The Effects of Limited Attention Through Linguistic Stimuli on Auditory Ensemble Coding

Francis Indaheng [1], Steve Jiang [1], Mantej Panesar [1], Andy Peng [1]
[1]Department of Psychology, University of California, Berkeley

ABSTRACT

Language, which plays a major role in our everyday lives, is perceived and processed subconsciously with greater prominence than random noises. This is due to our familiarity and comprehension of words. Thus, hearing familiar words will divert our attention. Less known, however, is how this diverted attention, through linguistic stimuli, will affect the ability to ensemble code auditory percepts. Previous studies have shown that attention is not a notable factor in one's ability to ensemble code (Alvarez & Oliva, 2009; Joo et al., 2009) - at least visually. Nevertheless, the effects of diverting attention using familiar language remain in question. To answer this, we designed an experiment in which random sequences of tones and simple words (such as cat, dog, and tree) were presented in each trial. Then, subjects were asked to either identify whether a test tone was higher or lower than the average pitch of tones in the sequence or recall how many times a certain word was played. Our findings suggest that limited attention does not substantially affect one's ability to ensemble code as subjects performed slightly better on the ensemble coding task when asked to focus their attention on the words.

KEYWORDS

Auditory attention; auditory percepts, auditory system, ensemble coding; linguistic stimuli.

INTRODUCTION

Language holds a strong presence in our everyday lives. Compared to trivial noises, words can communicate complex thoughts, ideas, and expressions that are perceived and processed--even if done so unintentionally. Because of this unbiased, unavoidable processing, one cannot help but divert attention towards familiar linguistic sounds upon audition. In the past, ensemble coding has been shown to play a role in auditory processing; tones of different frequencies can be averaged from a set of tones (Piazza et al., 2013). Furthermore, previous studies have shown that visual ensemble coding is not affected by reduced attention (Alvarez & Oliva, 2009; Joo et al., 2009). However, the relationship between hearing a language familiar to the listener and the listener's ability to ensemble code remains in question. Specifically, our goal was to investigate the following question:

Does the integration of an interfering auditory percept, specifically linguistic stimuli, affect the human auditory system's ability to ensemble code?

To test this, we ran an auditory attention task, in which sequences of tones were presented along with linguistic stimuli ("dog," "cat," and "tree") dispersed within the trial. The subjects were then randomly given a task to either identify if a test tone is higher or lower than the mean tone of the presented sequence or answer how many times a certain word was played. Subsequently, we analyzed the subject's performance on the auditory ensemble task to determine whether the presence of interfering auditory percepts affect our ability to ensemble code.

METHODOLOGY

GENERAL EXPERIMENT - Subjects had four practice trials in which they were given an auditory ensemble (mean tone perception) task and four additional practice trials with a word-counting task. They were then given 192 experimental trials in which both tones and words were played. Before beginning each trial, subjects were instructed to focus on a specific stimulus type (either tones or words). In each experimental trial, the subjects were then given either an auditory ensemble task or a word-counting task. In order to study the effects of limited attention on auditory ensemble coding, only 50% of the tasks matched the stimulus type that subjects were instructed to focus on.

OBSERVERS - Four subjects (undergraduates and members of University of California, Berkeley) gave consent to participate. There were three females and one male, with the mean age being 20.75 years, and all had normal (or corrected-to-normal) visual and auditory acuity. They were each tested individually in an isolated, quiet room. All were fluent in English.

STIMULI - In 196 trials (4 practice and 192 experimental), a series of four tones were played consecutively and in a random order, with the tones being played for 300 milliseconds. These frequencies were generated using MATLAB's built-in Musical Instrument Digital Interface (MIDI). To convert from MIDI values to tone frequencies, we used the following equation:

$$Tone\_Frequency = 440*2^{((Midi\_Value-69)/12)}$$

Using this tone frequency, we constructed a playable sinusoidal wave using the following equation:

$$\sin(2*\pi* Tone\_Frequency * Tone\_Length)$$

Finally, an onset and offset ramp, in the form of cosine waves, were created to play smoother tones without static. The following equations were used:

$$Onset = (1 + 2 * \pi * Frequency\_Ramp * Ramp\_Vector ./ fs - \pi/2)/2$$
$$Offset = (1 + 2 * \pi * Frequency\_Ramp * Ramp\_Vector ./ fs + \pi/2)/2$$

Each random sequence of four tones was produced by randomly sampling without replacement from the tones ±1, ±3, and ±5 semitone distances away from the mean. The mean semitone was randomly chosen from the MIDI value range 48 (C3) to 72 (C5), inclusive.

A series of three audio files were recorded on an iPhone 6s (we did not have access to official audio files, given our time constraint and the license required), and these recordings were edited on Audacity (an audio editing program). We recorded three English words -- "dog," "cat," and "tree." We employed English words as the counterpart to the tonal stimuli due to the subjects' fluency and familiarity in processing English. This way, we could test whether the mechanisms that subconsciously control humans' natural language processing, thus making a significant difference in one's ability to ensemble code.

These audio files were converted to ".wav" files and then the stimuli were read into MATLAB and then played using Psychport Audio, a package in Psychtoolbox (a program compatible with MATLAB). In the 196 trials in which a sequence of four words were played (4 practice and 192 experimental), the sequence was generated by randomly sampling with replacement from the three words available ("dog," "cat," and "tree").

PROCEDURES - Subjects were prompted to input their demographic information including: their first name, last name, gender, and age. Next, they were presented with following instructions:

You will listen to various audio tones. Pay attention to the various auditory stimuli. At the end of each trial, you will be asked to make an input based on a question asked. Press enter to continue.

Before the experimental trials, the subjects were given eight practice trials. For the first four trials, subjects listened to a sequence of four tones and were asked to identify if a test tone was higher or lower than the mean tone of the sequence. These trials were followed by an additional four practice trials in which a sequence of four words was played and the subject was tasked with answering how many times a certain word was played during the sequence. During the remaining 192 trials, subjects listened to a sequence of four tones and four words played sequentially (tone, word, tone, word...). During these 192 trials, the subjects were directed to focus on one type of auditory stimuli (either the tones or the words): for a random 96 of the 192 trials they were directed to focus on the stimuli which the question would refer to, and for the other 96 trials they were wrongly directed to focus on the stimuli which the question would not refer to. After the stimuli were played, the subjects were prompted with one of two tasks:

1. Either press the "H" key if the test tone played was of higher frequency than the mean tone, or the "L" key if the outlier tone was of lower frequency than the mean tone.

2. Give a user input for how many times a certain word ("dog," "cat," or "tree") was played.

Subjects were given unlimited time to read and understand general instructions for the tasks. Moreover, to prevent any potential confounding variables and subjects building bias, we counterbalanced the distance between the test and mean tones, the task given to the subject, and the instructions telling the subjects what to focus on. The participants completed a total of 200 trials (8 practice trials and 192 experimental trials).

RESULTS

We constructed psychometric curves using a logistic regression equation [Fig. 1], plotting the percentage of higher responses (y-axis) as a function of distance between the test tone and the mean tone (x-axis). The subjects were able to identify the mean tone under limited attention condition (average slope = 0.6374; average point of subjective equality (PSE) = 0.8235). In fact, they did so with greater accuracy than under the full attention condition (average slope = 0.5922, average PSE = 0.8235) [Fig. 2].

We also compared the percentage accuracies across all four conditions [Fig. 3]. The subjects performed worse on average on the word counting task when attention was limited (81.77% with low attention versus 89.58% with full attention), but they performed better on average on the auditory ensemble condition when attention was limited (88.54% with limited attention versus 86.98% with full attention).

DISCUSSION

Here, our data suggests that the human auditory system is able to ensemble code despite reduced attention, even if a familiar linguistic stimuli acts as the distractor. This finding further bolsters previous research claiming that attention does not play a notable role in ensemble coding by examining how linguistic stimuli and limited attention can affect auditory ensemble coding. Interestingly, we found that subjects actually ensemble coded with greater accuracy when their attention was focused on the words, albeit by only a marginal percent (1.56%). We theorize that this occurs due to subject's calling upon their subconscious ensemble coding abilities, rather than giving too much thought and overthinking the tones. However, further testing is required to make any legitimate assumptions about why and if, at all, limited attention can actually increase the ability to ensemble code auditory percepts.

It should be noted that this pilot study was concluded with limited data (only four participants and 800 trials), and thus should be tested in greater magnitude in future studies.

ACKNOWLEDGEMENTS

This study was made possible by the Daniel D. Sonderegger Perception Science Internship, and the faculty of the Whitney Lab at UC Berkeley, who provided invaluable assistance with all steps of the research process. A special thanks to Dr. Allison Yamanashi Leib who believed in us every step of the way, and Jin Jeon for his consistent willingness to help.

REFERENCES

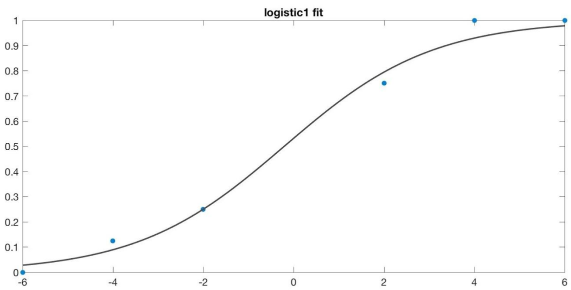
Alvarez G. A., Oliva A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proc. Natl. Acad. Sci. U.S.A.* 106 7345–7350. 10.1073/pnas.0808981106

Joo S. J., Shin K., Chong S. C., Blake R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *J. Vis.* 9 71–7.12. 10.1167/9.9.7

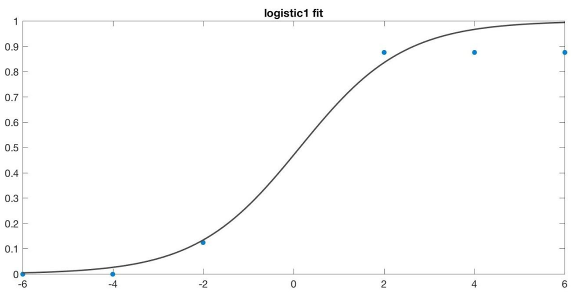
Piazza, Elise A., et al. “Humans Use Summary Statistics to Perceive Auditory Sequences.” Psychological Science, vol. 24, no. 8, Dec. 2013, pp. 1389–1397., doi:10.1177/0956797612473759.

FIGURES

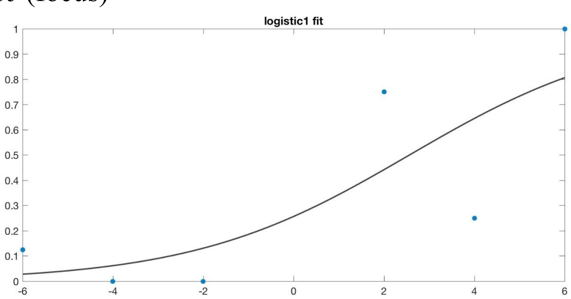
DP (focus)



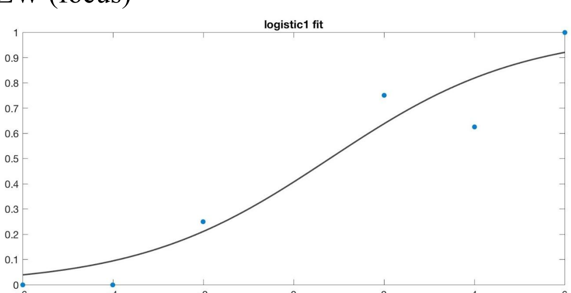
AD (focus)



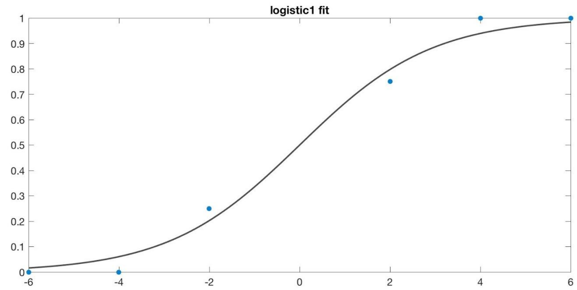
JJ (focus)



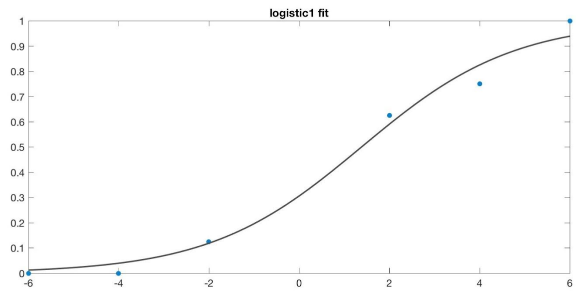
ZW (focus)



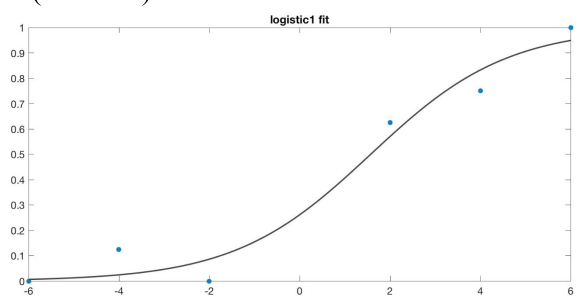
DP (no focus)



AD (no focus)



JJ (no focus)



ZW (no focus)

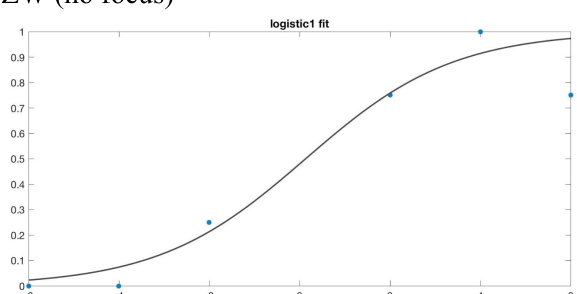


Figure 1

Psychometric curve analyses for all four subjects on both variants of the auditory ensemble perception task (one in which the subjects were told to focus on the tones and one in which the subjects were told to focus on the words). The number of "H" responses ("test tone is higher than mean tone") is plotted on the y-axis against the distance between the test tone and mean tone on the x-axis.

Initials	**Slope (focus)**	**PSE (focus)**	**Slope (no focus)**	**PSE (no focus)**
DP	0.6121	-0.2067	0.6852	0.0002
AD	0.8706	0.1372	0.5923	1.3824
JJ	0.4150	2.5629	0.6604	1.5730
ZW	0.4709	0.8004	0.6117	0.1331
Average	0.5922	0.8235	0.6374	0.7722

Figure 2

Data from the psychometric curve analysis for all four subjects. The second and fourth columns contain the slopes of the logistic curve fits for the auditory ensemble task condition in which subjects focused on the tones. The third and fifth columns contain the points of subjective equality (PSE) of the psychometric curves for the auditory ensemble task condition in which subjects focused on the words.

Initials	**Tones (focus)**	**Tones (no focus)**	**Words (focus)**	**Words (no focus)**
DP	0.8958	0.9167	0.8125	0.6250
AD	0.9167	0.8750	0.9167	0.8125
JJ	0.8125	0.8750	0.8750	0.8750
ZW	0.8542	0.8750	0.9792	0.9583
Average	0.8698	0.8854	0.8958	0.8177

Figure 3

Summary accuracy data for all four subjects. The second and third columns contain the accuracy on the auditory ensemble task for the tone-focus and word-focus conditions, respectively. The fourth and fifth columns contain the accuracy on the word counting task for the word-focus and tone-focus conditions, respectively.