MASSIVE AUDITORY LEXICAL DECISION: INVESTIGATING PERFORMANCE IN NOISY ENVIRONMENTS

Pearl Lorentzen, Filip Nenadić, Matthew C. Kelley, and Benjamin V. Tucker[1]

1 – University of Alberta
lorentze@ualberta.ca, nenadic@ualberta.ca, mckelley@ualberta.ca, bvtucker@ualberta.ca

ABSTRACT

Although most auditory lexical decision experiments are performed in a laboratory setting, humans tend to communicate in uncontrolled and noisy environments. We investigated, indirectly, the impact of noise and other distractions on lexical processing. The present study used a subset of words from the Massive Auditory Lexical Decision (MALD) database and ran a series of shorter auditory lexical decision experiments at a busy science museum, the Telus World of Science – Edmonton, as part of a science outreach program. As expected, the surroundings provided many distractions, as well as a larger variety in age and other participant characteristics in comparison to a laboratory study. In this environment, speed and accuracy of responses were both reduced, but the same patterns of results as those obtained in laboratory settings emerged.

Keywords: phonetics, auditory lexical decision, lexical processing, realistic environment, various ages

Acknowledgements: Research performed in collaboration with Telus World of Science - Edmonton LabQuest. This project was supported in part by the Social Sciences and Humanities Research Council of Canada [Insight Grant 435-2014-0678] and a grant from the Kule Institute for Advanced Studies both to the last author. Brant Harker, Michele Leung, Danielle Fonseca, Spenser Halfyard, Melina Sinclair, Shirley May, Theresa Nguyen, Ivy Mok, and Harveen Brar assisted in collecting data.

1. INTRODUCTION

In psycholinguistics, databases constructed using large-scale studies with responses to tens of thousands of words have become important resources for researchers investigating many topics, particularly lexical processing and representation. The benefits of creating large-scale databases are discussed in the paper describing the Massive Auditory Lexical Decision (MALD) database project [6]. The MALD database is a massive dataset investigating spoken word recognition, similar to other large auditory databases created for Dutch [3] or French [4], or for visual word recognition [1]. MALD aims to encompass various listener (e.g., native language, dialect, age), item (e.g., part of speech, frequency, morphological complexity), and situational factors that may influence listener performance. To date, MALD is the largest database in the auditory domain for English.

In this report, we present results from a branch of the MALD project, conducted at the TELUS World of Science – Edmonton (TWOSE) as part of a science outreach program called LabQuest. TWOSE is a science museum in Edmonton, AB. Unlike data gathered as part of our standard experimental procedures, TWOSE-MALD experiments were not performed in sound-attenuated booths in the lab, but on tablets in an often crowded TWOSE, with multiple participants being tested at the same time. Most participants had varying levels of distraction while performing the experiment. This environment is more ecologically valid than a laboratory experiment, as it more closely reflects real-life conditions in which lexical processing occurs.

The primary goals of this study were two-fold. The first goal was to engage in science outreach by bringing our research into the community and talking to people about the type of research we perform and the questions we are interested in. The second goal was to ascertain how participants perform in an auditory lexical decision task with realistic background noise and distractions.

2. METHOD

2.1. Participants

Employees, volunteers, and visitors at Telus World of Science – Edmonton (TWOSE) participated in the experiment. Responses were gathered between July 2017 and August 2018. Complete response data was recorded for 1,254 participants. In the analysis reported here we only consider the 1,099 native speakers of English. The actual number of recruited participants was much larger. There were many instances where we did not keep participant data, as anyone was welcome to participate but not all participants gave or could give consent for their data to be saved and included in the final analysis. Further, some did not finish the experiment either due to technical issues or because other things were more pressing and they left.

Participant age ranged from 4 to 86 (M = 26.74, SD = 17.94), with data missing from 15 participants. The number of declared males and females was roughly the same (50.77% male, 47.13% female, 1 participant selected ‘other’, and no response was provided by 22 participants).

2.2. Stimuli

Stimuli were a randomly selected subset of 2000 words and 2000 pseudowords from the MALD project [6], all recorded by one male speaker of western Canadian English. The stimuli were further randomly divided into 20 lists, each containing 100 words and 100 pseudowords, with no practice stimuli. Each participant was presented with a single list. The entire experimental session usually lasted between five and ten minutes. In a few exceptional cases some participants did the experiment multiple times over the course of the year, as they visited TWOSE multiple times in the period. Since no identifying information was retained it was impossible to identify these duplicate participants.

2.3. Procedure

The experiment was set up in various locations at TWOSE, but tables and chairs were always available for participants. The participants were recruited as they traversed the science center. If interested in participating, the participants were given a tablet (Samsung Galaxy Tab A, 177.7 mm display) with written instructions. Oral instructions by the experimenter were given as well, even if another participant was in the process of participation nearby.

If the participants agreed to share their data, they first completed a nine-item survey about their sociodemographic characteristics and language background. After that, the participants put the headphones on and the lexical decision task began. First, a fixation cross appeared for 500 ms, followed by an auditory word or pseudoword. Stimulus order was random. The participants responded by tapping the appropriate section of the screen: a blue square with a large “YES” for the “word” response, or a red square with a large “NO” for the “not a word” response. The participants were instructed to complete the task as quickly and as accurately as possible (but see Distractions below).

There were no incentives for participation in the study. However, as part of the science outreach component of the experiment participants were given the option of having their total accuracy and average response latency posted to an anonymous scoreboard for the day’s participants. The scoreboard plotted the participants of the day onto a histogram of the MALD data so that participants could see how they compared to the data previously collected and to other participants of the day. This was also taken as an opportunity to engage the participant in a discussion about the experiment or linguistics in general.

2.4. Distractions

Some participants completed the experiment alone or with a friend, with very little distractions. Most participants, however, had varying levels of distractions while performing the experiment. Most often these included noise coming from other visitors and activities in TWOSE. Specific examples of the distractions were an IMAX theatre, a Tyrannosaurus Rex mascot, hands on demonstrations, music from the gift shop, and other visitors. Additional noise was generated by other participants doing the experiment or talking to the experimenters, small children who accompanied adults wanting attention or asking for help in their own experimental task, and experimenters themselves performing their tasks. Finally, some participants were not skilled in handling the tablets, and many would perform the task slowly, using a single hand to respond to items instead of having one hand over the “YES” and the other over the “NO”.

3. RESULTS

In comparison to participants tested in a laboratory setting [6], native listeners collected as part of TWOSE-MALD were less accurate (a decrease from 87% to 79%) and had slower response latencies (an increase in mean latency from 1,017 ms to 1,262 ms). Subsequently, we investigated response latencies in more detail using generalized additive mixed effects modeling [8, 7]. For this analysis, 20.20% of the participants were excluded because they had accuracies under 55%. The final model included smoothed effects of age, trial number, stimulus duration in ms, logged COCA frequency [2], logged phonological neighborhood density, and phonological uniqueness point. Random effects for particular words and experiment lists, and random smooths of trial number per participant were included in the model as well.

Figure 1 illustrates that participants' reaction times became faster up to the age of 20. The reaction times seem to level off after the age of 20, with a general slowdown as age increases. We also find that higher frequency words were responded to more quickly than low-frequency words (Figure 2), and words with higher neighborhood density were responded to more slowly than words with lower neighborhood density (Figure 3).

Figure 1: Age as a predictor of response latency.

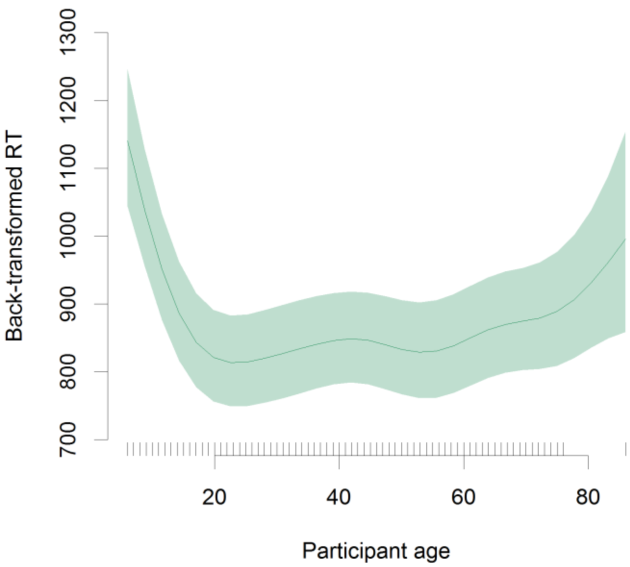
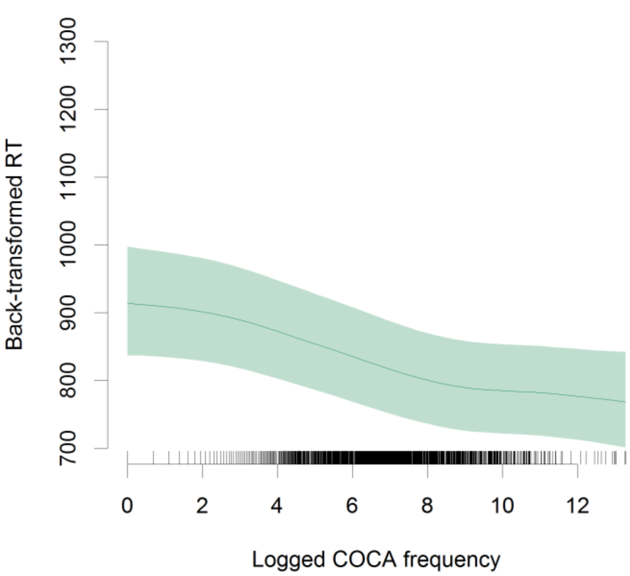


Figure 2: Frequency as a predictor of response latency.



Later phonological uniqueness point predicted faster responses (because word duration was included in the model). Figure 4 illustrates that participants had slower responses to longer words. This effect is likely due to the fact that response latencies were measured from word onset. Additionally, the participants responded more quickly as the experiment progressed.

Figure 3: Neighborhood density as a predictor of response latency.

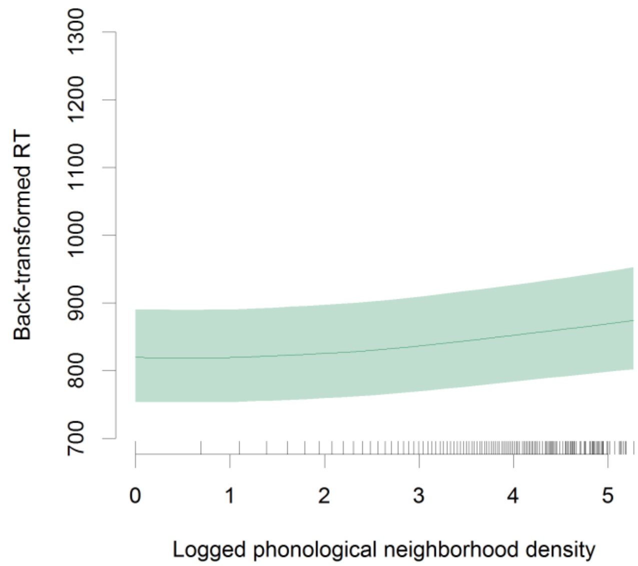
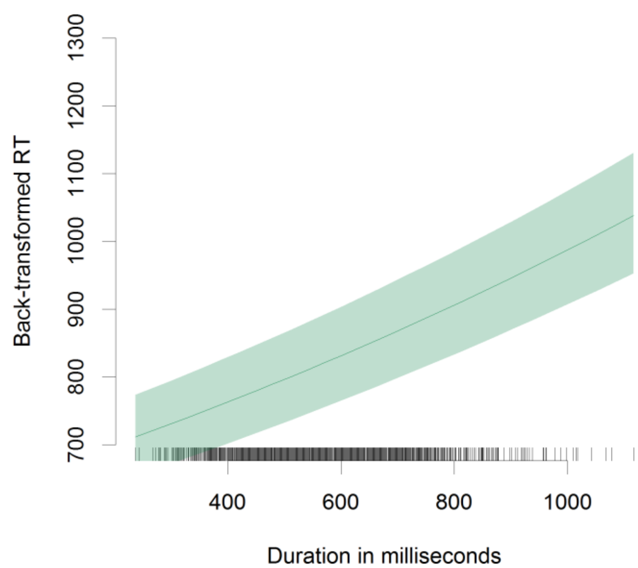


Figure 4: Duration as a predictor of response latency.



4. DISCUSSION

The performance decrease for participants at TWOSE indicates that in a more representative sample and distracting setting, the time required to recognize a word is longer than what laboratory experiments may find. The use of tablets instead of computers with specially calibrated button boxes may also have introduced additional latency to responses. Furthermore, the MALD database used mostly traditional-age university students (age 17 to 21) [6], while this experiment tested participant from 4 to 86 years of age. The larger age range was also a likely contributor to the difference in response time and accuracy between the two experiments. The length of the experimental session was also much shorter for TWOSE-MALD, which likely influenced response latency because response speed tends to increase as the experiment progresses.

However, the relations between predictors included in the model and response latency are similar to those obtained in laboratory lexical decision experiments [6]. We see the same effects of trial and stimulus duration. More importantly, both frequency and neighborhood density are affecting response latency in the same way they do in a laboratory setting. Therefore, even if overall accuracy and speed are lower, the process of spoken word recognition remains the same outside of the laboratory. Further investigation and a more direct in-depth analysis are required, but this finding corroborates the results of laboratory studies of spoken word recognition.

With the vast age range covered, TWOSE-MALD allows us to assess how age influences response latencies, which the main MALD experiment does not. (Note, however, that a dataset of responses from participants of various ages participating in the MALD experiment is also currently being created in a laboratory setting.) We find that response latencies decrease rapidly in the first approximately 20 years, after which response latencies seem to have a steady increase as participants' age increases. However, at the much older ages, there are fewer data points and more variability. This finding is in line with previous studies employing various tasks measuring response latencies, although the causes of the increase are a subject of debate [5].

In sum, running experiments in noisy and distracting environments may open venues to more versatile ways of data collection and easier access to a wider population of language listeners to answer questions that have previously not been as easily addressable with laboratory-run experiments.

REFERENCES

1. Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459.
2. Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990-2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14*(2), 159-190.
3. Ernestus, M., & Cutler, A. (2015). BALDEY: A database of auditory lexical decisions. *The Quarterly Journal of Experimental Psychology*, *68*(8), 14691488.
4. Ferrand, L., Méot, A., Spinelli, E., New, B., Pallier, C., Bonin, P., ... & Grainger, J. (2017). MEGALEX: A megastudy of visual and auditory word recognition. *Behavior Research Methods*, 1-23.
5. Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, *6*(1), 5-42.
6. Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., & Sims, M. (2018). The Massive Auditory Lexical Decision (MALD) database. *Behavior Research Methods*, 1-18.
7. Van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2016). itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version 2.2.
8. Wood, S. N. (2006). Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.