Lee, O. & Ahn, H. (2020). Faster and less clear L2 speech with more errors during a verbal working memory task but not during a spatial task. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Northern Arizona University, September 2019 (pp. 141–153). Ames, IA: Iowa State University.

FASTER AND LESS CLEAR L2 SPEECH WITH MORE ERRORS DURING A VERBAL WORKING MEMORY TASK BUT NOT DURING A SPATIAL TASK

Ogyoung Lee, Seoul National University
Hyunkee Ahn, Seoul National University

The present study tested a prediction, derived from the hypothesis of phonological-phonetic encoding, that speech planning and production is dependent on verbal working memory. We addressed whether overloading working memory impairs L2 speech production, specifically the sequencing and articulation of sound. Twenty Korean L2 learners of English spoke thirty-two English sentences during a verbal and spatial working memory task and the same sentences under a control no-load condition. In the load conditions, speakers were engaged in a task that taxed either verbal or spatial working memory while speaking. In the control condition, they completed mathematical equations before speaking and thus had no additional load while speaking. The results showed the L2 learners were disrupted by the verbal working memory task during speech production. They made more speech errors, spoke faster, produced less variable word durations, and articulated vowels less distinctively. They were distracted during the spatial task, making more errors, but the prosody remained intact. The results suggest that, contrary to L1 speech production (Lee & Redford, 2015), L2 speech production depends more on active encoding of phonological-phonetic information via verbal working memory than on direct retrieval from remembered articulatory templates. A speech production model is proposed for L1 vs. L2.

INTRODUCTION

Speech-language production is described as a multi-staged process in psycholinguistics (e.g., Levelt, Roelofs, & Meyer, 1999). In Levelt et al. (1999), the production process begins with conceptualization and the activation of a lexical concept to be expressed; e.g., PRODUCE (X, Y). The concept then triggers lexical selection, which provides syntactic information in the form of a lemma; e.g., *produce* tagged as a transitive with two arguments. Next, the relevant morphemes are retrieved; e.g., <produce> and <s>. At the same time, the metrical and segmental properties are retrieved; <produce> as an iambic foot and <s> as an extrametrical suffix; segments are laid out in sequence (/p/, /r/, /ə/, /d/, /j/, /u/, /s/, /ə/, /z/) and are syllabified (/prə.dju.səz/), which also transform the underlying phonemes into allophones. The syllabification procedure allows for the selection of appropriate articulatory routines from a mental syllabary. Once selected, the routines are passed to the articulatory buffer where they are held until a prosodic word is compiled for execution. The process from the retrieved word form through to syllabification is referred to as phonological encoding. The selection of articulatory routines through to the compilation of prosodic words is referred to as phonetic encoding. These are the processes that constitute speech production, which is the phonological-phonetic planning and implementation of morphosyntactic lexical forms (Fowler, 2010).

Insofar as phonological-phonetic encoding refers to the selection and manipulation of phonological material (see, e.g., Gathercole & Baddeley, 1993), it predicts working memory involvement in speech production. Working memory refers to the cognitive system responsible for the active maintenance, manipulation, and retrieval of relevant information for on-going cognition (Unsworth et al., 2009). Working memory is capacity limited, which means we can only effectively process a set amount of information at any given time and overloading working memory results in impaired performance (Engle, 2002). In spite of this, we are usually able to complete two unrelated tasks at the same time; for example, listening to the news while solving a jigsaw puzzle. It is much more challenging to complete two related tasks at the same time; e.g., telling a story while comprehending the news. According to Baddeley and Hitch (1974)'s multicomponent model, this is because different working memory components serve different types of information processing. Verbal working memory serves linguistic tasks; spatial working memory serves spatial relations. The embedded-processes model (Cowan, 1999) defines working memory as the selective activation of attentionally-focused memories within long-term memory. Poor performance arises from the limited capacity that our 'focus of attention' can hold up to four activated chunks.

Although the hypothesis of phonological-phonetic encoding also predicts verbal working memory involvement in speech planning and production, there is no direct evidence to support this prediction. Moreover, the indirect evidence suggests the opposite, that working memory is not relevant to speech planning and production (see, e.g., Gathercole & Baddeley, 1993: Ch. 4). This evidence is consistent with phonetically-informed theories of production that hypothesize that planning is based on the activation of word-sized chunks stored in long-term memory (Browman & Goldstein, 1992). These chunks abstractly encode relative timing information that guides articulatory movements over the course of word production, and are stored in association with the lexical concept. This association allows for their direct access, obviating the need for a phonological-phonetic encoding stage in production.

Research Questions and Predictions

The paper investigates whether working memory is involved in L2 speech production, while testing the phonological-phonetic encoding hypothesis against the retrieval hypothesis.
Following the literature that speaking uses verbal working memory resources, it tests working-memory load type effect during speech production. The research questions are:

1. Do speakers make more speech errors while they process a verbal task compared to when doing a spatial task?
2. If speakers produce correct sentences, is there any prosodic disruption from a verbal or spatial task while speaking?

The encoding hypothesis predicts that only verbal working memory effects on speech output. Disrupted speech is predicted under verbal load condition, distracted speech under spatial load due to divided attention, and normal speech under no-load.

The retrieval hypothesis predicts no (verbal vs. spatial) type effect. Because pre-stored phonological-phonetic chunks and the relevant articulatory movements are directly retrieved from long-term memory, no active manipulation of information is involved in the working memory system and thus there is no type effect.

METHODS

Participants

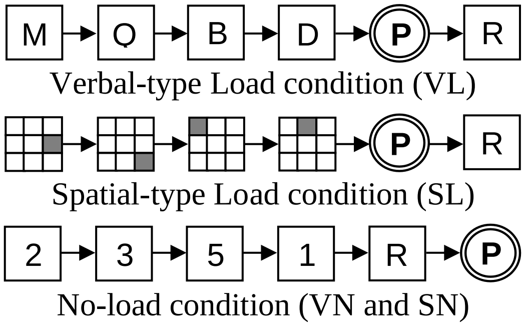
Participants were 20 (10 males, 10 females) adult (in their 20s) Korean EFL speakers, recruited at Seoul National University. All reported normal hearing and no history of speech-language therapy.

Speech Materials

The speech materials manipulated sentence structure in order to elicit differently prosodified sentences while controlling for phrase length. Thirty two sentences were designed around four structures (* eight sentences) manipulating the dependent relative clause (RC): the RC was either subject-extracted or object-extracted (e.g., *the smart shy boy that liked the quiet girl cut the cake* versus *the fat black cat that the mad dog hurt climbed the tree*); and, the RC was either embedded in the middle of the matrix clause or appended to the end of it (e.g., *the sly gray wolf bit the sheep that wore the gold bell*). Each sentence consisted of 12 monosyllabic words: three definite articles (*the*), three adjectives, three nouns, two verbs, and one relativizer (*that*).

Working Memory Manipulation

The task manipulated working memory load during speaking using a modified complex span task (adapted from Unsworth et al., 2009). Load type was manipulated to tax either verbal or spatial working memory. In the load condition, a speaker was required to hold onto a sequence of four letters (verbal) or four spatial locations (spatial) while speaking aloud one of the stimulus sentences. After speaking the sentence (primary task), the speaker completed the load task by choosing the correct sequence of either letters or spatial locations from among a set of 8 options (distractor task). In the control, no-load condition, participants were presented with a sequence of four numbers and asked to choose the correct sum from among eight options before speaking the sentence. The primary production task (P in Figure 1) came either between the serial presentation of to-be-remembered items and recall (R, load conditions) or after recall (no-load condition).



Verbal-type Load condition (VL)

Spatial-type Load condition (SL)

No-load condition (VN and SN)

Figure 1. Working memory manipulation during speaking. The sentence to be produced (P) was presented either before or after the eight options used to test recall (R).

During presentation, each letter, location, or number remained on the monitor for 800 milliseconds. Each sentence was displayed on a single line for eight seconds, as were the eight response options. All letters in the verbal condition were consonants, in non-permissible sequences according to English phonotactics; none formed acronyms; each letter was to be pronounced as single syllable in Korean.

Elicitation

Data was collected in a within-subjects design, with two fixed factors: Load Type (verbal, spatial) and Load Condition (load, no-load). All participants produced all the sentences and engaged in all levels of Load Type and Load Condition.

Prior to the main experimental task, participants were given as much time as they needed to read through the sentences. This was intended to control for effects of language planning.

Once participants verbally expressed confidence in their comprehension of all the sentences, they proceeded to the main task, which was blocked by Load Type and Load Condition. At the beginning of each block, participants were provided with practice, which was comprised of both the span and elicitation tasks. This practice used a simple sentence *a happy fish was swimming in the river*. Once participants had completed the practice session, they clicked on the monitor to proceed to the main task for that block.

For the main task, the 32 sentences were divided into two sets of 16 sentences with four sentences from each of the four syntactic structures. Sentence assignment to a particular set was randomized for each participant. Each set was then assigned to a Load Type (verbal or spatial) and elicited in random order during the associated load and no-load conditions. The order in which participants completed elicitation under Load Type and Load Condition was also randomized. The fixed sequences of letters, locations, numbers were also randomly paired with the 16 sentences. Accordingly, each speaker produced a total of 64 sentences. Speech was digitally recorded using a Tascam DR-100MKIII. The entire experiment took no more than 60 minutes to complete.

Measurement

A total of 1,280 sentential productions were collected from 20 speakers * 32 sentences * two productions. Thirty four sentential productions were excluded due to non-linguistic disruption during production (e.g., coughing); the remaining 1,246 sentential productions were measured and coded for analysis.

Error determination. Each utterance was first transcribed and then coded as correct or incorrect productions. Incorrect productions included disfluencies and/or speech errors. Disfluencies were defined (following three out of four types in Maclay & Osgood, 1959) as (i) filled pause (e.g., *whipped the poor* <*uh*>), (ii) false start (e.g., <*the wa-*> *the wild bad guy*), and (iii) repeat (e.g., *the nice large cow* <*cow*>). Speech errors were defined (following Shattuck-Hufnagel, 1979) as (i) addition (e.g., *we*<*f*>*t* for *wet* or *aunt cleaned* <*up*> for *aunt cleaned*), (ii) omission (e.g., *spot*() for *spots* or *the smart* () *boy* for *the smart shy boy*), (iii) substitution (e.g., {*th*}*ick* for *sick* or {*girl*} for *friend*), (iv) exchange (e.g., *g*{*lu*}*ped* for *gulped* or *that* {*the had*} for *that had the*), and (v) shift (e.g., from Shattuck-Hufnagel, 1979, *myn ow way* for *my own way* or *give youing* for *giving you*). To examine prosodic (i.e., rhythmic and intonational) differences (e.g., durational variability) and to avoid subjectivity (c.f., Maclay & Osgood, 1959, p. 24), durational variations that may traditionally be disfluencies were coded as correct, i.e., unfilled pauses (the fourth type in Maclay & Osgood, 1959) and lengthening without semantic change (e.g., *th*[*e*~], [*s*~]*ick*). Nonnative phonemic qualities that are part of the speaker's speech system (of competence errors) were not coded as errors because they were not errors that were affected by the manipulated working memory conditions. The determination was based on both the same speaker's other productions and the possibility of the incorrectly produced English phoneme falling into a single Korean phoneme. For example, if a speaker consistently produces a /p/-like /f/ as in *fat* /pæt/ and if /p/ and /f/ may be categorized as a single Korean consonant /p^h/, the production was considered correct. The same applied to *read* /lid/, *big* /big/, *cook* /kuk/, etc.

Acoustic measurement. All matched sentential pairs that were correctly produced by the same speaker were selected for acoustic measurement and analyses. They were acoustically segmented first into spoken chunks, then into vocalic intervals. Nine measures were obtained:

(i) sentence duration, in seconds including unfilled pauses;
(ii) articulation rate, as syllables per second excluding pauses;
(iii) duration variability, in word durations using the normalized pairwise variability index (nPVI; Low et al., 2000);
(iv) duration range: minimum word duration subtracted from maximum of the same sentence;
(v) pitch initial, sentence initial pitch in Hertz represented by the median F0 of vocalic interval in the first content word (or the second word) per sentence;
(vi) pitch mean, of medians across the vocalic intervals for each sentence;
(vii) pitch variability, calculated using the nPVI formula;
(viii) pitch range, max - min median pitch of the same sentence;
(ix) articulation clarity, in vowel space area (in F1 x F2).

For articulation clarity, the first three formant-values at the midpoint of content-word monothongal vowels were first extracted automatically from 85 words and 9 vowels per speaker. Tracking errors were hand-corrected. The frequency values in Hz were then converted to Bark using the formula $Z = [26.81/ (1 + 1960/f)] – 0.53$, where any Z values lower than 2 Bark were corrected using $Z' = Z + 0.15 (2 – Z)$ (as proposed in Traunmüller, 1990). Formant values were then normalized for vocal tract length using a modified Bark Difference Metric (Syrdal & Gopal, 1986). Bark-transformed F0 was subtracted from bark-transformed F1 (i.e., Z1-Z0) to model vowel height; bark-transformed F2 was subtracted from bark-transformed F3 to model tongue advancement (Z3-Z2). Mean height and tongue advancement for each vowel type produced by each speaker gave us per-vowel centroids for each speaker, blocked by load type and load condition. MATLAB (version R2019b 9.7.0.1319299) calculated the area of each vowel space by creating a polygon around the boundary points. A larger area was assumed to correlate with clearer articulation in that the vowels were farther apart and potentially more distinctive.

Statistical Analyses

Phonological encoding

Generalized linear mixed-effects logistic regression models with logit link function, implemented in the package lme4 (Bates et al., 2015) in R, version 3.6.3 (R Core Team, 2020) were used in combination with multimodel inference, implemented in the R package MuMIn (Bartoń, 2019) and likelihood ratio tests (via anova function in R). The best model justified by the data (e.g., ranked first with w_i of .092 in a multimodel inference) included the following: the dependent variable was presence of speech error in a sentence (henceforth speech error); fixed factors were Load Type (verbal, spatial) and Load Condition (load, no-load); random intercepts were speakers and sentences; within-unit random slopes were initially included for maximal random effects structure but were removed due to nonconvergence and singular fit. Linguistic factors RC Type ($\chi^2(1) = 1.03$, $p = .311 > .05$, in a likelihood ratio test) and RC Location ($\chi^2(1) = 0.55$, $p = .459$) were justified to be removed from the model. Neither RC Type nor RC Location predicted speech error ($b = -.54$, $se(b) = .32$, $p = .088$; $b = -.12$, $se(b) = .32$, $p = .701$) or interacted with each other ($b = .61$, $se(b) = .45$, $p = .180$) or with other factors ($p > .05$).

Phonetic encoding

A factorial multivariate analysis of covariance (MANCOVA, with manova syntax in SPSS version 25.0.0.1) evaluated the effects and interaction of Load Type and Load Condition (IVs) on weighted multivariate composite of speech production (DV) after the composite was adjusted by the sentences (CV). The nine acoustic measures described above constituted the composite. Significant influence of the covariate on the multivariate composite ($F(279, 5112) = 3.77$, $p = .000$) and on six univariate measures ($p < .006 = .05/9$, Bonferroni procedure) supported the importance of controlling for sentences. Homogeneity was assumed in the insignificant interactions between the IVs and the CV on the composite ($p = .779$; .145). Pillai's Trace (P) was adopted as test statistics for multivariate significance. Simple effects tests examined interaction effects, i.e., which IV groups impacted the multivariate and the univariate DVs. Alpha was adjusted (i.e., .05/2 = .025, Bonferroni procedure) to maintain the probability of Type I error at .05.

RESULTS

Phonological Encoding

32% (*N* = 399) out of 1,246 sentences were incorrectly produced with at least one disfluency and/or error. Speech error was statistically significantly predicted by Load Type, *b* = .48, *se*(*b*) = .17, *p* = .005 < .05, and by Load Condition, *b* = -.69, *se*(*b*) = .19, *p* < .001, with a significant interaction effect, *b* = -.58, *se*(*b*) = .26, *p* = .025 < .05. The working memory factors explained 5.2% (marginal R^2_{GLMM}, σ^2_ε, or 6.9%, σ^2_d) of the variance in speech error; the entire model including both the fixed and the random factors explained 12.5% (conditional R^2_{GLMM}, σ^2_ε, or 16.6%, σ^2_d) of the variance.

More sentences were produced incorrectly in the load conditions (41.8% = 260/622) than in the no-load conditions (22.3% = 139/624), and in the verbal type (34.2% = 214/625) than in the spatial type (29.8% = 185/621). As in Figure 2, speakers produced more incorrect sentences while multitasking than when only speaking: VL ≠ VN, *b* = -1.26, *se*(*b*) = .19, *p* < .001; SL ≠ SN, *b* = -.69, *se*(*b*) = .19, *p* < .001. They spoke even more sentences incorrectly when they had to memorize consonant sequences while speaking (47.1% = 147/312) than when remembering spatial locations while speaking (36.5% = 113/310), VL ≠ SL, *b* = .48, *se*(*b*) = .17, *p* = .005 < .025 = .05/2. Speaking different sentences, as in the two blocks of the no-load condition, did not increase error rate, VN = SN, 21.4% = 67/313, 23.2% = 72/311, *b* = -.11, *se*(*b*) = .20, *p* = .573.

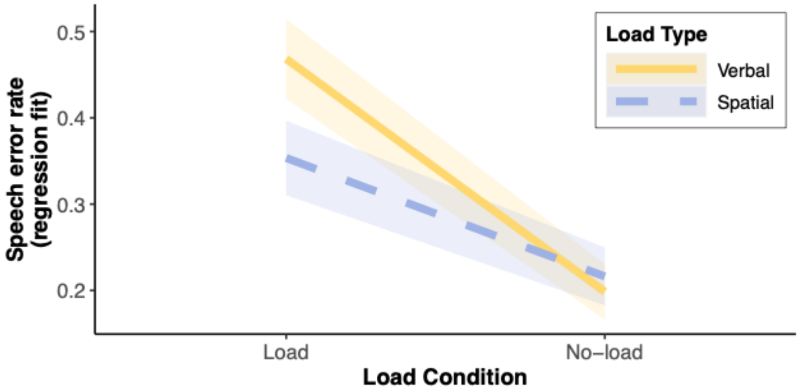


Figure 2. Increased speech error rate during a verbal task and during multitasking. The colored area around the regression lines denotes standard error of regression coefficient.

Phonetic Encoding

600 matched sentences were analyzed. Overall speech production patterns, as weighted multivariate combination of the nine dependent measures, varied systematically with Load Type, *P* = .03, *F*(9, 587) = 2.14, *p* = .025 < .05, η^2 = .01, and Load Condition, *P* = .09, *F*(9, 587) = 6.53, *p* < .001 < .05, η^2 = .08, with significant interaction, *P* = .03, *F*(9, 587) = 2.25, *p* = .018 < .05, η^2 = .03. The interaction accounted for 3.3% of the composite variance (Λ = .97). Standard discriminant function coefficients (*SDFC*) and structure coefficients (*r*) indicated articulation rate (*SDFC* = -.94, *r* = -.53) and articulation clarity (.79, .69) contributed primarily to distinguishing the working memory groups.

Multivariate simple effects tests indicated that only the verbal task statistically significantly changed the acoustic composite patterns of the produced speech. Speech produced during a verbal task exhibited significantly different composite scores (i) from speech during a spatial task, VL ≠ SL, *F*(1, 597) = 19.68, *p* < .001, and (ii) from speech produced without additional processing load, VL ≠ VN, *F*(1, 597) = 62.86, *p* < .001. By contrast, sentences spoken during a spatial task were not acoustically different from the same sentences produced without additional load, SL = SN, *F*(1, 597) = 3.23, *p* = .073. Likewise, different sentences did not show systematic acoustic differences, VN = SN, *F*(1, 597) = 1.16, *p* = .282. Figure 3 graphs the relations.

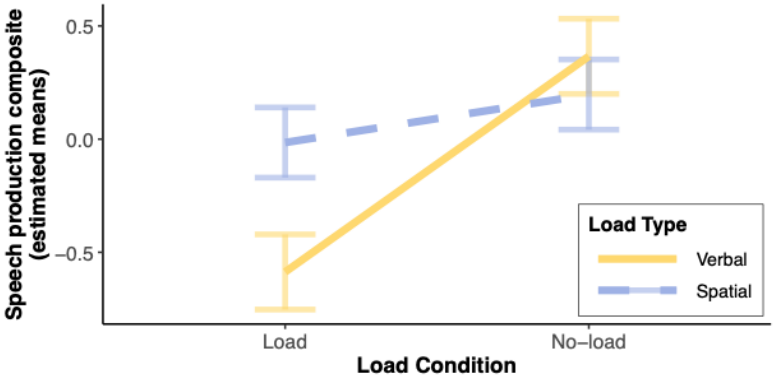


Figure 3. Speech production composite influenced only by verbal working memory load. Error bars indicate 95% Confidence Interval.

Univariate simple effects tests suggest that the significant effects of working memory load on speech composite came from durational aspects and vowel articulation. VL was different from SL in sentence duration, *F*(1, 597) = 9.44, *p* = .002 < .025, articulation rate, *F*(1, 597) = 7.11, *p* = .008, and articulation clarity, *F*(1, 597) = 8.55, *p* = .004. VL was also different from VN in sentence duration, *F*(1, 597) = 5.41, *p* = .020 < .025, articulation rate, *F*(1, 597) = 15.82, *p* < .001, duration variability, *F*(1, 597) = 8.24, *p* = .004, and articulation clarity *F*(1, 597) = 32.28, *p* < .001. Speakers completed speaking a sentence faster during a verbal working memory task (*M* = 3.88 seconds, *SD* = 0.63) than during a spatial task (4.12, 0.73) or than when without additional task (4.07, 0.60); they articulated more words (or syllables) per second (*M* = 3.34, *SD* = 0.42 vs. *M* = 3.21, *SD* = 0.41 or *M* = 3.15, *SD* = 0.41); their vowels were acoustically closer together (having a smaller vowel space, 8.15, 2.62 vs. 9.07, 2.62 or 9.83, 2.33). They also produced the same sentences with less variable word durations during a verbal task (*M* = 64.99, *SD* = 11.50) compared to the no-load condition (VN, *M* = 68.48, *SD* = 11.49). All other differences were statistically insignificant. Figure 4 exemplifies one durational aspect and articulation rate. Figure 5 demonstrates how speakers’ vowel space (from means of all speakers’ by-vowel centroids) got smaller due to verbal load during speaking.

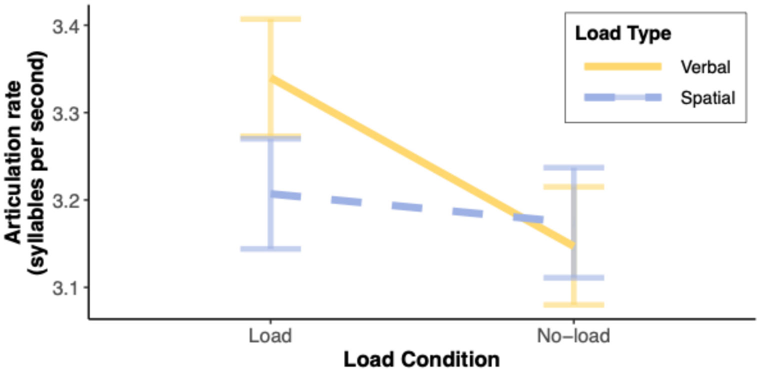


Figure 4. Articulation rate: Faster speech during a verbal task.

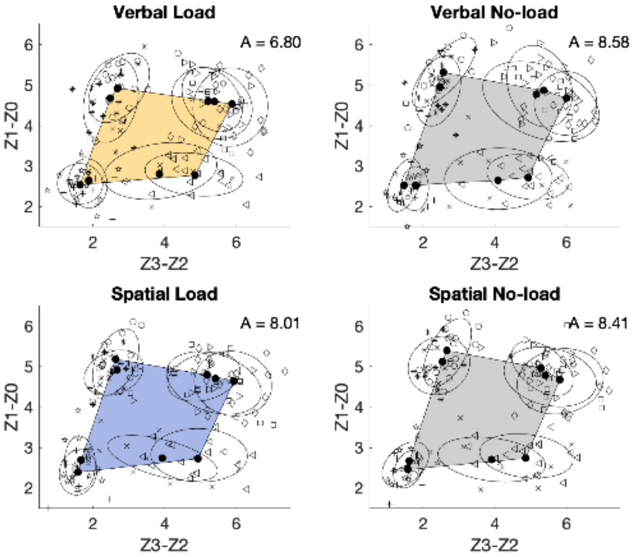


Figure 5. Articulation clarity: Smaller vowel space (area, A) and less distinctive articulation during a verbal task.

DISCUSSION

This paper contributes only a preliminary step to understanding the effects of working memory in speech production. Follow-up experiments may present the letters and locations simultaneously with the sentences, to make the tasks more like a concurrent multitask than a sequence of two tasks. Interaction with English-speaking proficiency may give better insight into this issue, such as whether advanced L2 learners pattern like L1 speakers.

This study, however, provides direct evidence of significant effects of working memory load type and condition on L2 speech planning and production. The L2 speakers made more speech errors when multitasking than when speaking without an additional task. They produced even more errors when the task was verbal than when it was spatial. Likewise, multitasking also impacted the acoustic patterns of the produced speech. However, the spatial task did not change the speech patterns statistically significantly. Only the verbal task significantly impacted the speech acoustics. While trying to memorize a sequence of English consonants, the Korean EFL speakers articulated the English sentences faster, in terms of both the total duration, including pauses to complete a sentence, and the articulation rate, excluding pauses. Their produced words were durationally less variable across sentences, and the articulated vowels were acoustically closer to one another indicating a possibility that the speakers failed to pay more attention to articulate the vowels more clearly and distinctively. These production differences may well be associated with faster speech, as partly indicated in significant correlation results among these measures. Faster speech may have resulted in more errors, smaller durational ratios, and less time to move articulators to hit articulatory targets. Speed-up due to added processing load was also reported for perception in Dronjic (2013), although load types were irrelevant. The L1 and L2 readers of English speeded up reading when they were to remember the result of a math calculation and concurrently judge morphological grammaticality.

Significant type effect in L2 production is consistent with the predictions of encoding hypothesis and multicomponent model of working memory. The morphosyntactic forms, displayed on a monitor, become input for phonological-phonetic encoding, when verbal working memory allows for planning the metrical and segmental specifications and executing the prosodic words. Thus, it may tax speakers' verbal working memory resources to engage in another linguistic task of remembering a letter sequence on top of speaking. The overloaded system results in malfunctioning. A spatial task does not overload verbal working memory because it is processed separately in the spatial component of working memory.

By contrast, if the phonological-phonetic information is retrieved from articulatory templates in long-term memory and executed automatically as overly practiced articulatory behaviour, we should not find significant type effects. The embedded-processes model does not account for type effect because focus of attention does not distinguish the type of information.

It may be true that L1 speech production is automatic and working memory is irrelevant to L1 speech production (Gathercole & Baddeley, 1993). A study on English L1 by Lee and Redford (2015) found significant effect of load condition, i.e., more errors and faster speech during a working memory task. However, different from the L2 speakers, the L1 speakers produced similar speech patterns regardless of the type of load. This dissociation between L1 and L2 implies dissociation of psycholinguistic processes underlying L1 and L2 speech planning and production. We accordingly propose a speech production model as in Figure 6.

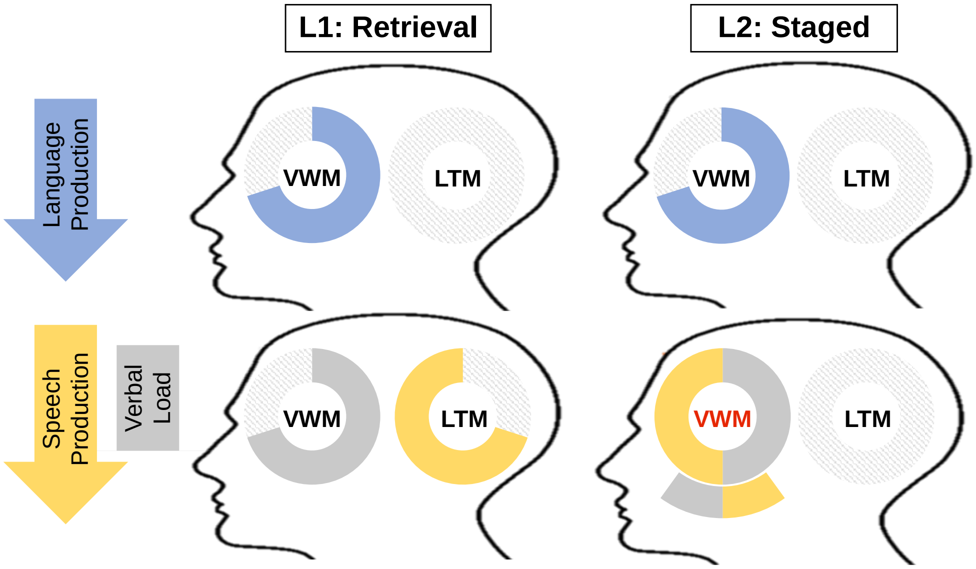


Figure 6. Proposed speech production model. L2 speech production requires active ongoing planning via verbal working memory, different from L1 production via retrieval from memory.

The paper suggests the following: L1 speech production uses the long-term memory resources by spontaneously retrieving the already-stored remembered articulatory gestures from memory; L2 speech production uses the working memory resources by actively manipulating the phonological-phonetic information and computing articulatory gestures. It accounts for the observed working memory overload and dissociation of load type effect between L1 and L2. Given the verbal load from letter sequences taxing verbal working memory, L1 speech production is not impacted as the letter sequences are the only task that uses working memory resources. By contrast, L2 speakers' verbal working memory is overloaded by holding and processing two verbal tasks.

ACKNOWLEDGMENTS

We thank Melissa A. Redford (University of Oregon) for valuable advice.

ABOUT THE AUTHORS

Ogyoung Lee is a PhD candidate in the Department of English Language Education at Seoul National University. Her research focuses on psycholinguistic and working memory processes underlying L1 and L2 speech production and acquisition.

Hyunkee Ahn is full professor in the Department of English Language Education at Seoul National University. His research interests include perception and production of English as a foreign language, L2 phonemic awareness, and L2 pronunciation instruction and learning.

REFERENCES

Baddeley, A., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York, NY: Academic Press.

Bartoń, K. (2019). *MuMIn: Multi-Model Inference*. R package version 1.43.15. https://CRAN.R-project.org/package=MuMIn.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. Doi:10.18637/jss.v067.i01.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155-180.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). New York, NY: Cambridge University Press.

Dronjic, V. (2013). *Concurrent memory load, working memory span, and morphological processing in L1 and L2 English* [Unpublished doctoral dissertation]. University of Toronto.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19-23.

Fowler, C. A. (2010). Speech production. In I. B. Weiner, & W. E. Craighead (Eds.), *The corsini encyclopedia of psychology, 4th edition* (Vol. 4, pp. 1685-1687). Hoboken, New Jersey: John Wiley & Sons, Inc.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. New York, NY: Psychology Press.

Lee, O., & Redford, M. A. (2015). Verbal and spatial working memory load have similarly minimal effects on speech production. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK. ISBN 978-0-85261-941-4.

Levelt, W. J. M., Roelofs. A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75.

Low, L. E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, *43*, 377-401.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*(1), 19-44.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295-342). Hillsdale, NJ: Lawrence Erlbaum Associates.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America*, *79*(4), 1086-1100.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of Acoustical Society of America*, *88*(1), 97-100.

Unsworth, N., Redick T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*(6), 635-654.