

Independent Contributions of Amplitude and Frequency Modulations to Auditory Perception: I. Consonant, Vowel, and Sentence Recognition

Kaibao Nie, Ginger Stickney, and Fan-Gang Zeng

Hearing and Speech Research Laboratory
Departments of Otolaryngology and Biomedical Engineering
364 Med Surg II, Irvine CA 92697
University of California, Irvine

knie@uci.edu fzeng@uci.edu

Previous studies have demonstrated that a high level of speech recognition in quiet can be achieved with primarily temporal envelope cues in normal hearing listeners (Shannon et al. 1995; Smith et al. 2002) and cochlear implant users (Wilson et al. 1991). However, background noise presents a particular challenge to speech perception with primarily temporal cues (Dorman et al. 1998). Although natural sounds such as speech and music contain both amplitude modulation (AM) and frequency modulation (FM) cues, it remains unclear how they are processed in the auditory system and whether they have different functional roles in auditory perception.

We have developed a signal processing strategy that can independently extract slowly-varying amplitude and frequency modulations within a frequency band with the number of bands as an independent variable. The independent contributions of AM and FM to auditory perception are systematically investigated using this new strategy. Normal-hearing listeners are presented with the original sounds and processed sounds including amplitude modulation only and both amplitude and frequency modulations. Here we introduce the implementation of the strategy first and then present results of consonant and vowel recognition in quiet and noise, as well as sentence recognition in quiet or with a competing speaker. Further results in tonal language, music and speaker recognition are presented in a companion poster.

The addition of the FM cue significantly improves phoneme and sentence recognition in quiet and in noise. The degree of improvement is the most for sentence recognition with a competing talker, suggesting the importance of the FM cue in speech perception under realistic listening environments. The FM cue may have allowed the listener to tell one talker (signal) apart from the other (noise). The FAME strategy may be used as an efficient method to encode fine structure information in cochlear implants. It can also be used for low-rate, high-quality speech coding and synthesis.