

Effects of prosody on the intelligibility, communication efficiency, and perceived naturalness of synthetic speech in augmentative and alternative communication

Jennifer M. Vojtech, Jacob P. Noordzij Jr., Gabriel J. Cler, Cara E. Stepp

Some individuals with motor speech disorders may require speech-generating devices in order to communicate. Synthesized speech enables these individuals to convey urgent needs and advocate for self-management; however, the resulting speech output is robotic, unnatural, and at times, unintelligible, as current text-to-speech augmentative and alternative communication systems do not integrate prosody. Consequently, these individuals must repeat their speech in order to be understood, and thus expend more effort to relay their needs, while listeners must use more cognitive effort to comprehend the intended message. Here, we assessed the impacts of modulating the prosodic cues of pitch and speech rate on the intelligibility, communication efficiency ratio (CER), and perceived naturalness of synthetic speech. We elicited orthographic transcriptions and judgments of naturalness from sixteen healthy listeners using synthetic speech stimuli with altered pitch and rate modulations. The stimuli were evaluated for perceived naturalness using the listener judgments, and for intelligibility and CER using the orthographic transcriptions. Results highlight the effects of prosodic cues on synthesized speech as a crucial step toward enabling individuals with severe motor speech impairments to communicate using fast, reliable synthetic speech.