Towards Robust Word Alignment of Child Speech Therapy Sessions

Manuel Sam Ribeiro, Aciel Eshky, Korin Richmond, Steve Renals,

Centre for Speech Technology Research, University of Edinburgh, UK

{sam.ribeiro, aeshky, s.renals, korin}@ed.ac.uk

Abstract

Developmental Speech Sound Disorders (SSDs) are a common communication impairment in childhood. These describe cases where children consistently exhibit difficulties in the production of specific speech sounds in their native language. SSDs have the potential to negatively affect the lives and the development of children. For example, self-awareness of disordered speech may lead to low-confidence in social situations or introduce communication barriers that lead to increased difficulty in learning and decreased literacy levels [1].

Clinical intervention is typically available for children with SSDs. However, current clinical methods for speech therapy are subjective and inaccurate [2]. Instrumented methods, such as spectrogram analysis or articulatory imaging, are useful, but require a large amount of manual effort from speech pathologists. In the Ultrax Speech Project (`www.ultrax-speech.org`), we explore objective methods that could alleviate manual processes undertaken by Speech and Language Therapists (SLTs) using audio and ultrasound. This paper lays out some of the major challenges for processing this type of data and presents initial results for the tasks of speaker labeling and word alignment.

We use the `UltraSuite Repository` [3], a collection of datasets of ultrasound and acoustic data collected from recordings of child speech therapy sessions. The repository contains one dataset of Typically Developing children and two datasets of children with SSDs. There are various challenges associated with this type of data. For example, the interaction between speech therapist and child, insertions and deletions with respect to the given prompt, mispronounciations, and the various challenges associated with child speech processing and disordered speech processing. These challenges are noticeable when force-aligning the audio with the expected prompt. Using baseline standard methods, we observe an f1-score of word recovery of 69% in Typically Developing children and 30% in diagnosis sessions of Speech Disordered children.

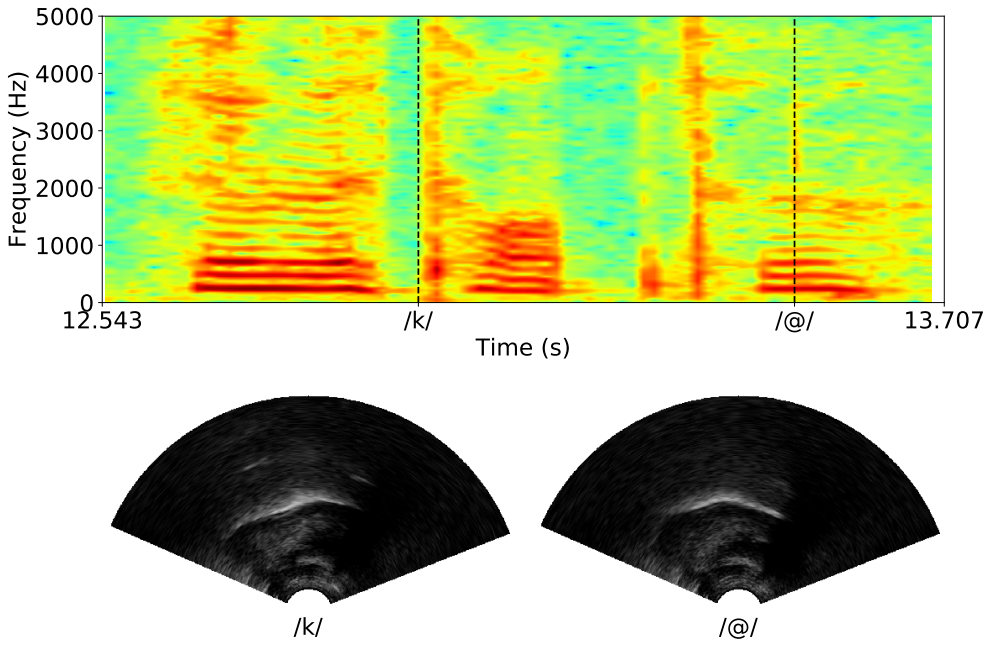

Figure 1: *Spectrogram for the word* helicopter *with two corresponding ultrasound frames, ellicited during a session with a six-year-old child diagnosed with velar fronting. Ultrasound frames show a mid-saggital view of the oral cavity with the tip of the tongue facing right.*

1. References

[1] S. McLeod and E. Baker, *Children's speech: An evidence-based approach to assessment and intervention*. Pearson, 2016.

[2] S. Howard and A. Lohmander, *Cleft palate speech: assessment and intervention*. John Wiley & Sons, 2011.

[3] A. Eshky, M. S. Ribeiro, J. Cleland, K. Richmond, Z. Roxburgh, J. Scobbie, and A. Wrench, *UltraSuite: A Repository of Ultrasound and Acoustic Data from Child Speech Therapy Sessions*. Manuscript submitted for publication, 2018.