Abstracts for the presentations at the *Campinas Workshop on Vocal Profile Analysis* (VPA) to be held at UNICAMP, April 4–8, 2016

"Everyone knows what voice is until they try to pin it down"

Anders Eriksson

Department of Linguistics, Stockholm University, Stockholm, Sweden

The title of my presentation is a quote attributed to Johan Sundberg and he should know. In my presentation I will discuss a variety of ways to define the concept of voice and show that all of them make sense in some given context – hence the conceptual confusion. At the same time we have all experienced recognition of familiar voices that, at least to some extent, seems to be independent of context or even language. Under optimal conditions such minimal utterances as a hesitation sound may be enough to recognize a familiar speaker. The naïve interpretation of "voice" is based on this kind of everyday experience. It suggests that there is some invariant quality in the speech of an individual speaker that we as human listener can decode. How can this be explained? I am not going as far as claiming to have a solution to the problem, but I will suggest ways of approaching a solution based on the *Modulation Theory of Speech*.

Vocal Profile Analysis (VPA) in Speaker Comparison: What protocols are used? Which labs are using VPA? How well does the VPA perform?

Anders Eriksson

Department of Linguistics, Stockholm University, Stockholm, Sweden

There is an extensive literature on perceptual voice description and numerous protocols have been suggested. Many of the protocols are designed for the description of pathological voices and not particularly well suited in forensic applications. Other protocols are extremely extensive and difficult to use. I will describe some of the more common protocols including the two currently used in forensic casework. Vocal Profile Analysis (VPA) is used by at least two laboratories in Europe – The *Bundeskriminalamt* (BKA) in Wiesbaden, Germany, and The *J P French Associates* in York, U.K. Both laboratories use a simplified form of the Laver protocol, albeit not identical versions. In addition a German version of the GRBAS protocol is used by the BKA. I will present the protocols, how they are used and how they may contribute to speaker comparison based on discussions I have had with representatives from the two labs. To find out if any other labs are using VPA, I have made a small, informal inquiry by mail of some of the other European forensic labs. I will report the results from this inquiry. A handful of experiments have been carried out to test the performance of VPA in casework-like situations. I will give a brief overview of these studies.

An acoustic-perceptual study on voice quality

Zuleica Camargo[1], Sandra Madureira[1], Nathalia dos Reis[1], Albert Rilliard[2]

[1]Integrated Acoustic Analysis and Cognition Laboratory-LIAAC. Graduate Applied Linguistics and Language Studies Program. Pontifical Catholic University of São Paulo, São Paulo, Brazil

[2] LIMSI - Université Paris-Sud

A great number of factors influencing voice quality patterns have not been addressed in studies focusing acoustic-perceptual correlations. The present study aims at investigating the power of some acoustic measures to predict supralaryngeal, laryngeal and tension voice quality settings based on the application of statistical methodological procedures. The corpus was composed by semi-spontaneous speech samples and repetitions of 3 key-sentences samples, recorded in a sound proof room by 74 speakers. Acoustic measures (minimum and maximum f0, A-weighted intensity and aperiodicity (jitter, shimmer and mean harmonics-to-noise) were extracted by means of some scripts running in the PRAAT software. The same samples were perceptually evaluated, using the Vocal Profile Analysis Scheme for Brazilian Portuguese: BP-VPAS (Camargo, Madureira, 2008). The acoustic measures and voice quality settings judgments were statistically analyzed in order to

investigate the validity of each acoustic parameter to predict voice quality settings related to neutral and non-neutral supralaryngeal, laryngeal and tension settings groups (discriminant analysis, canonic correlation and linear regression). The results are presented and discussed in terms of the relevance to establish acoustic-perceptual correlations and the demand for improving systems of voice recognition.

Camargo, Z. and S. Madureira (2008). Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). Proceedings of Speech Prosody 2008. Campinas, Brazil: 57–60.

Key-words: Speech Acoustics; Voice Quality; Auditory Perception

Pre-attentive speaker recognition: A realistic possibility or Science Fiction?

Lisa Gustavsson

Department of Linguistics, Stockholm University, Stockholm, Sweden

In my talk I will present a study on neural processing of voices. The aim was to investigate the possibility of using ERPs as a measure of recognition of a familiar voice. The methodology however raises questions concerning pre-attentive processing of voices. I will present the study on voice familiarity and discuss the typical MMN (Mismatch Negativity) that was found in relation to voices, but not to familiarization. Acoustic analysis of voice characteristics in the current study as well as follow up studies with controlled exposure and voice parameters will also be addressed. I would like to discuss these issues with you, and also the implications of a possible MMN to familiar voices.

Gustavsson, Kallioinen, Klintfors & Lindh (2013) Neural processing of voices – familiarity. Proceedings of Meetings on Acoustics (ASA).

The vocal profile analysis: a discussion of theoretical and applied issues

Sandra Madureira and Zuleica Camargo

Integrated Acoustic Analysis and Cognition Laboratory (LIAAC)
Graduate Applied Linguistics and Language Studies Program (LAEL)
Pontifical Catholic University of São Paulo (PUCSP)

The VPAS Voice quality is a prosodic element which has linguistic, paralinguistic and extralinguistic functions and is therefore an important index of the subjects' identity. For a long time in speech studies impressive labels were used to describe types of voice quality. Laver (1980) remarkably changed this panorama by introducing a descriptive phonetic model of voice quality analysis and a protocol the Voice Analysis Scheme (VPAS) to categorize settings of voice quality. The 2007 version of the VPAS has two main sections, one for describing voice quality settings and another for describing the prosodic aspects and temporal organization. The objective of this presentation is to discuss the characteristics of the 2007 version of the VPAS and its potentiality to identify voices. The issues to be discussed include: the comprehensiveness of the model, the comparison to a neutral setting, the language-dependentness issue, the need to take into account inherent phonetic characteristics of the speech segments, the 6- point scale, the need to incorporate other phonetic features, the contribution to finding similar vocal profiles and the training of the judges.

Key-words: Voice Quality; VPAS; Speaker identity.

Simplified Vocal Profile Analysis for Forensic Speaker Comparison: a preliminary proposal

Eugenia San Segundo

Department of Language and Linguistic Science, University of York, UK

Voice quality multidimensionality is often considered difficult to tackle in perceptual assessments (verbal voice profiling). This study presents a simplified version of the Vocal Profile Analysis (VPA) scheme (henceforth SVPA) with the aim of alleviating such difficulties particularly in Forensic Speaker Comparison. The original VPA protocol consists of 38 different settings (labial, velopharyngeal, voicing type, etc.) and 6 possible scalar degrees while the proposed SVPA only includes 10 major settings – with several subsettings – and no scalar degrees. This new protocol

enables a simple measurement of similarity distances between pairs of speakers. These distances take the form of simple matching coefficients (SMCs) ranging between 0 (greatest dissimilarity) and 1 (greatest similarity).

The measurement of inter- and intrarater agreement in the perceptual assessment of voices

Eugenia San Segundo

Department of Language and Linguistic Science, University of York, UK

In many scientific fields it is common to study agreement among ratings of multiple tests, experts, judges, etc. However, there is usually little consensus in the literature about what statistical method is best to analyse such agreement (including agreement between observers, diagnostic tests, ratings, etc.), which sometimes gives rise to lack of consistency and reliability. I will present an overview of how the question of inter- and intrarater agreement has been dealt with in perceptual assessment research, and will review the basic issues which the different methodological alternatives depend on (e.g. type of data, identification of explicit goal or research question).

The use of voice profiling in Forensic Speaker Comparisons conducted by the Brazilian Federal Police

Nerci Lino de Almeida Tonaco and Ronaldo Rodrigues da Silva

National Criminalistic Institute, Federal Police, Brazil

In our talk, we will present a brief history of the development of the Forensic Speaker Comparison (FSC) field in Brazilian Federal Police (DPF). Aspects related to the FSC specialization training courses adopted by the DPF, including its methodology, will be addressed as well. Furthermore, limitations in the use of voice profiling protocols in audio recording traces received for analysis, mainly due to its poor quality, quantity and linguistic diversity, will be discussed. Finally, some typical analysis currently included in FSC reports are shown.

A preliminary study on voice quality identification

Renata Vieira and Sandra Madureira

Integrated Acoustic Analysis and Cognition Laboratory-LIAAC
Graduate Applied Linguistics and Language Studies Program - LAEL
Pontifical Catholic University of São Paulo - PUCSP - São Paulo, Brazil

The present study shows a preliminary analysis of an acoustic and perceptual study in voice quality identification. This study has two aims: to investigate the applicability of a perceptive protocol to discriminate subjects in a group of speakers; to investigate a set of acoustic parameters to discriminate subjects in a group of speakers. In order to analyze voice quality perceptually, the Vocal Profile Analysis Scheme for Brazilian Portuguese: VPAS-PB was used. For the acoustic analysis of the data the ‘forensic data tracking’ script was used. It extracts the frequencies of the first two formants of the vowels, the fundamental frequency, duration units the size of the syllable and vowel, spectral emphasis, the fundamental base frequency, the rate of evolution of formants and standard deviation of duration of consonant intervals. The corpus was composed by semi-spontaneous speech samples. The subjects, in number of 5, are firemen from Rio de Janeiro. The speech samples were captured with a cell phone. The cell phone was positioned on their right ear. The phone calls were made from cell phone to cell phone both using the Vivo operator. Conversations were captured with an application (‘app’) installed in the cell phone that received the calls. Sampling frequency was 44100 Hz with .wav extension. The following equipment was used: I Phone 5S, Samsung Galaxy A5 cell phone, with Android system and 1.2 GHz processor Quad Core, and recorder app AutomaticCall Recorder.

Key-words: Speech Acoustics; Voice Quality; VPAS; Speaker identity.