Cognitive and Meta-cognitive Signatures of Memory Retrieval Performance in Spoken Word Learning

Keywords: Adaptive Learning, Cognitive Modeling, Memory Retrieval, Speech, Structural Equation Modelling, Prosody, Confidence

Background

Cognitive models of memory retrieval aim to capture human learning and forgetting over time. Such models have been applied in learning systems that aid in memorizing information by adapting to the needs of individual learners (e.g., Lindsey, Shroyer, Pashler, & Mozer, 2014). Adaptive learning systems track learning performance to provide personalized feedback or optimize item repetition schedules. The effectiveness of such learning systems critically depends on their ability to use behavioral proxies to estimate the extent to which learners have successfully memorized the materials. The present study examines cognitive and meta-cognitive indicators of memory strength that are present in the learners' recorded speech signal while studying vocabulary items by vocally responding to cues.

In most model-based learning systems, predictions of memory retrieval rely on the accuracy and response latency of retrieval attempts. In this project, we will focus on spoken responses to visually-presented retrieval cues, which contain prosodic speech features (PSFs). PSFs are high-level properties of units of speech such as syllables, words or sentences, and include intonation (pitch variations), loudness and speaking speed. PSFs can carry information that is not conveyed by grammar or vocabulary, such as the emotional state of the speaker, emphasis, or the form of the utterance (e.g., question versus statement/command) (Xu, 2011). A recent study by Goupil and Aucouturier (2021) demonstrated that both the *objective accuracy* of a response, and a speaker's *meta-cognitive* confidence in the response are deferentially reflected in speech. In their study, participants were instructed to complete a visual detection task, where they had to verbally choose which word they saw before from a number of alternatives and rate their confidence in the response. The results showed that *some* of the PSFs (speaking speed and pitch) were associated with the subjective confidence in a response, whereas the other PSF (loudness) was associated with objective accuracy.

In a recent study, Wilschut and colleagues (Wilschut, Sense, Scharenborg, & van Rijn, 2022), demonstrated that the above results generalize to a memory retrieval paradigm, and found that using PSFs on the current trial could increase the prediction accuracy for memory retrieval success on future learning trials. In the current project, we extend their work by further investigating the exact way in which PSFs are associated with both cognitive *and* meta-cognitive aspects of memory retrieval. Examining this question is important, as information about cognitive and meta-cognitive indices of memory performance—extracted from speech in real time—may be used to effectively inform models of memory retrieval and improve adaptive learning systems.

Methods

A total of 40 participants studied 30 Lithuanian-English vocabulary items. The first presentation of an item involved the visual presentation of both Lithuanian cue and the visual presentation of the English translation. Subsequent presentations of the item just showed the visual Lithuanian cue, and the participant was asked to utter the English translation. After this response was recorded, the participant was asked to rate their subjective confidence in the accuracy of the response using a slider-response scale, followed by corrective feedback. Participants cycled through the total list 4 times. At the start of a new cycle, the 30 items where split in to subsets of the first 15 and last 15 items, and both subsets where were shuffled. Speech features were extracted from the recorded data after the experiment, and all speech features were standardized within participants.

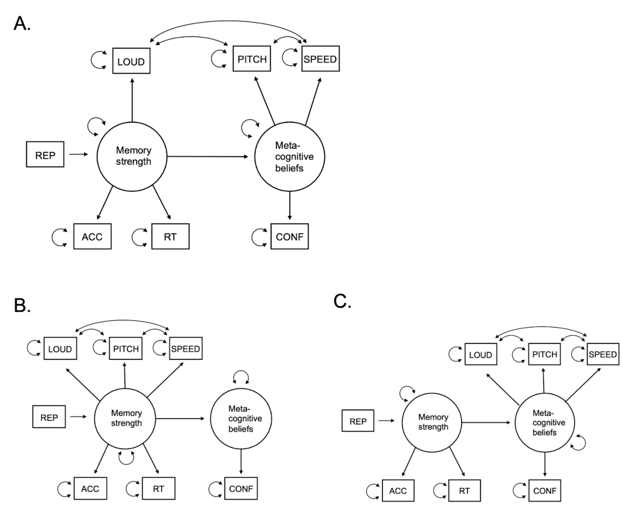

Figure 1: Structural equation models showing alternative possible relationships between latent factors memory strength and meta-cognitive beliefs as measured by accuracy, response times, confidence ratings, and PSFs. **A** shows the hypothesised model, **B** and **C** are alternative models.

To examine the relationship between PSFs and cognitive and meta-cognitive aspects of memory retrieval, we contrasted a hypothesized model to two alternative, competing models using structural equation modeling (Ullman & Bentler, 2012, (SEM)). All three models assume a relationship between latent variables memory strength and a learners' meta-cognitive beliefs about performance, with memory

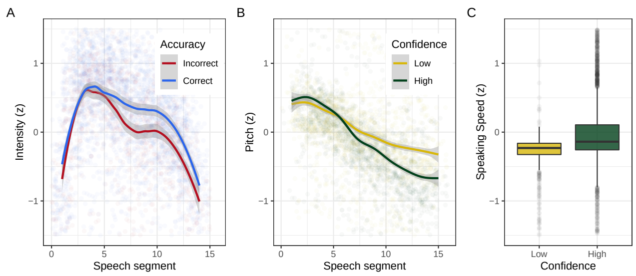

Figure 2: PSFs as a signature of memory retrieval performance. **A** shows that the average loudness over spoken retrieval attempts was higher for correct than for incorrect responses. **B** shows that average pitch slopes were higher for low versus high confidence responses, and **C** shows that speaking speed was lower for low than for high confidence responses. Shaded area's represent 95% confidence intervals.

strength measured by accuracy (ACC) and response times (RT), and meta-cognitive beliefs measured by subjective confidence judgements (CONF; see Figure 1). The hypothesized SEM is shown in Figure 1A, with two alternative models shown in Figure 1B and 1C. In the hypothesized model (A), in line with earlier research, meta-cognitive beliefs are measured by intonation and speaking speed, whereas memory strength is measured by loudness. The alternative models reflect two different underlying relationships: In the first alternative model (B), all PSFs directly reflect memory strength, and not meta-cognitive beliefs. In the second alternative model (C), all PSFs only indirectly measure memory strength, via a speaker's meta-cognitive beliefs about memory performance. We used Vuong's likelihood ratio test (Merkle, You, & Preacher, 2016) to compare the three models.

Results

The correctness of the responses was determined by Google's speech-to-text API, yielding sufficiently high transcription accuracy for the subsequent analyses. Figure 2A shows the average standardized loudness (intensity) over the duration of each utterance, for both correct and incorrect spoken retrieval attempts. Responses were, on average, louder for correct compared to incorrect responses. Figure 2B shows the average standardized pitch, separated for high subjective confidence scores (confidence scores above average for that participant), and low subjective confidence scores (below average for that participant). The figure shows a less negative averaged pitch slope for responses with low confidence than for responses with high confidence. Finally, Figure 2C shows that the average standardized speaking speed for high confidence retrieval attempts was higher than the average standardized speaking speed for low confidence retrieval attempts. These results underline and extend earlier findings, (Goupil & Aucouturier, 2021) and (Wilschut et al., 2022) by demonstrating that both cognitive (accuracy) and meta-cognitive (confidence) markers of memory performance are present in spoken word learning.

To compare the fit of the three SEM models outlined above, we used Vuong's likelihood ratio test. The hypothesised SEM model (A) fits the experimental data significantly better than both alternative models B and C (z = 7.177, $p < 0.001$; z = 2.980, p = 0.001, respectively). This supports the idea that meta-cognitive beliefs about memory retrieval are captured in different PSFs than the objective accuracy of a response.

Conclusion

This study examined which cognitive and meta-cognitive proxies of memory strength are present in the speech signal during spoken retrieval attempts. Participants studied vocabulary items using spoken retrieval practice. The results of the study are twofold. First, we demonstrate that it is possible to extract information about (1) the accuracy of a response and (2) a speaker's subjective confidence in a response from the speech signal. Second, we show that meta-cognitive beliefs about memory performance are measured mainly by variations in pitch and speaking speed, whereas the objective accuracy of a response is mainly measured by its loudness. The results of this study have theoretical and practical relevance. They contribute to a better understanding of the relationship between prosodic speech variations and (meta)memory processes and could facilitate the development of speech analyses as a new tool to explore open questions in learning research (e.g., about a learner's confidence in their responses. Second, as they demonstrate that the speech signal contains relevant information about memory retrieval performance, they may have important implications for the further development of models of memory retrieval used in adaptive learning systems. For example, extracting information about a speaker's confidence from the speech signal in real time may allow for improvement of predictions of future retrieval success—without the learner having to make explicit confidence judgments after each learning trial.

References

Goupil, L., & Aucouturier, J.-J. (2021). Distinct signatures of subjective confidence and objective accuracy in speech prosody. *Cognition*, *212*, 104661.

Lindsey, R. V., Shroyer, J. D., Pashler, H., & Mozer, M. C. (2014). Improving students' long-term knowledge retention through personalized review. *Psychological science*, *25*(3), 639–647.

Merkle, E. C., You, D., & Preacher, K. J. (2016). Testing nonnested structural equation models. *Psychological Methods*, *21*(2), 151.

Ullman, J. B., & Bentler, P. M. (2012). Structural equation modeling. *Handbook of Psychology, Second Edition*, *2*.

Wilschut, T., Sense, F., Scharenborg, O., & van Rijn, H. (2022). Beyond responding fast or slow: Improving cognitive models of memory retrieval using prosodic speech features. *Paper presented at In-Person MathPsych/ICCM 2022, Via mathpsych.org/presentation/858*.

Xu, Y. (2011). Speech prosody: A methodological review. *Journal of Speech Sciences*, *1*(1), 85–115.