

A lexical bottleneck in shadowing and translating of narratives

Jeroen van Paridon ^a, Ardi Roelofs^b and Antje S. Meyer^a

^aMax Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; ^bDonders Centre for Cognition, Radboud University, Nijmegen, The Netherlands

ABSTRACT

In simultaneous interpreting, speech comprehension and production processes have to be coordinated in close temporal proximity. To examine the coordination, Dutch-English bilingual participants were presented with narrative fragments recorded in English at speech rates varying from 100 to 200 words per minute and they were asked to translate the fragments into Dutch (interpreting) or repeat them in English (shadowing). Interpreting yielded more errors than shadowing at every speech rate, and increasing speech rate had a stronger negative effect on interpreting than on shadowing. To understand the differential effect of speech rate, a computational model was created of sub-lexical and lexical processes in comprehension and production. Computer simulations revealed that the empirical findings could be captured by assuming a bottleneck preventing simultaneous lexical selection in production and comprehension. To conclude, our empirical and modelling results suggest the existence of a lexical bottleneck that limits the translation of narratives at high speed.

ARTICLE HISTORY

Received 20 March 2018
Accepted 21 February 2019

KEYWORDS

Lexical bottleneck;
simultaneous interpreting;
speech comprehension;
speech production;
shadowing

Simultaneous interpreting, also known as conference interpreting, is the online oral translation of spoken language. Most often used at international conferences and institutions, this mode of translation provides a near instantaneous translation to the listener. While there is an extensive (and often contradictory) literature on cognitive differences between interpreters and non-interpreter bilinguals (e.g. Morales, Padilla, Gómez-Ariza, & Bajo, 2015; Woumans, Ceuleers, Van der Linden, Szmalec, & Duyck, 2015), the processes of speech comprehension and production occurring during simultaneous interpreting have not been studied in much detail. Behavioural studies have examined the linguistic skills involved in interpreting (Christoffels, De Groot, & Kroll, 2006; Christoffels, De Groot, & Waldorp, 2003), and neuroimaging has started to identify the neural bases of interpreting (Hervais-Adelman, Moser-Mercer, & Golestani, 2015; Hervais-Adelman, Moser-Mercer, Michel, & Golestani, 2015). However, no theory of interpreting exists that describes the time course of concurrent speech comprehension and production in simultaneous interpreting.

Even though professional interpreters are highly trained at concurrent listening and speaking, comprehension and production are still somewhat impaired by their temporal overlap during interpreting. More errors are

made during interpreting than during simple shadowing, and interpreting a speech leads to significantly worse recall than simply listening to that speech (Gerver, 1974). Additionally, interpreters cannot interpret at very high speech rates. In a seminal study, Gerver (1969) had six professional interpreters shadow recordings of diplomatic speeches played at different speeds, while six others interpreted the same recordings. For the materials used by Gerver, the maximum input speech rate for fluent French to English interpreting (with more than 90% of words being translated correctly) was around 110 words per minute on average, with performance declining linearly at higher input speech rates to less than 60% correct at 164 words per minute. Below the maximum input rate, interpreters can approximately match the output speech rate to the input speech rate, producing mostly complete and correct translations. At higher input rates, interpreters start to omit words and phrases, and produce in short, high speech rate bursts. The maximum interpreting rate lies well below the maximal speech rate interpreters can comprehend or produce when not interpreting, as evidenced by Gerver's shadowers, who were still fluent at 142 words per minute. These differences suggest that the limit on interpreting rate is not set directly by limits on the processes of speech comprehension or production separately, but rather by limits on the speech system as a

CONTACT Jeroen van Paridon  jeroen.vanparidon@mpi.nl

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/23273798.2019.1591470>.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

whole arising when comprehending and producing speech concurrently.

How this coordination is achieved in a fluent manner, why it breaks down at high input speech rates, and how the resulting error pattern comes about is not explained by any of the relatively few models of interpreting that have been put forward in over half a century of interpreting research. Models of simultaneous interpreting can be grouped into several categories. One type is the *effort* model proposed by Gile (1997), which poses that interpreting consists of four different types of *effort*: listening, production, memory, and coordination. These types of effort are assumed to be additive and to simultaneously require capacity. It is not apparent, however, how this model might be tested empirically. Another type of model is the *process* model that describes the organisation of processing in interpreting. This type of model tends to resemble a complex flowchart of processing steps, but none of the existing models makes falsifiable predictions about measurable indices of interpreting processes such as timing, error rates, or error types (Gerver, 1975; Mizuno, 2005; Moser, 1978). The only interpreting model that has been empirically tested is the *cognitive load* model by Seeber and Kerzel (2012), which makes predictions about physiological indices of cognitive load (i.e. pupil diameter) based on hypotheses about the processing demands of different types of linguistic input. Seeber and Kerzel found that translating German SOV (subject-object-verb) sentences into English SVO (subject-verb-object) sentences produced a marginally higher cognitive load than translating from SVO into SVO sentences. Their examples of SOV sentences include long-distance dependencies, however, which could explain the increased cognitive load regardless of task demands specific to interpreting. Despite this apparent confound, their model suggests that word order might play a role in interpreting performance, but does not explain the specific limits on interpreting speech rate and the associated error patterns.

A model of simultaneous interpreting of the type that Gerver (1975) suggested, that is a model that explains all of the linguistic and metalinguistic processes a professional interpreter relies on, cannot currently be specified in quantitative terms such as latencies or error rates. This is because we do not have a sufficiently detailed understanding of all the processes involved. However, a simpler, purely lexical model that explains only the simultaneous word comprehension and production aspect of interpreting can be generated by combining behavioural data of the type collected by Gerver (1969) with current psycholinguistic models of word production and comprehension. Such a model would not be

a complete model of interpreting but it could extend experimentally supported psycholinguistic models of word production and comprehension and describe the coordination of production and comprehension, which is one of the key elements of simultaneous interpreting. Showing that such a model predicts error rates in simultaneous production and comprehension would demonstrate that interpreting is subserved by normal language processing, albeit under abnormal task demands. More generally, the specific adaptations needed to simulate the error rates reported by Gerver (1969) could provide new insights into the way comprehension and production are coordinated when fluid and frequent transition between the two is required.

Prior work on spoken word production in a dual-task paradigm has demonstrated that semantic interference in a production task can cause delays in response selection for a second, unrelated task performed at the same time, whereas a phonological effect in the same production task does not always propagate to the second task. This suggests that central attention is required for response selection at the lemma level, but not (or less) at the phonological level (Cook & Meyer, 2008; Ferreira & Pashler, 2002; Piai, Roelofs, & Schriefers, 2014; Roelofs, 2008; Roelofs & Piai, 2011). Having to coordinate selections at the lemma level for both comprehension and production could conceivably create a lexical-selection bottleneck during interpreting. The present study examined whether a computational model of interpreting and shadowing that includes such a bottleneck could account for error rates in relevant behavioural data. This was done by adding a lexical bottleneck to the model of word production and comprehension proposed by Indefrey and Levelt (2004).

Of course, generally speaking, syntactic processing must be an important component of interpreting, especially where the source text and the correct translation differ in word order. However, in our texts the English and Dutch word order were mostly the same. Moreover, the Indefrey and Levelt (2004) model does not describe syntactic processes beyond the assumption that lemma selection affords access to the syntactic properties of a given word. Therefore, we chose, as a parsimonious starting point for the model, not to include an explicit processing cost for syntactic processing, but rather to test whether the lexical model suffices to simulate the relevant behavioural data.

To be able to test the model empirically, we first collected relevant behavioural data. To this end, we repeated Gerver's (1969) study comparing interpreting and shadowing performance at different speech rates, but with a more rigorously controlled design. The languages involved in the present study were English

and Dutch. Our design was a within-participants comparison of shadowing and interpreting performance with source texts presented at a range of speech rates. We recruited participants without prior interpreting training, to exclude the possible use of interpreting-specific processing strategies. The behavioural data were then used to fit our computational model. Note that the present work is concerned with switching between production and comprehension (either within the same or in different languages) as required in the shadowing and translation tasks, and not with the (code) switching between L1 and L2 speech production, which is often considered in studies of bilingualism. Switching between comprehension and production in interpreting is a type of switching between L2 and L1 that is not required during shadowing. But because in interpreting L2 is used exclusively for comprehension and L1 for production, there is no need for language-specific inhibition of response-selection in production, which is often hypothesised to be the cause of bilingual switch costs (e.g. Meuter & Allport, 1999).

Method

Participants

The participants were native speakers of Dutch, recruited from the participant pool at the Max Planck Institute for Psycholinguistics. To identify Dutch-English bilinguals with sufficient proficiency in English to perform the tasks, 215 participants were screened using LexTale, an online English vocabulary test, which correlates well with other measures of English proficiency (Lemhöfer & Broersma, 2012). From this group, participants with a LexTale score over 85% (the top 33% of test takers) were invited to participate in the study. Of the invitees, 20 agreed to take part in the study (13 female, mean age 22.3 years). Mean self-reported age of acquisition of English was 10.3 years ($SD = 1.1$ years, $n = 14$), approximately the age at which English education starts in Dutch primary schools. None of the participants had prior experience in shadowing or interpreting; their mean LexTale score was 91.4% ($SD = 5.1\%$).

Procedure and design

Participants performed two sessions of shadowing and interpreting, one week apart. Both sessions were recorded but the first session was meant solely to familiarise participants with the tasks and was not analysed. The second session consisted of two blocks of roughly 20 min: one shadowing block of five spoken texts presented at different speech rates and one interpreting

block of five spoken texts presented at different speech rates. The order of texts, tasks, and speech rates was counterbalanced across participants so that each text was presented to two participants at every speech rate and in both tasks, but texts and speech rates were not repeated within participant.

Materials

Stimuli were ten samples of around 300 words in length, taken from a variety of books for children between six and ten years. Children's books were selected because they feature few rare lexical items and few complex syntactic structures that would require extensive reformulation during translation. Using a teleprompter script, the sample texts were recorded by a male native speaker of English at a controlled rate of 150 words per minute. To produce the desired stimulus speech rates, the recordings were sped up or slowed down to 100, 125, 150, 175, and 200 words per minute using the Audacity audio editor (Version 2.0.6; Audacity Team, 2014). Because the digital speech rate manipulation produces audible distortions in the recordings, the stimulus texts were rerecorded by the same speaker while playing the digitally sped-up or slowed-down recordings over headphones as a continuous speech rate cue.

Analysis

Shadowing performance was scored by transcribing participant recordings and counting the percentage of words correctly reproduced from the source text. Interpreting performance could not be scored so straightforwardly; instead, recordings were transcribed by native speakers of Dutch and the percentage of words from the source text that was represented in the transcription was taken as the score. Scoring was double-checked by a second native speaker of Dutch.

Speech rate, task, and interaction effects on performance were analysed with a logistic mixed-effects model using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R (Version 3.3.3; R Core Team, 2017). Statistical inference for the coefficients was computed using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017).

Results

Figure 1 shows participants' performance. As expected, both shadowing and interpreting performance decreased as the source text speech rate increased. However, the difference in slopes of interpreting ($-.26\%$ per wpm) and shadowing ($-.17\%$ per wpm)

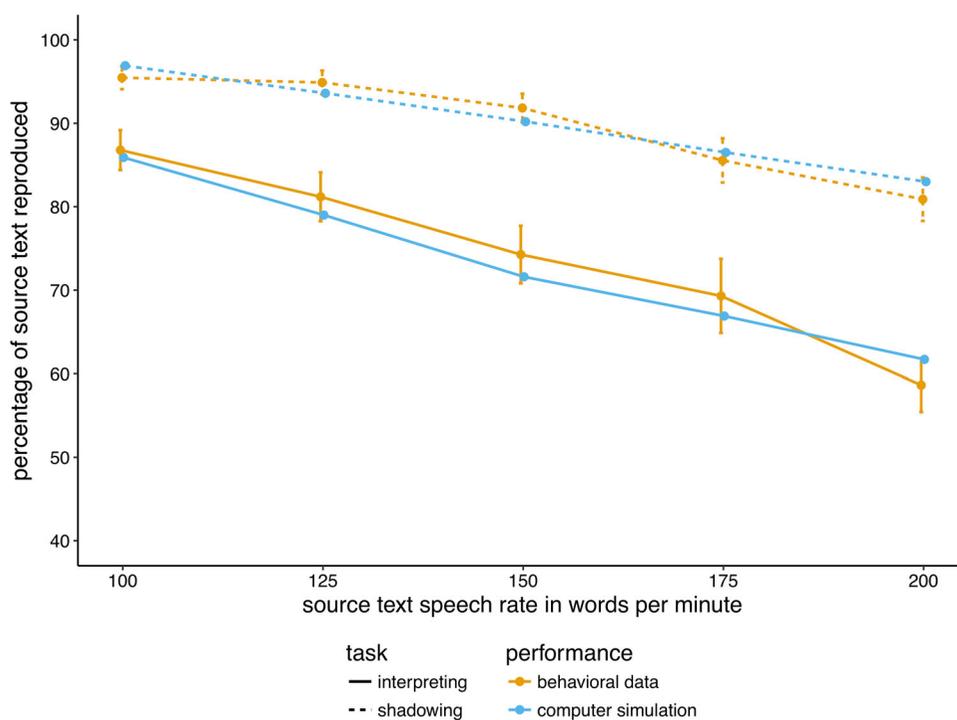


Figure 1. Mean percentage of source text reproduced while shadowing and interpreting across source speech rates in the behavioural experiment and model. Error bars represent standard error for the behavioural data.

performance across source speech rates indicates an interaction effect.

When controlling for random effects of participant and source text with random intercepts, there were significant effects of task ($\beta = 0.63$, $SE = 0.014$, $p < .001$), source speech rate ($\beta = -0.62$, $SE = 0.014$, $p < .001$), and the interaction between task and source speech rate ($\beta = -0.05$, $SE = 0.014$, $p < .001$). Controlling for random effects of participant and source text using a more elaborate random effects structure with random slopes was not possible due to the limited number of observations in each cell.

Discussion

Shadowing performance was at ceiling at 100 and 125 words per minute, which is consistent with the shadowing performance reported by Gerver (1969). Aside from the ceiling effect, the decrease in performance was roughly linear for both interpreting and shadowing. While there appears to be an interaction of source speech rate and task in the data reported by Gerver similar to the interaction in the present study, the deterioration in interpreting performance with increasing source speech rate was more severe in Gerver's data (from 95% correct to less than 60% correct for a 70% increase in speech rate) than in the present study (from 87% correct to 59% correct for a 100% increase

in speech rate). This difference may have been caused by the nature of the source materials used by Gerver, although this should be partially mitigated by the professional-level proficiency of the participants in that study. Another possible cause is the design used by Gerver: a between-participants design with only six participants performing each task. Such a design is likely to be underpowered and more susceptible to noise than the present within-participants design with 20 participants performing both tasks. Regardless of the quantitative differences between the present results and those reported by Gerver, the notion that there are different factors limiting interpreting and shadowing performance at high source speech rates is supported by the interaction of source speech rate and task in both studies.

What these factors that limit interpreting and shadowing are is not obvious from the behavioural data. Participants reported feeling that interpreting at high speech rates required alternately attending to input and producing output, as in task switching. If one of these tasks takes too long (e.g. a long sequence of words needs to be produced, but in the meantime new input is coming in) words are lost, often several at a time. At lower speech rates, participants reported that interpreting felt "more natural" or "automatic", which either reflects an ability to genuinely attend to both comprehension and production at the same time or more

fluent task switching that participants are not as aware of.

One difference between the shadowing and interpreting tasks that potentially modulates the effect of task on performance is the production language. During interpreting the participants spoke in their native language (Dutch), while during shadowing they spoke in their second language (English). Participants were screened for English proficiency, but most likely production was easier in their native Dutch. However, any native language advantage would only serve to increase performance in the interpreting task and therefore decrease or mask the task effect.

Computational model

The observed interaction between task and speech rate suggests a temporal coordination problem that causes shadowing and interpreting performance to differentially degrade with increasing speech rate. To identify the source of that problem, we constructed a simple computational model. Based on dual-task studies of

language production (Cook & Meyer, 2008; Ferreira & Pashler, 2002; Piai et al., 2014; Roelofs, 2008; Roelofs & Piai, 2011), we set out to test the assumption of a lexical-selection bottleneck. The model represents the combined model of speech comprehension and production presented by Indefrey and Levelt (2004), implemented as a chain of consecutive processing stages, as illustrated in Figure 2. It takes as input a sequence of words and their onset times. To replicate the behavioural paradigm as closely as possible, the recordings presented to the participants were used to generate these input sequences for the model. We used the WebMAUS automatic speech segmentation service to assign onset times to each word in an orthographic transcription of the recordings (Kisler, Reichel, & Schiel, 2017).

Each word starts at the first processing stage in the model and is then passed along after being processed for a specific duration. This was implemented computationally by representing each processing stage as a simplified linear ballistic accumulator; the simplification being that the rate of evidence accumulation was fixed

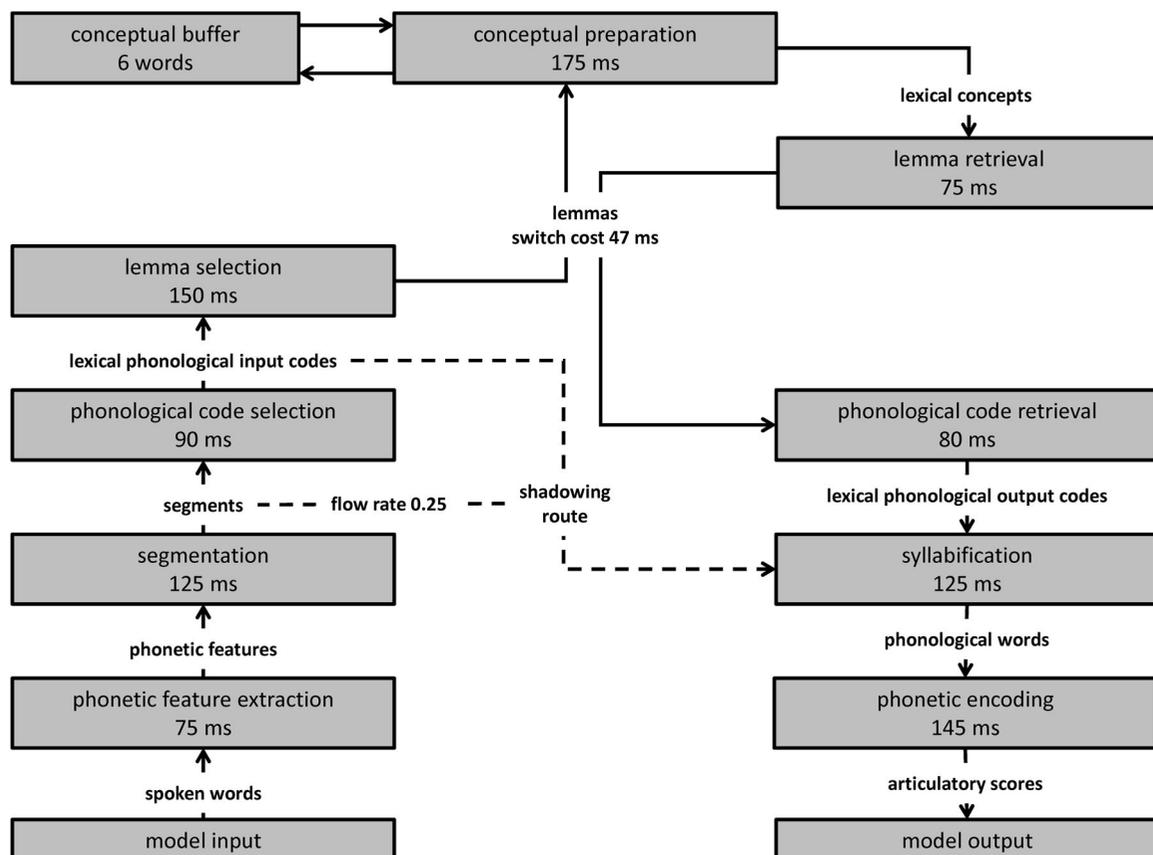


Figure 2. Structure and parameters of the computational model. The solid arrows are hypothesised to represent a route used in both simultaneous interpreting and shadowing, while in shadowing many words can also be reproduced along the route represented by the dashed arrow. Parameters set using particle swarm optimisation are conceptual buffer size, lemma switch cost, segmentation-syllabification accumulation rate, and function word accumulation rate.

to a rate at which the time to reach threshold matches the durations reported by Indefrey and Levelt (2004; Indefrey, 2011) instead of drawing the accumulation rates from a normal distribution as originally proposed by Brown and Heathcote (2008). This simulated processing does not comprise any sort of linguistic processing because the model operates only on the onset times of the words. Details of component processes were unimportant as only the latencies of the processes and their interdependencies mattered (cf. Schweickert, 1980). Professional interpreters likely use interpreting-specific strategies to facilitate processing, but because we were attempting to model the error rates of untrained interpreters we did not attempt to model these processing strategies. Modelling interpreting-specific strategies might also reduce the validity of the model for describing the coordination of language comprehension and production in contexts other than interpreting.

To account for the reduced processing demands of function words compared with content words, function words were assigned an increased rate of evidence accumulation. As an initial estimate, the rate of evidence accumulation was set to double that of content words, but this value was later adjusted in a parameter optimisation procedure described below. On average, 52.8% of the words in a stimulus text were classified as function words. For a complete list of the words classified as function words in the present study see Appendix A in the Supplemental Materials.

The duration of conceptual processing was difficult to derive, as Indefrey and Levelt (2004) based their estimates on picture naming and single word listening, instead of a task that involves sentences and combines both speech comprehension and production. One commonly used experimental paradigm that requires sequential comprehension and production is single word translation. However, the reported latencies for single word translation vary from roughly 800 ms when words were presented orthographically (La Heij, Hooglander, Kerling, & van der Velden, 1996), to as much as 1200 ms when words were presented auditorily (De Groot, 1992). As an initial approximation, therefore, we adopted the 175 ms estimate reported by Indefrey and Levelt, because even though that estimate is derived from picture naming experiments, it leads to an overall single word interpreting latency that roughly matches the latencies reported by De Groot.

The remaining component process durations were also based on the latencies reported by Indefrey and Levelt (2004), and the resulting model fit was measured as root-mean-square deviation (RMSD) from mean participant performance across texts for each speech rate in the behavioural experiment. The initial model was a

poor fit for the behavioural data (combined RMSD = 5.4%). After observing that the poor fit was caused in part by the model systematically underperforming in the shadowing task, we added an extra connection from segmentation to syllabification to improve shadowing performance. The extra connection reflects the unique affordance in shadowing of starting selection of an output phoneme directly after identifying an input phoneme because the output is identical phoneme-for-phoneme to the input (cf. Roelofs, 2004, 2014). The existence of such a low-level connection is supported by the short latencies found in previous shadowing experiments (e.g. Fowler, Brown, Sabadini, & Weihsing, 2003). The connection was implemented by allowing additional evidence accumulation for syllabification from the moment segmentation is completed. Setting the accumulation rate through this additional segmentation-syllabification connection to an initial value of .5 markedly improved the model fit (combined RMSD = 2.6%).

To improve the fit of the model performance in the interpreting task, we first introduced a conceptual buffer into the model to make it more closely resemble human processing of consecutive words. This addition was based on the observation that it is not only possible to conceptually combine the meanings of a set of words and to reorder them before production, but that this is required during interpreting. In our computational model, we capture the function of the conceptual workspace by assuming a buffer that holds concepts until they can be passed to lemma selection for production.

Next, we implemented the critical assumption of a lexical-selection bottleneck (Cook & Meyer, 2008; Piai et al., 2014; Roelofs, 2008). The bottleneck was implemented at the lemma level, which is assumed to be shared between production and comprehension (e.g. Levelt, Roelofs, & Meyer, 1999; Roelofs, 2004, 2014). Lemma selection for production was blocked while selecting a lemma for comprehension, and vice versa. In translating, a switch is required between comprehension in one language (English) and production in another (Dutch), which results in a switch cost (e.g. Monsell, 2015, for a review). This switch cost means that delay of access to the lemmas from the production stream can last for multiple words if new input words come into the comprehension stream close enough together to not allow time to switch back to lemma selection for production in the meantime. To determine the optimal switch cost and conceptual buffer size we implemented a parameter optimisation procedure. Function word accumulation rate factor and segmentation-syllabification accumulation rate were also entered into the parameter optimisation procedure.

To optimise our simulation of participant behaviour, we minimised the model's RMSD from the mean participant performance across texts for each speech rate in the behavioural experiment for both tasks by varying its free parameters using particle swarm optimisation implemented in the Optunity parameter optimisation library (Claesen, Simm, Popovic, Moreau, & De Moor, 2014). Particle swarm optimisation uses a swarm of communicating particles moving through the parameter space looking for an optimal parameter set. Particle swarm optimisation does not use a gradient for optimisation and is therefore well suited to the problem of optimising the parameters of this model (Kennedy & Eberhart, 1995).

From 1920 iterations (96 particles for 20 generations), we selected the parameter set that produced the performance closest to that of the participants. Optimal parameters were a function word evidence accumulation factor of 2.0, a segmentation-syllabification accumulation rate of .25, a buffer length of 6 words, and a lexical selection switch cost of 47 ms (combined RMSD = 1.9%). This parameter set, and the model's structure, is reported in Figure 2.

Figure 1 shows the mean percentage of source text reproduced while shadowing and interpreting across source speech rates in the best-fitting model. In the simulations, the interpreting and shadowing performance progressively degraded with increasing speech rate, which corresponds to the empirical data. This degradation was stronger for interpreting than for shadowing, as empirically observed. Thus, the computational modelling suggests that to account for the data, it suffices to assume a lexical-selection bottleneck that precludes concurrent selection of lemmas in comprehension and production, and an associated switch cost.

General discussion

In the present study, we first replicated and expanded Gerver's (1969) study of interpreting and shadowing performance at different speech rates. We then used these data to test a computational model of interpreting and shadowing. The model structure and parameters were derived from a meta-analysis of speech comprehension and production experiments (Indefrey & Levelt, 2004). Performance of the computational model on a combined interpreting and shadowing measure most accurately simulated behavioural data when lemmas could not be selected concurrently for production and comprehension, creating a lexical-selection bottleneck with an associated switch cost. This switch cost is a possible explanation for the emergence of task-switching type

speech patterns at high speech rates, while at low speech rates comprehension and production seem to be temporally overlapping. The model suggests that temporal overlap is possible for processes such as phonetic decoding/encoding, phonological encoding/decoding, segmentation, and syllabification. Only lemma selection for comprehension and production cannot happen concurrently due to a lexical-selection bottleneck. At low speech rates, the switch cost can be "absorbed" into the pauses between words and the redundant parts of words that come after the uniqueness point. Therefore it is not perceivable to a listener that parts of production and comprehension are happening consecutively instead of concurrently. At higher speech rates the pauses are shortened and can no longer absorb the switch cost which then becomes a bottleneck, causing the model (and the participants) to periodically miss input or forget output, making the task set switching audible in the form of alternating bursts of listening and speaking.

Modelling the processing stages of speech comprehension and production as simplified linear ballistic accumulators makes the model as a whole computationally feasible, but it also necessitates that each consecutive step is discrete. This may be sufficient to capture the contrast between shadowing and interpreting, but it is important to note that recent more detailed models of single word production and/or comprehension (e.g. Roelofs, 2014; Ueno, Saito, Rogers, & Lambon Ralph, 2011; Walker & Hickok, 2016) feature connectionist components that more plausibly simulate phenomena such as interaction and competition in the speech system. Implementing a plausible connectionist model and fitting its parameters was not feasible for this study. Given the large number of parameters present in such a model and the relatively few data points it would be fitted to, there is an inordinate risk of overfitting. The lack of interaction between lower-level processes likely causes the present computational model to not capture facilitation or interference effects of cognates and incidental temporal coincidence of phonologically or semantically related words in the production and comprehension streams. It is unclear whether the net effect of this simplification in the model causes an over- or underestimation of the error rate. However, while the model does not capture small facilitation and interference effects, its contribution is that it postulates a critical path for both simultaneous interpreting and shadowing and demonstrates the temporal consequences for performance in both tasks. Future developments of the model could integrate cognate status and other lexical factors to allow for more specific predictions such as latency at the word level that cannot be derived

from the current model. Computational models like the recent Multilink model proposed by Dijkstra et al. (2018) present estimates of the effects of lexical factors such as semantic equivalence of possible translations and cognate status for single word translation; such estimates could be incorporated into an interpreting model as well.

As our model is mostly blind to linguistic content (with the exception of the distinction between function words and content words) and has no knowledge of syntax, any kind of temporal clustering of errors is simply due to the time course of the input and the structure of the model. Assuming that the bottleneck is situated at the lexical level appeared to be sufficient to explain the data. The fact that the model still replicates the error rates observed in participants who have syntactic knowledge, and use it to reformulate English sentences into Dutch sentences, is striking. The lack of need for a syntactic component in the model suggests that in the present study syntactic processing did not impose a significant time cost, possibly be due to the simplicity of the stimulus texts and the close correspondence in word order between the two languages. Syntactic processing is thought to be largely incremental in nature, both in production (Konopka & Meyer, 2014; Levelt, 1989) and comprehension (Altmann & Mirkovic, 2009; Christiansen & Chater, 2016), and once the entire message of a phrase is conceptualised, reformulating the type of short, grammatically straightforward sentences found in children's books may be so trivial that it does not cause meaningful additional cognitive load or delay. The occasional differences in word order between English and Dutch might merely require some extra time during conceptual processing, reflected in the small increase in conceptual processing duration needed for optimal model fit, when compared to the values Indefrey and Levelt (2004) report for conceptual processing during picture naming. For syntactically more complex texts or structurally more different languages, these model components and parameter values may be insufficient to model the syntactic costs. Under certain conditions, such syntactic costs may even constitute another bottleneck. In the present study, however, the model suffices to demonstrate that one important bottleneck is located at the lexical level.

Conclusion

Simultaneous interpreting and shadowing performance progressively degrades with increasing speech rate. This degradation is stronger for interpreting than for shadowing. Computational modelling showed that to account for the data, it sufficed to assume a lexical-

selection bottleneck that precludes concurrent selection of lemmas in comprehension and production and causes the associated switch costs.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Jeroen van Paridon  <http://orcid.org/0000-0002-5384-8772>

References

- Altmann, G. T. M., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33, 583–609. <https://doi.org/10.1111/j.1551-6709.2009.01022.x>
- Audacity Team. (2014). Audacity: Free audio editor and recorder (Version 2.0.6) [Computer Software]. Retrieved from <http://audacity.sourceforge.net>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3), 153–178. <https://doi.org/10.1016/j.cogpsych.2007.12.002>
- Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39, e62. <https://doi.org/10.1017/S0140525X1500031X>
- Christoffels, I. K., De Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, 54, 324–345. <https://doi.org/10.1016/j.jml.2005.12.004>
- Christoffels, I. K., De Groot, A. M. B., & Waldorp, L. J. (2003). Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting. *Bilingualism: Language and Cognition*, 6(3), 201–211. <https://doi.org/10.1017/S1366728903001135>
- Claesen, M., Simm, J., Popovic, D., Moreau, Y., & De Moor, B. (2014). Easy hyperparameter search using Optunity. arXiv preprint. arXiv:1412.1114
- Cook, A. E., & Meyer, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 886–899. <https://doi.org/10.1037/0278-7393.34.4.886>
- De Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1001–1018. <https://doi.org/10.1037/0278-7393.18.5.1001>
- Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2018). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, 1–23. <https://doi.org/10.1017/S1366728918000287>

- Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(6), 1187–1199. <https://doi.org/10.1037/0278-7393.28.6.1187>
- Fowler, C. A., Brown, J. M., Sabadini, L., & Weising, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49(3), 396–413. [https://doi.org/10.1016/S0749-596X\(03\)00072-X](https://doi.org/10.1016/S0749-596X(03)00072-X)
- Gerver, D. (1969). The effects of source language presentation rate on the performance of simultaneous conference interpreters. In E. Foulke (Ed.), *Proceedings of the 2nd Louisville conference on rate and/or frequency controlled speech* (pp. 162–184). Louisville, KY: University of Louisville.
- Gerver, D. (1974). Simultaneous listening and speaking and retention of prose. *Quarterly Journal of Experimental Psychology*, 26, 337–341. <https://doi.org/10.1080/14640747408400422>
- Gerver, D. (1975). A psychological approach to simultaneous interpretation. *Meta: Translators' Journal*, 20(2), 119–128. <https://doi.org/10.7202/002885ar>
- Gile, D. (1997). Conference interpreting as a cognitive management problem. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 196–214). London: Sage Publications.
- Hervais-Adelman, A., Moser-Mercer, B., & Golestani, N. (2015). Brain functional plasticity associated with the emergence of expertise in extreme language control. *NeuroImage*, 114, 264–274. <https://doi.org/10.1016/j.neuroimage.2015.03.072>
- Hervais-Adelman, A., Moser-Mercer, B., Michel, C. M., & Golestani, N. (2015). fMRI of simultaneous interpretation reveals the neural basis of extreme language control. *Cerebral Cortex*, 25, 4727–4739. <https://doi.org/10.1093/cercor/bhu158>
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, 2, 255. <https://doi.org/10.3389/fpsyg.2011.00255>
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101–144. <https://doi.org/10.1016/j.cognition.2002.06.001>
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of the IEEE International Conference on Neural Networks*, 4, 1942–1948.
- Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, 45, 326–347. <https://doi.org/10.1016/j.csl.2017.01.005>
- Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, 73, 1–40. <https://doi.org/10.1016/j.cogpsych.2014.04.001>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- La Heij, W., Hooglander, A., Kerling, R., & van der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. *Journal of Memory and Language*, 35, 648–665. <https://doi.org/10.1006/jmla.1996.0034>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44, 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38. <https://doi.org/10.1017/S0140525X99001776>
- Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, 40, 25–40. <https://doi.org/10.1006/jmla.1998.2602>
- Mizuno, A. (2005). Process model for simultaneous interpreting and working memory. *Meta: Translators' Journal*, 50(2), 739–752. <https://doi.org/10.7202/011015ar>
- Monsell, S. (2015). Task-set control and task switching. In J. M. Fawcett, E. F. Risko, & A. Kingstone (Eds.), *The Handbook of attention* (pp. 139–172). Cambridge, MA: MIT Press.
- Morales, J., Padilla, F., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Simultaneous interpretation selectively influences working memory and attentional networks. *Acta Psychologica*, 155, 82–91. <https://doi.org/10.1016/j.actpsy.2014.12.004>
- Moser, B. (1978). Simultaneous interpretation: A hypothetical model and its practical application. In D. Gerver & H. W. Sinaiko (Eds.), *Language, interpretation and communication* (pp. 353–368). New York, NY: Plenum Press.
- Piai, V., Roelofs, A., & Schriefers, H. (2014). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 147–165. <https://doi.org/10.1016/j.jml.2011.01.004>
- R Core Team. (2017). R: A language and environment for statistical computing (Version 3.3.3) [Computer Software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>.
- Roelofs, A. (2004). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review*, 111, 561–572. <https://psycnet.apa.org/doi/10.1037/0033-295X.111.2.561>
- Roelofs, A. (2008). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6), 1580–1598. <https://doi.org/10.1037/a0012476>
- Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, 59, 33–48. <https://doi.org/10.1016/j.cortex.2014.07.001>
- Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology*, 2, 307. <https://doi.org/10.3389/fpsyg.2011.00307>
- Schweickert, R. (1980). Critical-path scheduling of mental processes in a dual task. *Science*, 209, 704–706. <https://doi.org/10.1126/science.7394529>
- Seeber, K. G., & Kerzel, D. (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism*, 16(2), 228–242. <https://doi.org/10.1177/2F1367006911402982>
- Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (2011). Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, 72(2), 385–396. <https://doi.org/10.1016/j.neuron.2011.09.013>

Walker, G. M., & Hickok, G. (2016). Bridging computational approaches to speech production: The semantic-lexical-auditory-motor model (SLAM). *Psychonomic Bulletin and Review*, 23, 339–352. <https://doi.org/10.3758/s13423-015-0903-7>

Woumans, E., Ceuleers, E., Van der Linden, L., Szmalec, A., & Duyck, W. (2015). Verbal and nonverbal cognitive control in bilinguals and interpreters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(5), 1579–1586. <https://doi.org/10.1037/xlm0000107>