Lexical stress in Spanish word segmentation

Alvaro Iturralde Zurita[1], *Meghan Clayards*[12]

[1] School of Communication Sciences and Disorders, McGill University, Canada
[2] Department of Linguistics, McGill University, Canada

`alvaro.iturraldezurita@mail.mcgill.ca, meghan.clayards@mcgill.ca`

Abstract

We explored the role of lexical stress in word segmentation as speech unfolds in time. We tested participants online with a Mouse Tracking listening experiment using temporarily ambiguous phrase pairs of the form "PAlo marron" vs. "paLOma roja". These pairs were segmentally ambiguous in the first three syllables but differed in the location of lexical stress. Thus, use of stress cues would allow participants to disambiguate the phrases more quickly. We also manipulated the presence of lexical stress correlates in two conditions (stress natural and stress neutral) and found that lexical stress has an early impact in Spanish word segmentation that can affect how quickly and efficiently the speech signal is processed.

Index Terms: word segmentation, lexical stress, speech perception.

1.   Introduction

Grouping sounds into meaningful units and being able to identify word boundaries is one key part of speech perception. This process, called word segmentation, is the cognitive process that draws imaginary lines to separate sounds belonging to one word from those belonging to the next or previous one. Previous studies on this phenomenon have shown that there are multiple types of information that can work as cues telling us where to draw word boundaries [1, 2, 3]. The way these cues are used and how important they are can change from language to language [4, 5, 6]. Thus, each language could present a different arrangement or hierarchy of how these cues are used by their listeners [7]. Mattys and Bortfeld [7] for example, argue that the knowledge that a group of segments form a particular word (lexicality) is more important for English word segmentation than lexical stress information. On the other hand, despite this proposal, [5, 6 and 8] have shown that English native speakers are perfectly capable of using lexical stress correlates in their L2 French segmentation. This raises more questions about the possible role lexical stress could play in word segmentation.

Indeed, most studies looking at the role of lexical stress in word segmentation do present a common limitation: results have been obtained using experimental designs that look mostly at participants' final decision [9]. In contrast, results such as [6], which used designs looking at the entire decision-making process as speech unfolds over time, have been more successful at finding effects of lexical stress correlates in word segmentation. However, studies such as [6] have faced other limitations, mainly when it comes to pinning down specific effects of segmentation cues to particular moments of the decision-making process. These limitations could be addressed with different statistical approaches allowing us to understand the detailed time-related effects of segmentation cues without major changes in the experimental design (see [10, 11]).

In the present study, we want to look deeper at the role of lexical stress in word segmentation as speech unfolds in time. At the same time, we are interested in exploring word segmentation in Spanish. Previous research has shown that mismatching stress placement on a prime word can interfere with Spanish listener's lexical activation, indicating stress is an important part of recognizing words in Spanish [12]. Here, we want to know how the presence or absence of lexical stress can inform listeners about word boundaries in the incoming speech.

Spanish is a language where lexical stress does not have a fixed position and it can differentiate word meanings [13]. However, the most common position for stress in Spanish is second to last (in 74% of nouns and adjectives) [14]. This seems to be reflected in listeners' preferences, who show a perception bias in favor of words with penultimate stress over other configurations when listening in noise [15]. Given the high percentage of 2 and 3 syllable words, second to last stress can be found in either the first or the second syllable of common nouns. This means that lexical stress in this position presents a variable distance relative to the left edge of the word and a fixed distance relative to the right edge of the word. Therefore, different from other non-fixed stress languages previously studied (such as Dutch or English) [8, 4], lexical stress in Spanish would arguably inform listeners about the word ending rather than the word beginning [16].

Given this structure, we believe an experimental procedure that presents listeners with temporarily ambiguous sentences pairing a target word/noun with an adjective, similar to [5], provides an adequate manner of testing the temporal effects of lexical stress in word segmentation. In the following sections, we report results of a Mouse Tracking [17] study where temporally ambiguous noun phrases were either naturally produced - containing stress cues - or acoustically neutralized - stress correlates were equated on all syllables of the target word. We further discuss the implications of our results in regard to Spanish word segmentation and the potential role of lexical stress for segmentation in general.

2.   Methods

2.1.   Participants

Twenty-four native speakers of Mexican Spanish (13 women, age 19-33) with self-reported normal hearing were recruited and tested online. Data from four additional participants was excluded because of experiment failure (N = 3) or because

mouse responses showed circular patterns[1] (N = 1). Participants were randomly assigned to one of 2 lexical stress conditions ("natural" or "neutral"). All participants reported being born and residing in Mexico and were recruited through the online platform Prolific [18].

2.2. Stimuli

For the test conditions, we selected 14 pairs of noun + adjective phrases of the form "PAlo marron" vs. "paLOma roja". These are temporarily ambiguous phrases where the first three syllables are always segmentally ambiguous. Segmental disambiguation starts in the fourth syllable of the sequence. Earlier in the noun, however, lexical stress information can potentially be used to disambiguate between "PAlo" with stress on the first syllable and "paLOma" with stress on the second syllable. We also created two control conditions. The first of these conditions is a Minimal pair condition, where we presented 14 pairs of phrases starting with minimal pair nouns such as "coco" in "coco duro" vs. "cono" in "cono dulce". The second is a baseline condition where we selected 14 pairs of phrases with no segment overlap such as "perro lindo" vs. "gato pequeño". All phrases were composed of real Spanish words[2].

All stimuli were recorded by a native Spanish speaker using a standardized accent (first author) in a sound-proof booth. Using Praat [19], all recordings were then modified to present the same average intensity and speaking rate and were further denoised. These were the final recordings for Baseline, Minimal Pairs and the Natural test condition. We then further modified the Natural condition in Praat to create a second set of recordings with neutralized stress (Neutral condition). We first hand annotated syllable and phoneme boundaries for the first three ambiguous syllables. Second, we modified intensity so that the three syllables had similar intensity envelopes. This was accomplished by modifying the intensity envelope of each syllable using the intensity tier, multiplying it to the original recording and then rescaling it using Modify Scale Intensity. Third, we flattened F0 using the PSOLA [20] method in Praat. Finally, we neutralized vowel length by equating each vowel's length in the ambiguous syllables to the average vowel length of the three ambiguous vowels, again using PSOLA in Praat.

2.3. Procedure

We used a mouse-tracking listening task. On each trial two words were presented in the top right and top left corners of the screen, a target (e.g., X) and a competitor (e.g., Y). At the bottom of the screen was a blue circle (figure 1).
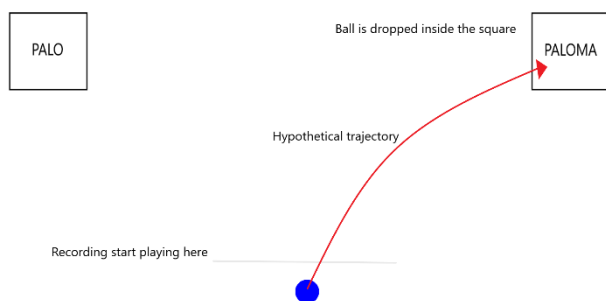

Figure 1: *Experimental set-up.*

The participants' task was to listen to phrases with different degrees of segmental overlap and identify the target word by dragging and dropping the blue circle on top of one of the words on screen. We tracked the mouse coordinates as participants made their responses. Because we wanted to observe mouse movements as participants were hearing the stimuli, we wanted them to start moving the mouse right away and not wait until they had heard the end [21]. To encourage early movements, we incorporated a movement initiation threshold to start the recording. This meant that participants needed to click on the circle and move it a centimeter up to listen to the recording. We used a circle that needed to be dragged (rather than just having mouse movement) to allow the experiment to be compatible with tablets and for data to be comparable when using touch screens as well as with a mouse. The experiment was coded in JavaScript using JsPsych [22] as the main frame for the experimental procedure. We created two plug-ins that run on JsPsych, one for mouse tracking trials and a second one that runs on touch-screen devices. After participants were recruited in Prolific, they were redirected to BRAMS-OTP [23], to be tested on their platform. The experiment started with a stereo sound test (to ensure participants were using headphones to complete the experiment), and a language background questionnaire before starting with the main experiment. The questionnaire was used to ensure that participants were native speakers of Mexican Spanish, born in Mexico and have lived most of their lives there. We also used it to make sure they would not be considered early bilinguals (e.g., have learned a different language before 4 years of age). Both preliminary tasks were coded using the ProsodylabExperimenter tool [24].

3. Results

For the data analysis, we excluded any trial with response times higher than 3 seconds after the beginning of the recording since decisions made in these trials may involve other types of strategy not related to the experiment's manipulations. This resulted in the exclusion of 5% of the total trials. For the analysis of the results, we will concentrate on three different measures. The first is response accuracy which is reflective of the effect of the predictors on the participant's final decision. The second is Reaction Times (RTs), measured from the moment the recording starts playing to the final moment when the circle is dropped on the word. This gives us a look into the general processing time required by each condition (the harder/less informative a condition, the longer it takes to process it). Finally, we will also analyze the changes on the x-coordinate over time. Given the position of the competitor and target word at the upper corners of the screen, the mouse position at a given point in time gives us more detailed information about the bias caused by our predictors at different moments of the trial.

We begin by looking at the filler conditions Baseline and Minimal Pair. For both no difference was found for between the two participant groups and baseline condition showed ceiling effects at 100% accuracy. Given the consistency found on these conditions between groups and that they are not theoretically relevant for the present study, we will continue the analysis only on the test conditions e.g., stress Natural and stress Neutral. For the following analysis we will use different types of Mixed Effects Models (MEM) with two theoretically important fixed

---

[1] Circular mouse patterns add a lot of extra noise to the trajectories and cannot be analyzed with the same methods as regular patterns.

[2] For a complete list of the stimuli see https://osf.io/c32x4/

effects, condition (stress Natural or stress Neutral), and number of syllables, which represents whether the target word had 2 syllables (PAlo) or 3 syllables (paLOma). All MEMs were run with the R package lmerTest [25] and P-values were calculated with Satterthwaite approximation.

Participants' accuracy shows a clear preference for 3 syllable words over 2 syllable ones. Regarding stress condition, it is important to notice that while it does not seem to affect the trials with 3 syllable words, it has a big effect on trials with 2 syllable words (see Table 1 and figure 2). We ran a Logistic MEM with condition, number of syllables and their interaction as fixed effects (neutral condition and 2 syllable words as reference categories respectively) and random intercepts per participant and word. We found significant effects of stress condition (β = 2.57, SE = 0.634, P < 0.001) indicating the 2-syllable natural condition was more accurate than the 2-syllable neutral condition, and number of syllables (β = 4.92, SE = 1.07, P < 0.001) indicating the 3-syllable neutral condition was more accurate than the 2-syllable neutral condition. We did not find a significant effect of the interaction as both natural and neutral 3 syllable conditions already present ceiling effects.
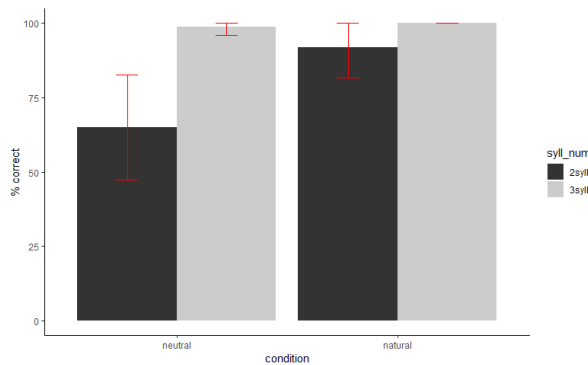

Figure 2: *Accuracy per stress condition and number of syllables.*

Table 1: *Results for Accuracy and RTs per conditions.*

Condition	%Acc	RTs[3] (sd)
Natural-2syll	91.9	1819 (481)
Natural-3syll	100.0	1435 (361)
Neutral-2syll	65.03	2307 (371)
Neutral-3syll	98.8	1843 (392)

For the analysis of RTs, we used only trials that participants answered correctly. From the data presented in Table 1 and Figure 3, we can observe a similar trend as the one presented in the accuracy analysis. Responses are faster on correct trials with natural stress than neutral stress and correct trials with 3 syllable words are responded to faster than 2 syllable ones. This time, the stress condition seems to have a clearer impact on 3 syllable words, but 2 syllable word trials are still the slowest. To analyze the data, we ran a Linear MEM with stress condition, number of syllables and their interaction as fixed effects (coded as before) and random intercepts per participant and word and random slopes for number of syllables per participant. We found significant and roughly equivalent effects of stress condition (β = -487.44, SE = 111.52, P < 0.001), and

number of syllables (β = -482.29, SE = 62.76, P < 0.001); and a non-significant interaction (β = 78.60, SE = 82.70, P = 0.351).

Figure 3: *RTs per stress condition and number of syllables.*
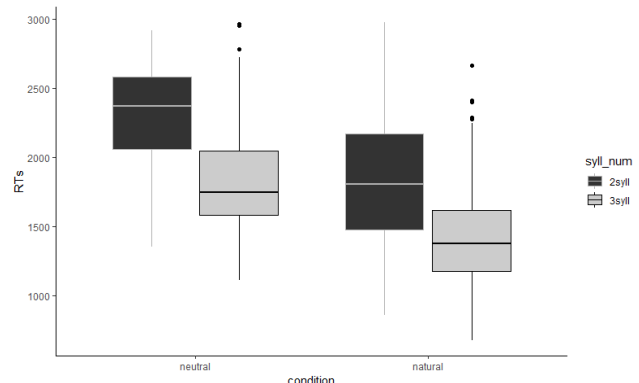
In order to analyze the mouse coordinate data, first, we centered the x-coordinates of each trial by subtracting the value of the first x-coordinate. This means that for each trial the first coordinate (e.g., the moment where the recording starts) is always 0. Then, for every trial where the correct word was assigned to the left of the screen the entire x-coordinate vector was multiplied by -1. This mirrors the mouse trajectories onto the other side of the coordinate system and makes the data of these trials comparable to the ones where the correct word was on the right side of the screen. Moreover, now positive values indicate movements towards the target word, and negative ones movements towards the competitor word. We used the method mt-time-normalize from the MouseTrap package for R [26] to normalize the time data into 100 even time windows.

As in the RT analysis, we only analyze trials that were answered correctly. To test the different effects of stress condition and number of syllables over time, we ran a series of Linear Mixed Effects Models on each time window and tracked the changes on β weights for each fixed effect[45]. Each model had fixed effects of stress condition, number of syllables and their interaction (coded as before), random intercepts per participant and word and random slopes for number of syllables per participant.

Figure 4 shows the evolution of each β-weight for each time window. Points represent significant effects (p <0.05). The evolution of the intercept shows that early in the trial (~30th window), participants showed a bias towards the competitor word (negative β-weights) for 2 syllable word trials with neutral stress. This bias changes towards the target word later in the trial (~70th window, positive β-weights). The effect of number of syllables tells us that for 3 syllable word trials with neutral stress, the bias towards the target word starts earlier than for the intercept (~50th window) and peaks around the 70th window. On the other hand, β-weights for stress condition show a pervasive effect of natural stress cues in 2 syllable word trials that favors movements towards the target word. This effect starts as early as the 30th window and extends past the 75th window. Finally, the interaction tells us that on trials with both

---

[3] The RT means correspond to trials answered correctly.
[4] For a complete table of the MEMs' coefficients please go to https://osf.io/atd35

[5] Given the particularities of x-coordinate data, mixed effects models presented singular fit warnings for the earliest and latest time windows. For consistency and clarity, we kept their structure, but see also [11].

3-syllables targets and natural stress cues, the combined bias of both factors towards the target is weaker than expected.
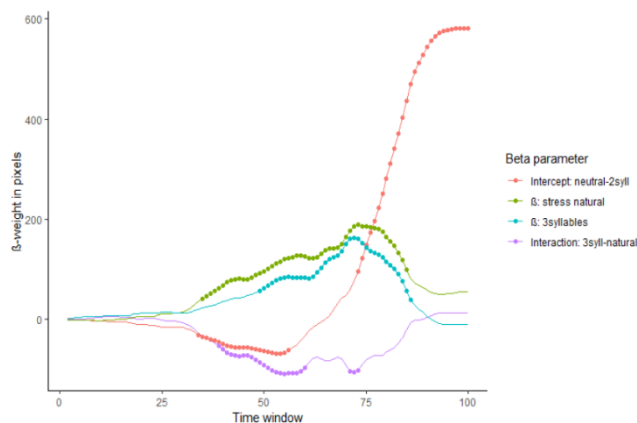

Figure 4: β weights over time per fixed effect. Points represent significant effects (p<0.05).

4. Discussion

The aim of the present study was to further explore the effects of lexical stress on word segmentation as speech unfolds in time. We used a Mouse Tracking experiment with temporally ambiguous phrases of the type "PAlo marron" vs. "paLOma roja" and analyzed three different types of measures: response accuracy, RTs, and x-coordinate changes over time.

The accuracy analysis, as an off-line measure, reflects the effect of the predictors on the final decision reached by the participants. On the one hand, we found a clear effect of the number of syllables of the target word that shows a bias of participants for the longer words "paloma" over the shorter ones "palo". Indeed, given that we matched the initial adjective syllable of the shorter word to be the same as the last syllable of the longer word (e.g., "ma" in "palo marron" was matched to "paloma"), it seems that participants tended to choose the longer word "paloma" as soon as they heard the third syllable "ma". On the other hand, although the effect of presenting natural stress for 2 syllable words was significant, it's estimated effect was not as large as the 3-syllable effect. Furthermore, 3 syllable words with neutral stress already showed ceiling effects and therefore no more added value was found for the presence of natural stress. Together, this could indicate a higher impact of lexicality (choose the word that best matches the segments you've heard) over stress cues on the final decision as argued by [7].

There could be two reasons for this asymmetry. The first is that participants do consider lexical stress cues in the unfolding speech, but these are overcome by a higher effect of lexicality (once the third syllable "ma" is heard); and the second is that lexical stress only plays a role much later in the trial if participants need to reevaluate their answers (in the case of 2 syllable word targets). The analysis of online measures should shed some light into the specific role of lexical stress.

The RT analysis mirrors the accuracy analysis in showing an effect of both syllable number, where 3 syllable trials are faster than 2 syllable ones; and stress condition, where the natural stress condition is faster than neutral condition for correct responses. Here the estimated effect sized for the two effects were nearly identical and the effect of stress cues is evident on 3 syllable trials. This means that even if it is not well reflected

in the final decision, the lack of lexical stress cues does make the stimuli harder to segment, even for 3 syllable word trials. This seems to support the first explanation outlined in the previous paragraph rather than the second one, e.g., that lexical stress does play a role as speech is unfolding rather than just coming into play if/when reevaluation is needed.

Finally, the x-coordinate analysis shows that the effect of lexical stress starts earlier than that of the number of syllables and the estimates are larger. This could be taken as further support for the processing and integration of lexical stress cues as speech unfolds in time. Under this view, as soon as lexical stress is presented, it is processed and integrated by the listener. The presentation of expected following segments (e.g., 'ma' after 'paLO' is consistent with 'paLOma') will result in fast and accurate segmentation of the speech signal. On the contrary, the presentation of contradicting segments (e.g., 'ma' after 'PAlo' is more consistent with the competitor word) will result in difficulties in the processing of the speech signal that need to be resolved by either waiting for disambiguating information, thus increasing reaction time; or by choosing segmental information over lexical stress and risking losing accuracy.

It is important to note that in the present experiment we present participants with a binary decision, the answer is either "palo" or "paloma", and that participants know in advance the word choices before they listen to the recordings. In real life speech comprehension, this is not the case. Although we acknowledge the role that semantic and syntactic context can play in shaping lexical expectation, we believe that it does leave expectations more open than the experimental design we used here. This leaves open the possibility of having a bigger effect of lexical stress in real life scenarios. It would be important for future work to test the current results in more naturalistic experimental designs. Furthermore, in the present study we only manipulated the presence or absence of lexical stress in Spanish. It is important as a next step to look at the specific role that different stress correlates can play in signaling word boundaries as it has been done in other languages, see [5,8]. Finally, it is also important to test the role of lexical stress in word segmentation as speech unfolds in time when comparing its usefulness to other segmental cues, such as allophonic variation.

5. Conclusions

In conclusion, we have found compelling evidence that lexical stress plays an important role in Spanish word segmentation. It goes beyond just being a secondary cue after lexicality and can potentially affect processing time and effort. Moreover, we were able to pin down the different effects of lexical stress and segmental information to specific moments in the decision-making process.

6. Acknowledgements

7. References

[1] S. L. Mattys, L. White, & J. F. Melhorn, "Integration of multiple segmentation cues: A hierarchical framework," *Journal of Experimental Psychology: General*, 134, 477–500, 2005.

[2] M. H. Banel, & N. Bacri, "On metrical patterns and lexical parsing in French," *Speech Communication*, 15, 115-126, 1994.

[3] O. Bagou, C. Fougeron, & U. H. Frauenfelder, "Contribution of Prosody to the Segmentation and Storage of 'Words' in the Acquisition of a New Mini-Language," Paper presented at *Speech Prosody*, Aix-en-Province, France, 2002.

[4] N. Cooper, A. Cutler, & R. Wales, "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," *Language and Speech*, 45, 207–228, 2002.

[5] A. Tremblay, C. E. Coughlin, C. Bahler, & S. Gaillard, "Differential contribution of prosodic cues in the native and non-native segmentation of French speech," *Laboratory Phonology*, *3*(2), 385-423, 2012.

[6] A. Tremblay, M. Broersma, C. E. Coughlin, & J. Choi, "Effects of the native language on the learning of fundamental frequency in second-language speech segmentation," *Frontiers in Psychology*, 7, 985, 2016.

[7] S. L. Mattys, & H. Bortfeld, "Speech Segmentation," in G. Gaskell & J. Mirkovic (Eds.), Speech perception and spoken word recognition (55–75). New York, NY: Routledge, 2016.

[8] A. Tremblay, M. Broersma, & C. E. Coughlin, "The functional weight of a prosodic cue in the native language predicts the learning of speech segmentation in a second language," *Bilingualism: Language and Cognition*, *21*(3), 640-652, 2018.

[9] N. A. MacMillan, & D. C. Creelman, "Detection Theory, A user's guide," In Mahwa, NJ: Lawrence Erlbaum Associates, 2005.

[10] T. B. Roettger, & M. Franke, "Evidential strength of intonational cues and rational adaptation to (un-) reliable intonation," *Cognitive science*, *43*(7), e12745, 2019.

[11] S. Scherbaum, & M. Dshemuchadse, "Psychometrics of the continuous mind: Measuring cognitive sub-processes via mouse tracking," *Memory & Cognition*, *48*(3), 436-454, 2020.

[12] S. Soto-Faraco, N. Sebastián-Gallés, & A. Cutler, "Segmental and suprasegmental mismatch in lexical access," *Journal of Memory and Language*, 45, 412-432, 2001.

[13] A. Quilis, "Caracterización fonética del acento español," *Travaux de Linguistique et de Littérature*,9, 53–72, 1971.

[14] N. Sebastian-Galles, M. A. Marti, F. Cuetos, & M. Carreiras, "LEXESP: Base de datos informatizada de la lengua española (Progress Report)," Departament de Psicologia Bàsica, Universitat de Barcelona, 1996.

[15] A. LaCrossa, J. Liss, B. Barragan, A. Adams & V. Berisha, "The role of stress and word size in Spanish speech segmentation," *The Journal of the Acoustical Society of America,* 140, EL484-490, 2016

[16] I. Oltra-Massuet, & K. Arregi, Stress-by-structure in Spanish. *Linguistic Inquiry*, 36(1), 43-84, 2005.

[17] R. Dale, C. Kehoe, & M. J. Spivey, "Graded motor responses in the time course of categorizing atypical exemplars," *Memory & cognition*, *35*(1), 15-28, 2007.

[18] Prolific, London, UK. Version: 09-2022. Available at: https://www.prolific.co/

[19] P. Boersma & D.Weenink, "PRAAT, a system for doing phonetics by computer. Report 132." 1996, Institute of Phonetic Sciences of the University of Amsterdam.

[20] H. Valbret, E. Moulines, & J. P. Tubach, "Voice transformation using PSOLA technique," *Speech communication,* 11(2-3), 175-187, 1992.

[21] P. J. Kieslich, M. Schoemann, T. Grage, J. Hepp, & S. Scherbaum, "Design factors in mouse-tracking: What makes a difference?" *Behavior Research Methods*, *52*(1), 317-341, 2020.

[22] J. R. De Leeuw, "jspsych: A javascript library for creating behavioral experiments in a web browser," Behavior research methods, vol. 47, no. 1, pp. 1–12, 2015.

[23] https://brams.org/category/online-testing-platform/

[24] M. Wagner, "ProsodylabExperimenter. a spreadsheet-in spreadsheet-out tool for running online experiments," 2021. [Online]. Available: https://github.com/prosodylab/prosodylabExperimenter

[25] A. Kuznetsova, P. B. Brockhoff, & R. H. B. Christensen, "Package 'lmertest'," *R package version*, *2*(0), 734, 2015.

[26] P. J. Kieslich, & F. Henninger, "Mousetrap: An integrated, open-source mouse-tracking package," *Behavior research methods*, *49*(5), 1652-1667, 2017.