Noise-band Vocoder Interferes with Auditory Statistical Learning in Adults

Tina M. Greco-Calub1, Hillary E. Snyder1, Paul N. Reinhardt1, Casey Lew-Williams2
1The Roxelyn & Richard Pepper Department of Communication Sciences & Disorders, Northwestern University, Evanston, IL
2Department of Psychology, Princeton University, Princeton, NJ

Abstract
The purpose of Experiment 2 was to identify potential differences in recognition of mono- vs disyllables in the clear and noise-band vocoded conditions. See Methods for the creation of disyllables.

Summary
We would like to thank Jenny Saffran for her assistance in stimulus creation. We would like to thank Katherine Simion for assistance in data collection. This study was funded by the National Science Foundation (BCS-1041624) and the National Institutes of Health (DC007335).

References

Noise-band vocoder interferes with auditory statistical learning

Background
Auditory statistical learning refers to the ability to extract structure (e.g., consistent syllable sequences) from patterned auditory input (e.g., language) and has been widely proposed as a foundation for early language development (Saffran, Aslin, & Newport, 1996). Critically, successful segmentation of speech into its constituent parts hinges on the ability to track these relations is currently unknown. This is a clinically relevant issue when considering that individuals who use cochlear implants must acclimate to spectrally degraded electric hearing and—in the case of infants—learn the rules of their native language. The purpose of this study is to determine the role of spectral resolution in auditory statistical learning and its results by manipulating an artificial language with noise-band vocoding.

Methods
Participants: Experiment 1: 48 young adults (32 female, 16-33 years old). Experiment 2: 26 young adults (female, 18-33 years old). All procedures were approved by the Institutional Review Board of Northwestern University.

Stimuli
Artificial language: Four nonsense trisyllabic words were concatenated to create a monophone (female), pause-free speech stream (3.25 min. duration). Rules of the language: 1. Two words occurred at a low frequency (35 times); 2. Three words occurred at a high frequency (70 times); 3. The high frequency words combined to generate two frequency-matched part-words (35 times); 4. Transition probabilities (TP) are the likelihood that syllable X will follow syllable Y. The high frequency words had a TP of 1. The frequency-to-word conversion had a TP of 0.2. 5. Two languages were used to counterbalance words versus part-words.

Table 1: Feature error patterns across the three conditions in Figure 2. Consonants were defined by manner of articulation, voicing, and place of articulation. Post-trial responses were compiled across subjects into a composite confusion matrix. Due to the low number of trials, errors from the confusion matrix were applied to the consonant feature definitions in order to derive total feature error counts. Error patterns (as in Figure 4) were developed for the clear (200-0.001), but not between the clear and 16-ch conditions (p = 0.193).

Syllable recognition accuracy decreases as spectral resolution decreases

The Greenwood function (Greenwood, 1990). The temporal envelope in each relative number of frequency bands (16 or 8, between 200-7000 Hz) using a one-way ANOVA showed a main effect of condition [F (2,45) = 7.7, p = 0.001]. Post-hoc comparisons revealed significant differences between the clear and 16-ch conditions (p = 0.015) as well as the clear and 8-ch conditions (p = 0.001). Performance in the 16-ch and 8-ch conditions was not statistically different (p = 0.10).

Figure 1: Word recognition accuracy per condition (mean ± S.D.). Adults identified words and part-words significantly above chance performance (50%, blue dotted line; chance performance level) in the clear (16-ch) and 8-ch noise-band vocoded conditions. A one-way ANOVA showed a main effect of condition (F [2,45] = 7.7, p < 0.001). Post-hoc comparisons revealed significant differences between the clear and 16-ch conditions (p = 0.015) as well as the clear and 8-ch conditions (p = 0.001). Performance in the 16-ch and 8-ch conditions was not statistically different (p = 0.10).

The purpose of Experiment 2 was to identify potential differences in recognition of mono- vs disyllables in the clear and noise-band vocoded conditions. See Methods for the creation of disyllables.