A time-aligned account of lingual and laryngeal gestures using ultrasound tongue imaging (UTI) and electroglottography (EGG)

Stefano Coretta

14/01/2017

Contents

1 Purpose

The combination of techniques described in the following sections allows a synchronous mapping of lingual gestures and phonation during speech. Such methodology enables a time-aligned account of the movements of the tongue and the concomitant configuration of the glottis that characterises phonation. These techniques employ ultrasound tongue imaging (UTI) and electroglottography (EGG) for the simultaneous acquisition of articulatory data from, respectively, the oral cavity and the glottis.

2 Methodology

2.1 Ultrasound tongue imaging

Ultrasound tongue imaging (UTI) uses ultrasonography for charting the movements of the tongue into a two-dimensional image. In medical sonography, ultrasound waves (sound waves at high frequencies, ranging between 2 and 14 MHz) are emitted from a probe in a fan-like manner, and travel through organic material (such as skin and muscles). When the surface of a material with different density is hit, the ultrasound waves are partially reflected, and such "echo" is registered by the probe. These echoes can then be plotted on a two-dimensional graph, where different densities are represented by different shades (higher densities are brighter, while lower densities are darker). The graph, or ultrasound image, will show high density surfaces as very bright lines, surrounded by darker areas. By positioning the ultrasound probe in contact with the submental triangle (the surface below the chin), sagittally oriented, it is possible to infer the cross-sectional profile of the tongue, which appears as a bright line in the resulting ultrasound image.

2.2 Electroglottography

Electroglottography (Fabre, 1957) is a technique that measures the size of contact between the vocal folds (the Vocal Folds Contact Area, VFCA). A high frequency low voltage electrical current is sent through two electrodes which are in contact with the surface of the neck, one on each side of the thyroid cartilage. The impedance of the current is directly correlated with VFCA, while its amplitude is inversely correlated. Thus, impedance increases with lower VFCA and decreases with higher VFCA. Conversely, amplitude decreases when the VFCA increases and it increases when the VFCA decreases. The EGG unit registers changes in impedance and it converts it in amplitude. Its output is a synchronised stereo recording which contains the EGG signal from the electrodes in right channel and the audio signal from the microphone in the left channel.

2.3 Equipment setup

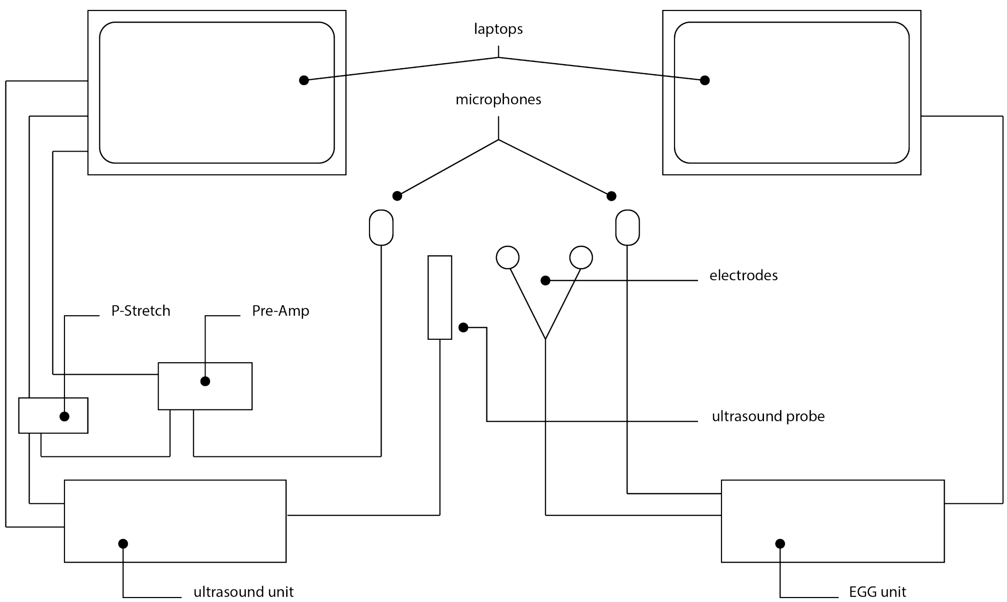

Figure 1: Equipment set-up scheme

Figure 1 shows the equipment set-up employed. The left part of the figure shows the ultrasound set-up, while the EGG set-up is shown on the right. Two separate laptops are used for the acquisition of the ultrasound and EGG recordings. The ultrasound unit is plugged into one laptop. A P-Stretch unit (used for signal synchronisation) and the ultrasound probe are directly connected to the ultrasound unit. The P-Stretch unit and a microphone feed a pre-amplifier system, which is plugged into the ultrasound laptop. A second microphone and the electrodes are connected to the EGG unit, which is plugged into the second laptop. A TELEMED Echo Blaster 128 system is used for ultrasonography and a Glottal Enterprises EG2-PCX2 unit for EGG. The subject wears a headset (a metallic ..., not shown in Figure 1) which holds the ultrasound probe in position (allowing free head movement by the subject) and a velcro strap with the EGG electrodes, located on each side of the thyroid cartilage, at the level of the glottis. The microphones are clipped to the headset on either side, at identical height.

2.4 Acquisition of ultrasound and EGG

2.5 Synchronisation

Since the signals from the ultrasound machine and the laryngograph are recorded simultaneously but separately, data from both machines need to be synchronised after acquisition. Synchronisation is achieved through the cross-correlation of the audio signals from both sources (Grimaldi et al., 2008). The cross-correlation method creates a new sound file from two audio files. The created new file is a convolution of the original files. The time of the maximum amplitude in the convoluted sound wave is the amount of off-set between the two original files. The off-set is trimmed from the beginning of the longer audio file, with the result that the files will be in sync. A measure taken at any particular time in the ultrasound source can thus be related to a measure taken at that same time in the laryngograph source.

3 Calculation of dEGG

Previous studies have shown that the mathematical first derivative of the EGG signal helps determine the moments of glottal closure and opening in each vibration cycle (glottal period). The first derivative of a signal is basically the velocity of the signal, in other words how fast the signal changes in time. The time of maximum velocity in the first derivative of the EGG signal (dEGG) roughly corresponds to the moment of glottal closure. The time of minimum velocity corresponds to the moment of glottal opening. Thus, glottal closure and opening for each glottal period can be extracted from the dEGG.

Herbst et al. (2010) describe a new technique, called electroglottographic wavegram, which displays the variations in the EGG and dEGG signals in a single graph. A wavegram contains temporal information on the x and y axis, while changes in the VFCA are rendered as different colour intensities on the z axis.

The extraction of dEGG maxima and minima has been implemented in this study using the PRAAT scripting language. The algorithm consists of the following stages:

1. detection of the glottal periods
2. calculation of the dEGG
3. extraction of absolute dEGG maximum (dEGG_{max}) and minimum (dEGG_{min}) for each glottal period
4. calculation of dEGG_{max} and dEGG_{min} relative to the glottal period

It is conventional to define a glottal period as the time between two consecutive moments of glottal closure, i.e. two consecutive dEGG maxima. However, since the maxima need to be identified in the first place, an arbitrary definition of glottal period is instead used. Glottal periods correspond to the intervals between two consecutive EGG minima [cf. ...]. First, the EGG signal is band-pass filtered (40Hz-10KHz) and smoothing is applied. A weighted sliding-average smoothing method (triangular smooth) is used, with smooth width $m = 11$. EGG minima are thus extracted from the smoothed EGG signal. The interval between any two consecutive minima constitutes a glottal period.

The dEGG is calculated with the formula $x'_n = x_{n+1} - x_n$, where x_n is the value of the EGG signal at the time n. After calculation, the resulting dEGG is smoothed with the same method as before (triangular smooth, $m = 11$). The algorithm then searches for dEGG maxima and minima within each glottal period (defined as two consecutive EGG minima). Finally, relative $dEGG_{max}$ and $dEGG_{min}$ are calculated as proportions of the respective glottal period. The resulting values are between 0 (beginning of period) and 1 (end of period).

3.1 dEGG tracegram analysis

As the authors note, the wavegram technique has the limitation of not being suitable for quantitative analysis. A new visualisation technique, based on wavegrams, is introduced here: electroglottographic tracegram. The tracegram method, even if it reduces the displayed dimensions, allows a statistical assessment of the varying $dEGG_{max}$ and $dEGG_{min}$, thus constituting a partial improvement over wavegrams. After the calculation of the relative $dEGG_{max}$ and $dEGG_{min}$, these values are plotted in a graph on the y axis at each time point which corresponds to the beginning of a glottal period. Since the values are restricted between 0 and 1 (being proportions), changes in glottal period (which corresponds to changes in fundamental frequency and hence pitch) are controlled for. The resulting graph, the tracegram, shows the traces of $dEGG_{max}$ and $dEGG_{min}$ as they change in time, in a way similar to the display of pitch contours. Statistical analysis is then performed on both traces separately using a Smoothing Spline ANOVA (SSANOVA).

References

Fabre, P. 1957. Un procede electrique percutane d'inscrition de l'accolement glottique au cours de la phonation: glottographie de haute frequence. Premiers resultats. *Bulletin de l'Académie nationale de médecine* 141. 66.

Grimaldi, Mirko, B. Gili Fivela, Francesco Sigona, Michele Tavella, Paul Fitzpatrick, Laila Craighero, Luciano Fadiga, Giulio Sandini & Giorgio Metta. 2008. New technologies for simultaneous acquisition of speech articulatory data: 3D articulograph, ultrasound and electroglottograph. *Proceedings of LangTech* 1–5.

Herbst, Christian T., W Tecumseh S. Fitch & Jan G. Švec. 2010. Electroglottographic wavegrams: A technique for visualizing vocal fold dynamics noninvasively. *The Journal of the Acoustical Society of America* 128(5). 3070–3078.