



Recall of Own Speech Following Interaction With L2 Speakers: Is There Evidence for Fuzzier Representations?

Frances Baxter, Ghada Khattab*, Andreas Krug and Fengting Du

Speech and Language Sciences Section, School of Education, Communication and Language Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

OPEN ACCESS

Edited by:

Joseph V. Casillas,
Rutgers, The State University of New
Jersey, United States

Reviewed by:

Laura M. Morett,
University of Alabama, United States
Stefano Coretta,
University of Edinburgh,
United Kingdom

*Correspondence:

Ghada Khattab
ghada.khattab@ncl.ac.uk

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 20 December 2021

Accepted: 16 June 2022

Published: 27 July 2022

Citation:

Baxter F, Khattab G, Krug A and Du F
(2022) Recall of Own Speech
Following Interaction With L2
Speakers: Is There Evidence for
Fuzzier Representations?
Front. Commun. 7:840041.
doi: 10.3389/fcomm.2022.840041

The aim of this study was to test claims that speakers of a first language (L1) incur cognitive and linguistic processing costs when interacting with second language (L2) speakers. This is thought to be due to the extra cognitive effort required for mapping incoming L2 speech signals onto stored phonological, lexical and semantic representations. Recent work suggests that these processing costs may lead to poorer memory of not only the L2 speech, but of one's own produced speech during an interaction with an L2 speaker. Little is known about whether this is also moderated by working memory (WM) capacity and/or the L2 interlocutor's proficiency. In a partial replication study of Lev-Ari et al., 54 healthy L1 English participants performed a WM test and then read a story and answered inference questions about it from a confederate in one of three conditions: the confederate was either a) a fellow L1 speaker; b) a Chinese L2 speaker of English with advanced proficiency or c) a Chinese L2 speaker of English with intermediate proficiency. Following a distractor task, participants were asked to recall their own answers in a surprise response-recognition questionnaire. Participants recognized their responses more accurately after interacting with the L1 speaker compared with the advanced L2 speaker but not compared with the intermediate L2 speaker. WM capacity correlated with higher accuracy when interacting with the L1 speaker, but with lower accuracy when interacting with the intermediate L2 speaker. These results suggest that effortful processing of input may lead to fuzzier lexical and/or semantic representations of one's own produced speech. However, the lack of significance in recall accuracy between the L1 and the intermediate L2 condition suggests other factors may be at play. Qualitative analyses of the conversations provided insights into strategies that individuals adopt to reduce cognitive load and achieve successful communication.

Keywords: speech processing, communication with L2 speakers, accent perception, L2 proficiency, working memory, recall

INTRODUCTION

Speech perception is a complex process which involves all levels of the grammar in a dynamic and graded manner. Portions of the speech signal need to be mapped on to stored lexical (and in some models pre-lexical) forms, which in turn trigger semantic and syntactic representations (e.g., McClelland and Elman, 1986; Gaskell and Marslen-Wilson, 1997; Luce and Pisoni, 1998; Meyer and Schiller, 2011). When the acoustic signal is degraded, listeners take longer to map what they hear onto stored representations; in terms of word-form processing, increased activation of competitors may take place, leading to more competition between words, weaker activation of target and/or longer time for selection of best match. The influence of the quality of the acoustic signal on perception has mostly been tested under controlled situations using noise or manipulated phonetic detail (e.g., Connine et al., 1993; Andruski et al., 1994; Norris et al., 1995). Listening to second language (L2) speech has not typically been considered in psycholinguistic models of speech processing, but some of the same characteristics of input which varies from the listener's representations described above apply to L2 speech. L2 speakers¹ typically make use of resources from their dominant language to scaffold second language production, resulting in phonetic (and other linguistic) patterns which deviate from those of first language (L1) speakers (Iverson et al., 2003; Wolter, 2006; Lev-Ari and Keysar, 2012). These patterns include phonetically similar material as well as distant targets which may be influenced by the orthography, false friends, and a host of other linguistic factors. When processing L2 speech, L1 listeners may encounter high variability in the acoustic realization of words, potentially leading to lexical competition and "effortful listening" (Van Engen and Peelle, 2014, p. 2), requiring more cognitive resources for successful perception (Munro and Derwing, 1995b; Clarke and Garrett, 2004; Lev-Ari and Keysar, 2012; Van Engen, 2015; Lev-Ari et al., 2018). Under such circumstances, listeners may rely more on top-down processing to comprehend the message, considering the interaction as a whole, and understanding the gist in order to maximize interpretation of linguistic structures (Newman and Connolly, 2009; Goslin et al., 2012; Lev-Ari, 2015).

Input from speakers with lower L2 proficiency is expected to lead to more mismatches between the acoustic signal and L1 listeners' stored lexical representations, potentially leading to lower comprehension and requiring more cognitive resources to encode what is heard into memory (Munro and Derwing, 1995a; Van Engen and Peelle, 2014; Van Engen, 2015). Working memory (WM) is the cognitive system where incoming and stored information are integrated during online speech perception and memory encoding in conversation. In general, the poorer the intelligibility of the speech, the more listeners rely on working memory (WM) for encoding and comprehension (Francis and Nusbaum, 2009). Therefore, WM is more active when the speech signal is degraded or when acoustic mismatch increases

in situations such as listening to L2 accented speech. This is thought to lead to encoding of less detailed semantic and conceptual representations into long-term memory (Rönnberg et al., 2008, 2013). Lev-Ari and Keysar (2012) tested this in an L2 speech processing context by investigating whether participants were better able to detect word changes in a story when listening to an L1 than an L2 speaker. They found that listeners remember fewer details of what an L2 speaker says compared with an L1 speaker due to their expectation of lower competence of the L2 speaker. This leads to increased reliance on contextual cues to deduce content, a process modulated by WM capacity: listeners with high WM increased their reliance on context and were subsequently less accurate at detecting word changes when the story was told by the L2 speaker. Lev-Ari (2015) suggested that participants with high WM are better able to adapt their language processing and can rely more on top-down processes to aid understanding of L2 speech.

The demands in terms of language processing as a result of listening to an L2 accent may be attenuated after more exposure to the accent, even after a couple of minutes. Clarke and Garrett (2004) exposed L1 English listeners to sentences spoken by Spanish and Chinese L2 speakers of English, as well as by L1 English speakers. Reaction time was at first longer for L2 speech but L1 listeners adapted quickly and any deficit in comprehension attenuated, even after listening to L2 sentences for just 1 min. This suggested that increased interactions between L1 and L2 speakers may support both parties to better compensate for the "processing costs" when listening to accented speech.

Expectations by the listener regarding the language proficiency of an L2 interlocutor can help listeners predict phonetic/phonological, semantic and syntactic features and adapt to these in order to maximize the success of an interaction and the recall thereof (Hailstone et al., 2012; Hanulíková et al., 2012). However, one possible methodological confound relates to social factors which may cloud one's perception of the difficulty in processing L2 speech, or one's willingness to attend to it. For instance, attitudes toward an L2 accent and stereotyping can negatively impact listeners' comprehension and linguistic processing (Fuentes et al., 2002; Lindemann, 2002; Dunton et al., 2011; Lippi-Green, 2012). Sociolinguistic research suggests that listeners with negative attitudes about L2 speakers may not accept the burden of communication during an interaction (Lindemann, 2002). They may also unconsciously perceive L2 speech as less able to reliably convey information and therefore rely more on the context of the interaction (Lev-Ari and Keysar, 2012; Lev-Ari et al., 2018).

While the above work focuses on processing interlocutor speech, a recent contribution to this research suggests that the memory of one's own spoken responses may also be impacted when interacting with an L2 speaker. In a unique study, which forms the basis of the experimental procedure developed here, Lev-Ari et al. (2018) constructed interactions between L1 and L2 speakers of English and tested if an L1 listener's recall of their own produced utterances was influenced by speaker condition. Participants read a story and were interviewed by an L1 or L2 confederate with inference questions about the story. Afterwards, participants performed a multiple-choice memory recognition

¹The bulk of the research we review here typically refers to L2 speech as "non-native" or "foreign-accented", but we make a concerted effort to avoid these terms given their negative connotations.

questionnaire of their own responses. Participants had a better memory for their own responses after an interaction with a fellow L1 speaker and were more likely to remember all their responses if they were interacting with an L1 rather than an L2 speaker. Participants were also more likely to choose a distractor response to represent their own answer or a “false alarm” when interacting with L2 rather than L1 speakers. Lev-Ari et al. (2018) argue that this impact is due to the more effortful integration of the incoming speech signal with the listeners’ stored lexical and semantic representations and of their sociolinguistic expectations of the L2 speaker. However, WM capacity was not directly tested in this procedure.

In sum, processing an L2 accent has been shown to incur processing costs in terms of the lexical accuracy and semantic detail of recall, with mediating factors such as familiarization, (perceived) linguistic proficiency of the L2 speaker, and attitudes toward L2 speakers. More recent research suggests that this cost may extend to recall of an L1 interlocutor’s own speech, but this line of enquiry is still in its infancy. The present study extends this work by examining whether L2 proficiency plays a role in L1 listeners’ recall of their own produced speech when interacting with L2 speakers, and the role of WM in such recall. In particular, we seek to answer the following questions:

- 1) Does interacting with an L2 speaker of English have a negative impact on the recall of L1 speakers’ own produced speech during an L1-L2 interaction? This part will be done through a conceptual replication of Lev-Ari et al. (2018) study.
- 2) Does a lower proficiency in the L2 have a more negative impact on the recall of L1 interlocutors’ own produced speech? This is an additional factor which was not part of Lev-Ari et al. (2018) study.
- 3) Does WM mediate an L1 speaker’s ability to recall their own answers after an interaction with L1 or L2 speakers? This factor was explored in an earlier study by Lev-Ari (2015).

Our working hypothesis is that participants who engage in an interaction with an L2 speaker will process fewer lexical and/or semantic details of their own speech, and hence recognize fewer of their own responses compared to when interacting with an L1 speaker. This is due to the cognitive effort involved in processing L2 acoustic input that does not match own stored phonetic details for the intended lexical target or which makes it harder to identify the intended target. This may shift attention away from own answers during the interaction. Any phonetic accommodation to the L2 speaker that is achieved in the fly may also make it harder to subsequently retrieve the message if the acoustic output does not match stored forms. The effect is predicted to be greater in the intermediate proficiency condition, due to the expected greater distance between the phonetic detail of the input and stored representations. We also predict more “false alarms” to be selected in the L2 intermediate condition than in the L2-advanced condition, with the fewest in the L1 condition. Further, participants’ Working Memory scores are predicted to show a negative correlation with their recall score when interacting with L2 speakers, and a positive one when interacting with L1 speakers. This is based on previous findings (Lev-Ari, 2015) which show that individuals with high WM increase their

reliance on context when interacting with L2 speakers compared with individuals with low WM, thereby remembering less lexical detail of what was said.

MATERIALS AND METHODS

Participants

The participants were 54 L1 English speakers who were studying speech and language therapy and linguistics-related degrees at a university in the north-east of England. They were all females aged between 19 and 30 years with no history of speech, language or communication needs and had no knowledge of this experimental procedure before the debrief.

Confederates

Three confederates were selected through an interview by the first two authors and remunerated for their participation. They were informed of the true aims of the experiment, were offered training on the experimental procedure, and were instrumental in the deception strategy. They were matched for gender (all female), age range (18–30) and education with each other and with participants (the confederates were also students at the same university). One confederate was a speaker of English as a first language (L1) and the other two were Mandarin L2 English speakers, with average scores of 6.5 and 8 out of 9 respectively on the International English Language Testing System (IELTS, 2007). The IELTS overall score represents the aggregate results of speaking, listening, reading and writing skills and is presented in bands from 5 to 9; band 6 demonstrates effective command of language with some inaccuracies and misunderstandings, while band 8 represent fully operations command of the language with only occasional inaccuracies. The two participants were hence regarded as having intermediate proficiency (L2_I) and advanced English proficiency (L2_A), respectively. Recruiting L2 confederates from the same L1 language background ensured that differences between them were in the degree rather than nature of L1 influence on the L2 since the characteristics of L2 accents are relatively consistent across speakers from the same L1 backgrounds (Bradlow and Bent, 2008). The choice of Mandarin as the L2 was that of convenience, due to the large population of Mandarin speakers in and around the university where recruitment took place.

Procedure

Seventeen participants were randomly matched with the L2_I confederate, 19 with the L2_A confederate and 18 with the L1 confederate. Unbeknownst to them, each participant was scheduled to arrive at the authors’ research lab at the same time as their matched confederate and was made to believe that both were participants in the study. After initial instructions given by the first author, each participant and their confederate were seated in front of a computer to complete a WM test (Section Working Memory Testing Phase). Once the WM task was completed, the participants had a short break and moved on to the experimental task. These were recorded using an Edirol R-09 recorder with a sampling rate of 44,100 Hz and 16-bit amplitude resolution.

In order to keep testing instructions constant during the experiment and to ensure replicability of procedures, all instructions to participants were standardized as a script that was rehearsed and delivered by the first author. After giving instructions to the participants about the tasks that they would engage in, the researcher left to an adjoining room with an observation window so that the proceedings only focused upon the confederate-participant interaction and not that of researcher-participant. The tasks are described in chronological order below.

Tasks and Scoring

Working Memory Testing Phase

The RSPAN (Automated Reading Span Test) test (Daneman and Carpenter, 1980) was first conducted and administered using Millisecond software in Inquisit and took 15–20 minutes to complete. In order to protect the status of confederates, they were instructed to complete the RSPAN at the same time as their matched participant during each trial. Briefly, the RSPAN consists of a series of sets, each set alternating between the presentation of a sentence to participants which they have to judge on plausibility and a letter after every sentence which they need to memorize and recall in order at the end of each set. The score included in the analyses is the absolute RSPAN, a measure the number of perfectly recalled sets in terms of letters and their order. For example, if a participant correctly recalled 2 letters in a set size of 2, the absolute score would be 2; otherwise, they would get 0 absolute score. This was used as a latent variable for WM capacity since it requires the ability to integrate different sources of information in a set amount of time.

Reading Comprehension and Surprise Memory Phase

The participant and confederate were then informed that they would silently read a story and pick a “random” color out of a box to decide who asks questions about it and who answers these. The experiment was set so that the confederate always asked the questions.

The 200-word text (Table 1) which was adapted for the story comprehension activity during the test was sourced from the Discourse Comprehension Tests, set B (Brookshire and Nicholas, 1993). This is a highly readable, clearly structured narrative, from which inference questions could be developed for the questions.

In order to ensure consistency in the linguistic content of the questions and limit differences in delivery to accent, a script with the questions was provided to all confederates ahead of the experiment and they had the chance to practice these. Seven inference questions based on the text were provided for confederates to ask participants after they finished reading the text (Table 2). The participants were free to respond to each question at any length.

All participants then completed a five-minute distractor picture-puzzle task which served to intercept the instant memory of their responses. The task consisted of 16 sets of pictures (four in each set); the participants were asked to examine each set and write down a three-letter word that best describes the four pictures within it. After this, participants were given 5 min to complete a surprise memory questionnaire (inference questions

TABLE 1 | Story used for the reading comprehension task.

George Smith was a quiet French bookkeeper. None of his friends believed him when he told them he was going to walk across Niagara Falls on a tightrope. But here he was, one spring day, looking at the rope which was stretched 50 meters above the falls. George's wife stood beside him trying to convince him not to try such a foolish stunt. She told him to think of his family. George just shook his head stubbornly and told her that this was something he had to do. Then he began to practice for the crossing by walking back and forth on a narrow wooden beam. By the time George was ready to begin the crossing, almost a thousand people had gathered to watch him. George stood uncertainly at the edge of the river. The long rope swayed slightly in the breeze. Slowly he set out across the rope toward the Canadian side of the falls. Twenty minutes later, a television reporter came up to him and asked, “What are you going to do next?” George thought for a moment and then answered, “I guess I'll walk back across the rope. I left my car on the other side.”

TABLE 2 | Inference questions.

1. Why didn't any of George's friends believe him when he said he was going to cross Niagara Falls?
2. Why did his wife try to convince him not to do it?
3. Why was George determined to do it?
4. Why did George hesitate uncertainly at the edge of the river?
5. Why was the TV reporter interested in approaching George?
6. Why did people gather to watch him?
7. Why did he hesitate before answering the reporter?

TABLE 3 | Example inference question with potential responses.

Why didn't any of George's friends believe him when he said he was going to cross Niagara Falls?

- He'd never done anything like that before.
- It was something really extreme.
- His friends didn't see him as adventurous.
- A quiet person would not be expected to do that.
- A bookkeeper would be unlikely to do that.
- He was an unlikely character for dangerous stunts.
- His friends thought he was joking/wasn't serious.
- It's the kind of stunt people joke about doing.
- His friends did not think he was trained to do it.

with possible answers). Here, the same seven questions that were asked of participants during the reading task were shown to them again with potential responses (Table 3). They could choose more than one response if necessary. Before experimental procedures began, an informal pilot study was conducted on peers of a similar demographic to the target sample, which informed the range of possible answers in constructing the memory questionnaire. Participants were asked to circle the responses which best represented their spoken answer in the interview. If participants were outside of three standard deviations from the overall mean number of responses or false alarms, their data were excluded from the analysis. Using measures such as the mean and standard deviation on the count data of the current study is not unproblematic but effectively identified two participants who circled, on average, five answers per question, compared to the overall mean of fewer than two responses. These participants

TABLE 4 | Scoring example.**During the reading comprehension task:**

- Confederate: Why did his wife try to convince him not to do it?
- Participant: Because she thought it was dangerous.

During the surprise memory task, the participant then selected the following three answers on the response sheet:

(* means it is not what they said during the interview)

- Because it was dangerous.
- She was trying to protect him. *
- Because he could have died. *

were excluded. The data from another participant was also excluded because they had retained the story text from the task and read from it, rendering the recall measure void.

Scoring of Inference Questions With Potential Responses

The first author calculated participants' recognition of responses they gave during the interview by comparing the responses they selected in the surprise recall test with the responses they provided during the recorded interview. Participants could only get 100% when the answers they selected perfectly matched what they said during the interaction with the confederate. Partial scores were awarded in other scenarios. For instance, if the participant chose one answer when there was scope to select another, then only 50% was awarded to that question; or if the participant gave one answer which agreed with the response along with two others which deviated from it (false alarms) they scored 33% (e.g., **Table 4**). False alarms were of particular interest because they offered a window into whether participants had fuzzier linguistic representations of what they said. If participants had answered a question in the reading task with "I don't know", this question was subsequently not used in the inference questions. Scores obtained were then compared with a second iteration of scoring on 100% of the sample (Intra-Rater) and with 20% of recordings from an independent coder (Inter-Rater). Cohen's (1960) was used for testing Intra- and Inter-rater reliability, with the first yielding a value of 0.88, suggesting high degree of agreement; the smaller dataset for Inter-rater reliability (10 observations) did not render it optimal for Cohen's Kappa testing as the two sets of scores were too similar, with the highest disagreement in scores being 7%. However, Pearson's *r* for the Inter-rater reliability was at 0.96, suggesting high agreement between the first author and the independent coder.

Language Background and Debrief

After the main tasks, participants completed a language background questionnaire which also included a self-rating measure of how often participants communicated with Mandarin L2 speakers of English.

To keep in line with ethical procedures, participants were given a verbal debrief accompanied with an explanation sheet after they had completed the experiment. All participants reported to have been successfully misled by the role of the

confederate and were given the opportunity to ask any questions about the research.

STATISTICAL ANALYSES AND RESULTS

Recall

This subsection addresses the participants' recall of their own speech, as measured by the percentage score from the surprise memory phase of the experiment. Lev-Ari et al. (2018) used mixed effects simple logistic regression in their study to identify significant effects. For the current dataset, the original intention was to conduct the analysis *via* mixed effects ordinal logistic regression to allow for a more nuanced coding of recall because the response variable in ordinal logistic regression can have more than two levels. However, both ordinal and simple logistic regression models resulted in issues of singular fit, particularly for the random effect of participant on recall. This suggests that the dataset in its current shape was insufficient for models that included random effects. Rather than fitting underpowered models, we decided to aggregate the data by participant. As a result, rather than having seven potential data points for each participant (one per question), we used one average recall score per participant. We acknowledge that aggregating the data results, first, in the loss of information on within-participant variation and, second, the necessity to interpret the results with caution due to increased type I and type M/S error rates. However, we considered it a practical solution to interrogate the data without fitting models that are too complex for the dataset.

Since the independent variable, that is the average recall score per participant, was bounded in the interval $0 \leq y \leq 1$, a beta regression rather than a linear regression was fitted to the data. Beta regressions are commonly used for proportion data, such as the one in the current dataset, because the data have natural limits (0 and 1) and often do not follow a normal distribution. For beta regressions, the response variable usually cannot take the extreme values $y = 0$ and $y = 1$. However, some participants in our study scored perfectly for all answered questions, resulting in an average recall score of $y = 1$. Therefore, following Smithson and Verkuilen (2006), all recall scores were transformed with the following formula, which included the sample size $n = 54$:

$$y_{transformed} = \frac{(y_{raw} \cdot (n - 1) + 0.5)}{n}$$

The model included condition, WM and response length as fixed effects. An interaction term between condition and WM was also included. The coding of each of these factors, the rationale behind including them and the predictions for their effect on the recall of the participants' own speech are provided below:

Condition

Condition is a categorical variable with three levels that correspond to whether the participants interacted with an L1 confederate, an L2_A or an L2_I confederate. We predicted that participants' recall of their own responses would follow this pattern: L1 > L2_A > L2_I.

TABLE 5 | Beta regression model output for recall of participants' own speech.

Coefficient	Estimate	Std. error	Z-value adjusted	P-value adjusted
Intercept	1.24	0.19	6.64	< 0.001
Condition				
L1 vs. L2_I	-0.24	0.26	-0.90	0.368
L1 vs. L2_A	-0.59	0.25	-1.98	0.048
L2_I vs. L2_A	-0.35	0.26	-0.94	0.348
Working memory	0.02	0.01	1.38	0.168
Response length	-0.02	0.02	-0.74	0.460
Condition (L1 vs. L2_I) x working memory	-0.04	0.02	-2.26	0.024
Condition (L1 vs. L2_A) x working memory	-0.01	0.01	-0.01	> 0.99
Condition (L2_I vs. L2_A) x working memory	0.03	0.02	-1.62	0.105

The bold values indicate signals with significant result.

WM

WM is a continuous predictor and corresponds to the participants' absolute RSPAN score. This score adds up the number of letters from all sets that were perfectly recalled by the participants. To allow for a more sensible interpretation of the model, the RSPAN score was centered before it was added to the model. The predictions for WM are less straight forward. On the one hand, higher WM usually results in better recall, which is evident from how WM is measured in the RSPAN procedure. However, Lev-Ari and Keysar (2012) found that their participants' recall of L2 speech was worse if their WM was higher. They argued that the adjustment to L2 speech required cognitive resources and, thus, was only possible for participants with higher WM. However, the participants' adjustment was found to lead to less-detailed representations and, as a result, worse recall. It is not yet clear if this finding also extends to the recall of one's own speech. Based on these considerations, it was predicted that, in interactions with L1 confederates, higher WM would be beneficial and correlate with better recall. In interactions with L2 confederates, higher WM would result in fuzzier representations of one's own spoken output. To test the potentially differential effect of WM on the L1 vs. L2 conditions, an interaction between condition and WM was included.

Response Length

Response length is a continuous predictor and corresponds to the number of words that the participants used to respond to a question. It was centered. We predicted that participants would recall their responses better if they gave shorter responses because there would usually be fewer items to recall. Words that were used by the participants to ask for a repetition of the question were not counted. Hesitation markers, such as *um* and *erm*, were also disregarded for the word count.

Table 5 provides the output of the beta regression model. Beta regression outputs the model coefficients in log odds or logits, which can be transformed into probabilities. For example, the

estimate for the intercept in **Table 5** ($x_0 = 1.24$) refers to the reference level of condition (i.e., L1 confederate) with WM and response length being kept at constant levels. Since WM and response length were centered, these constant levels refer to the mean WM and the mean response length. Thus, the probability for an overall perfect recall score when participants interacted with an L1 confederate, had average WM and an average response length is:

$$\frac{\exp(1.24)}{1 + \exp(1.24)} \approx 0.776 = 77.6\%$$

The logits and the corresponding probabilities for the other conditions can be calculated by adding the model estimates to the intercept. For instance, the probability of an overall perfect recall score for participants who interacted with an L2_A confederate and had average WM as well as average response length is:

$$\frac{\exp(1.24 - 0.59)}{1 + \exp(1.24 - 0.59)} \approx 0.657 = 65.7\%$$

The pairwise comparisons between conditions from **Table 5**² show one significant effect. Participants' recall of their own speech is worse when they interact with an L2_A as compared to an L1 confederate. Additionally, there one of the interaction terms was significant. Participants' recall is mediated by WM in that, when comparing interactions with an L1 vs. L2_I confederate, increasing WM has a detrimental effect on recall after interactions with L2_I confederates. These two effects are addressed in more detail below.

The mean percentage scores across conditions are shown in **Figure 1**. On average, participants' recall scores are 76.1, 65.6, and 70.2% for the L1, L2_A and L2_I condition, respectively. This corresponds to the significant difference between the L1 and L2_A conditions.

Figure 2 helps to better understand the interaction effect between condition and WM for L1 vs. L2_I confederates. The participants' RSPAN score is plotted against their average recall score. Condition is coded by color. Each point in the plot represents one participant. Lines of best fit were added to show the relationship between WM and condition. The range of available RSPAN scores varied between conditions. Therefore, the horizontal span of the lines is not equally large across the three conditions. As can be seen, WM is beneficial for recall in the L1 condition. In the L2_I condition, however, higher WM resulted in worse recall of the participants' own speech.

False Alarms

False alarms were defined as answers that the participants recalled as their own speech during the interview phase although they had not given these answers previously. False alarms

²The model coefficients for the pairwise comparison between the L2_I and L2_A conditions were taken from a model with the reference level L2_I for condition. P-values (and the corresponding z-values) were adjusted via Bonferroni-Holm corrections to account for the three pairwise comparisons.

are already incorporated into the recall measure presented in Subsection Recall. For example, a recall score of 50% could encode a question for which a participant should have circled one answer only but in reality circled an additional answer (i.e., gave a false alarm). In addition to the above analyses, false alarms are considered here separately because they

encode how fuzzy a specific participant's lexical and semantic representations were. If a participant gave one or several false alarms for a question, their memory representations were likely fuzzier.

Mixed effects simple logistic regression models were used to analyse the false alarms in the recall task. Since there were only relatively few cases with more than one false alarm (23 out of 360 responses), the data were coded in a binary fashion, with responses either containing false alarms or not. The models did not result in any singular fit or convergence issues, which suggested sufficient power and did not warrant for an aggregation of the data. The fixed effects in the models were condition, centered WM and centered response length. An interaction term between condition and WM was also added to the model. In addition to these fixed effects, the models included two random effects:

Participant

Since the occurrence of false alarms might vary beyond the fixed predictors specified above, random intercepts were included in the models. By-participants slopes were not appropriate because of the between-participants design of the study.

Question

The same seven questions were used for all participants across the three conditions. Therefore, by-question random intercepts were fitted as well as, initially, by-question random slopes for the fixed effect of condition. These random slopes were later dropped as the random effects structure proved too complex for the dataset.

The predictions for the occurrence of false alarms were in line with the predictions for the recall score in Subsection Recall. False alarms were predicted to be more prominent when the

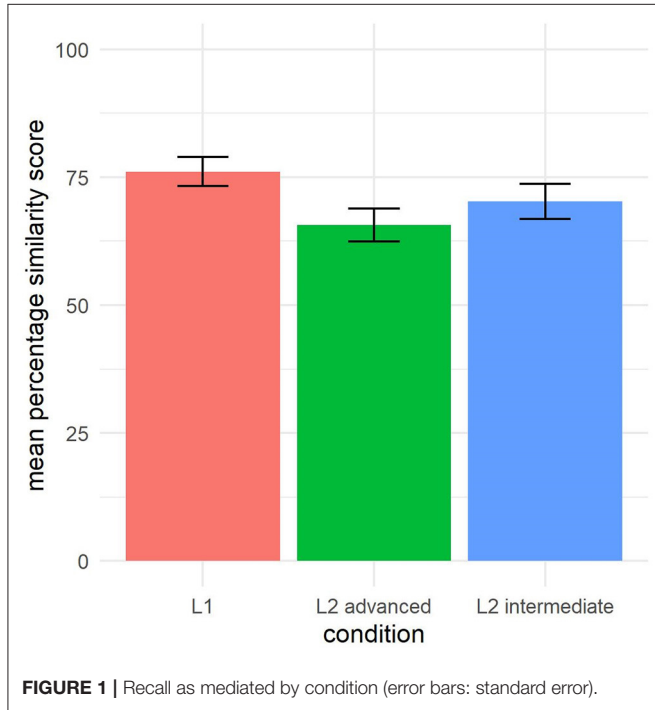


FIGURE 1 | Recall as mediated by condition (error bars: standard error).

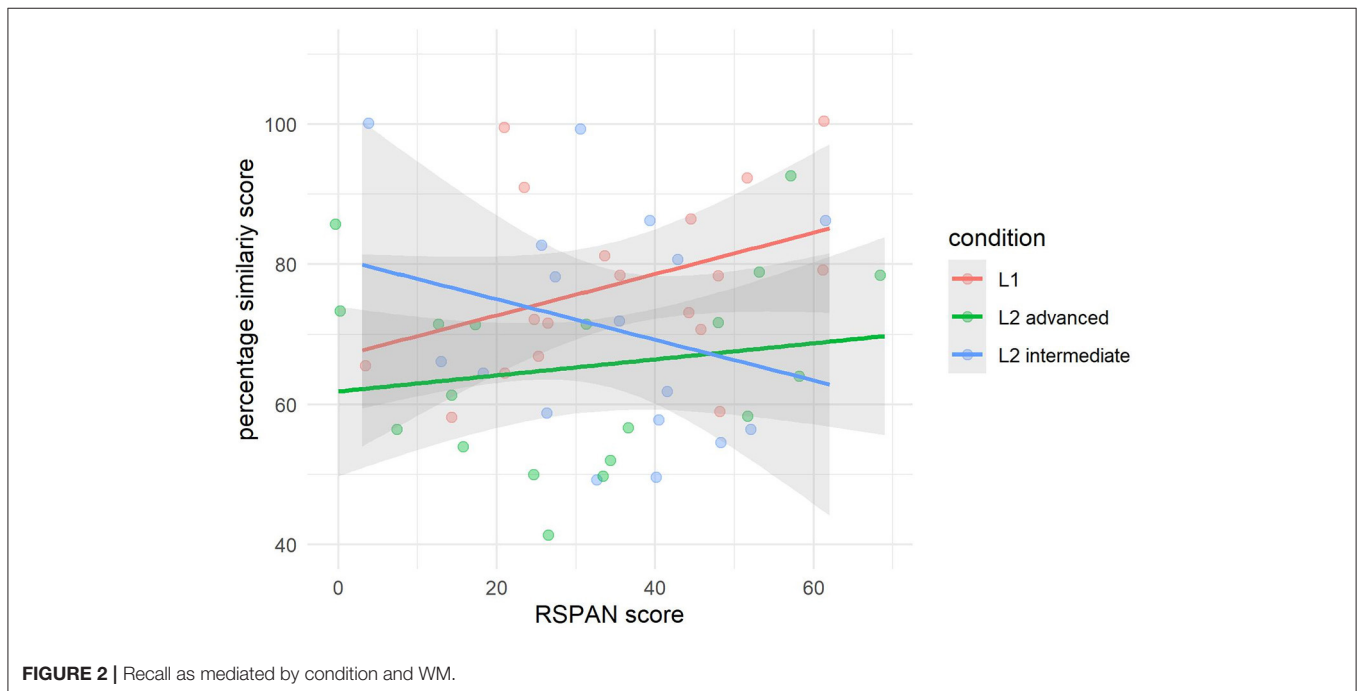


FIGURE 2 | Recall as mediated by condition and WM.

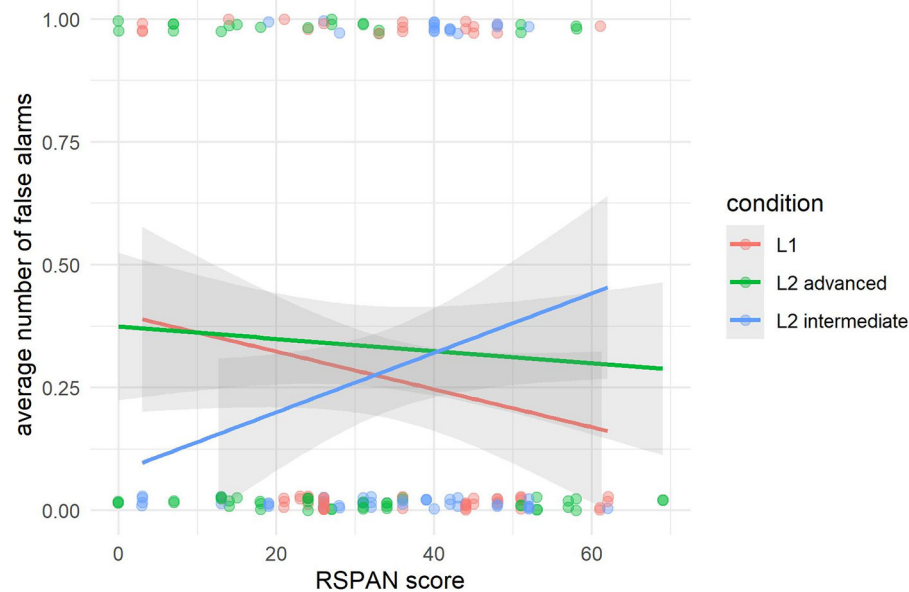


FIGURE 3 | Presence of false alarms as mediated by condition and WM.

participants interacted with L2_I as compared to L2_A and L1 confederates. The effect of condition was predicted to be mediated by WM, with higher WM improving recall in the L1 condition but decreasing it in the L2 condition, especially if the confederate had an intermediate command of English. Longer responses were predicted to result in a higher probability that false alarms would occur.

Likelihood ratio comparisons were used to identify significant effects. The full model was systematically reduced by one fixed effect or interaction term. Model comparisons then showed if the effect or interaction in question had a significant effect on the occurrence of false alarms³. The model comparisons showed a significant effect of the interaction between condition and WM ($p = 0.031$). The other effects in the model did not reach significance: condition ($p = 0.507$), WM ($p = 0.733$) and response length ($p = 0.228$).

Figure 3 shows the significant interaction effect in further detail. WM, as measured by the RSPAN score, is shown on the horizontal axis. Each point in the graph represents one of the seven questions. Because of the binary coding of false alarms, a participant's response to each question either did or did not contain a false alarm. Some jitter was added to the points so that they would not overlap for each participant. Lines of best fit are shown in different colors, one for each condition. The interaction effect stems from the different effects of increasing WM for L2 confederates on the one hand, and L1 as well as L2_A confederates on the other hand. Participants with higher WM are more likely to give false alarms in the L2_I than in the other

two conditions. This indicates worse recall of one's own speech in interactions with an L2_I confederate, provided that WM is high.

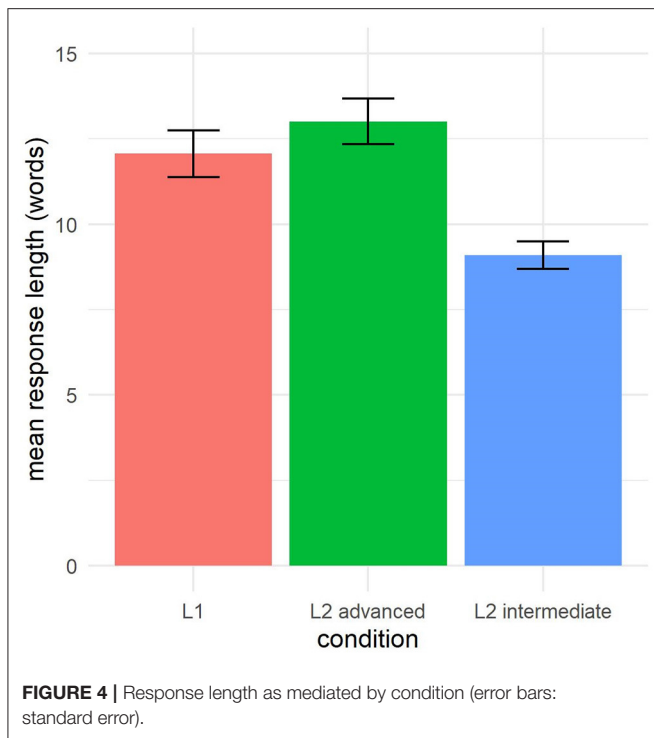
Response Length

Following an informal observation that participants gave comparably shorter answers to the L2 confederate with intermediate proficiency than to the other two confederates, an exploratory analysis was carried out to quantify this observation and report on a potential structural difference in interactions with L2 speakers. Response length was operationalised as the number of words in a participant's response to an inference question (see Subsection Recall).

To see if the participants' response length varied significantly across the three conditions, linear mixed effects models were used. Condition was added as a fixed effect with three levels (L1, L2_A and L2_I). Based on the observations during the experiment, it was predicted that responses would be shorter, that is contain fewer words, in interactions with the L2_I confederate. The models further included the random effects specified for the models in Subsection Recall.

No significant effect of condition on response length was found through the model comparisons ($p = 0.080$). However, the average number of words per response per conditions, as shown in **Figure 4**, displays a trend in the data that is in line with the qualitative comment above. The average response length was 8.8 words ($sd = 4.3$ words) for participants who interacted with the L2_I confederate, 11.7 words ($sd = 7.4$ words) for participants who interacted with the L1 confederate and 12.5 words ($sd = 7.6$ words) for participants who interacted with the L2_A confederate. Although the difference between the conditions is not significant, the shorter responses to the L2_I confederate are informative

³P-values for fixed effects that were included in the interaction (i.e. condition and WM) were identified by comparing a model without the interaction and a model without the interaction and without the fixed effect in question.



and will be discussed in the following, along with the preceding results.

DISCUSSION

This study investigated the potential effect of interacting with L2 speakers on the recall of one's own speech. The aim was to explore whether there are additional processing costs when listening to L2 accented speech as described in previous research (e.g., Van Engen and Peelle, 2014; Van Engen, 2015; Lev-Ari et al., 2018), and whether these are modulated by each of L2 proficiency of the speaker and/or the WM of the L1 speaker. The study found a cognitive disadvantage for participants who interacted with L2 compared with L1 confederates, but only in the L2_A condition. In other words, participants remembered their own responses more accurately, with their recall score higher in the L1 interlocutor condition than in the L2_A condition, but not when compared with the L2_I condition. There was therefore no general across-the-board effect of L2 interaction on recall. These results only partially replicate Lev-Ari et al. (2018) results and are somewhat surprising, since one would have expected lower English proficiency of the confederate to lead to more effortful processing for L1 participants and therefore fuzzier lexical and/or semantic representations, leading to worse recall. They suggest that other factors may have been at play during the task, which affected the communication and degree of orientation to the L2_I speaker. Given that the participants were answering questions about a passage they had just read ahead of their interaction with the confederates, their processing did not only solely consist of bottom-up processing of the linguistic signal; rather, they

will have been able to use contextual cues from the passage. Individual differences may also have played a role, since speech processing does not only involve processing of the linguistic signal; the use of extralinguistic and contextual information (e.g. knowledge about the world, expectations from particular situations) is commonly incorporated into the listening process, influencing how individuals come to understand the same discourse (Garman, 2012).

The WM results show a significant interaction between WM capacity and each of recall and false alarms in the participants recall of own produced speech, but with opposing effects depending on the language background of the confederate: higher WM led to better recall and fewer false alarms following communication with the L1 confederate, but worse recall and increased false alarms following communication with the L2_I confederate. These results suggest that speakers with high WM can benefit from integrating social-indexical information in their processing of an accent in the familiar/more compatible L1 condition (Drager, 2011), but this integration is more effortful in the L2 condition and leads to fuzzier lexical and semantic representations of one's own responses. This is the first study to extend previous WM findings to the less detailed recall of one's own produced speech. While the ability to use WM resources in challenging listening conditions enables listeners to orient their attention to their interlocutor and recall more of what they hear (Van Engen et al., 2012), this might have adverse effects on one's own memory of their speech. Another reason for the fuzzier recall of one's own spoken utterances may be due to speakers also adapting their own speech to that of their interlocutor, leading them to remember their own responses less accurately due to the greater mismatch between the acoustics of the response they produced and their own lexical representations (Akeroyd, 2008; Rönnerberg et al., 2008, 2013; Lev-Ari and Keysar, 2012). While we did not analyse the speech of our participants in their interactions with confederates, a large proportion of the participants were speech and language therapy trainees who are expected to be particularly skilled at orienting their speech to the listener. While this will have improved their recognition of what the interlocutor said, it may have adversely affected their recall of the detail in the L2_I condition. Importantly though, there was no main effect for language condition on recall.

Conceptual and semantic representations of language have been suggested to be less detailed after listening to an L2 speaker, leading to adverse effects on lexical access both in terms of interlocutor speech and one's own speech (Rönnerberg et al., 2008, 2013; Lev-Ari and Keysar, 2012; Lev-Ari, 2015; Lev-Ari et al., 2018). In this study we do not find strong evidence for the latter; while processing a less familiar accent may indeed be more effortful, strategies that both L1 and L2 speaker adopt during the interaction may help mitigate this effect. It is also important to note that, while research in this area has focused on L2 or so-called "foreign" or "non-native" accents, any difficulty that is due to unfamiliarity and lower intelligibility of an accent could equally apply to L1 interactions between speakers of different regional accents of the same language (Goslin et al., 2012; Lev-Ari and Keysar, 2012). It is important to disentangle subjective expectations

relating to the perceived difficulty of processing L2 accents from the more general increased cognitive load that may be required when processing an unfamiliar accent. The underlying sources of this load offer an interesting window into how we store and represent speech; storing social-indexical information together with lexical information during communication with other speakers is advantageous (Goldinger, 1998) but can also incur a 'cost' when processing unfamiliar speech.

The degree to which interlocutors are at ease in this unfamiliar setting and the strategies they adopt can either alleviate or compound the effortful communication. In this study qualitative observations of the communication between our participants and the confederates suggested a naturalistic and conversational style used by the L2_A, but a more mechanical, less relaxed interaction style by the L2_I confederate despite both receiving the same training. This may either be due to differences in proficiency or in personality and may have influenced the participants' conversation style too. For instance: 1) during the communication with the L1 and L2_A confederates, both interlocutors maintained eye contact throughout the interview, more often than during the communication with the L2_I confederate; 2) the L2_A confederate acted relaxed and laughed before the first question began, while the L2_I confederate did not; 3) the L2_A confederate used interjections before asking the questions, which may have increased the naturalness of the conversation and given the participants time to get ready for the question; the L2_I confederate tended to ask questions directly. There was a tendency for L2_I confederate's answers to be shorter, but this did not prove significantly different from the answers that the other two confederates gave; 4) the L1 and L2_A confederates were more interactive in the interview, smiling at or nodding to their participants, while the L2_I confederate was more task-oriented and less interactive with their participants; 5) participants in the L2_I condition asked the confederate to repeat their question more often than in the L2_A and L1 condition. The combined effect of these differences may have led to more entrainment between the participants and L2_A speakers, albeit with an increased cost to the participants' recall of their own speech. On the other hand, participants in the L2_I condition may have attended more to the task, and conversely remembered more of their own responses to the questions. The confederates' accents in this case may have been less likely to impact on participants' encoding of their own responses into memory.

It is important to note that, regardless of the differences in recall scores in the L1 and L2_A condition, recall scores were relatively high across all three conditions. The generalization of these results needs to be considered with caution for two reasons. First, the participants in this study were mainly SLT trainees who may have already possessed the skills to be attentive to the

needs of the interlocutor in order to maximize communicative success and may therefore have been more adept at adapting to the needs of the situation. Second, only one confederate was used in each condition, which might have resulted in speaker-specific effects. Nevertheless, what this suggests is that the success of communication between L1 and L2 speakers, or interactions between speakers who may not be familiar with each other's accent more generally, should not be the onus of one party, typically the speaker of the less-dominant accent. Attention and conversational strategies on the part of both interlocutors can overcome communicative challenges and ensure the success of the interaction, albeit with increased cognitive processing load and possible initial toll on the detail of the lexical and/or semantic representations of own and others' speech. Increased exposure to L2 accents has also been shown to improve the processing of these and other unencountered accents (Baese-Berk et al., 2013) in turn increasing listeners' trust in what L2 speakers say (Boduch-Grabka and Lev-Ari, 2021). This demonstrates that familiarization with diverse accents rather than expecting L2 speakers to reduce their "foreign" accent is a more equitable way forward in improving L1-L2 communication. This can be achieved on a large scale if various industries such as the media, education, and the arts made an effort to give more platform to speakers of non-dominant and non-standard varieties.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

FB and GK co-produced the literature review and background to the study and co-designed the methodology. FB oversaw the data collection and carried out the initial analyses. AK and FD supported the data analyses and carried out the statistical analyses. All authors co-wrote the discussion, contributed to the article, and approved the submitted version.

ACKNOWLEDGMENTS

We are grateful to the participants for taking part in this study and the confederates for their patience throughout the training and implementation of the research. We are also deeply grateful to Shiri Lev-Ari for her advice and insights in the early stages of the study and for generously sharing with us the story used in the Lev-Ari et al. (2018) study, as well as information on scoring and analysis.

REFERENCES

- Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int. J. Audiol.* 47, S53–S71. doi: 10.1080/14992020802301142
- Andruski, J. E., Blumstein, S. E., and Burton, M. (1994). The effect of. Mem. Lang. subphonetic differences on lexical access. *Cognition*. 52, 163–187. doi: 10.1016/0010-0277(94)90042-6
- Baese-Berk, M. M., Bradlow, A. R., and Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *J. Acoust. Soc.* 133, EL174–EL180. doi: 10.1121/1.4789864

- Boduch-Grabka, K., and Lev-Ari, S. (2021). Exposing individuals to foreign accent increases their trust in what nonnative speakers say. *Cogn. Sci.* 45:11,e13064. doi: 10.1111/cogs.13064
- Bradlow, A. R., and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*. 106, 707–729. doi: 10.1016/j.cognition.2007.04.005
- Brookshire, R. H., and Nicholas, L. E. (1993). *The Discourse Comprehension Test*. Tucson, AZ: Communication Skill Builders. A Division of The Psychological Corporation.
- Clarke, C. M., and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.* 116, 3647–3658. doi: 10.1121/1.1815131
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104
- Connors, C. M., Blasko, D. G., and Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition?. *J. Mem. Lang.* 32, 193–210. doi: 10.1006/jmla.1993.1011
- Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verbal Learning Verbal Behav.* 19:4, 450–466. doi: 10.1016/S0022-5371(80)90312-6
- Drager, K. K. (2011). Sociophonetic variation and the lemma. *J. Phon.* 39, 694–707. doi: 10.1016/j.wocn.2011.08.005
- Dunton, J., Bruce, C., and Newton, C. (2011). Investigating the impact of unfamiliar speaker accent on auditory comprehension in adults with aphasia. *Int. J. Lang. Commun. Disord.* 46, 63–73.
- Francis, A. L., and Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attent, Percept, Psychophys.* 71, 1360–1374. doi: 10.3758/APP.71.6.1360
- Fuertes, J. N., Potere, J. C., and Ramirez, K. Y. (2002). Effects of speech accents on interpersonal evaluations: implications for counseling practice and research. *Cult. Divers. Ethn. Minor. Psychol.* 8, 346–356. doi: 10.1037/1099-9809.8.4.347
- Garman, M. (2012). *Psycholinguistics*. Cambridge: Textbooks in Linguistics Series. Available online at: <https://doi-org.libproxy.ncl.ac.uk/10.1017/CBO9781139165914> (accessed December 15, 2021).
- Gaskell, M. G., and Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Lang. Cogn. Process.* 12, 613–656. doi: 10.1080/0169096973866646
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Rev.* 105, 251. doi: 10.1037/0033-295X.105.2.251
- Goslin, J., Duffy, H., and Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain Lang.* 122:2, 92–102. doi: 10.1016/j.bandl.2012.04.017
- Hailstone, J. C., Ridgway, G. R., Bartlett, J. W., Goll, J. C., Crutch, S. J., and Warren, J. D. (2012). Accent processing in dementia. *Neuropsychologia*. 50, 2233–2244. doi: 10.1016/j.neuropsychologia.2012.05.027
- Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., and Weber, A. (2012). When one person's mistake is another's standard usage: the effect of foreign accent on syntactic processing. *J. Cogn. Neurosci.* 24, 878–887. doi: 10.1162/jocn_a_00103
- IELTS. (2007). *The IELTS Handbook*. Cambridge: University of Cambridge Local Examinations Syndicate, The British Council, IDP Australia.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*. 87, B47–B57. doi: 10.1016/S0010-0277(02)00198-1
- Lev-Ari, S. (2015). Comprehending non-native speakers: theory and evidence for adjustment in manner of processing. *Front. Psychol.* 5, 1–12. doi: 10.3389/fpsyg.2014.01546
- Lev-Ari, S., Ho, E., and Keysar, B. (2018). The unforeseen consequences of interacting with non-native speakers. *Top. Cogn. Sci.* 10, 835–849. doi: 10.1111/tops.12325
- Lev-Ari, S., and Keysar, B. (2012). Less-detailed representation of non-native language: Why non-native speakers' stories seem more vague. *Discourse Process.* 49, 523–538. doi: 10.1080/0163853X.2012.698493
- Lindemann, S. (2002). Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society*. 31, 419–441. doi: 10.1017/S0047404502020286
- Lippi-Green, R. (2012). *English With an Accent: Language, Ideology, and Discrimination in the United States*. London: Routledge. doi: 10.4324/9780203348802
- Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hear.* 19, 1–36. doi: 10.1097/00003446-199802000-00001
- McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0
- Meyer, A. S., and Schiller, N. O. (2011). *Phonetics and Phonology in Language Comprehension and Production: Differences And Similarities*. Phonology Phonetics. Available online at: <https://doi-org.libproxy.ncl.ac.uk/10.1515/9783110895094> (accessed December 15, 2021).
- Munro, M. J., and Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 45, 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x
- Munro, M. J., and Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Lang. Speech.* 38, 289–306. doi: 10.1177/002383099503800305
- Newman, R. L., and Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: evidence for bottom-up and top-down effects on spoken word processing. *Biol. Psychol.* 80, 114–121. doi: 10.1016/j.biopsycho.2008.04.008
- Norris, D., McQueen, J. M., and Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *J. Exp. Psychol. Learn Mem. Cogn.* 21, 1209. doi: 10.1037/0278-7393.21.5.1209
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The ease of language understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7, 1–17. doi: 10.3389/fnsys.2013.00031
- Rönnerberg, J., Rudner, M., Foo, C., and Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). *Int. J. Audiol.* 47:2, 99–105. doi: 10.1080/14992020802301167
- Smithson, M., and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol. Methods*. 11, 54–71. doi: 10.1037/1082-989X.11.1.54
- Van Engen, K. J. (2015). Downstream effects of accented speech on memory. *J. Acoust. Soc. Am.* 137, 2210. doi: 10.1121/1.4920046
- Van Engen, K. J., Chandrasekaran, B., and Smiljanic, R. (2012). Effects of speech clarity on recognition memory for spoken sentences. *PLoS ONE*. 7, e43753. doi: 10.1371/journal.pone.0043753
- Van Engen, K. J., and Peelle, J. E. (2014). Listening effort and accented speech. *Front. Hum. Neurosci.* 8, 577. doi: 10.3389/fnhum.2014.00577
- Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: the role of L1 lexical/conceptual knowledge. *Appl. Linguist.* 27, 741–747. doi: 10.1093/applin/aml036

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Baxter, Khattab, Krug and Du. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.