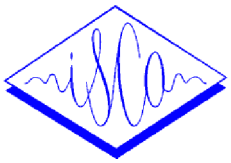
Proceedings of DiSS-LPSS Joint Workshop 2010,
25-26 September 2010, Tokyo, Japan, p. 1.



Infant Speech Database for Longitudinal Analysis of Spoken Language Development

Shigeaki Amano

Faculty of Human Informatics, Aichi Shukutoku University
psy@asu.aasa.ac.jp

1. Background

Both longitudinal and cross-sectional speech databases are used in research on the development of the spoken language. However, previous longitudinal speech databases (e.g., Hamasaki database and Miyata database in CHILDES project [1]) were limited in terms of the recording period or the number of utterances. To promote a developmental research, a large-scale longitudinal infant speech database has been developed from longitudinal recordings.

2. Longitudinal recording

Five infants [A(male), B(male), C(female), D(female), and E(female)] and their parents participated voluntarily in the recording. They were all Japanese. All the infants were born and raised in Tokyo or in Kanagawa prefecture, which adjoins Tokyo. They had no symptoms of disorder with respect to speech perception or speech production.

Utterances were recorded in a room in the participants' houses with a PCM recorder and a stereo microphone. Neither the infants nor parents were required to undertake any particular task while their natural spontaneous utterances were being recorded. Although infants B and E, and infants C and D are both siblings, their recordings were almost always obtained separately.

Recording started soon after the infant's birth. The recording time was about 1 hour per month for all the infants, but it was more or less than this in some cases. The duration and the number of recordings varied, because the recordings were conducted when the infant was feeling good. The total recording time was 541 hours.

3. Database development

An infant speech database was developed from the longitudinal recordings. Each recording was converted to a session file in the WAV format. Every utterance in the session file was automatically segmented and stored as an utterance file in the database. There were 269,467 utterance files.

By listening to the utterance files, utterance was transcribed and stored in a transcription file. Along with the transcription, speaker and utterance direction (infant directed or adult directed) were identified. Subjective levels of background noise and utterance loudness were also identified. These pieces of information were stored in a property tag file. By viewing a waveform of the utterance, its start and end times were identified in millisecond, and they were stored in a time record file.

Fundamental frequency of the utterance was estimated at every millisecond by using the "Ripple Enhanced Power Spectrum (REPS)" method [2], and it was stored in a fundamental frequency file. Along with the fundamental frequency, voiced and unvoiced parts of the utterance were estimated with a dominance spectrum method called "DominAnce Spectrum based Harmonics extraction (DASH)" [2], and they were stored in a voiced/unvoiced label file.

Trained phoneme labelers specified the phonemic segments in an utterance by listening to the utterance file and looking at its waveform and spectrogram, and they gave phoneme labels to the segments. The phoneme labels were stored in a phoneme label file.

Precise information about the database development can be found in [3]

4. Database release

The infant speech database with its custom-made search software is released by The Speech Resources Consortium (http://research.nii.ac.jp/src/eng/index.html) at a price of 85,500 yen. All database files except for the session and utterance files are provided in a text format so that users can read the database contents with a text editor or a computer program. A waveform editor tuned to the database is available from Arcadia Corporation (http://www.arcadia.co.jp/). This waveform editor can show an utterance waveform and its spectrogram with a trajectory of fundamental frequency and phoneme labels which are registered in the database.

5. Conclusion

An infant speech database was developed from 5 years of recordings of utterances of five Japanese infants and their parents. It contains 269,467 utterances with various types of their information. This large-scale database would be useful for a research of spoken language development in many aspects such as prosody (e.g., [4]), disfluency, errors, repairs and hesitations. It would be also useful for linguistic developmental analysis of vocabulary and syntax, and automatic speech recognition of spontaneous utterance of children.

6. References

[1] MacWhinney, B. "The CHILDES Project: Tools for Analyzing Talk," (3rd ed.) Lawrence Erlbaum Associates, Mahwah, NJ, 2000.

[2] Nakatani, T., Amano, S., Irino, T., Ishizuka, K., and Kondo, T. "A method for fundamental frequency estimation and voicing decision: Application to infant utterances recorded in real acoustical environments," Speech Communication, 50, 203-214, 2008.

[3] Amano, S., Kondo, T., Kato, K., and Nakatani, T. "Development of Japanese infant speech database from longitudinal recordings," Speech Communication, 51, 510-520, 2009.

[4] Amano, S., Nakatani, T., and Kondo, T. "Fundamental frequency of infants' and parents' utterances in longitudinal recordings," Journal of the Acoustical Society of America, 119, 1636-1647, 2006.