, as the only evidence-based work on this topic. The meta-analysis found the cepstral metric to be the most promising and most robust acoustic measure of dysphonic severity (1). It should be developed further for clinical use.

#### **Formant extractions**

Formants are boosts of overtones in the vocal tract. The vocal tract refers to the area from the edge of the lips and nostrils to the larynx. The formants are generated at four to six different places in the vocal tract, and different vowels make different spaces of resonance.

From the acoustics research and industry's point of view, a formant may be described as 'the spectral peaks of the sound spectrum  $|P(f)|^2$ ' (11). It is considered to be one of the most important features in speech signals, and may be used for various applications, for example speech characterization and synthesis (12). For voice analysis purposes, the formant frequencies need to be extracted. Formant extracting methods can be divided into 1) spectral peak picking, 2) root extraction, 3) analysis by synthesis (13, 14). These methods have low computational complexity and are hence widely used, but these methods also have some limitations: for example, the spectral peak picking may result in merging problems, where two formants may be identified as a single formant. Recent research has provided an alternative way of formant extraction, combining pathological spectral peak picking and root polishing (12). In order to achieve more accurate

analyses of the voice, is it crucial that formant extraction methods are constantly revised based on evidence in the voice clinics.

### **Phonetograms**

The purpose of a phonetography examination is to chart the clients' extent of voice: how loud / soft a patient is able to reproduce the note and pitch.

Phonetograms have been described by Shutte et al (15) for standard setups, using an intensity-meter and a tone generator. Wendler and Seidner (16) have developed some software with a singing formant analysis in the phonetogram. Pedersen et al (17) has developed area calculations in semitones times decibel in a standard software. Pabon has developed (18) the setup for online measure of phonetograms with formants. On the market there is now a new product called Lingwaves (19) which is currently being tested.

### **Perspectives**

The newer understanding of the (meta-analytic) statistical approach for randomized prospective controlled studies is as necessary in bioengineering as it is in nano- and biotechnology. (20) The biological situations in voice pathology are extremely varying and statistical evidence is possible in an even better way with meta-analysis comparing several randomised controlled trials (RCT) with well defined baselines of studies as well as power calculations. This has not been done in voice research on vocal cords and laryngopharyngeal reflux studies (20,21,22). It is not till now that it is discussed whether it is possible to define a beautiful voice with classic, jazz and belting techniques (23). Still, as it was shown, the measures of voice parameters presented are the updated measures for the evaluation of the so called well functioning voice only: high speed films, fundamental frequencies, phonetograms.

The problem related to biotechnology is also that the voice is not "alone" in the larynx, the whole swallowing process and respiration are always involved in intonation (24).

Nano-technology and biotechnology are new evidence based sciences preferred for research for the voice clinicians. Nano-diagnostics of multi functional imaging on a molecular level will be a help for diagnostics and to understand treatment effects of medication in the larynx (25). The bioengineers should refine the technical documentation of their products so that transparency and evidence is part of the technical industry related to voice, as demanded in the pharmacological industry.

### **Conclusion**

Clinical voice practise is renewed with understanding of the interaction between patient demands and advanced technology combined with evidence based statistics.

### **Acknowledgements**

Thanks to Julie Pedersen, Christian Larsen, Maria Dalby Pedersen, Anders Jönsson, Mike Ellington, Martin Eeg, Sarah Øst Jensen, Randi Engelberth, Line Jönsson, Umbreen Yousaf, Shahzleen Rajan and Anna Beneni.

## References:

- 1) Maryn Y, Roy N, De Bodt M, Van Cauwenberge P, Corthals P. (2009). Acoustic measurement of overall voice quality: A meta-analysis, *J Acoust. Soc. Am.*; 126: 2619-34
- 2) Pedersen M, Barney A, Clinical criteria for benign laryngeal mucosal evaluation and treatment, Congress report, American Pacific Voice Conference (2010) UCLA.
- 3) Schuster M, Lohscheller J, Kummer P, Eysholdt U, Hoppe U (2005). Laser projection in high-speed glottography for high-precision measurements of laryngeal dimensions and dynamics. *Ann. Otol. Rhinol. Laryngol*, 117(6):413-424.
- 4) Deliyski D (2010). Laryngeal High-Speed Videoendoscopy in the Examination of Vocal-Fold Vibration. *Cost Action 2103 Summer School Report: Voice Analysis: from Acoustics to High-Speed Videoendoscopy*.
- 5) Moore P (1938). Motion picture studies of the vocal folds and vocal attack. *Journal of Speech and Hearing Disorders*, 27:165-171.
- 6) Werner-kukuk E, von Leden H (1970). Vocal initiation: high speed cinematographic studies on normal subjects. *Folia Phoniatica*, 22:107-116.
- 7) Mehta DD, Deliyski D, Zeitels SM, Quatieri TF, Hillman RE (2010). Voice production mechanisms following phonosurgical treatment of early glottis cancer. *Annals of Otolology, Rhinology, and Laryngology*, 119:1-9.
- 8) Liu H, Auger J, Larson C (2010). Voice fundamental frequency modulates vocal response to pitch perturbations during English speech. *J. Acoust. Soc. Am.* 127 (1)
- 9) Pedersen M, (2008) Normal Development of Voice in Children. *Springer Verlag. Berlin Heidelberg*.
- 10) Heman-Ackah YD, Michael DD, Goding GS Jr. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *J Voice* 16(1);20-7.
- 11) Fant, G (1960). Acoustic Theory of Speech Production. *Mouton & Co, The Hague, Netherlands*.
- 12) Kim C, Seo K, Sung W (2006). A robust formant extraction algorithm combining spectral peak picking and root polishing. *Hindawi Publishing Corporation, EURASIP Journal on Applied Signal Processing 2006*, 1-16.
- 13) McCandless S (1974). An algorithm for automatic formant extraction using linear prediction spectra. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol 22, no. 2, 135-141.
- 14) Weiling L, Ney H (1998). Formant estimation for speech recognition. *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 1, 36-48.
- 15) Schutte H, Seidner W. Recommendation by the Union of European Phoniaticians (UE:P) Standardizing Voice Area Measurement/Phonetography. *Folia Phoniatr* 1983;35:286-8.
- 16) Wendler, J. & Seidner, W. (1987) Entwicklung der Stimme, in Lehrbuch der Phoniatrie. Thieme Verlag: 164-69.
- 17) Pedersen MF, Møller S, Krabbe S, Munk E, Bennett P, Kitzing P (1984). Change of voice in puberty in choir girls. A cross-sectional study of electroglottographically measured fundamental frequency in reading compared with phonetograms, pubertal stages and the endocrine analysis. *Acta Otolaryngol Stockholm*; S412: 46-9.
- 18) Pabon J, Plomp R. (1988) Automatic Phonetogram Recording Supplemented. *J Speech Hear Res.*1988; 31: 710-722.
- 19) lingWAVES Phonetogram Plus (2010).
- 20) Pedersen M, McGlashan J (2007) Surgical versus non-surgical interventions for vocal cord nodules (Cochrane review), John Wiley & Sons Ltd.

- 21) Hopkins C, Yousaf U, Pedersen M. (2006). Acid Reflux. Treatment for Hoarseness (Cochrane review), John Wiley & Sons Ltd.
- 22) Cochrane Handbook for Systematic Reviews and Interventions (2009). Oxford.
- 23) Mathelitsch L, Verovnik I, Ist die Schönheit einer Singstimme physikalisch messbar? *Vox Humana* (2011); 6: 35-48.
- 24) Birchall, MA, Baily M, Gutowska-Owsiak D et al. Immunological Response of the Laryngeal Mucosa to Extrapharyngeal Reflux. *Annals of Otology, Rhinology & Laryngology* (2008); 117: 891-895.
- 25) Kim, BYS, Rutka, JT, Chan WCW. *Nanomedicine. N Engl J Med.* 363; 25: 2434-43.