Multimodal Corpus of Speech Production: Work in Progress

Einar Meister and Lya Meister

Institute of Cybernetics at Tallinn University of Technology, Estonia

einar@ioc.ee, lya@phon.ioc.ee

Introduction

The multimodal corpus of speech production is aimed at studying dynamic articulatory patterns and acoustic-articulatory mapping in native Estonian speech; a further goal is to investigate the use of articulatory data in speech technology. The paper gives an overview of the instrumental methods of multimodal data collection and describes the work-in-progress on an articulatory-phonetic corpus of Estonian speech production.

Instrumental Techniques

Electroglottography - EGG

A non-invasive technique to register laryngeal movements during speech production. It provides the most accurate data on the physical measures of voice fundamental frequency and larynx movements.
Our system: Laryngograph Processor by Laryngograph Ltd, UK.

Electropalatography - EPG

A method to study the timing and location of tongue contact with the hard palate during continuous speech. During the data capture a speaker has to wear an artificial palate; on the surface of the palate are located 62 silver contacts which register the tongue-palate contact during articulation (Figure 1).
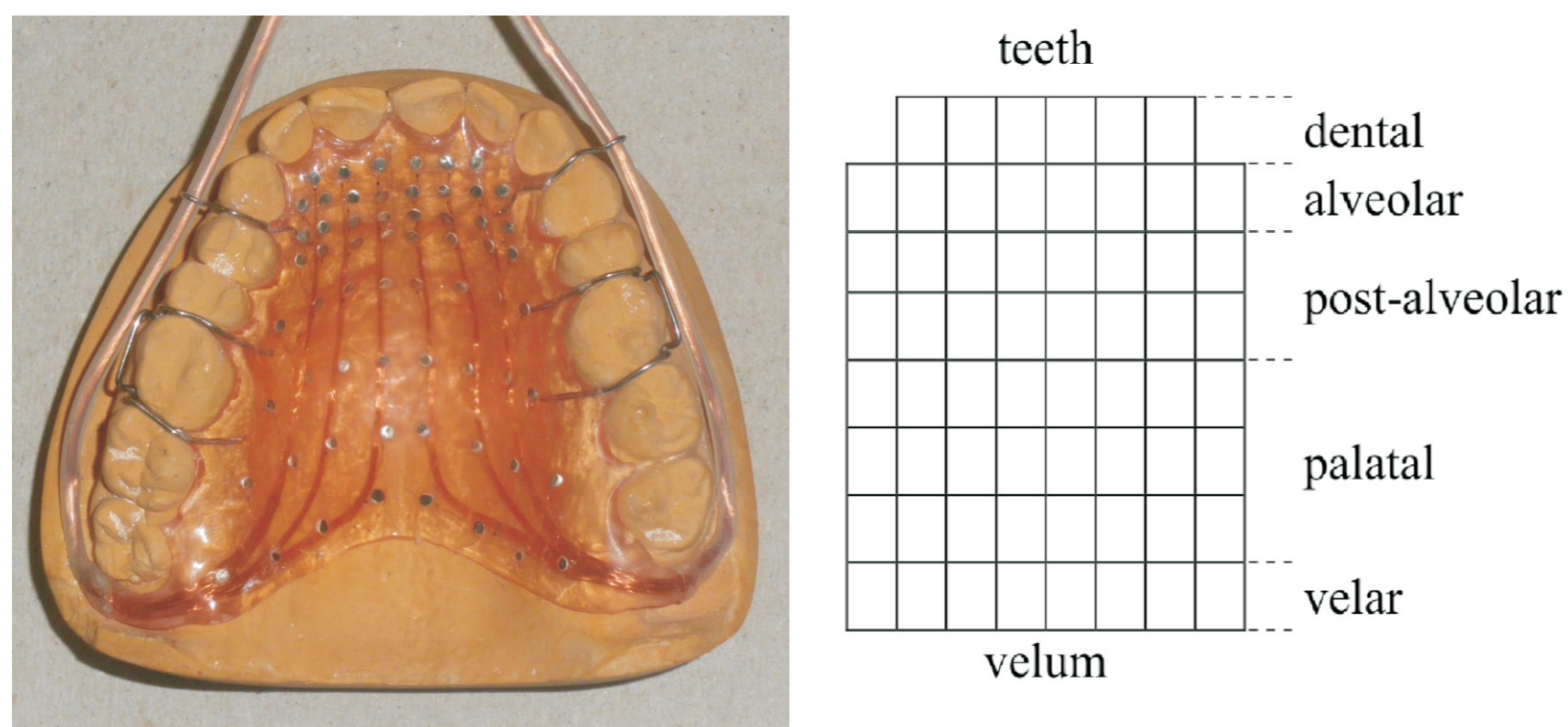Our system: WinEPG by ArticulateInstruments Ltd, UK.

Figure 1. The EPG palate and the tongue contact areas.

Electromagnetic Articulograph - EMA

A non-line-of-sight motion capture system specifically designed for tracking speech related articulatory movements and articulatory kinematics. It enables real-time 3D data capture from 16 sensors in a measurement volume of 50x50x50 cm, along with synchronized audio. In articulatory studies the sensors are typically glued onto the tongue, the lips, the front teeth and the jaw.
Our system: Wave Speech Research System by Northern Digital Inc, Canada.

Corpus Design and Recording Set-Ups

The text corpus compiled for data collection includes CVCV nonsense words (where V and C represent all Estonian vowels and consonants) and short sentences. Two recording set-ups have been used:
(1) EGG + EPG + audio (Figure 2, left), (2) EMA + audio (Figure 2, right)
In both set-ups the same text corpus has been read by two subjects. The audio data is recorded with a close-talk microphone at a sampling frequency of 22.1 kHz.

Figure 2. Left: the subject wearing electrodes of the EGG system, the EPG palate, and a close-talking microphone in the recording set-up 1; right: the position of EMA electrodes in the recording set-up 2.

Examples of Multimodal Data Analysis

EGG Data

EFXHIST program is used to analyse synchronously recorded speech and laryngograph signals (Figure 3). It finds the locations of every pitch period in the recording and calculates statistics of voicing, fundamental frequency, closed quotient, jitter, shimmer, etc.
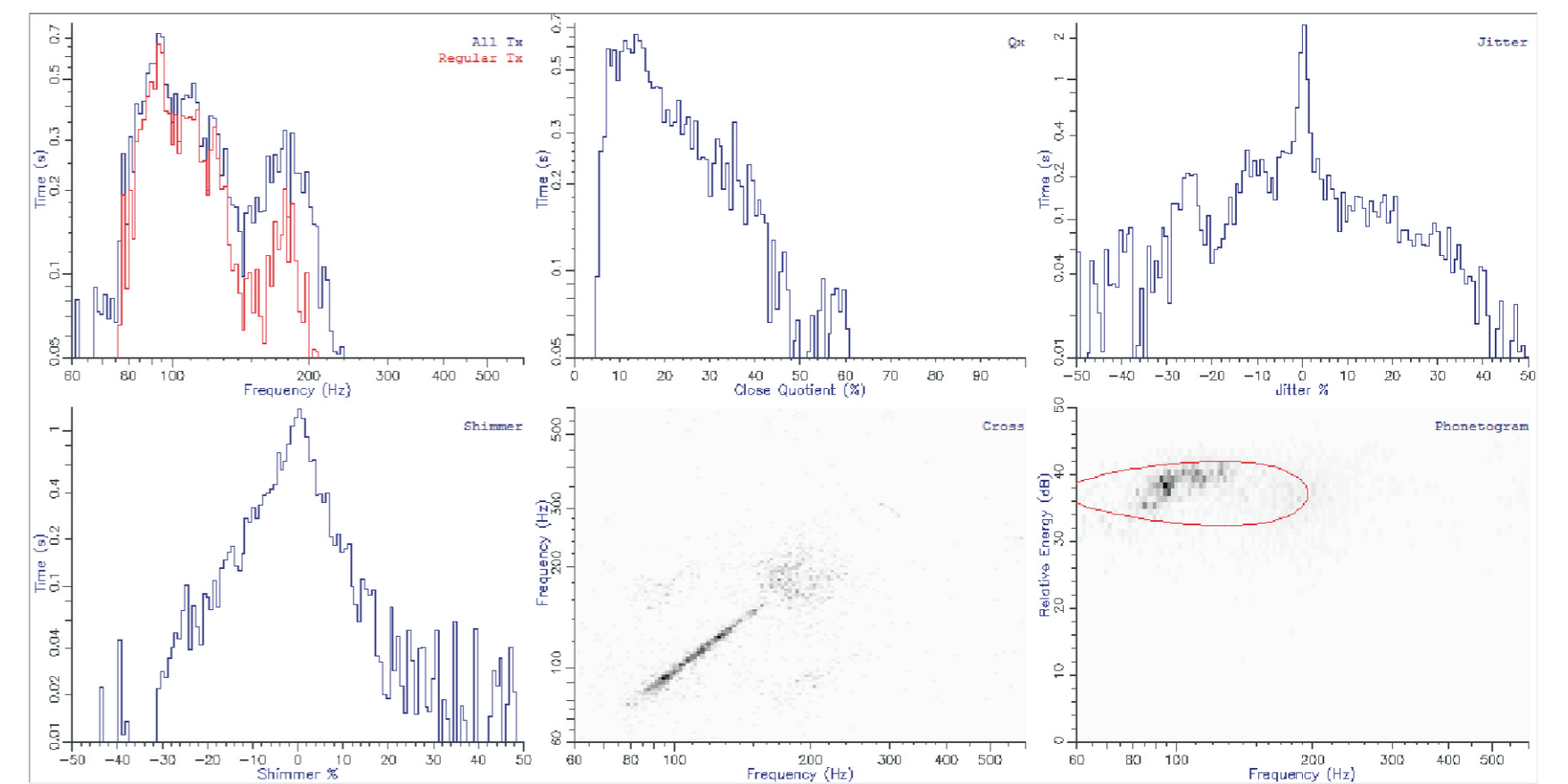

Figure 3. Different measures of voice quality calculated for one minute read speech of a male speaker. Top-left: Tx – distribution of all (blue) and regular (red) pitch periods; top-mid: Qx – a histogram of the closed quotient values; top-right: Jitter – a histogram of jitter values; bottom-left: Shimmer – a histogram of shimmer values; bottom-mid: Cross – a scatter plot of pairs of adjacent pitch periods; bottom-right: Phonetogram – a scatter plot of the frequency and energy.

EPG Data

EPG provides essential data for the comparing of palatalised and non-palatalised consonants revealing that the place of articulation of palatalised counterparts is characterised by a larger front and lateral contact area (Figure 4).
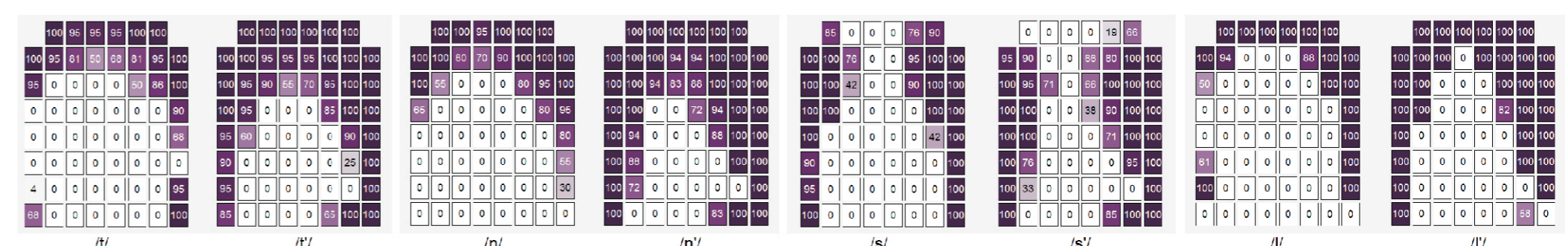

Figure 4. Palatograms of non-palatalised and palatalised Estonian consonants in intervocalic position in CVCCV nonsense words

EMA Data

EMA data recorded synchronously with audio provides information on the movements of the main articulators – tongue, lips and jaw.
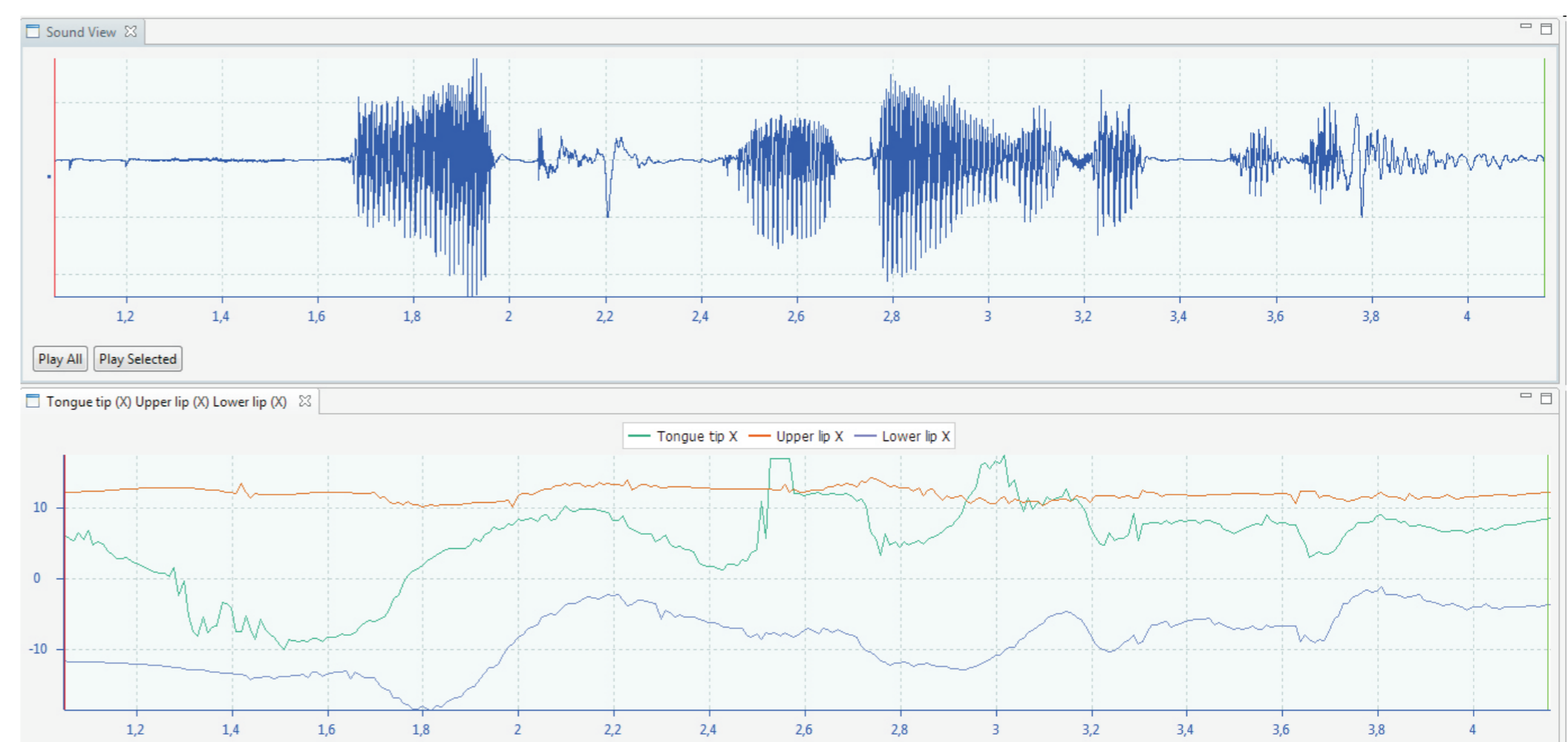

Figure 6. An illustration of EMA data in the sentence "Aeg on teele sättida". Speech signal in top window, the movement trajectories of the upper lip, the tongue tip and the lower lip in bottom window.

Exploiting Speech Production Data

In phonetics:
- description of articulatory patterns of Estonian speech sounds,
- studies of coarticulation,
- studies of acoustic-articulatory mapping.

In speech technology:
- audiovisual speech recognition,
- audiovisual speech recognition.