The Biasing of Auditory Working Memory by a Set of Similar Distractors

Sagarika Alavilli

Committee:

Dr. Anastasia Kiyonaga, Dr. Tim Brady, Jonathan Keefe

Department of Cognitive Science

University of California, San Diego

2020 - 2021

Abstract

It has been demonstrated that ensemble statistics are collected in both visual and auditory working memory. Further incorporation of the ensemble average into the memory representation of individual items has been shown in visual working memory but not auditory working memory. Thus we presented subjects with a series of tones and looked to see if there was biasing of the representation for the first tone in the series towards the ensemble average of the series. When prompted on memory for individual tones, subjects did demonstrate a consistent bias towards the average pitch of the series. These results support the model that there is systematic biasing of memory for individual tones towards the ensemble average in auditory working memory. This suggests that both auditory and visual working memory integrate ensemble information into their representation for individual items.

Introduction

Our working memory allows us to temporarily hold and manipulate the items that we perceive. Oftentimes we may even process multiple items as groups (i.e., ensembles) within our working memory, rather than just individual objects.

When processing ensembles it has been shown that people often take averages of feature information called summary statistics or ensemble averages. In visual working memory this means being able to pull out the average of features such as size, shape, color, and facial expressions from an ensemble. For example, when presented with a set of circles of different sizes, people are able to extract the average circle size. Further, Brady & Alvarez (2011) demonstrates that when presented with a set of circles of different sizes, subjects' memory for individual circles is pulled towards the average circle size (i. e. biased). This demonstrates that summary statistics are not only collected but can potentially interfere with memory for items currently being held in working memory, suggesting that summary statistics can play an important role in our processing of ensembles.

Ensemble processing has also been demonstrated in auditory working memory (McDermott et. al., 2013). In Piazza et. al. (2013), subjects were presented with a series of 6 pure tones of varying pitches. Listeners were then presented with a test tone that they were to indicate was the average pitch of the series or not. The mean pitch was manipulated to test the subjects' ability to extract the ensemble average. These subjects consistently demonstrated an ability to extract and correctly identify the average pitch of the series played, despite the average pitch never being included in the ensemble. This demonstrates that extracting summary statistics such as ensemble average, is not unique to visual working memory and can also occur in auditory working memory.

However it is important to note that visual and auditory working memory do not appear to process ensembles in the same way. Visual working memory summates over a spatial scene, such as a set of circles presented on screen at once, whereas auditory working memory typically summates over fixed temporal windows, such as a series of tones (McWalter & McDermott, 2018). This difference is likely due to the lower-level perceptual differences between the two modalities. Localization is more intuitive and done earlier in visual processing, while it is more complex and typically done later in the auditory processing pathway.

Due to the importance of summary statistics in ensemble processing for visual working memory and the differences between visual and auditory working memory, we are interested in seeing if biasing for individual items by the ensemble average is present in auditory working memory. If this biasing is present in auditory working memory it would demonstrate an important structural similarity in representations between the two modalities. To test this question, we presented subjects with a tone that they were asked to remember over a delay. During this delay period, a series of irrelevant distractor tones were presented. Critically, the relationship between the average frequency of these distractor tones and the target was manipulated to test for the potential biasing of the subject's target representation in working memory. Following this delay, subjects were presented with two tones and asked to indicate which was the target: either the original target tone or an incorrect foil. If the ensemble average of the distractor tones biases memory for individual tones in auditory working memory similarly to visual working memory, then we expected to observe changes in memory performance based upon the congruence of the average frequency of the distractor tones and the foils. This finding would suggest that biasing of individual items by the ensemble average is present across modalities.

Methods

Subjects

50 participants (average age: 21.16 years; 29 female, 18 male, 1 nonbinary, 2 unreported gender) from the University of California, San Diego community were included in the final sample of the half hour-long online study and received course credit for their efforts. Each participant provided informed written consent prior to the study's start, as per University of California, San Diego Institutional Review Board guidelines. On average, listeners had 3.021 years of musical training (SD = 3.542) and 2 reported having perfect pitch or being able to name individual tones in the task. Prior to performing the main task, participants performed a brief task meant to check that they were using headphones (Woods et. al., 2017). Data from 18 subjects were excluded from the final sample for failing to pass this headphone check - answering fewer than 2/3rds of the headphone check trials correctly. Data from 21 participants were excluded for poor task performance (d' < 0.25).

Stimuli

Stimuli were pure tones generated in MATLAB along the logarithmic tone space and played through headphones. Target tones were selected from the range of C3 to B4 (130.81 - 498.88 Hz), and each tone was played with equal probability. The possible distractor tones ranged from G2 to E5 (98 - 659.25 Hz). For the spacing between tones we chose a logarithmic scale (i.e., semitone, or musical) to imitate the natural auditory discrimination of humans (Moore, 2003).

In each trial, the distractor tones were always chosen as being either ±5, ±3, or ±1 semitones from the initial target tone (Piazza et. al. 2013). This ensured that there were never any two tones that were an octave apart, eliminating the possibility of octave confusion. At the end of each trial the target was always repeated as a response choice along with the incorrect foil tone which was ± 2 semitones away from the target. Further, apart from the target, no two tones were ever repeated within a trial. In the baseline condition, distractor tones were always ±5 and ±3 semitones from the target. When the distractor average was lower than the target the distractor tones were -5, ±3, and +1 semitones from the target. On the other hand when the distractor average was greater than the target the distractor tones were +5, ±3, and -1 semitones from the target. The direction of the distractor tone average, the direction of the foil, and the order of the response choice tones were all counterbalanced across trials.

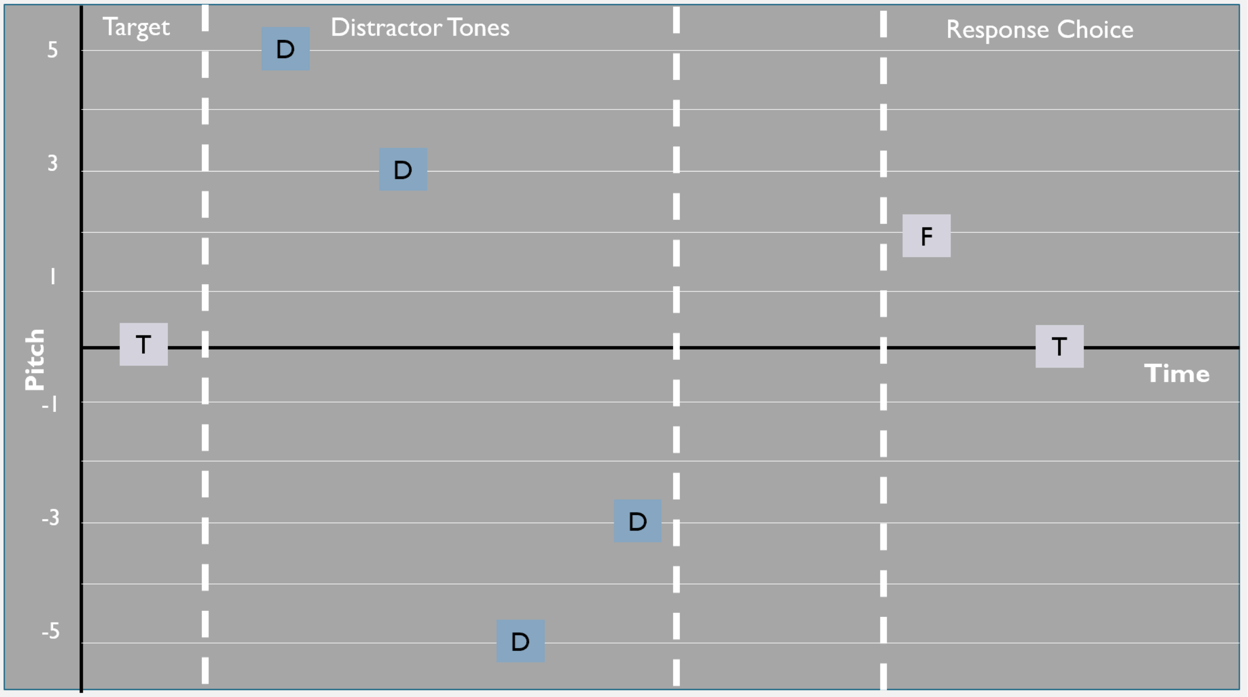

Figure 1. The above figure is a visual representation of a sample trial that is presented to subjects. The x -axis demonstrates the passage of time, while the y-axis represents the pitch, relative to the target, in semitones. Each trial is built around the target tone, with the distractors and foil positions varying based on the condition. Each rectangle in the figure represents a tone that is played. The purple rectangles represent tones that subjects are instructed to actively listen to, while the blue rectangles represent tones that listeners are instructed to ignore.

Procedure

Each trial was initiated by a button press and consisted of the target tone followed by 4 distractor tones. Subjects were instructed to remember the first (target) tone and ignore the distractor ensemble. They were then presented with a choice of two tones at the end of each trial, one of which was the target tone and the other of which was an incorrect foil. Subjects indicated which of the choice tones matched the initial target tone by pressing either the '"v" or "n" keys. All tones were played for 300 ms with a 2000 ms gap between the target and distractor tones and 300 ms between distractor tones. There was a 2000 ms pause between the last distractor and the first choice tone. The order of distractor tones was randomized in each trial. The direction of the distractor average and foil relative to the target was also counterbalanced across trials. Subjects completed a total of 4 blocks with 24 trials.

Analysis

Within each trial we manipulated both the direction of the average pitch of the distractor tone ensemble and foil relative to the initial target tone. The average pitch of the distractor tones could be either higher, lower, or equal to the target tone. The foil could be either higher or lower than the target.

In the baseline condition where the distractor average is equivalent to the target, the direction of the foil should not impact performance. We expect the lowest performance in the same direction condition where the distractor average and foil are in the same directions, which would include the condition where both the foil and distractor average are higher than the target tone and when both the foil and distractor average are lower than the target. If subjects are biased by the distractor ensemble average then their representation of the tone would be pulled closer to the foil, making subjects more likely to choose the foil than in the same average condition. This should lead to lower d’ in these trials. However, in the opposite direction conditions - where the foil is higher and the distractor average is lower or where the foil is lower and the distractor average is higher - we expect to see better performance than in the same average condition. This is because the foil is in the opposite direction of the distractor average in this condition, meaning that subjects should be less likely to select the foil over the target as the distractor average would bias their representation away from the foil.

We will be using d prime to measure performance for each condition. We chose d’ as it takes into account both hit rate (i.e., correct response rate) and false alarm rate (i.e., incorrect response rate) and can therefore serve as a more robust measure for accuracy. Similar to accuracy, a higher d’ indicates better performance.

Results

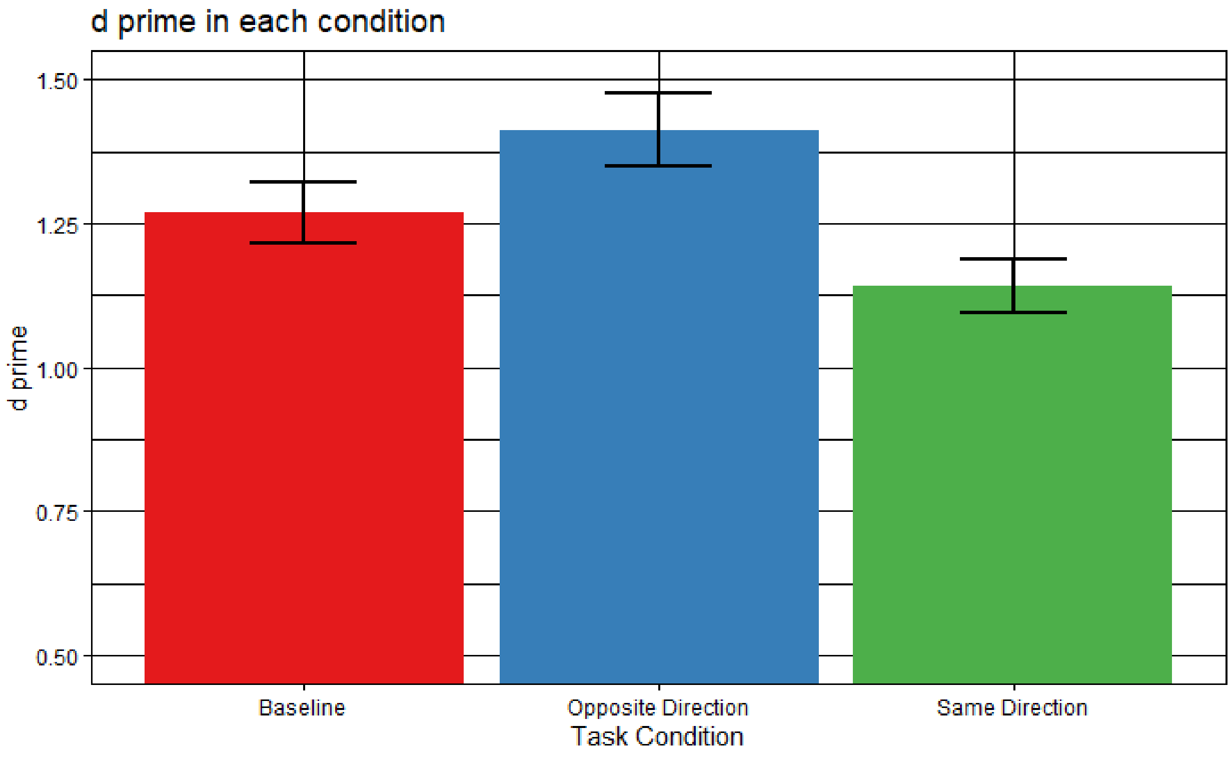

Figure 2. This figure plots the average d prime within subjects by each task condition. The error bars show the standard error of the mean. The y-axis of the begins at a d' of .50 so that any differences between conditions can be better seen. In this task, the highest performance was observed in the opposite direction condition, while the same direction condition show the lowest d prime. This pattern is consistent with the idea that if there is ensemble biasing subjects would be more likely to select the foil in the same direction condition decreasing performance. While in the opposite direction condition subjects would be less likely to select the foil resulting in the highest performance for this condition.

First we measured overall d prime performance for participants and found that participants were able to perform the task above chance (d' mean: 1.27, sd: 0.81). Any subject that performed with a d' below .25 was excluded, though most participants performed well above that.

To test the main hypothesis we compared performance across the 3 main within-subjects conditions of the task. In the baseline condition, where the distractor ensemble average was equivalent to the target, we collapsed across both foil directions. This was meant to be the baseline for comparison to the other conditions. Subjects had an average d' of 1.27 (sd = 0.37) in this condition. The next condition was the opposite direction condition, where the distractor ensemble average was in the opposite direction of the foil. This condition showed the highest average, with a d’ of 1.41 (sd = 0.45). These results are consistent with our hypothesis that if biasing is present subjects would be least likely to select the foil in this condition, thus having a higher d’. Finally, there was the same direction condition, where the ensemble average is in the same direction as the foil. In this condition we expected the lowest performance as biasing, if present, would make subjects much more likely to select the foil than in any other condition. As predicted, we see the lowest performance in this task, with an average d’ of 1.13 (sd = .33). As can be seen in figure 2 the pattern of results are as predicted if biasing by the ensemble average is present.

In order to test whether the observed differences between the conditions was statistically significant, we performed a one-way ANOVA with the distractor condition (baseline, same direction, opposite direction) as a factor. We found that there was a significant effect of the distractor condition upon performance, $F(2, 98) = 6.21$, $p = 0.003$, $\eta_p^2 = 0.113$. Follow-up t-tests demonstrated that there was a significant difference between the baseline condition and the same direction, $t(49) = 2.05$, $p = 0.046$, $d = 0.16$. There was also a significant difference between the same direction condition and the opposite direction condition, $t(49) = 3.41$, $p = 0.001$, $d = 0.34$. However, there was only a trend towards a significant difference in performance between the baseline and opposite direction conditions, $t(49) = 1.65$, $p = 0.10$, $d = 0.17$, although the results do still follow the expected pattern.

We also conducted a one-way ANOVA looking at response times with the distractor conditions as a factor. We found no significant effect for response times between the 3 distractor conditions, $F(98, 2) = .026$, $p = .974$, $\eta_p^2 = 0.001$. This indicates that the observed results are not a result of a speed-accuracy tradeoff. Overall the pattern of results seems highly consistent with the model of biasing by the ensemble average described.

Discussion

Overall, we have found that subjects are biased by the ensemble average when representing a set of tones in auditory working memory. When the ensemble average was in the same direction as the foil, subjects were much more likely to select the foil as the initial tone they heard. This supports the idea that the target representation is pulled towards the ensemble average, which in this condition is in the same direction as the foil. Thus subjects' internal representation of the target would be closer to the foil than the internal target tone, making it a more attractive option at the time of response. In the opposite condition, where the ensemble average is in the opposite direction of the foil, subjects were shown to select the foil less often. This result fits within the model to show that subjects' internal representation to the target is likely pulled towards the ensemble average, however this is away from the foil in this condition. Thus subjects are less likely to select the foil as it sounds further from their internal representation than the repeated target tone. The overall pattern of the relative highest performance occurring in the opposite direction condition and the lowest performance occurring in the same direction condition supports this model of biasing by the ensemble average described above.

These results are consistent with the visual working memory model of ensemble statistics as well. When presented with an ensemble in visual working memory, observers have been shown to collect and integrate ensemble statistics into their representation of individual items. For example when presented with a set of circles of different sizes, memory for individual circles will be biased towards the average circle size (Brady & Alvarez 2011). This aligns with the effect we see in our results where memory for individual pitches are biased towards the average pitch in a series of distractors. This similarity between the two models suggests that information is represented similarly following encoding across modalities. Similar to visual working memory, our results indicate that items in auditory working memory are not represented completely independently of each other. Along with individual representations, there is an ensemble (i.e., group) representation in auditory memory similar to visual working memory. Thus, it is likely that observers encode ensemble information (i.e. averages) along with individual information about items. Further, it seems that this ensemble information is important as it is collected and integrated into individual representations.

If ensemble information is taken into account, it raises the further question of when feature information is integrated. There is evidence from visual working memory literature that features are represented separately before being integrated later on in processing, resulting in a hierarchical effect (Brady & Alvarez, 2011). For example, when subjects are presented with a set of circles of different colors and sizes and prompted on their memory of an individual circle's size, they demonstrate both a biasing effect towards the average size of all circles as well as an effect towards the average size of circles of the same color. This hierarchical encoding effect has not been demonstrated in auditory working memory, however there has been evidence for representational grouping based on other features such as timbre and loudness (McDermott & Simoncelli, 2011; Semal & Demany, 1991; Popham et. al., 2018). As both visual and auditory working memory seem to integrate ensemble information into representations for individual items, it would be interesting to look at the impact of different features on this encoding. Based on that, a natural follow up would be to see if a similar hierarchical biasing effect can be present in auditory working memory.

One alternative explanation for this finding is that these results might be caused by a repulsion effect of the closest distractor tones rather than from biasing by the entire ensemble average. Previous visual working memory research has demonstrated that memory representations of very similar objects are encoded as being more different from each other than they actually are (i.e., repulsed; Chunharas et. al., 2019). For example, an orange item is reported as more red than it actually is when presented alongside a concurrently encoded yellow item. In the auditory modality, this would be akin to a target being reported as higher in pitch than it actually is when presented close in time to a similar tone of a slightly lower pitch. Though we think it unlikely, it is possible that this repulsion is responsible for the effect we observe. In the same and opposite conditions of the present study, the closest distractor tones were 1 semitone away from the target, which is within the range of highest interference (Deutsch, 1974). That means that in the same direction condition subjects could be selecting the foil because the semitone within the distractor ensemble is causing repulsion rather than because of a biasing effect from the overall average. In the opposite direction condition, this could similarly be a result of repulsion, as the closest tone causing repulsion would push the target representation in the same direction as if ensemble biasing were occurring. Thus subjects would be more likely to select the target due to their internal representation of the target being repulsed from the closest semitone within the distractor set. Based on the pattern described, the current experimental design is not able to differentiate the results of repulsion from that of biasing resulting in the same pattern for results for both. As both models would demonstrate the same pattern of results we observed, further exploration with a different paradigm would be needed to fully reconcile these ideas. However the biasing model is more consistent with our experiment as repulsion is typically shown with smaller set sizes (Chunharas et. al., 2019). Further, repulsion effects are primarily present when subjects are actively encoding the entire set, which while possible, is not instructed in our experiment. Therefore, we find this distractor repulsion account unlikely.

In conclusion we found that subjects' auditory working memory representations of items can be biased by the ensemble statistics of a set. This suggests that people integrate ensemble information into their representation for individual items. While this biasing has been previously demonstrated in visual working memory, the presence of this effect in auditory working memory indicates deeper representational similarities between the two modalities.

References

Brady, T., & Alvarez, G. (2011). Hierarchical Encoding in Visual Working Memory: Ensemble Statistics Bias Memory for Individual Items. *Psychological Science*, *22*(3), 384–392.

Chunharas, C., Rademaker, R. L., Brady, T. F., & Serences, J. (2019, February 4). Adaptive memory distortion in visual working memory. https://doi.org/10.31234/osf.io/e3m5a

Diana Deutsch (1974) Generality of interference by tonal stimuli in recognition memory for pitch, The Quarterly Journal of Experimental Psychology, 26:2, 229-234, DOI: 10.1080/14640747408400408

McDermott, J. H., & Simoncelli, E. P. (2011). Sound Texture Perception via Statistics of the Auditory Periphery: Evidence from Sound Synthesis. *Neuron*, *71*(5), 926–940. https://doi.org/10.1016/j.neuron.2011.06.032

McDermott, J., Schemitsch, M., & Simoncelli, E. (2013). Summary Statistics in Auditory Perception. *Nature Neuroscience*, *16*(4), 493–498.

McWalter, R., & McDermott, J. H. (2018). Adaptive and Selective Time Averaging of Auditory Scenes. *Current Biology*, *28*(9), 1405-1418.e10. https://doi.org/10.1016/j.cub.2018.03.049

Moore, B. C. J. (2003). An introduction to the psychology of hearing (5th ed.). San Diego, CA: Academic Press.

Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological science*, *24*(8), 1389–1397. https://doi.org/10.1177/0956797612473759

Popham, S., Boebinger, D., Ellis, D. P. W., Kawahara, H., & McDermott, J. H. (2018). Inharmonic speech reveals the role of harmonicity in the cocktail party problem. *Nature Communications*, *9*(1), 2122. https://doi.org/10.1038/s41467-018-04551-8

Semal, C., & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *The Journal of the Acoustical Society of America*, *89*, 2404–2410. https://doi.org/10.1121/1.400928

Woods, K., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, perception & psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2