The effect of isolated words on segmenting noise-band vocoded speech

Katherine M. Simeon¹, Hillary E. Snyder¹, Casey Lew-Williams² & Tina M. Grieco-Calub¹

¹The Roxelyn & Richard Pepper Department of Communication Sciences & Disorders, Northwestern University, Evanston, IL
²Department of Psychology, Princeton University, Princeton, NJ

Background

According to the theory of statistical language learning, humans segment speech and detect word boundaries by tracking transitional probabilities (TP) of different syllable combinations (Saffran, 2003; Singh et al., 2012). TP refers to the likelihood that syllable Y will follow syllable X. Successful segmentation of spoken language assumes that listeners have full access to the spectral and temporal components of the speech signal. The robustness of statistical learning under conditions of degraded speech, however, has yet to be tested. The purpose of this study is test the ability of adults to segment speech that has been spectrally degraded by noise-band vocoding. Experiment 1 (Exp. 1) tests the hypothesis that statistical language learning is dependent on the spectral fidelity of the signal. Experiment 2 (Exp. 2) tests the hypothesis that statistical language learning of degraded speech can be facilitated with temporal cues (i.e., silence) around target words.

Methods

Participants: Exp. 1: 48 young adults (32 female, 16-33 years old); Exp. 2: 42 adults (31 female, 18-34 years old). All participants were native English speakers, with no prior history of hearing loss or speech and language services. All procedures were approved by the Institutional Review Board at Northwestern University.

Stimuli

Artificial language: Four nonsense trisyllabic words were concatenated to create a monotone female, pause-free speech stream (Duration: 3.25 minutes for Exp. 1, 3.52 minutes for Exp. 2). Rules of the language:

1) Two words occurred at a low frequency (35 times);
2) Two words occurred at a high frequency (70 times);
3) The high frequency words combined to generate two frequency-matched part-words (35 times);
4) Transitional probability (TP) is the likelihood that syllable Y will follow syllable X. The low frequency words had a TP = 1. The frequency-matched part-words had a TP = 0.5;
5) Two languages were used to counterbalance words versus part-words.

Listening conditions

The artificial language and test words were either:
1) Unprocessed (full spectral resolution);
2) Noise-band vocoded with 16 frequency channels (16-Ch);
3) Noise-band vocoded with 8 frequency channels (8-Ch).

Procedures

Familiarization: Participants were seated in a room and were familiarized with the stimuli. The stimuli were presented in either the 16-channel vocoded or 8-channel vocoded. For Exp. 1, N = 16 per listening condition. For Exp. 2, N = 14 per listening condition.

Two-Alternative Forced Choice (2-AFC): Participants heard a word pair and were instructed to select the word that sounded more “familiar.” There were 24 trials and trial order was counterbalanced for each participant. The measurement of interest was task accuracy.

Comparing the results of Experiments 1 & 2

A 2 x 3 ANOVA was used to test main effects of experiment (No temporal gap, temporal gap) and condition (16-Ch, 8-Ch) on 2-AFC performance. Results revealed a main effect of experiment (F (1,84) = 4.39, p < 0.05) and a main effect of listening condition (F (2,84) = 9.19, p < 0.05). These results show that, on average, speech segmentation improves with greater spectral fidelity (F (1,84) = 4.39, p < 0.05). Post-hoc comparisons revealed a statistically significant difference between 16-Ch and 8-Ch conditions (p < 0.05) and a marginal difference between 8-Ch and unprocessed conditions (p = 0.059).

Figure 1: Accuracy (%  0) on the 2-AFC task by listening condition. Adult segmented words significantly above chance performance (50%, red dotted line) in the unprocessed listening condition. Performance was not statistically different from chance for the 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 < 0.05). Post-hoc comparisons revealed statistically significant differences between unprocessed and 16-Ch conditions (p < 0.05) as well as the unprocessed and 8-Ch conditions (p < 0.05).

Figure 2: Accuracy (%  0) on the 2-AFC task by listening condition. Adults segmented words significantly above chance performance (50%, red dotted line) across all listening conditions. A one-way ANOVA revealed a main effect of listening condition (F (2,36) = 4.524, p < 0.05). Post-hoc comparisons revealed a statistically significant difference between 8-Ch and 16-Ch conditions (p < 0.05) and a marginal difference between 8-Ch and unprocessed conditions (p = 0.059).

Summary & Conclusions

1. Noise-band vocoding disrupts speech segmentation (Figure 1). This suggests that listeners who have limited access to spectral fidelity of speech, like those with hearing loss or who use cochlear implants, may have impaired speech segmentation. Because speech segmentation has been promoted as a strategy used by infants in the early stages of language development, these data suggest that infants who hear with a cochlear implant may be particularly at risk for language delays.

2. Although there continued to be a main effect of listening condition, temporal gaps improved speech segmentation abilities (Figure 2). The benefit of temporal gaps appeared to be largest for the 16-ch condition. This finding suggests that when spectral fidelity is high (unprocessed condition), temporal gaps do not add benefit. Conversely, when spectral fidelity is very poor (8-ch condition), temporal cues may be limited in their ability to facilitate performance. The 16-ch might represent a particularly good balance of spectral and temporal cues that results in fairly good speech segmentation abilities.

Future Directions

Data collection for Exp. 2 is still in progress in order to achieve equal sample sizes across both experiments; however, the preliminary results of Exp. 1 & 2 motivate potential follow up studies. One question is how do degraded listening conditions contribute to speech segmentation in individuals with hearing loss? When noise-band vocoding simulates a cochlear implant that processes the signal, we are unsure of how a poor sound signal leads to a damaged auditory system. A logical next step would be to test cochlear implant users in a spectral speech, without any band vocoding. To see how they process language through their implant. Additionally, other cues, such as prosody and speech rate, may contribute to speech segmentation. Moreover, if spectral cues are present in natural speech will contribute to current knowledge of how information in the speech signal can compensate for a lack of spectral resolution.

References


Acknowledgements

We would like to thank Sarah Schaefer for assistance in data collection. Additionally, we are grateful to David Dorsch, Silvia Lam, and Kristi Ward for comments on this poster.