

Understanding Children's Speech Productions: Man Versus Machine

Madeleine Yu

University of Toronto, Toronto, Ontario, Canada

Amrita Bagga

University of Toronto, Toronto, Ontario, Canada

Elizabeth Johnson

University of Toronto, Mississauga, Ontario, Canada

Abstract

Young children's speech pronunciations deviate systematically from adult forms. For example, onsets are often simplified (e.g., stop becomes top), unstressed syllables frequently deleted (e.g., spaghetti becomes getti), and certain segments are commonly replaced with other ones (e.g., rice becomes wice). The current study examined how well adults and a popular automatic speech recognition system (i.e., Siri) deal with these deviations. The same 12 children were recorded producing 32 words in isolation at three ages: 2.5, 3.5, and 5.5 years. 12 adults were also recorded. These recordings were presented to 48 young adults, 7 mothers, and Siri for transcription. All listeners performed worst with 2.5-year-old productions, and humans outperformed Siri with all ages ($p < 0.001$). Mothers demonstrated the highest accuracy with 2.5-year-old productions (86%). Additionally, Siri made distinctive transcription errors with children's speech. These errors may reflect the system's lack of training with young children's voices.