The influence of syllable onset complexity on tongue movements: An ultrasound study

The characteristics of syllables of different complexity in speech have been mainly studied in terms of temporal differences. The general conclusions are that syllable duration is mainly affected by the syllable's stress and the number of segments, with stressed syllables and those with more segments being longer. However, syllables with a greater number of segments are additionally affected by a compression of segments and by coarticulation.. As for the parts of the syllable, onset and nucleus are supposed to be more temporally stable than the offset.

Less data is available on the articulatory properties of differently structured syllables, especially data collected by directly investigating the movement of the articulators. Although acoustic analysis can reveal temporal properties and formant patterns of speech it only provides an indirect view of the articulatory movements. A better understanding of the nature of the movements would be achieved by using imaging techniques which enable direct observation of the articulators. One of the most appealing techniques is ultrasound study of tongue movements.

Ultrasound has been used in speech research for about twenty years and is becoming increasingly popular, especially with the development of ultrasound machines with faster frame rates (up to 100 Hz) and the software that allows temporal alignment of ultrasound and audio data, automatic tongue contour tracking and extraction of the data from the image. Ultrasound imaging is safe, noninvasive, relatively inexpensive and it can give a view of the entire tongue contour. However, as with every technique, it has its limitations as well. The biggest one is that it does not image a raised tongue tip because of the air pocket created below it which causes the reflection of the ultrasound waves at the tissue-to-air boundary before reaching the tip. Another limitation is that it is not possible to reliably detect parts of the tongue since there are no reliable orientation points on the tongue or on its position in the oral cavity.

To obtain an ultrasound image of the tongue, the probe has to be placed under the speaker's chin. Depending on the orientation of the probe, the resulting image can show either midsagittal or coronal view of the tongue. To achieve the same view over the recording session it is important to stabilise the probe and the head in such a way to allow speech movements but prevent any others. This can be achieved by fixing a special support system or a helmet to the speaker's head.

The probe emits ultrasound pulses which travel onwards until they reach a boundary of differently dense media at which they are reflected. In the case of tongue imaging, the majority of ultrasound waves is reflected at the tongue-air boundary. Because the ultrasound waves hit this boundary more or less perpendicular, they are reflected back to the probe and detected by it which allows the creation of an image representing the lower tongue edge. Most ultrasound machines record at the rate of 30 frames per second meaning that an image is created about every 33 ms.

The main aim of this study is to investigate tongue movements of normal English-speaking adults in terms of the distance the tongue travels over target C(C(C))V syllables, the time needed and the overall speed. The results will be used in the future study of tongue movements in developmental apraxia of speech.

In addition to the increase of duration and the distance the tongue travels it is expected that the speed will increase with the increasing number of onset segments and that the change of speed will be more affected by the number than the type of the segments. The region where the greatest distance is travelled is expected to be at the CV transition.

Speech material was selected with the future study of developmental apraxia of speech in mind, where it is expected that speakers will have more difficulties articulating complex than simple syllables. It consists of six word initial CV combinations: /peI/, /seI/, /leI/, /pleI/, /sleI/, /spleI/. All targets are monosyllabic real words (pay, say, lay, play, slay, splay) and are uttered in a carrying sentence "Say ______ again" to secure the same tongue starting position for the target words. Ten adult native speakers of English without any speech, language or hearing impairments are going to participate in the study and each speaker will repeat the utterances five times.

The data will be analysed using Articulate Assistant and Matlab software. Articulate Assistant enables ultrasound and acoustic alignment, automatic tracking of the tongue surface and spline fitting, and export of splines as a sequence of (x,y) coordinate points for quantitative analysis.

The distance between splines is measured for every consecutive spline pair (spline 1 and spline2, spline 2 and spline 3...). The splines of the pair are first cut at the same minimum and maximum x values and interpolation is used to obtain an equal number of points on both splines. In the next step, the distance between the points of a pair of splines is measured using the nearest neighbour technique. The distance between the consecutive splines is calculated as the average of all the nearest neighbour distances, and the total distance travelled for the whole syllable is the sum of the distances between all the spline pairs.

In addition to the distance, duration of the target syllable will be measured to observe temporal differences between the syllables and to allow calculating the overall speed of the tongue movement (total distance travelled over the syllable divided by the syllable's duration).

Measuring tongue movements in the proposed way enables investigating where in the syllable is the region of the greatest change (in duration and in speed), and how this is affected by differently structured onsets. Later on, it will allow quantifying the differences in the tongue movements of speakers with developmental apraxia of speech.