Arsenault, J., He, Y., Bidelman, G., & Alain, C. (2014). The impact of context on the perceptual organization of speech. *Journal of the Canadian Acoustical Association, 42*, 72-73.

THE IMPACT OF CONTEXT ON THE PERCEPTUAL ORGANIZATION OF SPEECH

Jessica S. Arsenault[1,2], Yu He[1], Gavin M. Bidelman[3], and Claude Alain[1,2]
[1]Rotman Research Institute, Toronto, Ontario, Canada
[2]University of Toronto, Toronto, Ontario, Canada
[3]University of Memphis, Memphis, Tennessee, United States

1 Introduction

One of the major functions of the human auditory system involves auditory scene analysis – separating a noisy environment into distinct auditory representations. Streaming is a well-studied laboratory task that is meant to model real-world object recognition, and typically involves the presentation of tonal stimuli in an ABA- type galloping pattern, with low A tones and high B tones[1]. When the frequency difference between A and B is small, participants typically perceive the tones as a single integrated source (ABA-ABA-…). When the frequency difference between A and B is large, participants tend to perceive two separate streams, known as streaming, with A repeating at twice the rate of B (A-A-A-A and B---B---…). At an intermediate frequency difference, perception is more ambiguous, though a series of recent studies found that both prior stimulus and prior perception have a large impact on how individuals classify ambiguous sequences [1,2].

Snyder and colleagues showed that when participants hear a non-ambiguous sequence (i.e., either one or two streams) followed by an ambiguous sequence, they are likely to classify the ambiguous sequence as the opposite of the initial sequence (impact of prior stimulus); for example, when presented with one stream followed by an ambiguous sequence, participants are likely to classify the ambiguous sequence as two streams. Interestingly, when two ambiguous sequences are presented in a row, the opposite effect is observed; if the first ambiguous sequence is classified as one stream, the next ambiguous sequence is more likely to also be classified as one stream (effect of prior perception [2]). Electrophysiological evidence suggests that the mechanisms through which recent stimuli/perceptual states influence the formation of auditory streams are at least partially independent from those influencing the current auditory stimuli [3].

Given the importance of verbal stimuli in auditory scene analysis, the objective of the current study was to extend the findings of Snyder and colleagues by using speech sounds instead of pure tones. Using more complex and ecologically valid stimuli, we gain a better understanding of auditory stream segregation and how we interact with the auditory world.

2 Methods

Sixteen healthy young adults (M_{age} = 23.25yr, SD = 4.39; 8 females) were recruited from the Baycrest participant database. All participants were right-handed except for one who was left-handed and all were fluent English-speakers with no known neurological or psychiatric issues and no history of hearing or speech disorders. Participants gave informed written consent according to guidelines established by Baycrest's Research Ethics Board.

Stimuli consisted of the vowel sounds /i/ (as in s*ee*) and /ae/ (as in c*a*t), henceforth referred to as "ee" and "ae". The vowels were presented in an ABA- pattern as ee-ae-ee-, with mainly first formant (f_1) frequency differences between A (ee) and B (ae). The stimuli represented one of three conditions: large f_1 difference between A and B (Δf_1 = 285Hz; typically perceived as two streams), small Δf_1 between A and B (Δf_1 = 47Hz; typically perceived as one stream), or intermediate Δf_1 (Δf_1 = 110Hz; ambiguous perception). Each trial consisted of an adaptation phase, which could have either a small, intermediate, or large Δf_1, as well as a test phase, which was always intermediate. Both phases were 7s each in duration, and were separated by a 1.44s inter-stimulus interval. In each phase, 14 repetitions of the ABA- triplets were presented sequentially, after which the participant made a response via button box indicating whether the previous sequence was perceived as one or two streams.

Participants were seated in a comfortable chair in a sound-attenuated chamber for the duration of the study. The testing session began with two hearing tests – the pure tone thresholds audiometry and the QuickSIN (speech-in-noise recognition). The order of the two tests was counter-balanced across participants. The concept of streaming was explained to participants and a brief practice session was given in order to familiarize participants with the stimuli and task. Participants were encouraged to keep their eyes fixated in a comfortable position and listen to the sounds. Participants completed five blocks of 30 trials each for a total of 150 trials, with each adaptation condition (small, intermediate, or large Δf_1) being presented 50 times throughout the study.

3 Results

During the adaptation phase, participants were more likely to report hearing two streams (streaming) when the formant difference between A and B was intermediate or large than when it was small (see Figure 1). There was also a difference in the ambiguous test trials based on which condition was presented at adaptation; participants were

[1] For an interactive example, please go to the following link: http://auditoryneuroscience.com/topics/streaming-galloping-rhythm-paradigm

significantly less likely to perceive intermediate test sequences as two streams as adaptation shifted from small to large Δf_1 [$F(2,14) = 10.479$, $p < 0.001$). These results demonstrate an effect of current/prior stimulus using speech sounds similar to previous psychophysical findings observed with tonal stimuli [1].

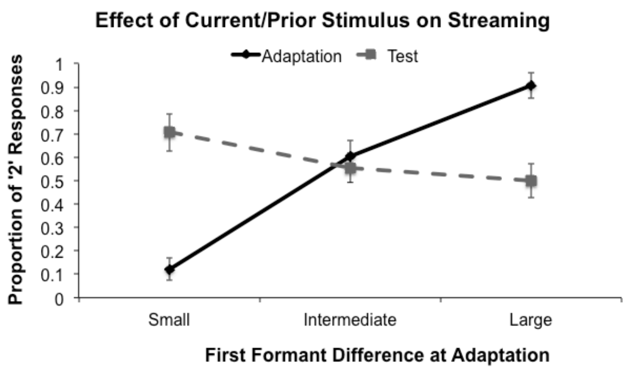

Figure 1. Effects of first formant differences on perception of streaming during the adaptation and test phase. Error bars represent standard error of the mean.

In order to investigate the impact of prior perception on subsequent classification, we looked at the likelihood of reporting streaming at test based on the perception of intermediate adaptation sequences. We found an increase in the perception of streaming at test when streaming was also perceived during intermediate adaptation sequences [$F(1,15) = 21.420$, $p < 0.001$; see Figure 2].

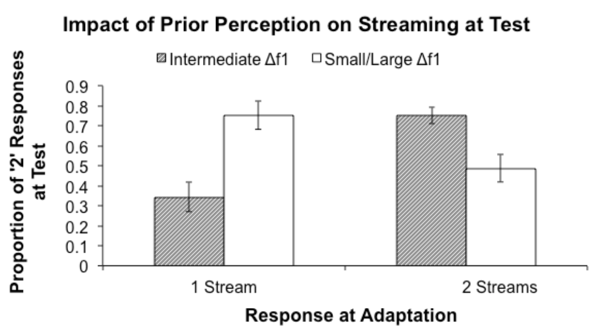

Figure 2. Effect of prior perception during adaptation on perception at test for ambiguous (intermediate Δf_1) and non-ambiguous (small/large Δf_1) adaptation sequences. Error bars represent standard error of the mean.

Importantly, this is the opposite of the prior stimulus pattern that was observed when the adaptation sequence had either a small or large Δf_1, as described above.

4 Discussion

Consistent with previous research using pure tones, we found that large differences in first formant between vowels (similar to frequency differences in tones) elicited more reports of streaming than a small difference between first formant transitions. Further, prior stimulus presentation seemed to bias current perception *away* from the stimulus that was just heard, while prior perception of ambiguous stimuli seemed to prime current perception *towards* the stimulus just perceived. Using tonal stimuli, Snyder and colleagues manipulated lags and intertrial intervals in order to precisely measure the time course of context effects, and concluded that it is unlikely that these effects are largely driven by response bias. Instead, it is suggested that different levels of neural representations reflect stimulus-related (i.e., Δf_1) and perception-related (i.e., 1 stream vs. 2 streams) processes, and that the effects of both processes build up over time [1,3]. Further research is required to determine whether similar mechanisms are responsible for speech sound segregation as tonal segregation, as well as whether the streaming of speech sounds is also affected by factors such as attention and knowledge [4].

One of the fundamental processes of the human auditory system is to organize sounds into meaningful elements, such as separating a police siren from the music playing through your car radio, or identifying and attending to your friend's voice in a noisy room. The findings of the current study support the notion that auditory stream segregation of speech sounds is impacted by context. A crucial next step would be to obtain electrophysiological data to better understand the neural mechanisms responsible for such processes. Also of interest would be the effects of aging on streaming of speech sounds, as previous research has shown age-related deficits in the sequential streaming of vowels [5].

Using complex, ecologically valid stimuli, we have replicated patterns of streaming previously only observed with pure tones. This study adds to the rich volume of literature characterizing the phenomenon of streaming and the effects of context on auditory scene analysis.

Acknowledgments

This research was supported by a Canadian Institutes of Health Research (CIHR) grant held by CA.

References

[1] Snyder, J. S., Carter, O. L., Lee, S.K., Hannon, E. E., & Alain, C. (2008). Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1007–16. doi:10.1037/0096-1523.34.4.1007.

[2] Snyder, J. S., Holder, W. T., Weintraub, D. M., Carter, O. L., & Alain, C. (2009). Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology*, *46*(6), 1208–15. doi:10.1111/j.1469-8986.2009.00870.x.

[3] Snyder, J. S., Carter, O. L., Hannon, E. E., & Alain, C. (2009). Adaptation reveals multiple levels of representation in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1232–1244. doi:10.1037/a0012741.

[4] Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin*, *133*(5), 780–99. doi:10.1037/0033-2909.133.5.780.

[5] Hutka, S. A, Alain, C., Binns, M. A, & Bidelman, G. M. (2013). Age-related differences in the sequential organization of speech sounds. *The Journal of the Acoustical Society of America*, *133*(6), 4177–87. doi:10.1121/1.4802745.