

IpSC36. Acoustic modeling of spontaneous speech of Japanese preschool children. Izumi Shindo, Tobias Cincarek, Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano (Nara Inst. of Sci. and Technol., 8916-5 Takayama-cho, Ikoma-shi, Nara 630-0192, Japan)

In recent years, there is an increasing demand for speech recognition of children. However, the recognition of children's speech, especially preschool children (2 to 5 years of age), is very difficult. For example, recognition accuracy using a children's acoustic model provided by the Japanese Dictation Toolkit is only 21.4%. Many different variations of child speech with palatal sounds and pronunciation error decrease recognition performance. This paper proposes a recognition method that investigates the characteristics of preschool children's speech using experimental data and considers phonetic changes. Mapping between standard and altered pronunciations of words is determined. In experiments, a large amount of spontaneous child speech (2 to 15 years of age) was collected with the speech-oriented public guidance system, "Takemaru-kun," which is currently available. Recognition performance increases to 49.2% by acoustic model adaptation of preschool children's speech. When allowing multiple pronunciation variations per word during recognition, further improvement to 52.0% is achieved.