Ultrasound image segmentation

Sherman Charles

Indiana University (USA)

sdcharle@indiana.edu

Summary: A basic technical understanding of ultrasound as well as speech anatomy forms part of the foundation needed in order to interpret ultrasound images of the tongue. The formation and steering of an ultrasound beam, and its interaction with structures along its path, are critical characteristics that define the resulting images. Identification of visible anatomical landmarks provides context and can help differentiate between the tongue surface and other structures. This Master Class will focus on 1) understanding how ultrasound images are generated, 2) reviewing relevant vocal tract anatomy, and 3) exploring pitfalls and solutions for tongue surface segmentation. The Master Class will benefit from the availability of 3D ultrasound data, while emphasizing practical applications for 2D ultrasound tongue segmentation.

Ultrasound Signals: Ultrasound signals rely on reflections to generate images. Reflections are created by differences in density between the oropharyngeal structures and the medium that fills the oral and pharyngeal cavities. Objects of more similar densities create soft reflections but allow the signal to continue through to more distant objects. Adjacent objects with greater density differences create loud reflections and block superior objects from view.

Vocal Tract Anatomy: The oropharyngeal structures are comprised of a variety of tissues of different densities (e.g. soft tissues like epithelia, intrinsic and extrinsic muscles of the tongue, tendons; hard tissues like the mandible and hyoid bone). Because the soft tissues are collectively about the same density as water, the ultrasound signal can reflect off of most of the structures and continue through to other superior structures. Inversely, bone is much denser than the surrounding tissues. These structures appear as dark areas with hard edges on the surface closest to the probe. Most structures beyond the hard tissues are invisible in the ultrasound image. Similarly, the density difference between the surface of the tongue and the air above it causes a loud reflection, which in turn blocks superior structures from view and appears as bright areas in the image. See Figure 1 for examples. Knowledge of the anatomical structures within view and the tissues from which they are formed can be extremely informative when interpreting an ultrasound image.

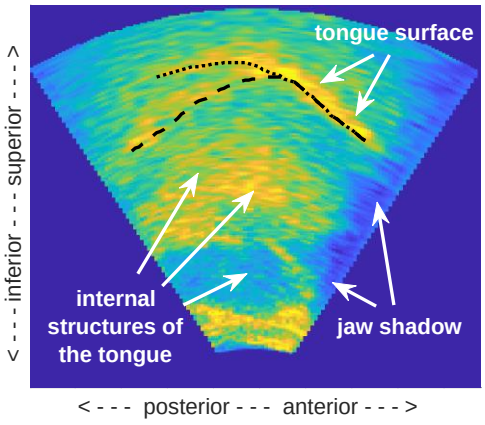

Figure 1: *Example of ultrasound reflections and ambiguous tongue surface identification*

Tongue Surface Segmentation: The orientation of objects in relation to the direction of the beam also impacts detectability. Density differences that run perpendicularly to the probe are detected more clearly than those that run in parallel. Edges of structures that are vertical in relation to the probe are left undetected. This could help explain why the tongue lamina is not visible for some coronal sounds (see Figure 2) as well as the tongue root for some high and front sounds.

Ultrasound beams have a thickness. In other words, they are not detecting objects along a single plane, but rather through a volumetric slice. As a result, nearby objects can bleed into the image. This phenomenon might create the appearance of multiple tongue contours (Figure 1) or a large bright blotch (Figure 2) where the tongue surface is expected. Stone (2005) established a general guideline that in the case of ambiguous images, one should select the most inferior. However, with the introduction of 3D/4D ultrasound to speech research, we are uncovering that this might not always be the most accurate assumption. Context from surrounding sagittal, coronal, and transverse slices as well as preceding and following frames can provide the necessary information for disambiguating difficult images. We have also learned that most speech sounds, both vocalic and consonantal, are produced with a groove along the midline of the tongue. For high sounds, the groove is primarily along the root of the tongue, and for low sounds it is along anterior portion (dorsum to apex) of the tongue. Grooves can cause the appearance of double tongue surfaces due to the close proximity of the trough and peak. With a prior expectation of grooves, segmenting ultrasound images can become easier.

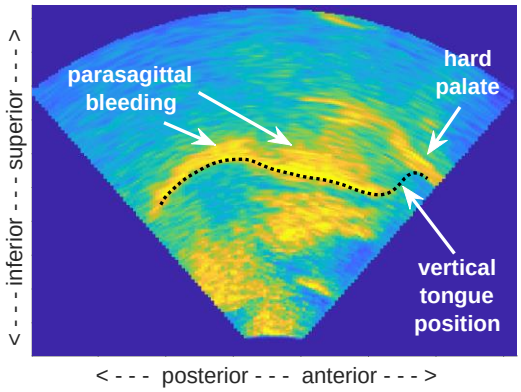

Figure 2: *Example of vertical tongue position and parasagittal bleeding*

Three-dimensional or even cross-plane images are not available to everyone using ultrasound for speech research. Spatial context is limited in two-dimensional data. However, temporal information can still provide some support for segmenting challenging images. Taking the time to properly orient the probe along the midsagittal plane before collecting data can also aide in the data analysis process.

References

Lulich, S. M., Berkson, K. H., & de Jong, K. (2018). Acquiring and visualizing 3D/4D ultrasound recordings of tongue motion. *Journal of Phonetics*, *71*, 410–424.

Stone, M. (2005). A Guide to Analyzing Tongue Motion from Ultrasound Images. *Clinical Linguistics and Phonetics*, *19*(6/7), 455–502.

Keywords: ultrasound, speech anatomy, tongue segmentation