Automatic Speech-to-Phoneme Alignment of the Spoken British National Corpus

Sergio Grau, Ladan Baghai-Ravary, Greg Kochanski and John Coleman

Phonetics Laboratory, University of Oxford, U.K.

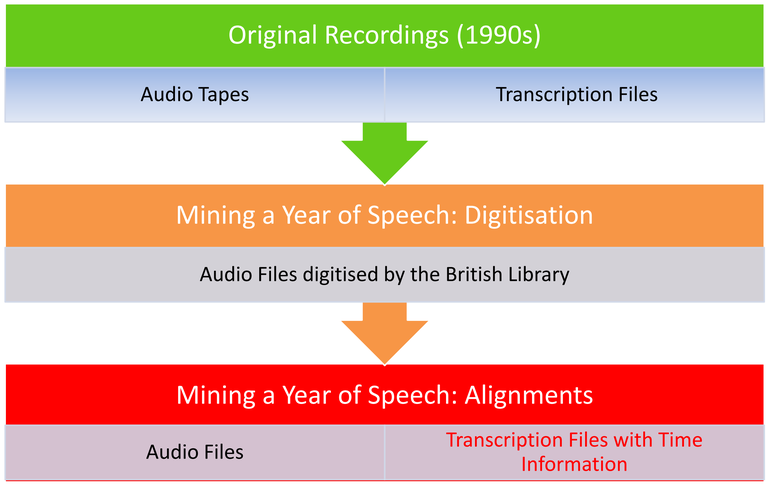
Introduction

This poster describes the creation of an automatic word and phoneme alignment between the audio recordings of the **Spoken British National Corpus** (BNC) [1] and their corresponding word-level transcriptions.

The work presented here is part of the **"Mining a Year of Speech"** project [2] which aim is to produce automatic speech-to-phoneme alignments of an approximately one year of audio recordings. The Spoken BNC recordings consist of unscripted, spontaneous speech conversations in different recording conditions, accents and background noises. The range of topics covers from radio programs to family conversations, council meetings or chemistry courses. The Spoken BNC was originally recorded on analogue cassette tapes between 1991 and 1994. These tapes have been recently-digitised by the **British Library**. The resulting dataset is composed of approximately 2,000 digital audio files with an average duration of 45 minutes and their associated word-level transcriptions.

This poster describes the goal of the project, the dataset and the automatic alignment method.

Goal

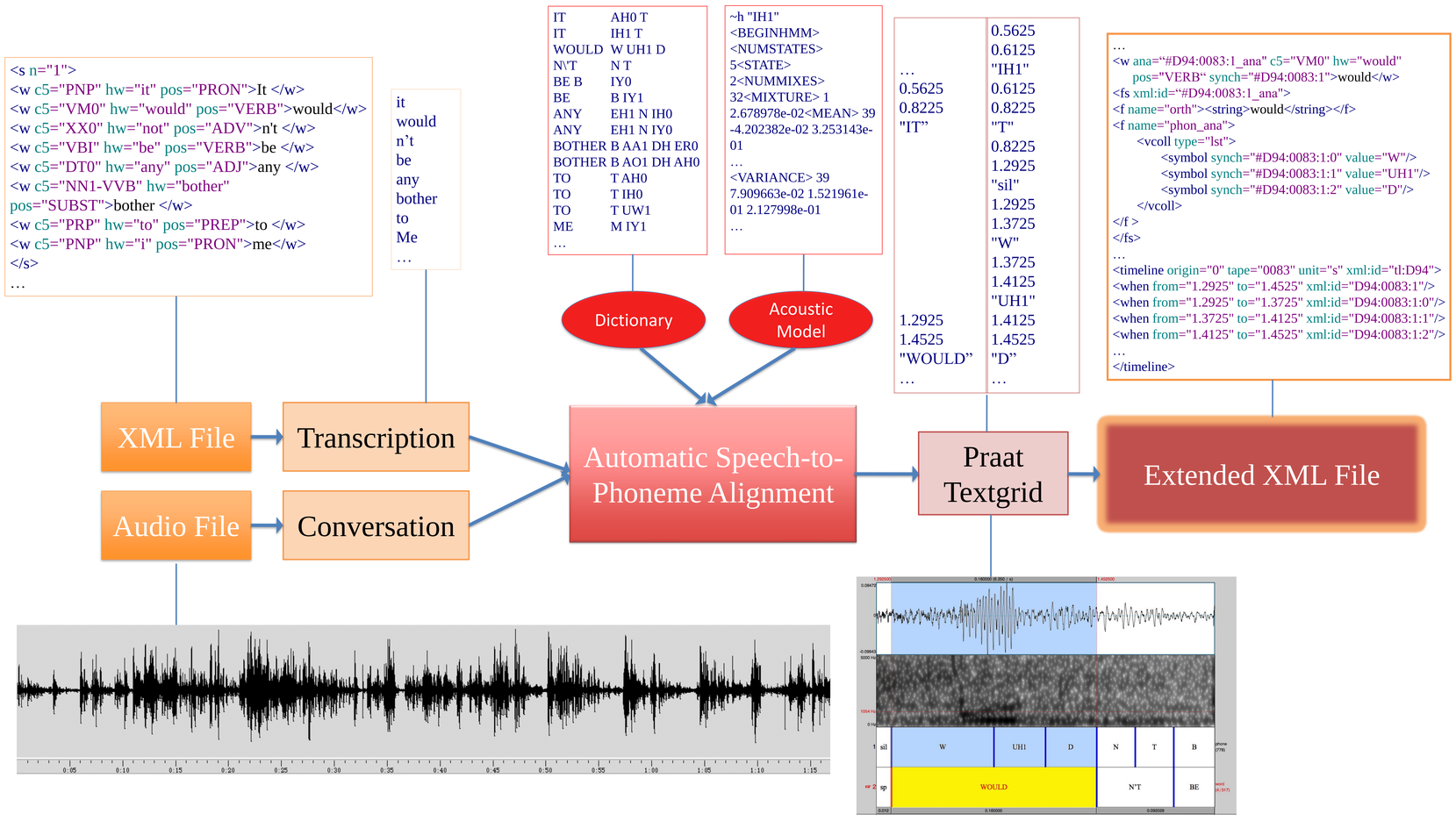


Dataset

- **Audio:** around 2,000 audio files. Each track of the original cassette recordings has been digitised to a single file, so the data to be aligned is generally just over 45 minutes long

- **Transcriptions:** 908 XML transcription files

Method



Summary

- **Automatic Speech-to-Phoneme Aligner:** Penn Phonetics Lab Forced Aligner, P2FA [3]. P2FA is an automatic phonetic aligner based on HTK [4], and developed at the Phonetics Laboratory of the University of Pennsylvania.

- **Dictionary:** The current CMU Pronouncing Dictionary [5] was extended to include all the out-of-vocabulary words and to include a range of common British English word pronunciations. This extension was performed using semi-automatic methods by experienced phoneticians.

- **Alignment Quality:** How do identify regions of bad alignment? Semi-automatic evaluation of the alignment of large speech corpora [6]

References

- [1] "The British National Corpus", http://www.natcorp.ox. ac.uk/
- [2] Coleman, J., Liberman, M., Kochanksi, G., Burnard, L., and Yuan. J., "Mining a Year of Speech", New Tools and Methods for Very-Large Scale Phonetics Research Workshop, University of Pennsylvania, January 28-31 ,2011.
- [3] Yuan, J. and Liberman, M., "Speaker identification on the SCOTUS corpus", in Proc. Acoustics'08, pp. 5687-5690, 2008.
- [4] Young, S. J.; Kershaw, D.; Odell, J.; Ollason, D.; Valtchev, V. & Woodland, P., "The HTK Book Version 3.4" , Cambridge University Press, 2006.
- [5] Carnegie Mellon University Pronouncing Dictionary, available from http://www.speech.cs.cmu.edu/cgi-bin/ cmudict
- [6] Baghai-Ravary, L., Grau, S., Kochanski, G., "Detecting gross alignment errors in the Spoken British National Corpus", New Tools and Methods for Very-Large Scale Phonetics Research Workshop, University of Pennsylvania, January 28-31, 2011.

Acknowledgments

We thank JISC (in the UK) and NSF (in the USA) for their support of Mining a Year of Speech, under the Digging into Data programme. This work is also partly supported by the UK ESRC (awards RES-062-23-2566, RES-062-23-1172, and RES-062-23-1323).