The effect of regular rhythm on the perception of linguistic and non-linguistic auditory input

Outhmane Rassili[1,*], Mikhail Ordin[1,2,*]

[1]BCBL – Basque Centre on Cognition, Brain and Language, Mikeletegi 69, 20009, San Sebastián, Spain

[2]Ikerbasque – Basque Foundation for Science, Maria Diaz de Hro 3, 48013, Bilbao, Spain

[*]address for corresponding authors: Basque Centre on Cognition, Brain and Language, Paseo Mikeletegi 69, San Sebastian, 20009 Spain.
outhmanerassili@gmail.com
m.ordin@bcbl.eu

Keywords: rhythm, rhythmicity, isochrony, attentional cycles, rhythmic cognition

Abstract

Regular distribution of auditory stimuli over time can facilitate perception and attention. However, such effects have to date only been observed in separate studies using either linguistic or non-linguistic materials. This has made it difficult to compare the effects of rhythmic regularity on attention across domains. The current study was designed to provide an explicit within-subject comparison of reaction times and accuracy in an auditory target-detection task using sequences of regularly and irregularly distributed syllables (linguistic material) and environmental sounds (non-linguistic material). We explored how reaction times and accuracy were modulated by regular and irregular rhythms in a sound- (non-linguistic) and syllable-monitoring (linguistic) task performed by native Spanish speakers (N=25). Surprisingly, we did not observe that regular rhythm exerted a facilitatory effect on reaction times or accuracy. Further exploratory analysis showed that targets that appear later in sequences of syllables and sounds are identified more quickly. In late targets, reaction times in stimuli with a regular rhythm were lower than in stimuli with irregular rhythm for linguistic material, but not for non-linguistic material. The difference in reaction times on stimuli with regular and irregular rhythm for late targets was also larger for linguistic than for non-linguistic material. This suggests a modulatory effect of rhythm on linguistic stimuli only once the percept of temporal isochrony has been established. We suggest that temporal isochrony modulates attention to linguistic more than to non-linguistic stimuli because the human auditory system is tuned to process speech. The results, however, need to be further tested in confirmatory studies.

Introduction

The perceptual system does not process continuous sensory input equally at all times: some elements of the input are more attended than others (Landau & Fries, 2012). Attention samples the continuously changing environment in discrete chunks, which correspond to the periods of neural oscillations in the 4-8Hz frequency band (VanRullen, 2018). Perception thus operates on these chunks of sensory information, while the phase of neural oscillations modulates attentional intensity, and thus the probability of perceiving a certain element in the environment.

The central (Doelling, Arnal, Ghitza, et al., 2014; Ghitza, 2013) and peripheral (Greenberg & Ainsworth, 2004) auditory neural systems are also sensitive to rhythmic patterns in the environment. This sensitivity plays an important role in processing auditory information, including both linguistic and non-linguistic auditory input. Some segments in the input are better attended because their occurrence is predicted by a repetitive rhythmic pattern. Several theories (e.g. Dynamic Attentional Theory, Jones, 1976; The Attentional Bounce Hypothesis, Shields, McHugh, & Martin, 1974, discussed below) have been put forward to explain how these rhythmic patterns might modulate attention allocated for processing sensory input. Attentional rhythms (most likely based on neural oscillations; Fiebelkorn & Kastner, 2018; Haegens & Golumbic, 2019; Hickok et al., 2015; VanRullen, 2016) and environmental rhythms can become synchronized, such that more attentional resources are allocated to the expectancy cues provided by regularly occurring events in the environmental input (Obleser et al., 2017). Regular metrical patterns in environmental input facilitate establishing and maintaining synchronization between environmental and attentional oscillations, which may lead to entraining even spontaneously occurring neural oscillations to environmental rhythms. Neural rhythms sample the environment rhythmically, and the effect of environmental rhythms, which entrain neural rhythms and thus lead to resampling the world based on what the environment is like in particular circumstances, and drawing attention to more relevant aspects of the world at a particular moment. This, in turn, allows for faster and more accurate processing of continuous sensory inputs (Jones, 1976).

The coupling between neural and environmental rhythms can also enhance processing of auditory linguistic information which contains more regularly distributed salient acoustic events (stressed syllables, vowel onsets). Attention is drawn to stressed syllables more than their unstressed counterparts (Cutler, 1977; Cutler & Foss, 1976), and mispronounced phonemes are more likely to be perceived as deviant in stressed positions (Bond & Garnes, 1980; Cole & Jakimik, 1980). Shields et al. (1974) asked people to monitor a particular phoneme in connected speech. Target phonemes were detected faster and more reliably in stressed syllables than in unstressed syllables. Moreover, the facilitatory effect of stress was not observed when the same words containing the target phoneme were embedded in nonsense sentences. The authors suggested that stressed syllables in meaningful sentences are temporarily predictable and thus attract more attentional resources and facilitate target detection. In meaningless sentences, expectancy cues do not exist (participants cannot predict upcoming words and stressed syllables). Thus, it is the expectancy cues rather than acoustic correlates or the perceptual salience of stressed syllables that enhance phoneme detection in speech. Their results led to the Attentional Bounce Hypothesis: attention locks onto the quasi-isochronous distribution of stressed syllables and moves from one stressed syllable to another (Shields et al., 1974). The longer the preceding rhythmic pattern leading to the target, the better the percept of temporal isochrony is established, and the stronger expectancy and its facilitatory effect (Pitt & Samuel, 1990). In line with this, Ordin et al. (2019) showed that in AX discrimination experiments, regular rhythm in the A stimulus (first stimulus in a stimulus pair) led to faster and more accurate responses, regardless of whether the X stimulus (second stimulus in a stimulus pair) was rhythmically similar to or different from the A stimulus. They suggested that rhythmic regularity in the first stimulus enhances attention and thus results in better performance.

A succession of stressed syllables creates a metrical grid that is used to facilitate speech processing. The perceptual system can rely on this grid to predict the occurrence of the next stressed syllable, allocating more resources to these syllables so as to process input more efficiently. In addition to metrical expectancy, which relies on the metrical grid created by the occurrence and predictability of stressed syllables, Quené and Port (2005) explored the effect of timing expectancy on the processing of linguistic input. Participants performed a phoneme monitoring task while listening to a sequence of isolated words. Timing expectancy was manipulated by variation in the inter-word duration: regular inter-word intervals provided stronger expectancy cues for word onsets than variable inter-word intervals. Metrical expectancy was achieved by modifying the stress patterns in the preceding word sequence. A similar stress pattern (iambic or trochaic) across all the words in a sequence allowed for expectancy cues to emerge, whereas varying stress patterns within a sequence disrupted metrical expectations. For timing expectancy, the resulting reaction times were shorter in the regular than in the irregular condition, showing a clear timing expectancy effect when inter-word durations were constant. On the other hand, metrical expectancy did not exert an effect on reaction times. These results suggest that speech perception may be more affected by timing than by metrical patterns.

Although the facilitatory effect of regularity and expectancy on auditory perception has been observed on linguistic (Pitt & Samuel, 1990; Quené & Port, 2005; Shields et al., 1974) and non-linguistic (Jones, 1976) materials, there are important differences in how linguistic and non-linguistic sounds are processed (Warren, Obusek, Farmer et al., 1969). Speech is segmented into phonemes which belong to classes defined by fine-grained boundaries, while non-linguistic sounds are not (Hendrickson, Walenski, Friend, et al., 2015). Humans engage in attentive processing of linguistic sounds in contrast to non-linguistic sounds (Warren et al., 1969). Even if the latter are heard on a daily basis, they are mostly processed unconsciously by the auditory system; humans tend to filter out and ignore passive sounds they are accustomed to hearing constantly, but which hold no significance for them (e.g., birds chirping outside the house).

The human auditory system may be better honed for processing linguistic than non-linguistic acoustic material. Thus, the magnitude of the expectancy effect on attentional rhythms might vary for linguistic and non-linguistic material. The current study provides an explicit within-subject comparison of reaction times and accuracy in an auditory target-detection task in a sequence of regularly and irregularly distributed syllables (linguistic material) and environmental sounds (non-linguistic material). We hypothesized that rhythmic expectancy modulates attention to linguistic more than to non-linguistic stimuli. To test this hypothesis, we set up a syllable-monitoring (linguistic sounds) and sound-monitoring (non-linguistic sounds) task for 25 native Spanish participants. They were instructed to listen for auditory targets within a series of auditory sequences, characterized by either regular or irregular inter-stimulus intervals.

Methodology

Participants

Twenty-five native Spanish speakers (mean age = 22.72 years, median age = 22 years; 17 women) were recruited. None of them reported any speech or hearing problems. The experiment was approved by the BCBL ethical review board. All subjects signed written consent in accordance with the Declaration of Helsinki and received €8 as financial compensation.

Material

The linguistic material included 25 distinct consonant-vowel syllables recorded by a native Spanish speaker. The syllables were composed of 5 consonants (l, m, n, s, f) and 5 vowels (a, o, i, e, u), all of which exist in Spanish. We avoided plosive consonants because their durations cannot be manipulated without compromising naturalness. Each syllable was recorded in a separate audio file. The durations of consonants and vowels were manipulated such that all consonants lasted 100ms and all vowels lasted 200ms (300ms per syllable). All of the audio files were equalized in terms of their average level of intensity (with the upper threshold set to 80dB), to ensure that none of the syllables in *relatively* louder (i.e., has relatively higher intensity level) than other syllables.

The non-linguistic stimuli comprised 25 distinct transient sounds (e.g. a drop of water). These sounds were carefully picked to differ in terms of spectral characteristics and to be clearly and easily discriminated by the participants. These sounds were equalized in terms of duration (300ms) and average intensity (with an upper threshold of 80dB).

We created sequences of 8 syllables or sounds to be used as experimental trials. Each sequence started and finished with a 140ms period of silence. We counterbalanced sounds/syllables across sequences. The inter-syllable/sound intervals were manipulated to generate rhythmic differences: constant inter-syllable/sound durations for regular rhythm and jittered inter-syllable/sound durations for irregular rhythm stimuli. In the regular rhythm condition, all 7 inter-syllable/sound intervals were constant at 140 ms. For the irregular condition, we used seven interval values: 50ms, 80ms, 110ms, 140ms, 170ms, 200ms, 230ms. Within each stimulus, these values occurred in randomized order, with each value used only once, such that the total duration of each trial sequence was 3660ms. In total, 300 stimuli were created, 75 per each of the 4 conditions defined by rhythm type (regular and irregular) and stimulus type (linguistic and non-linguistic).

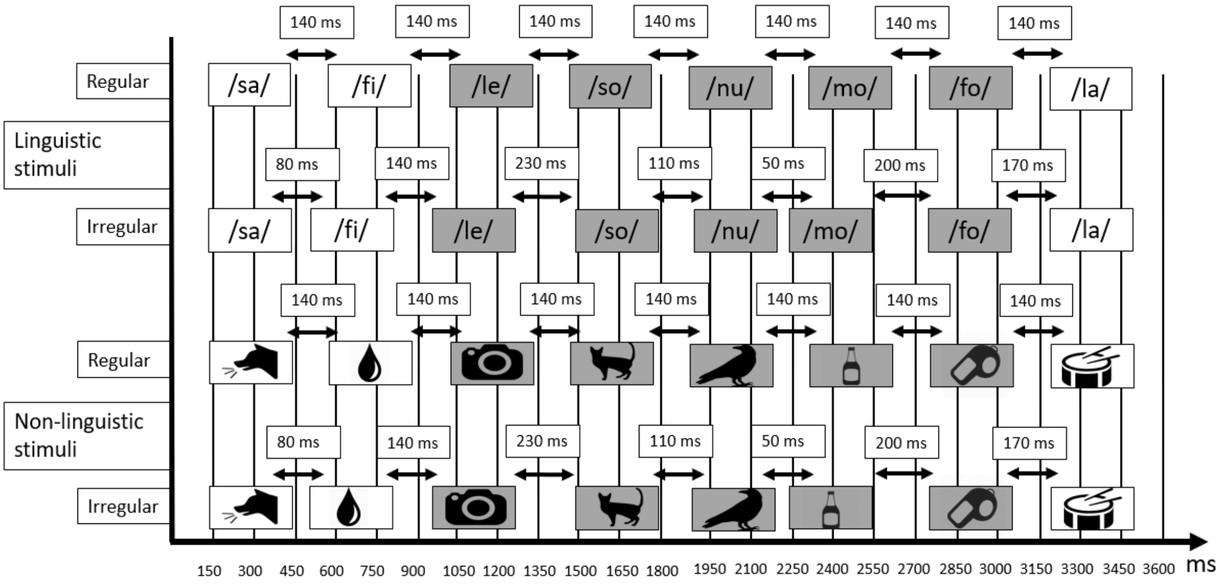

Figure 1: *Schematic overview of stimuli in the regular and irregular rhythm stimuli. Each box represents a syllable/sound. Possible positions of target syllables/sounds are marked by a darker colour.*

Procedure

The experiment consisted of a syllable- (for linguistic stimuli) and sound- (for non-linguistic stimuli) monitoring task. The participants were seated in front of a screen in a soundproof room. The stimuli were presented via headphones in PsychoPy.

On each trial, participants heard the sentence “Now listen for X” (where X stands for the specific target sound or syllable used); after a 1-second pause a sequence of syllables or sounds was played. On each trial, a target was embedded in the presented sequence. The target positions varied between the third, fourth, fifth, sixth, and seventh positions within the sequence. (**Figure 1**). Syllables/sounds were counterbalanced to appear as targets the same number of times in each position. The targets and their positions within the sequences were counterbalanced across trials, with every single syllable/sound and position being selected three and fifteen times respectively in each condition.

Participants were instructed to press the button on a response box as soon as they heard the specified target sound/syllable. The trial was interrupted when the button was pressed. The participants manually initiated every trial at their own discretion. Nothing was presented on the screen when the sounds were being played. The order of trials with linguistic and non-linguistic material and with regular and irregular rhythms, was randomized for each participant. Prior to the experiment, 4 practice trials were initiated as a training session, in order to familiarize participants with the task and to allow them to adjust the volume to a comfortable level. Volume adjustment changed the loudness of the stream overall, but due to the intensity normalization procedure the *relative* loudness of the sounds/syllables in sequences was kept constant.

Results

Reaction time was calculated as the delay between the onset of the target syllable/sound and the time when the participant pressed the button to signal that the target had been detected. Listeners failed to detect the targets 72 times (0.96% of all trials lacked responses; a response was considered to be missing if it was not given within 1000ms after the end of a sequence). On 212 trials (2.82% of all trials), a response was given before the target was presented. Missing responses and premature responses amounted to a total of 284 errors (3.7%) trials in the entire experiment. The number of errors across rhythm types was the same. In the linguistic stimuli, there were exactly 43 errors for both the regular and irregular rhythms, while in the non-linguistic stimuli, there were exactly 99 errors for both the regular and irregular rhythms (**Table 1**).

Table 1. *Errors categorized by type, rhythm, and stimuli*

Errors in linguistic stimuli				**Errors in non-linguistic stimuli**			
Regular		**Irregular**		**Regular**		**Irregular**	
43		43		99		99	
Early Response	**No Response**	**Early Response**	**No Response**	**Early Response**	**No Response**	**Early Response**	**No Response**
31	12	36	7	75	24	70	29

For premature responses, the effect of *stimulus type* (linguistic vs. non-linguistic) on the number of errors was significant and strong, $F(1.24)=8.977$, $p=.006$, $\eta_p^2=.272$, while the effect of *rhythm* (regular vs. irregular), $F(1,24)<.0005$, $p=1.0$, $\eta_p^2<.0005$, and the interaction between *rhythm* and *stimulus type*, $F(1,24)=.623$, $p=.438$, $\eta_p^2=.025$ were not significant. For missing responses, the effect of *stimulus type* was significant and strong, $F(1.24)=11.352$, $p=.003$, $\eta_p^2=.321$, while the effect of *rhythm*, $F(1,24)<.0005$, $p=1.0$, $\eta_p^2<.0005$, and the interaction between *rhythm* and *stimulus type*, $F(1,24)=1.263$, $p=.272$, $\eta_p^2=.05$ were not significant. This reveals a significantly and substantially larger number of errors on non-linguistic than on linguistic material, and this pattern is not modulated by the rhythm implemented in the stimuli.

Data screening was performed on remaining trials to detect the outlying RT values defined as the values exceeding 2SD from the mean in each rhythm*stimuli combination (22 trials or 0.3% of all trials). Including trials discarded due to errors, this brings the total percentage of discarded trials to 4% (306 out of 7.500 trials).

The remaining reaction time values were averaged for each participant per rhythm (regular/irregular) and stimulus type (linguistic/non-linguistic) and included in the statistical analysis. A repeated measures ANOVA with *rhythm* (regular vs. irregular) and *stimulus type* (linguistic vs. non-linguistic) as factors did not reveal a significant effect of *rhythm*, F(1,24)=.542, p=.508, η_p^2=.018 or *stimulus type*, F(1,24)=3.577, p=.071, η_p^2=.13. The interaction between *rhythm* and *stimulus type* was also not significant, F(1,24)=.386, p=.368, η_p^2=.015. Thus, the data does not confirm the original hypothesis: we did not find evidence that regular rhythm modulates performance more on linguistic than non-linguistic stimuli. Moreover, to our surprise, we did not observe any effect of rhythm type: the data did not provide evidence that metrical regularity facilitated performance on monitoring tasks. Therefore, we decided to explore the data further in an exploratory study.

We expected to find a significant effect of rhythm based on earlier studies and theoretical assumptions. The absence of this effect might indicate that it takes time to establish the percept of isochrony between syllable or sound onsets. This would result in either a weak or non-existent effect on targets closer to the beginning of the trial sequences. Therefore, we decided to explore the effect of *rhythm* in early and late positions. The third and fourth positions were considered the *early target positions*, while the fifth, sixth, and seventh positions were considered *late target positions.* We calculated the average reaction time for each participant separately for *early* and *late* positions per stimulus type and rhythm (**Figure 2**), and introduced *position* (early vs. late) as a factor in the model (in addition to the effects of *stimulus type* and *rhythm)*. The analysis revealed a significant and substantial effect of *position*, F(1,24)=166.24, p<.0005, η_p^2=.874. Targets in late positions were detected with shorter reaction times than targets in early positions (**Table 2)**. However, the effect of *stimulus type*, F(1,24)=2.994, p=.096, η_p^2=.111, and the effect of *rhythm*, F(1,24)=.919, p=.347, η_p^2=.037 were not significant. None of the interactions were significant.

Table 2. *Averaged reaction time per stimuli type, rhythm, and position*

Condition	Mean (in ms)	SD	95% confidence interval around M
Linguistic Irregular Early Position	.4928	.044	.48:.51
Linguistic Irregular Late Position	.4446	.033	.43:.46
Linguistic Regular Early Position	.5054	.044	.49:.52
Linguistic Regular Late Position	.4353	.035	.42:.45
Non-Linguistic Irregular Early Position	.508	.065	.48:.54
Non-Linguistic Irregular Late Position	.4483	.052	.43:.47
Non-Linguistic Regular Early Position	.5122	.071	.48:.54
Non-Linguistic Regular Late Position	.4529	.039	.44:.47

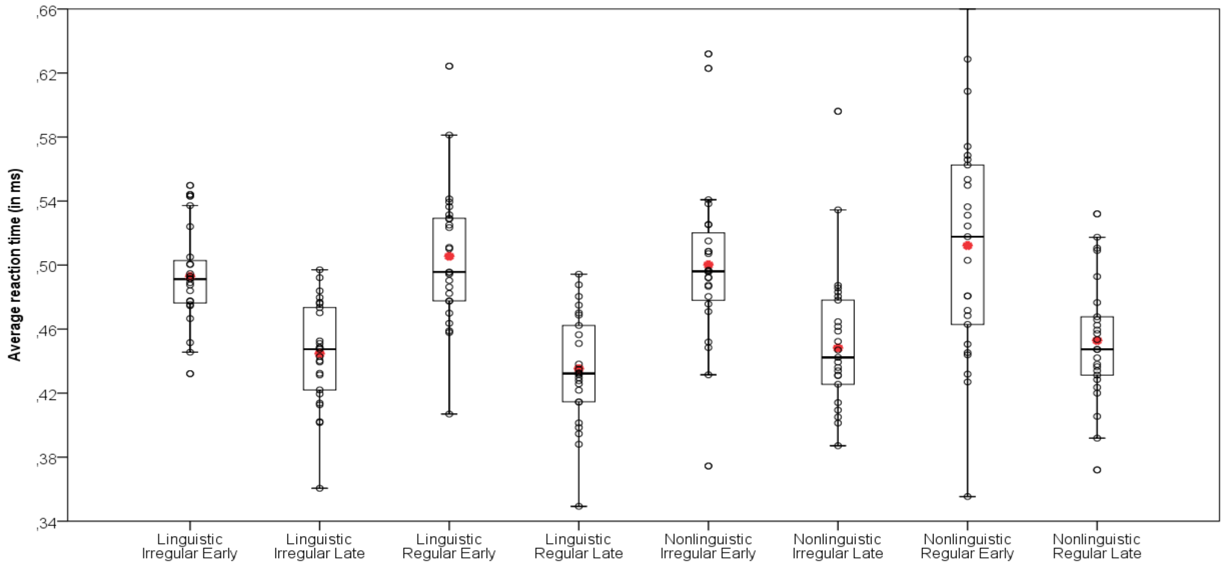

Figure 2. *Average reaction times categorized by stimulus type (linguistic or non-linguistic), rhythm (regular or irregular), and position (early or late). Each circle stands for the data from an individual participant, red dots represent the means and horizontal lines represent the medians. Boxes contain 50% of the datapoints, downward and upward whiskers each span 25% of datapoints.*

Following the pattern, we compared the effect of *rhythm* on reaction times for linguistic and non-linguistic stimuli, separately, using paired t-tests (two-tailed, assumption of normality verified by the Kolmogorov-Smirnov tests, which showed no significant deviations in the data distribution from the normal distribution in any sample, all p values >.9). On linguistic material, targets in late positions were identified faster in the regular than the irregular condition, t(24)=2.49, p=.02 (corrected by the Bonferroni method), mean difference M=.009(SD=0.19), SE=.004, 95%CI of the difference [.002: .017], d=.28. On non-linguistic stimuli, however, the effect of rhythm in late positions was not significant, t(24)=.688, p=.498, M=.005(0.34), SE=.007, 95%CI of the difference [-.019: .001], d=.1. To directly test the hypothesis that the difference in RTs in late position between regular and irregular conditions was larger for linguistic than non-linguistic material, we ran a paired t-test, t(24)=1.958, p=.031, d=.506. This pattern suggests that the effect can only be observed on linguistic stimuli, and at least four inter-syllable intervals are required for the effect of isochrony to emerge – possibly because it takes this long for the percept of regularity to become established.

To further understand this relationship, we ran Spearman's correlations between the positional order when the target was detected (all positions were considered, without splitting them into bins) and reaction times, separately for linguistic regular (ρ=-.635), linguistic irregular ρ=-.535), non-linguistic regular (ρ=-.489) and non-linguistic irregular (ρ=-.459) stimuli. The fact that all these correlations were negative shows that reaction time decreases as ordinal position increases (people are faster towards the end of a sequence). However, the only significant difference between the correlation coefficients was between correlations for regular linguistic and regular non-linguistic stimuli, z=1.7, p=.045 (2-tailed). This shows that regularity is associated with a sharper decline in reaction times in linguistic stimuli than non-linguistic stimuli as a sequence progresses (and the ordinal number of the target position increases). This is in line with the earlier conclusion that regularity has a stronger effect on linguistic than on non-linguistic materials. We did not observe differences in correlation strengths in other tests (Table 3).

Table 3. *Comparing strengths of non-parametric correlations between the ordinal position of the target and reaction times in stimuli with isochronous (regular) and non-isochronous (irregular) distribution of speech-like syllables (linguistic stimuli) and non-linguistic sounds (non-linguistic stimuli).*

Condition	z	p	Interpretation
Linguistic Regular vs. Linguistic Irregular	1.192	.117	Tests linguistic materials to ascertain if regular rhythm is associated with a sharper decline in reaction times than irregular rhythm
Non-Linguistic Regular vs. Non-Linguistic Irregular	.282	.389	Tests non-linguistic materials to ascertain if regular rhythm is associated with a sharper decline in reaction times than irregular rhythm
Linguistic Regular vs. Non-Linguistic Regular	1.7	.045	Tests whether regular rhythm is associated with a sharper decline in reaction times for linguistic compared to non-linguistic stimuli
Linguistic Irregular vs. Non-Linguistic Irregular	.79	.215	Tests if irregular rhythm is associated with a sharper decline in reaction times for linguistic compared to non-linguistic stimuli

Discussion

Surprisingly, we did not observe any effect of temporal regularity on performance (accuracy and reaction times) in the sound-monitoring task, which we used as a proxy for online attention. Our exploratory analysis showed that targets towards the end of the sequences were detected faster. The way we randomized inter-sound/syllable intervals might explain why we observed better performance both on regular and irregular stimuli in late positions. To create irregularity, we used a set of 7 possible values, ranging from 50ms to 230ms (with 30-ms steps) as inter-syllable and inter-sound intervals, with each value used only once per sequence. Consequently, participants might have figured out that the inter-sound intervals were not completely random, since with each passing sound, the degrees of freedom for the remaining values reduced, increasing the predictability of the duration of the next inter-sound interval. That is, at the beginning of any sequence, there are seven possible interval durations, but for late positions, only three possible durations remain; if participants kept track of durations that had already occurred as the sequence progressed, they could better predict the onset of the following sound/syllable, and prepare to make a behavioral response if a target was detected. This predictive effect might mask any effect of rhythmic regularity, if the latter only emerges in late positions. In early positions, a regularity effect has not yet been established in the regular condition, nor has a predictability effect been established in the irregular condition, thus explaining why no significant differences between conditions or material types were observed. A similar phenomenon has also been reported in the visual perceptual modality, when targets were embedded in a sequence of isochronously and non-isochronously presented sequences of images (Coull, 2009). Reaction times decreased over time because the conditional probability that a target would appear given that it has not yet appeared increases over time, thus alerting participants' attention towards the end of the sequence. In future experiments, the position of the target could be kept constant, while manipulating the presence vs. absence of the target. This would control for differences in attention at the beginning and the end of the sequences due to the increase of the conditional probability that the target will occur, given that it has not yet occurred.

Further tests showed that in late positions, targets were detected faster when syllables – in linguistic stimuli – were distributed with isochronous inter-syllable intervals. Importantly, significant differences in reaction times for detecting target sounds – in non-linguistic material – were not observed. The difference in reaction times between regular and irregular conditions for linguistic stimuli was significantly bigger than the difference in reaction times between regular and irregular conditions for non-linguistic stimuli. This suggests that a modulatory effect of regularity can only be observed on ***linguistic*** material ***later*** in sequences (which is in line with Pitt & Samuel, 1990, who observed a stronger facilitatory effect of regular rhythm in longer than in shorter sequences). However, due to the exploratory nature of this analysis in our study, a confirmatory study using a more targeted design should be conducted.

It is not clear why we found no effect of regularity on non-linguistic material even in late positions. We can suggest several factors that could interfere with the modulatory effect of rhythm on the sound monitoring task. Individual differences in familiarity with sounds could be an interfering factor; some participants might have experienced more exposure to certain sounds than others. Another factor might be the problem of classifying non-linguistic sounds. Such sounds might not be concretely classified in the mental lexicon, and thus make it more difficult to exactly determine their identity, resulting in semantic misinterpretations (Hendrickson et al., 2015). Also, humans interact with linguistic and non-linguistic sounds differently. They allocate more attention to sounds which are deemed more important; linguistic sounds are likely to be accorded more importance as they are a means for communication.

Spontaneous neural oscillations lead to rhythmic sampling of the world (VanRullen, 2016; Fiebelkorn & Kastner, 2018). However, acoustic rhythms in environmental input can modulate excitability patterns in the auditory system, entraining neural rhythms to environmental rhythms (Hickok et al., 2015; Lakatos et al., 2019; Obleser et al., 2017). This entrainment could change how the world is sampled, and what aspects of the environmental stimuli are zoomed in on and selected for attention. Because of the functional importance of speech, it is also possible that speech-like stimuli lead to stronger modulatory effects on neural oscillations, producing better entrainment of neural oscillations than non-linguistic stimuli. If the initial non-linguistic sound in a regular condition in our experiment was aligned with a phase of the perceptual cycle where the probability for perception was low, the following sounds in the sequence, which are distributed at equal temporal intervals, might also have been aligned with that same phase, where they were unlikely to be perceived (VanRullen, 2016). By contrast, syllables, which are quickly recognized as functionally important linguistic inputs, could lead to a more rapid re-adjustment of the phase of neural oscillators. Even in cases where initial syllables in a sequence were aligned with a phase in the perceptual cycle that afforded only a low probability for perception, syllables later in the sequence could become aligned with a phase with a higher probability of perception. Slower phase resets for non-linguistic stimuli and faster phase resets for syllables might explain why we found better task performance for linguistic than non-linguistic material towards the end of the sequence. This interpretation, however, calls for additional empirical testing.

Whether the effect of linguistic stimuli would be observed in the visual modality presents an interesting question for further investigation. As an evolutionarily ancient phenomenon, spoken language might involve finely tuned perceptual and cognitive mechanisms specifically adapted for speech processing and comprehension. Regular rhythm (in the auditory modality) not only affects auditory processes related to speech processing but also higher-level mechanisms involved in language comprehension, including lexico-semantic integration (Rothermich, Schmidt-Kassow, & Kotz, 2012). Writing is a relatively recent cultural innovation and, in the visual unlike the auditory modality, linguistic stimuli might not exert a stronger effect on rhythmic attention than non-linguistic stimuli. This hypothesis, if verified, might throw light on how the speech and language faculty influence general cognitive processes in humans.

Acknowledgements

This research was supported by the Spanish State Research Agency through BCBL's Severo Ochoa excellence accreditation SEV-2015-0490. The authors are thankful to Prof. Arthur Samuel (Stony Brook University, USA) and to Dr. Efthymia Kapnoula, Dr.

Marina Kalashnikova, and Dr. Antje Stoehr (BCBL, Spain), and also thank the reviewers for useful comments, fruitful discussions, and valuable suggestions.

Competing Interests

The authors have no competing interests to declare.

Authors contributions

OR and MO conceived the study, OR prepared and ran the experiment and analyzed the data, OR and MO interpreted the results and wrote the manuscript.

Data accessibility

Data will be made available on the figshare repository.

References

Ballas, J., & Howard, J. (1987). Interpreting the Language of Environmental Sounds. *Environment and Behavior 19(1)*, 91-114. doi: 10.1177/0013916587191005

Bond, Z.S., & Garnes, S. (1980) Misperceptions of fluent speech. In Cole, R.A. (Eds). P*erception and production of fluent speech* (pp. 115–132.). Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 115–132.

Bregman, A. (1978). The formation of auditory streams.*Attention and Performance 7*, 63-76.

Cole, R., & Jakimik, J. (1980). How are syllables used to recognize words? *The Journal of the Acoustical Society of America* 67, 965-70. DOI: 10.1121/1.383939.

Coull, J.T. (2009) Neural Substrates of Mounting Temporal Expectation. PLoS Biol 7(8): e1000166. doi:10.1371/journal.pbio.1000166

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics* 2*0*(1), 55-60. doi: 10.3758/bf03198706

Cutler, A., & Foss, D. (1977). On the Role of Sentence Stress in Sentence Processing. *Language and Speech 20*(1), 1-10. doi: 10.1177/002383097702000101

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage* 85(2), 761-768. DOI: 10.1016/j.neuroimage.2013.06.035

Fiebelkorn, I.C., & Kastner, S. (2018). A Rhythmic Theory of Attention. *Trends in Cognitive Science 23(2)*, 87-101.

Ghitza, O. **(**2013). The theta-syllable: a unit of speech information defined by cortical function. *Frontiers in Psychology 4:138*. doi: 10.3389/fpsyg.2013.00138

Greenberg, S., & Ainsworth, W. (2004). Speech Processing in the Auditory System: An Overview. In Greenberg, Steven, Ainsworth, William A., Fay, Richard R. (Eds.) Speech Processing in the Auditory System. (pp. 1-62). Berlin: Springer Verlag.

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. Neuroscience & Biobehavioral Reviews, 86, 150–165.

Hendrickson, K., Walenski, M., Friend, M., & Love, T. (2015). The organization of words and environmental sounds in memory. *Neuropsychologia 69*, 67–76. DOI: 10.1016/j.neuropsychologia.2015.01.035

Hickok, G., Farahbod, H., & Saberi, K. (2015). The Rhythm of Perception. Psychological Science, 26(7), 1006–1013.

Jones, M. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review 83*(5), 323-355. DOI: 10.1037//0033-295x.83.5.323

Lakatos, P., Gross, J., & Thut, G. (2019). A New Unifying Account of the Roles of Neuronal Entrainment. Current Biology 29(18), R890-R905.

Landau, A.N., & Fries, P. (2012). Attention Samples Stimuli Rhythmically. *Current Biology 22*, 1000–1004.

Obleser J, Henry M., & Lakatos P (2017) What do we talk about when we talk about rhythm? PLoS Biol 15(9): e2002794.

Ordin, M., Polyanskaya, L., Gomez, D., & Samuel, A., (2019). The role of native language and the fundamental design of the auditory system in detecting rhythm changes. *Journal of Speech, Language and Hearing Research*, 62(4):835-852.

Pitt, M., & Samuel, A. (1990). The use of rhythm in attending to speech. *Journal Of Experimental Psychology: Human Perception And Performance 16(3)*, 564-573.

Quené, H., & Port, R. (2005). Effects of Timing Regularity and Metrical Expectancy on Spoken-Word Perception. *Phonetica 62*(1), 1-13. doi: 10.1159/000087222

Rothermich, K., Schmidt-Kassow, M., Kotz, S. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia 50(2)*, 232-244. Doi: 10.1016/j.neuropsychologia.2011.10.025

Shields, J., McHugh, A., & Martin, J. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal Of Experimental Psychology 102*(2), 250-255. doi: 10.1037/h0035855

VanRullen, R. (2016). Perceptual Cycles. *Trends in Cognitive Science 20(10)*, 723-735.

VanRullen, R. (2018). Perceptual Cycles. *Neuron 99*, 632-634.

Warren, R., Obusek, C., Farmer, R., & Warren, R. (1969). Auditory Sequence: Confusion of Patterns Other Than Speech or Music. *Science 164*(3879), 586-587. doi: 10.1126/science.164.3879.586.

FIGURE CAPTIONS

Figure 1: Schematic overview of stimuli in the regular and irregular rhythm stimuli. Each box represents a syllable/sound. Possible positions of target syllables/sounds are marked by a darker color.

Figure 2. Average reaction times categorized by stimulus type (linguistic or non-linguistic), rhythm (regular or irregular), and position (early or late). Each circle stands for the data from an individual participant, red dots represent the means and horizontal lines represent the medians. Boxes contain 50% of the datapoints, downward and upward whiskers each span 25% of datapoints.