Crossing the Pond: Extending Automatic Alignment Techniques to British English Dialect Data

Tools for automatic phonemicization and alignment of speech are gaining currency in sociolinguistic research (e.g. Labov et al., 2013). These tools convert an orthographic transcription into phonemes, then automatically time-align words and phonemes to the speech signal. However, these processes rely on a standard dictionary of reference pronunciations, and the extent to which they may be used effectively with data from non-standard dialects is not yet known. In this paper, we evaluate the performance of three different forced alignment tools on data from several dialects of British English.

Our data consist of interviews with speakers representing the dialects of Essex, Manchester, Liverpool, and Received Pronunciation. All interviews were transcribed and submitted to three different forced alignment procedures: FAVE-align (Rosenfelder et al., 2011), Prosodylab-Aligner (Gorman et al., 2011), and SPPAS (Bigi, 2012). Alignment was performed using a British pronunciation dictionary (Robinson, 1997).

To evaluate aligner performance, we identified five vowels which differentiate the dialects under study: FOOT, STRUT, BATH, and TRAP, which constitute two lexical splits differentiating northern and southern England (Wells, 1982); and GOOSE, which is undergoing fronting (Harrington et al., 2008). Tokens of these five vowels in stressed position were manually segmented by the authors, and these measurements became the 'gold standard' against which the output of the three aligners was compared.

Overall, the systems perform well: 70% of machine-placed vowel onsets and 71% of offsets are within 20 ms of the human annotator's placement. Figure 1 breaks down these results by dialect and by vowel for two of the aligners. Problem areas for alignment include /l/-vocalisation, glottalisation, and phoneme inventory differences; these contribute to varying performance across dialects.

FAVE-align's performance slightly exceeds the other aligners', with 74% of boundaries within 20 ms of the manual placement. However, FAVE-align uses American English acoustic models; accordingly, it requires transcriptions to be coerced into American vowel classes. We weigh its accuracy against this drawback and also compare the aligners on other metrics, such as ease of use. We conclude by addressing the effects of inaccurate alignment on automatic vowel measurement and providing recommendations for researchers working with similar data.

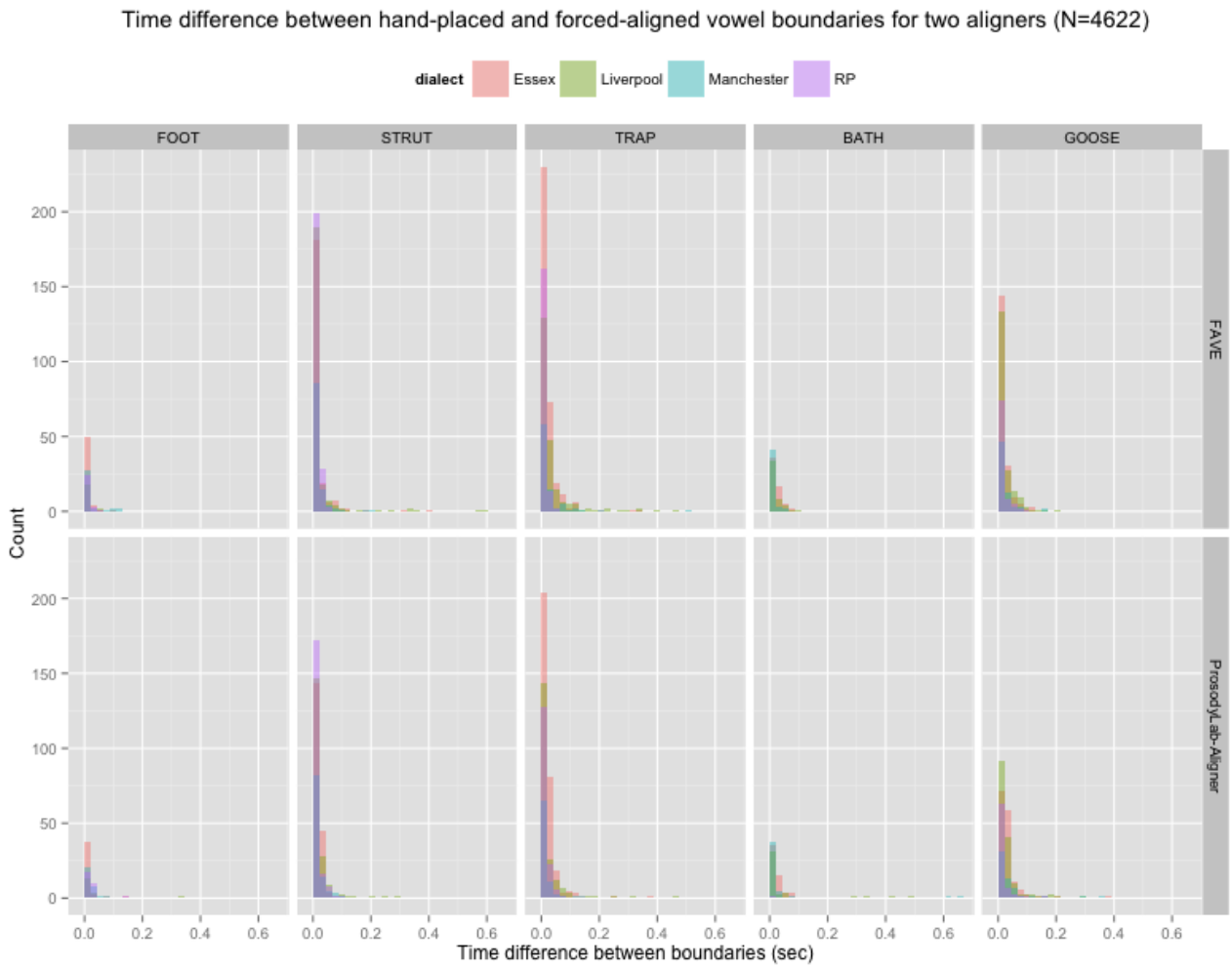

Figure 1. Histograms of alignment errors.

References

Bigi, Brigitte. 2012. SPPAS: a tool for the phonetic segmentations of speech. The Eighth International Conference on Language Resources and Evaluation, Istanbul, May 2012.

Gorman, Kyle, Howell, Jonathan, and Wagner, Michael. 2011. Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Proceedings of Acoustic Week in Canada '11*.

Harrington, Jonathan, Kleber, Felicitas, & Reubold, Ulrich. 2008. Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America* 123:2825–2835.

Labov, William, Rosenfelder, Ingrid, and Fruehwald, Josef. 2013. One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language* 89:30–65.

Robinson, A. J. 1997. British English Example Pronunciation (BEEP). Available from ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/dictionaries.

Rosenfelder, Ingrid, Fruehwald, Josef, Evanini, Keelan, and Yuan, Jiahong. (2011). FAVE (Forced Alignment and Vowel Extraction) program suite. Available from http://fave.ling.upenn.edu.

Wells, John C. 1982. *Accents of English.* Cambridge: Cambridge University Press.