Say It Like You Mean It: Can Prosody Predict Objective Accuracy and Subjective Confidence in Memory Tasks?

Anne F.H.J. Lukassen

S4806166

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor thesis (engels)

Group 19

Supervisors: Dr. T. Sarampalis and MSc. T.J. Wilschut

Second evaluator: Dr. O. Dimigen

In collaboration with: Y.B. Daskalova, E. Fernández-Amurrio, A.A. Jarczynska, T.I. Peters, E.C.C. Velthuizen

June 27, 2025

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but it does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned.

Acknowledgments

First of all, I would like to thank Dr. T. Sarampalis and Msc. T. J. Wilschut for designing the project methodologically, and guiding me throughout. You both provided me with valuable knowledge and expertise, for which I am truly grateful. Additionally, I am especially grateful for my group members, Y.B. Daskalova, E. Fernández-Amurrio, A.A. Jarczynska, T.I. Peters, E.C.C. Velthuizen, with whom I experienced a very pleasant collaboration. Lastly, I'd like to thank Dr. O. Dimigen in advance for taking the time to read and grade my bachelor's thesis.

Abstract

In memory research, reaction time and accuracy are commonly used as indicators of memory strength. Previous studies have shown that prosodic speech features (PSFs) can add to existing models by revealing how speech prosody's role conveys cognitive and metacognitive states. Therefore, we examined whether prosodic speech features (average pitch, pitch change, speaking speed, and average intensity) can predict subjective confidence and objective accuracy in memory retrieval in simple S-V sentences. Furthermore, we compared these relationships between L1 and L2 in a within-subject design. Forty-eight Dutch native speakers were asked to complete a learning task in which they studied and verbally recalled simple sentences in Dutch and Italian. Subsequently, the participants rated their confidence. Results showed that faster and louder speech was most strongly associated with confident and correct responses. These responses were also more likely to have a lower pitch. In the Dutch condition, responses that were correct and confident had a rise-fall pitch trajectory, whereas unconfident or inaccurate responses showed a rising pitch at the end of an utterance. Overall, most relationships were stronger and more consistent in the Dutch condition, with speaking speed as an exception. Our findings suggest that prosodic information can improve predictions of memory strength beyond response time and accuracy, offering potential benefits for adaptive and inclusive learning. However, stronger and more consistent effects were observed in the native language, implying that adaptive learning systems may require language-specific adjustments to accurately recognize prosodic cues in second-language learning contexts.

Keywords: adaptive learning system, objective accuracy, memory strength, prosody, speech-based learning, subjective confidence

Learning by Speaking: Can Prosody Predict Objective Accuracy and Subjective Confidence in Memory Tasks?

"One important feature of oral communication is that the voice conveys a rich variety of information beyond the content of a message (Guyer et al., 2021, p.481). This quote emphasizes how something is as important as what is said. For example, studies have shown that emotions can be accurately identified by focusing on the linguistic and paralinguistic features of speech in audio fragments (Guyer et al., 2021; Kraus, 2017; Pell et al., 2009). Prosody, a general term for a range of suprasegmental acoustic characteristics present in natural speech, contains both linguistic and paralinguistic features. For this paper, the prosodic speech features (PSFs) are described in three categories; intonation, rhythm, and stress. Intonation refers to pitch variation across an utterance, rhythm to its timing patterns, and stress to the emphasis (loudness) on specific syllables (Jackson & O'Brien, 2011; Wilschut et al., in press). These PSFs do not only support comprehension, (Hoyte et al., 2009), but they also convey emotion, attention, intent and, mental states (Hellbernd & Sammler, 2016; Xu, 2011a, 2011b), raising the question: can prosody help estimate memory strength?

Understanding how behavior reflects memory strength plays a key role in facilitating memory research and improving adaptive learning systems. Adaptive learning systems tailor instruction and feedback to individual performance (Chen et al., 2017; Lindsey et al., 2014). To estimate the strength of a memory trace, adaptive learning systems often use response latency and response accuracy (Mettler et al., 2011). These findings suggest that stronger memory traces are recalled faster (Van Rijn et al., 2009; Wilschut et al., 2021).

Van Rijn et al. (2009) present an adaptive learning model based on the spacing effect and the testing effect. The spacing effect, first described by Ebbingehaus (1913), shows that learners remember items better when they learn them with spaced repetition. The key idea of the testing effect is that retrieving knowledge from memory (i.e., taking a test) improves long-term memory retention more than restudying the material (Delaney and Verkoeijen, 2010).

Expanding on this, MemoryLab develops SlimStampen, a learning tool based on the ACT-R model. This model uses the spacing effect and speed of forgetting to adapt itself to the learner's abilities. An item's speed of forgetting is estimated using response accuracy and response latency (MemoryLab, 2025). By prioritizing items with a higher forgetting rate, SlimStampen balances the advantages of the testing effect combined with repetition. During learning, the adaptive learning system constantly updates each item's rate of forgetting and creates individual optimal repetition schedules. This is important since not all items are encoded with similar memory strength, resulting in the items being forgotten at different rates (Van den Broek et al., 2016)

Although accuracy and reaction time have long been used as measures of memory strength in typing-based algorithms, newer research explores speech-based algorithms and prosody as possible complementary predictors. Wilschut et al. (2021) examine whether speech-based learning can benefit from such adaptive techniques. They show that vocabulary learning through speech and typing has comparable learning effects. In both cases the adaptive algorithms can improve recall performance, suggesting that the benefits of such systems are not limited to typing-based responses. In addition, lower response times and higher accuracy averages, show that speech-based RT algorithms are more efficient for learning than the traditional flashcard (Leitner) method.

Previous studies propose that spoken retrieval efforts reveal how well a learner has retained an item. Wilschut et al. (2023), showed that learners who recalled items correctly tended to speak louder, faster, and with a falling pitch. Conversely, incorrect replies were associated with low speaking speed, low vocal volume, and a rising pitch. Confirming these findings, Gustafsson et al. (2022), discovered a similar pattern in an eyewitness task: correct answers had a higher pitch, greater vocal energy in the lower frequency range of the voice, a higher speech rate, and shorter pauses. These findings show that PSFs like speed, loudness, and pitch indicate retrieval success, supporting use in adaptive learning. According to Wilschut et al. (2023), “some prosodic speech features are associated with accuracy and response latency for retrieval attempts, and speech feature-informed memory models make better predictions of future performance than models that only use accuracy and response latency” (p.255). This suggests that prosody can enhance the predictive power of adaptive learning models beyond traditional measures.

PSFs also entail information about the speaker’s confidence. Knowing a person’s trust in their recall is crucial for memory research since it offers valuable insights in predicting accuracy and understanding the involved memory processes. However, this relationship is not foolproof. People do not always sound certain when they are right, and they do not always sound unsure when they are wrong. This pattern fits with earlier findings of Goupil and Aucouturier (2021) and Wilschut et al. (2023), who claim that confidence and accuracy are interrelated but influenced by different prosodic features. Goupil and Aucouturier (2021) reported that confidence was expressed via specific intonation patterns (rise-fall pitch dynamics) and faster speech, whereas accuracy was predicted mainly by loudness. Moreover, they found that confidence and accuracy occur at different times during a response. Even when speakers were unaware of being correct, their voices already revealed it. Likewise, Jiang and Pell (2017) demonstrated that confident speech had a higher pitch range, higher amplitude, and faster speaking rate, while unconfident speech was characterized by a higher mean pitch, more pauses, and a slower tempo. Altogether, PSFs reflect both accuracy and confidence. Analyzing this relationship is especially useful for optimizing adaptive learning systems by basing repetition not only on accuracy but also on how certain a learner's response sounds.

Second language learners often struggle with native prosodic patterns, affecting speech comprehension. Pálvölgyi (2025) discovered that Hungarian learners of Spanish tend to overuse rising intonation at the end of utterances. These melodic rises are atypical for a native Spanish speaker, while the use of this pattern is common in Hungarian learners of Spanish. Jackson and O'Brien (2011) similarly showed how American L2 learners of German can use native-like prosodic features - like pauses, pitch accents, and word duration - to communicate meaning, but inconsistently. Because of these, native listeners often perceived their speech as hesitant or incomplete. According to Liu and Xu (2007), this was not surprising since questions typically have a rising intonation pattern while statements have falling contours. When L2 learners apply rising intonation incorrectly, it can unintentionally indicate insecurity or doubt. By highlighting that the perceived prosodic systems for L1 and L2 do not always align, speech-based AL models must adapt to this prosodic gap accordingly. Therefore, this study investigates the prosodic differences between L1 and L2 in accuracy and confidence.

Overall, adaptive learning systems use reaction times and accuracy as main indicators of memory strength. However, recent research suggested that subjective confidence and prosody offer complementary predictive value in this process. By verbally learning words, speech-based AL systems may be able to recognize speech prosody associated with confidence and accuracy and accurately estimate the strength of the memory representation. While this relationship has already been investigated on a word level (Wilschut et al., in press), the present study extends this to a sentence level. This study looks at PSFs (average pitch, pitch change, speaking speed, and average intensity) to see if they can be reliable predictors of subjective confidence and objective accuracy in simple S-V sentences. Essentially, the more information available about a response, the more accurately its memory strength can be estimated. The second aim explores whether there are differences between native (L1) and non-native (L2) speech. By examining PSFs and language differences, the study aims to improve adaptive learning systems by adding prosodic information for the MemoryLab model.

Optimizing speech-based learning systems has several advantages for education, especially for individuals with learning or visual impairments. Previous studies have shown that by reducing the error rate and improving recall and speed, for example, learners with dyslexia could experience great benefits (McTear et al., 2000; Wilschut et al., 2024). Where typical learners would score better than those with dyslexia during typing, this gap disappears when reacting verbally. These findings suggest that speech-based learning might offer a promising alternative to tying-based methods, particularly while trying to make education more effective and inclusive. Besides this, in typical learners, speech-based learning would have practical benefits, allowing for multitasking in hands-free contexts (e.g., during a workout). Learning a new language could occur while driving or exercising, hands-free. Understanding prosody's link to memory and recognition can improve adaptive learning designs.

Method

2.1 Participants

Forty-eight Dutch native speakers (M= 19.9, SD =1.632) aged 18-25 years participated after giving informed consent. The participants included 34 women and 14 men, who all were students from the University of Groningen, and reported no formal Italian knowledge and no speech or hearing impairments. In this experiment, a non-probability sampling called self-selection sampling was used. Individuals could self-enroll for the study on SONA, a participant recruitment and management platform. This experiment was approved by the Ethics Committee of the University of Groningen (PSY-2223-S-0257). A power analysis, based on the prior effect sizes and $\alpha = 0.05$ with power = 0.80, suggested a required sample size between 30 and 67.

2.2 Materials and Apparatus

Apparatus

Human Research.

The experiment was programmed in JavaScript and HTML5 and run online on a JATOS-MindProbe server (Lange et al., 2015). The experiment was completed in a computer laboratory setting using an Iiyama ProLite G2773HS monitor and Nedis Xyawyon GHST100BK headphones. Participants's responses were recorded by a Google Web Speech API (https://webaudio.github.io/web-speech-api/) and analyzed using Praat 6.2.07 (Boersma, 2006).

Materials

Description of Stimulus Materials.

Initially, a list of 40 simple Subject-Verb (S-V) phrases in the simple present tense was created (e.g., *the sun shines*). Afterward, the 40 phrases were translated into Dutch and Italian (see appendix). Using Narakeet (https://www.narakeet.com/), a platform generating text-to-speech items, the spoken versions of the test stimuli were generated from this text. All sentences were converted with the voice Vittorio for the Italian condition and Famke for the Dutch condition, and downloaded as .wav files. Each sentence used unique nouns and verbs with meaningwise logical structures. Additionally, words that had similar roots in L1 and L2 were avoided. The 40 generated sentences were separated into one set of five practice sentences, and two sets of 15 testing sentences. The other five sentences were excluded from the experiment. The sets of sentences were based on the order in which the sentences were written down.

2.3 Design and Procedure

The present study aimed to examine memory strength by investigating the connection between learning and speech prosody. The main independent variable was language, implemented in two within-subject conditions: a Dutch condition (L1) and Italian condition (L2). By the manipulation of L1 and L2, we measured the influence on speaking speed, pitch change, average pitch, intensity, accuracy, confidence, and reaction time in order to assess the strength of memory traces.

Before starting the experiment, participants read an information letter outlining the study's aims, the research team, and the possible risks. Next, they filled in an informed consent form and a background questionnaire about the participant's age, gender, language knowledge, and speech/hearing ability.

Thereafter, the participants were guided to the computer cubicles, where the experiment took place. The experiment consisted of study and test trials, in which the participants learned both Italian and Dutch sentences. Figure 1B illustrates the structure of these study and test trials. For example, the participant had to learn the translation of the sentence *"Il bambino mangia"*. In a study trial, along with its correct translation - *"De baby eet"* - the sentence appeared on the screen. Simultaneously, the correct pronunciation was played, after which the participants were asked to repeat the translation out loud.

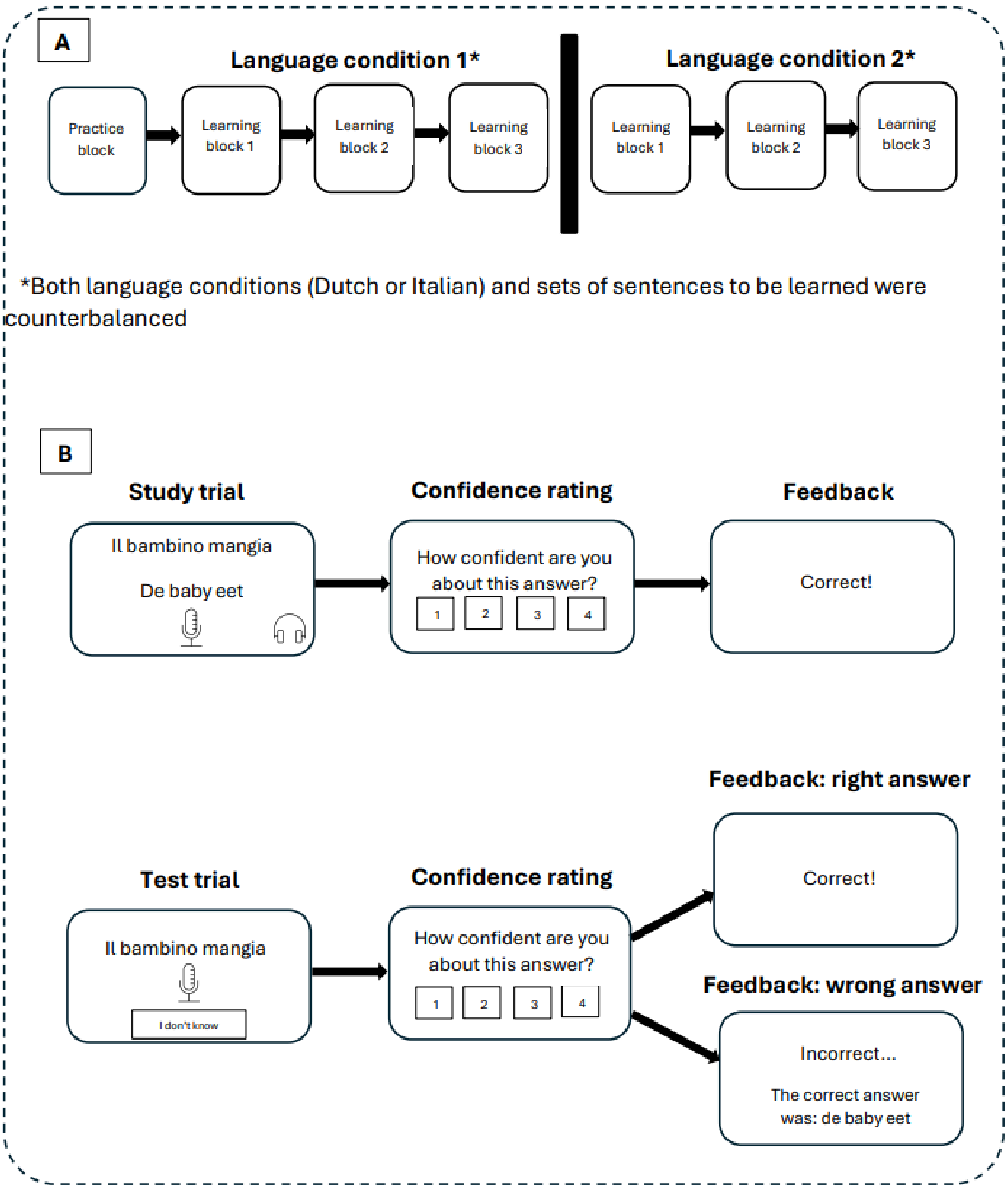
Following each response, participants rated their confidence on a 4-point Likert scale (*'1 = not confident', '2 = slightly confident', '3 = moderately confident', '4 = confident'*). Using Google Web Speech API, the participants' responses were transcribed to provide real-time feedback. If the answer was incorrect, the Google Web Speech API showed the right answer on the screen. After completing the study trials, the testing trials started, where neither the pronunciation nor the translation was provided. During each trial, the participants' answers were recorded.

In total, the experiment consisted of 7 blocks: one practice block followed by six learning blocks. In the practice block, all participants learned the same four phrases in L1 or L2, depending on which condition they were assigned to first. Each sentence was repeated 3 times: once as a study trial and twice as a testing trial. Each learning block contained five sentences, each repeated five times: one study trial and four testing trials. The six learning blocks consisted of the two language conditions, with three blocks per condition (see Figure 1A). Between conditions, the participants were allowed to take a break. The three blocks in the second condition were structured the same as in the first condition. To score an item's accuracy, a cut-off score was established. A response was considered correct if it differed from the correct transcription by no more than two characters.

The present study had a within-subject design where each participant was exposed to both conditions and both sets of sentences. However, the order of I-D & D-I and which set of sentences was studied first, was counterbalanced. This resulted in four unique experimental conditions a participant could be assigned to. The order in which the sentences were presented stayed consistent. The experiment was conducted entirely by an online platform, with no researcher intervening in the process of task completion.

Figure 1

The experimental design



Note. The concept plots show the trajectory of the experiments: panel A shows how the experiment is built, whereas panel B shows the experiment trials.

Results

API Confidence

Before examining the research aims, it was established how well the Google Web Speech API transcribed the study trial responses per participant per condition to see how certain it is it transcribed the item correctly. Assessing the transcription reliability of the APIs was crucial for ensuring valid interpretation of the learner responses. Because the Italian API demonstrated a higher average confidence (96.1%) than the Dutch API (80.0%), it might have distinguished between correct and incorrect responses more accurately.

Descriptive Statistics

Seven 2 (Language: Italian, Dutch), x 2 (Trial Type: Study, Testing) RM-ANOVA were conducted to investigate the effects of trial type and language on the participants' reported subjective confidence, reaction time, accuracy, speaking speed, intensity, pitch and, pitch change. Assumption checks for the RM-ANOVA showed that the Q-Q plots were all approximately normal. The analyses were conducted using Jasp (Version 0.19.3.0; JASP Team, 2025) and R (v.4.3.1, R Core Team, 2021).

For **Accuracy** (Panel 2A), mean scores were highest in the Dutch study trials ($M = .92$, $SD = .09$) and lowest in the Italian test trials ($M = .24$, $SD = .15$). Participants were more accurate in their native language, where performance dropped by 11 percentage points, from .92 in study to .81 in test trials. In contrast, in the Italian condition the drop was way steeper, with a drop of 51 percentage points from .75 to .24. This is confirmed by the significant main effects for language, $F(1, 47) = 595.28$, $p < .001$, $\eta^2_p = .927$, and trial type, $F(1, 47) = 315.38$, $p < .001$, $\eta^2_p = .870$ and the significant interaction effect, $F(1, 47) = 235.56$, $p < .001$, $\eta^2_p = .834$.

Panel 2B reveals that **Reaction Time** was highest in the Italian study trials (*M* = 2,869.27, *SD* = 616.17) and lowest in Dutch testing trials (*M* = 1,758.14, *SD* = 323.29). During study to testing trials the reaction times significantly dropped, $F\ (1, 47) = 52.41$, $p < .001$, η^2_p =.527. Additionally, speech onset was faster in the Dutch condition, $F\ (1, 47) = 59.64$, $p < .001$, $\eta^2_p = .559$, with a steeper improvement, $F(1, 47) = 37.93$, $p < .001$, $\eta^2_p = .447$.

The average expressed confidence was higher in the Dutch trials than in the Italian trials (Figure 2C). During study trials, a difference of 0.96 points between the Dutch (*M* = 3.83, *SD* = .26) and Italian (*M* = 2.87, *SD* = .73) condition was observed. This gap increased slightly in test trials, where the Dutch average was 3.72 points (*SD* = .36) and the Italian 2.50 points (*SD* = .73). This is supported by the significant main effect for language, $F(1, 47) = 204.68$, $p < .001$, η^2_p = .813, trial type, $F(1, 47) = 18.05$, $p < .001$, $\eta^2_p = .278$, and interaction effect, F(1, 47) = 6.20, p = .016, $\eta^2_p = .116$,

Responses were fastest in the Italian trials with no differences between study (*M* =1.71, *SD* = .30) and test trials (*M* = 1.71, *SD* = .31). Conversely, there was an increase of .21 syllables per second from study (*M* = 1.36, *SD* = .24) to test (M = 1.57, *SD* = .21) trials in L1. This was supported by the significant main effect of trial type, $F(1, 47) = 11.59$, $p = .001$, $\eta^2_p = .198$, language, F(1, 47) = 64.09, p < .001, $\eta^2_p = .577$, along with the interaction effect, $F(1, 47)$ = 16.76, $p < .001$, $\eta^2_p = .263$ (Figure 2D)

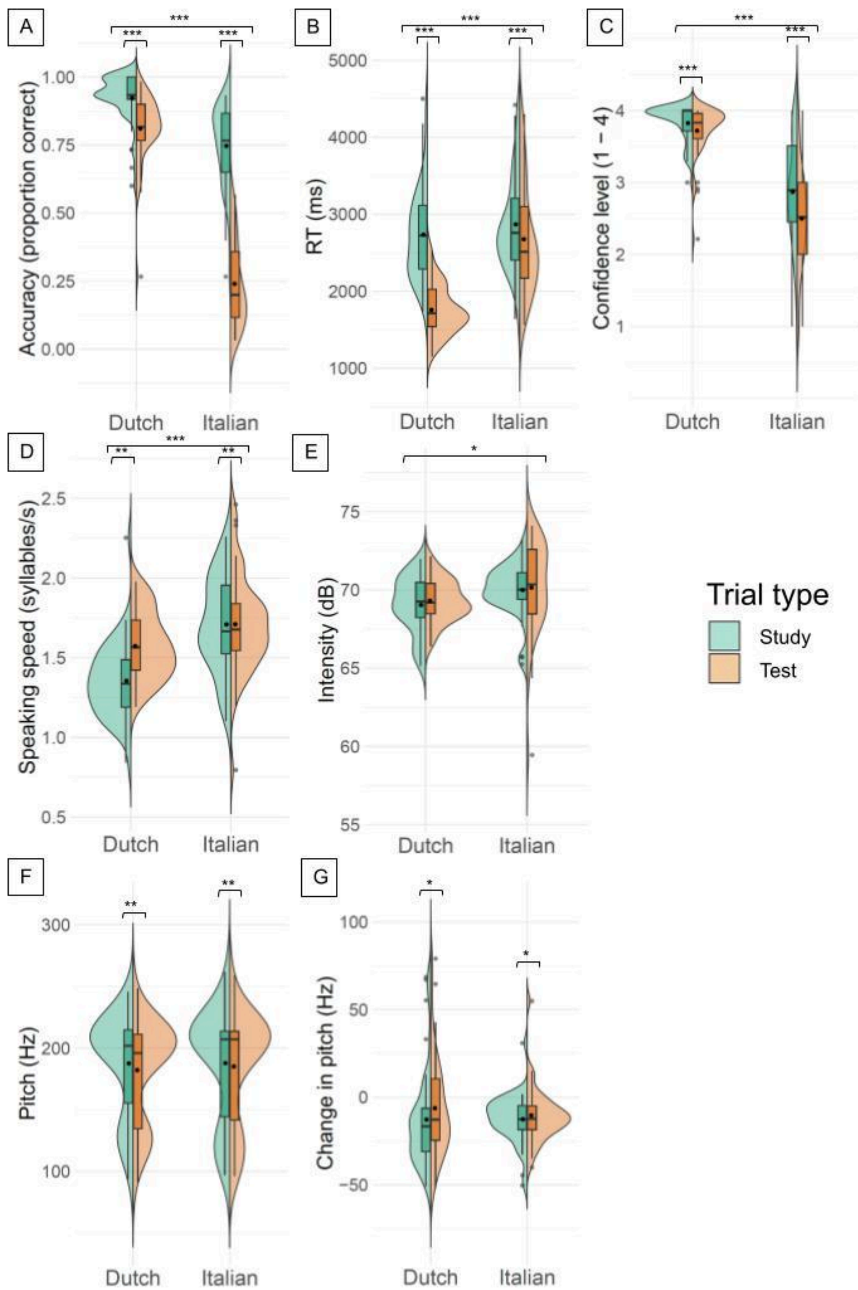
Panel 2E illustrates that responses were frequently louder in L1 than in L2, with no variation across trial types. On average in study trials, Italian speech was 70.00 dB (*SD* = 1.74), and in Dutch speech 69.04 dB (*SD* = 1.83). In the testing trials, Italian responses (*M* = 70.15, *SD* = 2.97) remained louder than Dutch responses (*M* = 69.31, *SD* = 1.36). The 2 x 2 RM-ANOVA confirmed the significant main effect for language, $F(1, 38) = 5.09$, $p = .030$, $\eta^2_p = .118$, but no effect for trial type and interaction was found.

Generally, pitch was higher in study trials than in testing trials. This decline is demonstrated by the significant main effect for trial type, F(1, 38) = 8.04, p = .007, $\eta^2_p = .175$, where in L1 pitched dropped by 5.43 Hz, in L2 this difference was 2.75 Hz (Figure 2F). Additionally, no main effects of language, F(1, 38) = 0.00, p = .961, $\eta^2_p < .001$ and interaction, F(1, 38) = 0.93, p = .341, $\eta^2_p = .024$ were found.

Pitch change was computed as the difference between the average pitch of the last five and the first five segments of each response. Especially in the Dutch condition, pitch change was bigger in study trials compared to test trials. The significant main effect of trial type, $F(1, 38) = 4.19$, $p = .048$, $\eta^2_p = .099$, showed that in L1 pitch change reduced with 6.49 Hz, from study ($M = -12.71$, $SD = 28.36$) to test ($M = -6.22$, $SD = 28.34$). In Italian, this drop was 2.24 Hz (study: $M = -12.55$, $SD = 14.08$; test: $M = -10.31$, $SD = 16.15$). Subsequently, no main effects of language, $F(1, 38) = 0.38$, $p = .540$, $\eta^2_p = .010$ nor interaction, $F(1, 38) = 1.41$, $p = .242$, $\eta^2_p = .036$, was observed (Figure 2G).

Figure 2

Descriptive Boxplots of all Dependent Variables Split by Trial Type and Condition



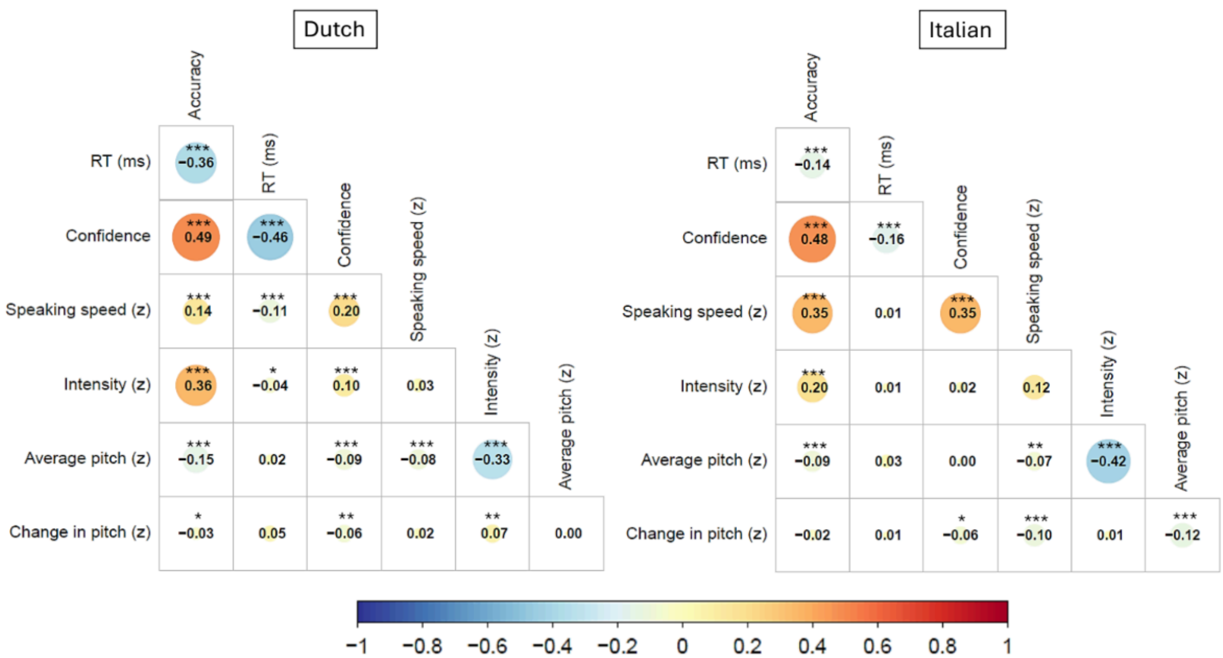
Note. For each dependent variable, the means were calculated and added to the data set. Split violin plots were computed using data from correct answers only, except for accuracy. Accuracy plots were computed using both incorrect and correct answers. Split violin plots display kernel density estimates of the distribution. Violin shapes may slightly extend beyond the theoretical bounds for accuracy (0 – 1) and confidence (1 – 4), due to smoothing. The figure shows the significant effects of language or trial type (*$p < .05$; **$p < .01$, ***$p < .001$).

Main Analysis

Dutch and Italian Testing Trials Correlations

Figure 3

Pairwise Correlation Matrix of Dutch Test Trials



*Note.**$p < .05$, **$p < .01$, ***$p < .001$. The matrix in Figure 3 shows the results of the pairwise correlation analysis between the dependent variables for both the Dutch and Italian test trials, including both correct and incorrect trials. Before the analysis was performed, the values of the acoustics were standardized. This was especially important for the pitch features to account for sex differences since in general male voices are lower than female voices.

Dutch Condition.

First, a significant negative correlation between reaction time and both subjective confidence (*r* = -0.46, $p < .001$) and accuracy (*r* = -0.36, $p < .001$) was observed, indicating responses with a quicker onset often were more likely to be confident and correct. Additionally, subjective confidence has a moderate positive correlation with accuracy (r = 0.49, $p < .001$).

Across trials, reaction time tended to be lower when responses were faster (*r = -0.11, p < .000)*. Even so, lower reaction times were weakly associated with louder responses (r = -.04, p =.039). No significant correlations were found between reaction time, average pitch (*r* = .02, *p* = .385), and pitch change (*r* = .05, *p* =.019).

On the contrary, higher confidence in responses was associated with a faster pace(*r* = .20, *p* = .000) and increased loudness (*r* = .10, *p* = .000). Additionally, significant correlations were found between subjective confidence and pitch features. Higher confidence was negatively associated with a lower pitch (r = -.09, p = .000) and less pitch variation (*r* = -.06, *p* =.005).

The prosodic features of speaking speed (*r* = .14, *p* = .000) and intensity (*r* = .36, *p* = .000) are significantly correlated to accuracy. Average pitch was negatively significantly correlated with accuracy, with (*r* = -.15, *p* =.000), with incorrect responses being associated with higher pitch. However, a nonsignificant association was found between accuracy and pitch change (*r* = -.03, *p* =.181). Overall, these findings suggest that accurate responses were louder, faster, and lower-pitched.

Italian Condition.

The strongest significant correlation was revealed between subjective confidence and accuracy (*r* = .48, $p < .001$). Reaction time was negatively correlated with both subjective confidence (*r* = –.16, $p < .001$), and objective accuracy (*r* = –.14, $p < .001$), indicating faster responses were more likely to be correct and confident. There was no consistent relationship between participants' response time, and their speaking speed (*r* = .001, *p* = .680), loudness (*r* = .011, *p* = .657), average pitch (*r* = .034, *p* = .176) and pitch change(*r* = .008, *p* = .750).

Confidence, however, did have a significant positive correlation with speaking speed (*r* = .35, *p* = .000). Moreover, faster responses tended to have higher confidence ratings. A weak but significant effect was seen for change in pitch (*r* = -.06, *p* = .046). Subjective confidence did not have significant correlations with average intensity (*r* = .02, *p* = .437) and average pitch (*r* = -.00, *p* = .842)

Correct responses were positively associated with a higher speaking speed (*r* = 0.35, *p* = .000), and intensity (*r* = .20, *p* = .000). A significant negative correlation was also found between accuracy and average pitch (r = -.09, p = 1e-04), indicating that responses with a higher pitch often reflected incorrectness.

Comparison Prosodic Predictors in L1 and L2

Overall, prosody may be a more reliable predictor in the native language, as shown by the broader range of significant correlations between prosodic variables and behavioral measures. In both languages speaking speed and intensity showed significant positive correlations with confidence and accuracy. However, the strength of these relationships differed. Speaking speed had stronger correlations with confidence (*r* = .35, *p* = .000) and accuracy (*r* = 0.347, *p* = .000) in L2, compared to L1 (*r* = .200, *p* = .000; *r* = .144, *p* = .000, respectively).

This suggests that in L2 speaking speed would be a better marker of subjective confidence and objective accuracy than in L1. Nevertheless, for intensity, a stronger significant correlation between subjective confidence (*r* = .104, *p* = .000) and objective accuracy (*r* = .360, *p* = .000) was seen, while in the Italian condition, only a significant correlation was found for objective accuracy ($r = .195$, $p = .000$). Therefore, loudness may be more informative in the native language.

Next in both L1 ($r = -.148$, $p = .000$) and L2 ($r = -.088$, $p = 1e-04$) a negative correlation between average pitch and accuracy was seen. As the average pitch in the voice goes down, the accuracy tends to go up. Noticeably, the effect is stronger in the Dutch condition but the pattern is similar. A moderate correlation in L1 ($r = -.33$, $p = .000$) and L2 ($r = -.42$, $p = .000$) between intensity and average pitch was also observed.

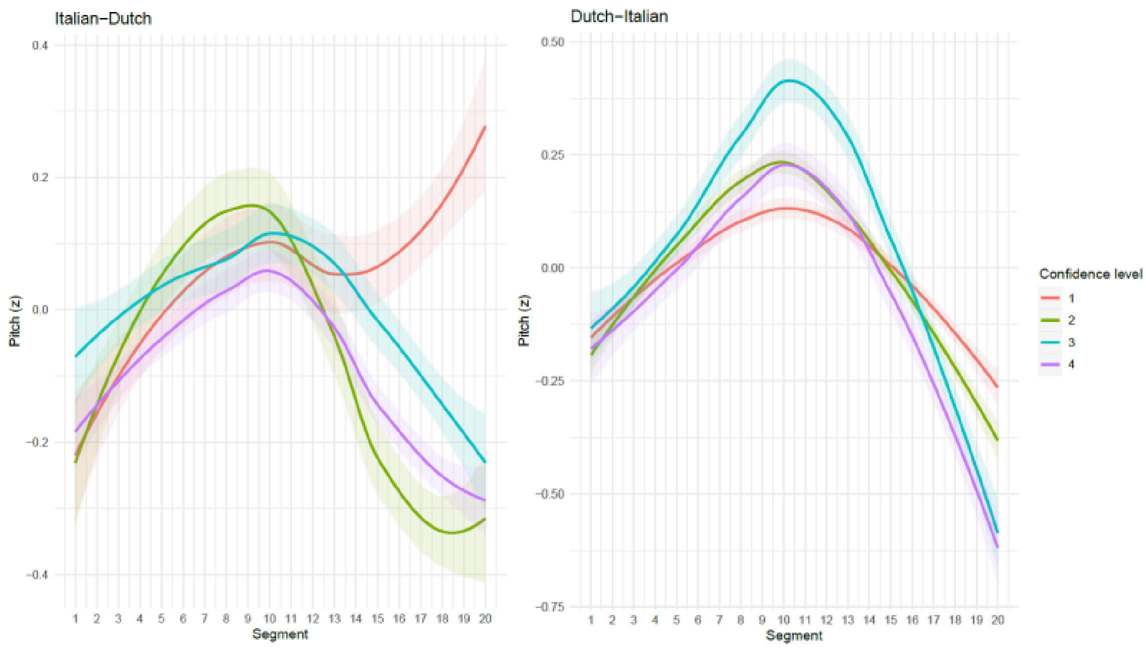
Pitch Segment Analysis

Figures 4 and 5 visualized the pitch trajectory of utterances across the confidence and accuracy levels. The plot included all the correct and incorrect answers in the testing trials in both conditions. In Figure 4, the general course of an utterance had a rise-and-fall contour for both the Dutch and Italian conditions. Dissimilarly, participants who reported the lowest confidence level in the Dutch condition showed a rising pitch in their response, a pattern often associated with uncertainty (Liu and Xu, 2007). In the Italian condition, the pitch trajectories of level 3 and 4 confidence were similar. However, the pitch of confident responses (level 4) was lower than moderately confident responses (level 3)

Figure 5 illustrates the average pitch between correct and incorrect responses. Correct responses in L1 followed a clear rise-fall pattern and had a lower overall pitch, while incorrect responses ended with a rising pitch. In L2, both correct and incorrect responses had a rise-and-fall pattern, but correct responses had a higher pitch peak and steeper decline. Incorrect responses remained relatively flat.

Figure 4

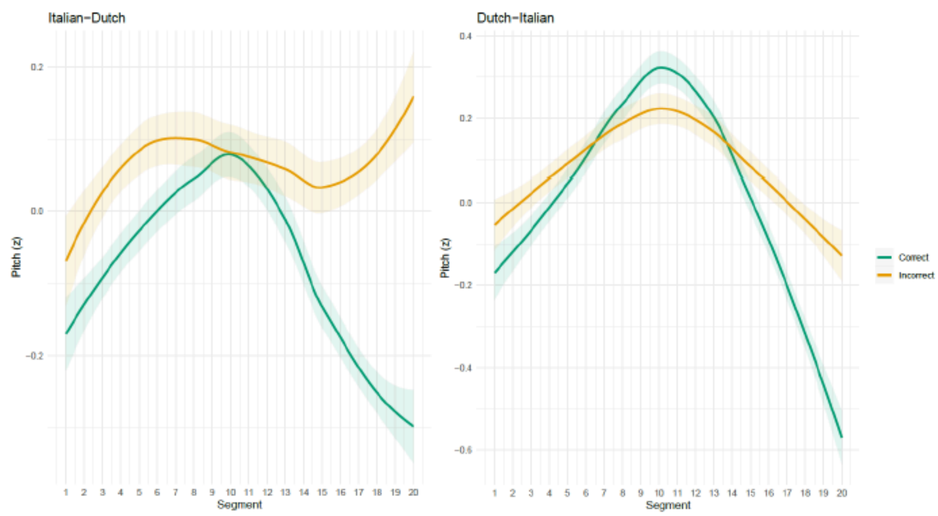
Standardized pitch across segments by condition (Italian-Dutch, Dutch-Italian) and confidence level (1-4)



Note. The panels show the standardized pitch for Italian and Dutch responses, respectively. Data from the study and testing trials were included.

Figure 5

Standardized pitch across segments by condition (Italian-Dutch, Dutch-Italian) and response (correct vs. incorrect)



Note. The panels show the standardized pitch for Italian and Dutch responses, respectively. Data from the learning and testing trials were included.

Discussion

While previous studies commonly focused on accuracy and reaction time in memory research (Pavlik & Anderson, 2008; Sense et al., 2021; Van Rijn et al., 2019), recent studies added prosody to see whether it can provide additional information about the strength of a memory trace. Based on research by Wilschut and colleagues (in press), the present study aimed to replicate and extend their findings by investigating whether PSFs - more specifically, average pitch, speaking speed, pitch change and intensity - could predict both accuracy and confidence in simple S-V sentences when verbally recalled. Our first research aim examined the relationship between prosodic prosodic-behavioral relationships in simple S-V sentences, and the second aim focused on comparing these across the two language conditions (L1 vs. L2).

Consistent with the findings of Wilschut et al. (in press), our study found that louder responses were more likely to be correct and confident, to a lesser extent. Importantly, average pitch was moderately negatively correlated with intensity. This indicates that louder responses were also more likely to be lower in pitch. Given that intensity was more related to accuracy, the co-occurrence of lower pitch and louder speech reflects a prosodic indicator of successful memory access. Next, we found that speaking speed was positively correlated with confidence and accuracy. Despite speaking speed having a stronger correlation with confidence than pitch slope, Wiltschut et al. (in press) suggested that pitch slope would be a better predictor when examining unique contributions in structural equation modeling. Overall, this suggests that speaking speed and pitch slope both convey confidence, while intensity is most consistently associated with accuracy. However, Wilschut et al. (in press) found higher correlations between pitch slope, confidence, and accuracy compared to the present study. This disparity may be due how pitch slope was calculated in the studies, potentially impacting the strength of our correlations.

According to Jiang and Pell (2017) and Wilschut et al. (in press), correct responses tend to have a lower pitch and more stable pitch trajectory. This was supported by our findings in which we observed a similar pattern. While incorrect responses were associated with a higher pitch, a lower pitch was associated with correct responses. Even though this negative relation was small, its significance in both conditions indicates that it can function as a reliable marker of accuracy. This was also revealed in our pitch segment analysis where more confident and correct responses exhibited a lower pitch overall. Additionally, the pitch segment analysis (Figures 4 and

5) illustrates the pitch slope trajectory of different levels of confidence and accuracy. Similarly to what Liu and Xu (2007) reported, the lowest confidence and inaccurate responses displayed a rising pitch at the end of a response, especially in the Dutch condition. In a study by Goupil and Aucouturier (2021), they found that especially rise-fall intonation patterns were indicative of high confidence, while loudness and duration reflected accuracy. Complementing this, Jiang and Pell (2027) highlighted that confident responses showed greater pitch variability, a faster speech rate, and louder speech, whereas unconfident responses were marked by a slower speech rate, more pauses and a higher mean pitch. These findings align with the current research and suggest the dual nature of prosody; a means for expression and a marker of underlying mental processes.

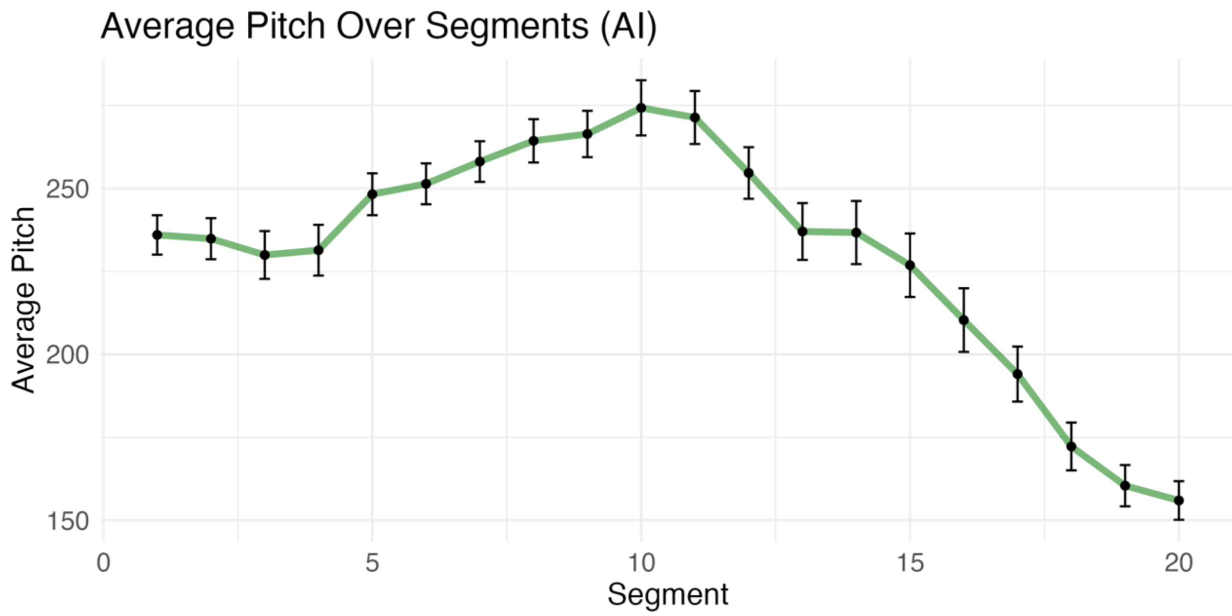
Regarding speaking speed, however, our results differed from those of Wilschut et al. (in press). Whereas they found that slower speech was associated with greater accuracy and confidence, the current study found the contrary: faster speech was associated with higher confidence. This difference between the two studies is most likely due to differences in response style (e.g., single words vs. simple sentences) rather than clashing findings. Together with the fact that the present study also showed stronger correlations between the measures, this indicates that there is more prosodic variation in simple sentences compared to single words. Thus, by adding prosodic dependent variables a more complete picture of the responses emerges since distinct PSFs were associated with accuracy and confidence. Speech-based adaptive learning systems could benefit from integrating prosody for memory estimation (Wilschut et al., 2023).

Secondly, we examined whether the obtained prosodic-behavioral relationships differed between L1 and L2. Overall results showed that particularly in the native language the relationship between the behavioral measures and prosodic speech features was broader and stronger. This suggests that native-language prosody more reliably reflects accuracy and confidence. Prior research offers useful insights into this difference. Learners of a second language often have problems with producing native-like rhythmic and intonational patterns partly as a result of transfer effects from their native language (Van Maastricht, 2018; Van Maastricht et al., 2016; Pálvölgyi, 2025; Swerts et al., 2002). In particular, the Italian language is a non-plastic language, whereas the Dutch language is plastic, indicating that the different languages differ in syllable stress patterns. As a result, prosodic mismatches are created, contributing to diminished comprehensibility (Munro & Derwing, 1999). Radzikowski et al. (2021) argue that automatic speech recognition (ASR) systems sometimes have difficulties in recognizing speech from a non-native speaker due to the aforementioned. This implies that L2 speech, particularly its prosodic features, is less robust about cognitive and metacognitive outcomes. However, neural style transfer can account for this gap by modifying non-native speech so that it will more accurately resemble native speech (Radzikowski et al., 2021).

Yet, in the current study, API confidence was higher for the Italian language. This possibly indicates that the weaker results for L2 also can have a different cause. We compared the average pitch trajectory of the Italian audiofiles (Figure 6) to the responses of the participants (Figures 4 and 5) in the Italian condition. The three graphs illustrate a similar pattern, indicating that participants mimicked the original audiofile. Therefore one could argue that participants' dependence on imitating the audio model is a plausible cause of the low prosodic variation in the Italian condition and high API confidence. Additionally, this also could explain the steep drop in accuracy from study to test trials. Because in study trials, participants could rely on imitation, whereas in the test trials, they had to fully generate the prosody themselves. The difficulty of the Italian condition could also be a confounding factor.

Figure 6

Pitch Trajectory of Generated Sentences by a Text-To-Speech Generator



Note. Figure 5 shows the pitch trajectory of the self-generated responses by a text-to-speech generator.

Remarkably, the only variable with a stronger correlation in the Italian condition compared to the Dutch condition was speaking speed. Participants spoke faster when they were confident and correct in the non-native language. But why is the correlation stronger in L2? According to Peters et al. (2024), speaking speed and fluency are more sensitive to task difficulty in L2 than in L1. This was likely because speaking a foreign language requires more mental effort, whereas in the native language, this process is largely automatic. Speaking speed may thus react more directly to accuracy and confidence. It can therefore be viewed as a more behaviorally informative and sensitive measure for speech-based adaptive learning systems in L2 than in L1.

While this study provides valuable insights into the topic of memory research, some limitations also had to be addressed as they may impact the generalizability, validity, and reliability of the results. First of all, the memory task had a fixed item scheduling. This structure contrasts with the fundamental ideas of adaptive learning, which modifies repetition and content based on the skills and needs of the participants (Van der Velde, 2023). Thus, in our study all presented items were tested the same amount of times and in the same order during the memory task. Especially in the Dutch condition, participants mentioned that they knew which word they had to recall by the order in which they were presented. Rather than remembering the right translation, they relied on the order of the sentences. This downside of the fixed schedule might be reflected in our confidence and accuracy scores. Even though the items were longer in the present study, higher accuracy and confidence were found compared to the study of Wilschut and colleagues (in press).

Secondly, our participant pool solely consisted of students. Regarding generalizability, this could pose some problems since students often tend to be younger and more educated than nationwide samples. Memory processing and speech processing can be influenced by these demographic characteristics. This shows the limited generalizability of our results to other backgrounds. We also observed that the API confidence on average was lower in the Dutch condition. In the Dutch condition more often it was the case that the transcription of the sentences was incorrect even though the participant said them correctly. This might have interfered with our data's accuracy scores.

Future research may aim to create a more balanced level of difficulty across conditions and account for prosodic L2 limitations by implementing gradual scaffolding. Introducing participants to single-word items first, and then gradually progressing to simple S-V sentences, may help reduce performance discrepancy between conditions. Additionally, this correlational study cannot infer causality.

Conclusion

How can prosody add to traditional measures like reaction time and accuracy in speech-based adaptive learning models? The present study explored this by examining whether PSFs, specifically speaking speed, average pitch, pitch change, and average intensity can predict confidence and accuracy in memory tasks involving simple S-V sentences. Expanding on research by Wilschut et al. (in press), these relationships were examined at a sentence level and included a L2 condition to compare prosodic patterns between one's native and acquired language. In conclusion, we found that on a low to moderate level faster, louder, and lower-pitched speech was associated with both more accurate and more confident utterances. In both language conditions, the most reliable predictors were speaking speed, and intensity. While the majority of the correlations were stronger in the native language, speaking speed was the exception. In L2, speaking speed showed in both accuracy and confidence the strongest association. This suggests that speech-based adaptive learning models might need to take language-specific prosodic patterns into account, possibly due to L1 prosodic transfer. Furthermore, the consistent louder-lower speech pattern observed in accurate responses suggests a stable prosodic marker of accurate retrieval. These findings support adding prosody to adaptive models to improve predicting power. Future research should focus on optimizing these models by improving response evaluation, personalized feedback, and supporting broader target groups.

References

Boersma, P.: Praat: doing phonetics by computer (2006). http://www.praat.org/

Chen, Y., Li, X., Liu, J., & Ying, Z. (2017). Recommendation System for Adaptive Learning. *Applied Psychological Measurement*, *42*(1), 24–41. https://doi.org/10.1177/0146621617697959

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 53, pp. 63–147). Elsevier. https://doi.org/10.1016/S0079-7421(10)53003-2

Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology. In *Teachers College Press eBooks*. https://doi.org/10.1037/10011-000

Goupil, L., & Aucouturier, J. (2021). Distinct signatures of subjective confidence and objective accuracy in speech prosody. Cognition, 212, 104661. https://doi.org/10.1016/j.cognition.2021.104661

Gustafsson, P. U., Laukka, P., & Lindholm, T. (2022). Vocal characteristics of accuracy in eyewitness testimony. *Speech Communication*, *146*, 82–92. https://doi.org/10.1016/j.specom.2022.12.001

Guyer, J.J., Briñol, P., Vaughan-Johnston, T.I. *et al.* Paralinguistic Features Communicated through Voice can Affect Appraisals of Confidence and Evaluative Judgments. *J Nonverbal Behav* 45, 479–504 (2021) https://doi.org/10.1007/s10919-021-00374-2

Hellbernd, N., & Sammler, D. (2016). *Prosody conveys speaker's intentions: Acoustic cues for speech act perception.* Journal of Memory and Language, 88, 70–86. https://doi.org/10.1016/j.specom.2017.01.011

Hoyte, K. J., Brownell, H., & Wingfield, A. (2009). Components of Speech Prosody and their Use in Detection of Syntactic Structure by Older Adults. *Experimental Aging Research*, *35*(1), 129–151. https://doi.org/10.1080/03610730802565091

Jackson, C. N., & O'Brien, M. G. (2011). The Interaction between Prosody and Meaning in Second Language Speech Production1. *Die Unterrichtspraxis/Teaching German*, *44*(1), 1–11. https://doi.org/10.1111/j.1756-1221.2011.00087.x

JASP Team (2024). JASP (Version 0.18.3)[Computer software]

Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, *88*, 106–126. https://doi.org/10.1016/j.specom.2017.01.011

Pálvölgyi, K. B. (2025). 7 The prosody of word stress realization as a possible sign of insecurity in the spontaneous speech of Hungarian learners of Spanish. In *De Gruyter eBooks* (pp. 157–174). https://doi.org/10.1515/9783111248196-008

Kraus, M. W. (2017). Voice-only communication enhances empathic accuracy. *American Psychologist*, *72*(7), 644–654. https://doi.org/10.1037/amp0000147

Lange, K., Kühn, S. & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online

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. https://doi.org/10.1177/0956797613504302

Liu, F., & Xu, Y. (2007). Question intonation as affected by word stress and focus in English. *ICPhS*. http://www.homepages.ucl.ac.uk/~uclyyix/yispapers/Liu_Xu_ICPhS2007.pdf

McTear, M., Conn, N., & Phillips, N. (2000). Speech recognition software: A tool for people with dyslexia. *6th International Conference on Spoken Language Processing (ICSLP*

2000), vols. 4, 81-84-0. https://doi.org/10.21437/ICSLP.2000-756

MemoryLab. (2025, 16 mei). *How it Works | MemoryLab*. https://www.memorylab.nl/how-it-works/

Mettler, E., Massey, C., & Kellman, P. J. (2011). Improving Adaptive Learning Technology through the Use of Response Times. *Cognitive Science*, *33*(33). https://escholarship.org/content/qt2xs4n8wz/qt2xs4n8wz.pdf?t=op2jw

Munro, M. J., & Derwing, T. M. (1999). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, *49*(s1), 285–310. https://doi.org/10.1111/0023-8333.49.s1.8

Narakeet - Easily Create Voiceovers and Narrated Videos Using . . . (n.d.). Narakeet. https://www.narakeet.com/app/text-to-audio/?projectId=6edb06e4-c1f1-47e0-9530-3ce8aa834702

Pálvölgyi, K. B. (2025). 7 The prosody of word stress realization as a possible sign of insecurity in the spontaneous speech of Hungarian learners of Spanish. In *De Gruyter eBooks* (pp. 157–174). https://doi.org/10.1515/9783111248196-008

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal Of Experimental Psychology Applied*, *14*(2), 101–117. https://doi.org/10.1037/1076-898x.14.2.101

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing Emotions in a Foreign Language. *Journal Of Nonverbal Behavior*, *33*(2), 107–120. https://doi.org/10.1007/s10919-008-0065-7

Peters, J., Frank, M., & Rohloff, T. (2024). Effects of task type and task difficulty on oral fluency in native and non-native speech. *Speech Prosody*, 364–368.

https://doi.org/10.21437/speechprosody.2024-74

Radzikowski, K., Wang, L., Yoshie, O., & Nowak, R. (2021). Accent modification for speech recognition of non-native speakers using neural style transfer. *EURASIP Journal On Audio Speech And Music Processing*, *2021*(1). https://doi.org/10.1186/s13636-021-00199-3

R Core Team. (2024). R: A Language and Environment for Statistical Computing [Computer software]. R Foundation for Statistical Computing. hhttps://www.R-project.org/

Sense, F., Van der Velde, M., & Van Rijn, H. (2021). Predicting university students' exam performance using a model-based adaptive fact-learning system. *Journal of Learning Analytics*, *8*(3), 155–169. https://doi.org/10.18608/jla.2021.6590

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. Journal of Phonetics, 30(4), 629–654. https://doi.org/10.1006/jpho.2002.0178

Van den Broek, G., Takashima, A., Wiklund-Hörnqvist, C., Wirebring, L. K., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the "testing effect": A review. *Trends in Neuroscience And Education*, *5*(2), 52–66. https://doi.org/10.1016/j.tine.2016.05.001

Van der Velde, M. (2023, June, 19). Adaptive Learning in Educational Practice: More than an Engineering Challenge. MemoryLab .https://www.memorylab.nl/blogs/adaptive-learning-in-educational-practice-more-than-an-engineering-challenge/

Van Maastricht, L. (2018). *Second Language Prosody : Intonation and rhythm in production and perception*.

https://pure.uvt.nl/ws/files/25772689/Van_Maastricht_Second_09_05_2018.pdf

Van Maastricht, L., Krahmer, E., & Swerts, M. (2016). Prominence Patterns in a Second Language: Intonational Transfer From Dutch to Spanish and Vice Versa. Language Learning, 66(1), 124–158. https://doi.org/10.1111/lang.12141

Van Rijn, D., Van Maanen, L., & Van Woudenberg, M. (2009). Passing the test: Improving Learning Gains by Balancing Spacing and Testing Effects. In A. Hoses, D. Peebles, & R. Cooper (Eds.), *Proceedings of the 9th International Conference on Cognitive Modeling* (pp. 108-114). Article 187.

Wilschut, T., Sense, F., Scharenborg, O., & Van Rijn, H. (2023). Improving Adaptive Learning Models Using Prosodic Speech Features. In N. Wang, G. Rebolledo-Mendez, N. Matsuda, O. C. Santos, & V. Dimitrova (Red.), *Artificial Intelligence in Education* (Vol. 13916, pp. 255-266). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9_21

Wilschut, T., Sense, F., & Van Rijn, H. (2024). Modality Matters: Evidence for the Benefits of Speech‐Based Adaptive Retrieval Practice in Learners with Dyslexia. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12769

Wilschut, T., Sense, F. & Van Rijn, H. (in press) The sound of recall: Cognitive and Metacognitive Markers of Memory Retrieval Performance in Speech Prosody.

Wilschut, T., Sense, F., Van Der Velde, M., Fountas, Z., Maaß, S. C., & Van Rijn, H. (2021). Benefits of Adaptive Learning Transfer From Typing-Based Learning to Speech-Based Learning. *Frontiers in Artificial Intelligence*, *4*. https://doi.org/10.3389/frai.2021.780131

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

Xu, Y. (2011b). The robustness and complexity of speech communication: Functional contributions of the jaw, tongue, and lips. *Journal of the Acoustical Society of America*, *130*(2), 1330–1344. https://doi.org/10.1121/1.3609777

Appendix

Sentence Stimuli Used in the Experiment

The following sentences were used in the experiment. In the table, you see the English sentences that we translated into Dutch and Italian. The simple sentences were learned by the participants in a memory task. Additionally, the number of syllables and frequency of words were calculated.

ID	Dutch Sentence	Number of Syllables	English Translation	Italian Sentence	Number of Syllables	Frequency
1	De nicht glimlacht	4	The niece smiles	La nipote sorride	7	229,594 + 24,926 = **254,520**
2	De baby eet	4	The baby eats	Il bambino mangia	6	1,517,971 + 115,986 = **1,633,957**
3	De kikker springt	4	The frog jumps	La rana salta	5	82,324 + 115,712 = **198,036**
4	Het meisje leest	4	The girl reads	La ragazza legge	6	3,827,400 + 254,2 9 = **4,081,690**

ID	Dutch Sentence	Number of Syllables	English Translation	Italian Sentence	Number of Syllables	Frequency
5	De ober kookt	4	The waiter cooks	Il cameriere cucina	8	100,161 + 72,034 = **172,195**
6	De broer rijdt	3	The brother drives	Il fratello guida	6	2,456,241 + 488,001 = **2,944,242**
7	De trein wacht	3	The train waits	Il treno aspetta	6	731,544 + 8,342,895 = **9,074,439**
8	De zus loopt	3	The sister walks	La sorella cammina	7	1,310,331 + 1,370,702 = **2,681,033**
9	De hond speelt	3	The dog plays	Il cane gioca	5	1,686,507 + 869,666 = **2,556,173**

ID	Dutch Sentence	Number of Syllables	English Translation	Italian Sentence	Number of Syllables	Frequency
10	De schilder tekent	5	The painter draws	Il pittore disegna	7	104,964 + 88,727 = **193,691**
11	De nacht begint	4	The night starts	La notte inizia	6	2,044,390 + 1,242,413 = **3,286,803**
12	De brandweerman spreekt	5	The firefighter speaks	Il pompiere parla	6	50,538 + 726,742 = **777,280**
13	De jongen rent	4	The boy runs	Il ragazzo corre	6	4,357,021 + 111,595 = **4,468,616**
14	De man liegt	3	The man lies	L'uomo mente	4	14,038,145 + 595,937= **14,634,082**

ID	Dutch Sentence	Number of Syllables	English Translation	Italian Sentence	Number of Syllables	Frequency
15	Het dier reist	3	The animal travels	L'animale viaggia	6	281,046+81,638= **362,684**
16	De vrouw komt	3	The woman comes	La donna viene	5	8,216,664+13,319,178 = **21,535,842**
17	De oma bakt	4	The grandma bakes	La nonna inforna	6	728,571+28,585 =**757,156**
18	De prijs verandert	5	The price changes	Il prezzo cambia	5	866,007 +511,784= **1,377,791**
19	De deur kraakt	3	The door creaks	La porta scricchiola	6	2,474,764+14,864= **2,489,628**
20	Het bot breekt	3	The bone breaks	L'osso si rompe	5	145,211+ 242,171= **387,382**

ID	Dutch Sentence	Number of Syllables	English Translation	Italian Sentence	Number of Syllables	Frequency
21	De klok tikt	3	The clock ticks	L'orologio ticchetta	7	23,897+ 38,418= **62,315**
22	De boom bloeit	3	The tree blooms	L'albero fiorisce	6	20,810+16,694= **37,504**
23	De dochter schreeuwt	4	The daughter shouts	La figlia grida	5	2,193,031+ 91,472= **2,284,503**
24	De vriend lacht	3	The friend laughs	L'amico ride	5	4,914,311+212,900= **5,127,211**
25	De vogel zingt	4	The bird sings	L'uccello canta	5	322,666+166,936= **489,602**
26	De groep wint	3	The group wins	Il gruppo vince	5	747,323+478,854= **1,226,177**
27	De haai zwemt	3	The shark swims	Lo squalo nuota	5	94,444+ 42,992= **137,436**

ID	Dutch Sentence	Number of Syllables	English Translation	Italian Sentence	Number of Syllables	Frequency
28	De zon schijnt	3	The sun shines	Il sole splende	5	686,723 + 414,366 = **1,101,089**
29	De familie betaalt	6	The family pays	La famiglia paga	6	3,041,202 + 484,571 = **3,525,773**
30	De leeuw slaapt	3	The lion sleeps	Il leone dorme	6	147,041+ 570,097 = **717,138**
31	De maan bestaat	4	The moon exists	La luna esiste	6	420,998 + 929,123 = **1,350,121**
32	De stoel valt	3	The chair falls	La sedia cade	5	512,012 + 1,529,862 = **2,041,874**
33	De muis verdwijnt	4	The mouse disappears	Il topo scompare	6	111,367 + 200,780 = **312,147**

ID	Dutch Sentence	Number of Syllables	English Translation	Italian Sentence	Number of Syllables	Frequency
34	De kat snurkt	3	The cat snores	Il gatto russa	5	528,477 + 17,837 = **546,314**
35	De buurman begrijpt	5	The neighbour understands	Il vicino capisce	7	132,634 + 1,026,083 = **1,158,717**
36	Het hout brandt	3	The wood burns	Il legno brucia	5	235,768 + 161,447 = **397,215**
37	De plant groeit	3	The plant grows	La pianta cresce	5	116,855 + 158,475 = **275,330**
38	De docent vermenigvuldigd	8	The teacher multiplies	L'insegnante moltiplica	8	34,073 + 3,659 = **37,732**
39	Het vat explodeert	5	The barrel explodes	Il barile esplode	7	19,049 + 24,011 = **43,060**

ID	Dutch Sentence	Number of Syllables	English Translation	Italian Sentence	Number of Syllables	Frequency
40	Het schip zinkt	3	The ship sinks	La nave affonda	6	1,152,771 + 28,585 = **1,181,356**