Lloyd-Kelly, M., Gobet, F., & Lane, P.C.R. (2016). Under pressure: How time-limited cognition explains statistical learning by 8-month old infants. Proceedings of the 38th Annual Meeting of the Cognitive Science Society.

Under Pressure: How Time-Limited Cognition Explains Statistical Learning by 8-Month Old Infants

Martyn Lloyd-Kelly (martynlk@liverpool.ac.uk)
Fernand Gobet (fernand.gobet@liverpool.ac.uk)
Department of Psychological Sciences, University of Liverpool,
Bedford Street South, Liverpool, L69 7ZA, UK

Peter C. R. Lane (peter.lane@bcs.org.uk)
School of Computer Science, University of Hertfordshire,
College Lane, Hatfield, AL10 9AB, UK

Abstract

In a classic experiment, Saffran, Aslin, and Newport (1996) used a headturn preference procedure to show that infants can discriminate between familiar sequences of syllables ("words") and new sequences of syllables ("non-words" and "part-words"). While several computational models have simulated aspects of their data and proposed that the learning of transitional probabilities could be mediated by neural-net or chunking mechanisms, none have simulated the absolute values of infants' listening times in the different experimental conditions. In this paper, we used CHREST, a model based on chunking, to simulate these listening times. The model simulated the fact that infants listened longer to novel words (non-words and part-words) than familiar words. While the times observed with the model were longer than those observed with infants, we make a novel finding with regard to phonological store trace decay. We also propose how to modify CHREST to produce data that fits closer to the human data.

Keywords: chunking; CHREST; headturn preference procedure; transitional probability; word segmentation

Introduction

Much of learning occurs implicitly, without conscious intention or awareness of the knowledge learned (Reber, 1993). Examples include learning in naturalistic environments (development of a first language and the acquisition of expertise) and controlled laboratory experiments (learning of artificial grammars and serial reaction time tasks). Particularly strong evidence is provided by first language acquisition: children become proficient in vocabulary, syntax, semantics and pragmatics without explicit instruction.

A standard example of implicit learning is word segmentation. To master language, infants must segment speech into words yet, word boundaries are unclear. For example, short silences in speech provide unreliable cues since they sometimes occur not only between words, but also within words. It is thought that children use a combination of several cue types to segment speech, including prosody, allophonic variation, phonotactic regularities, transitional probabilities, semantics, and words occasionally occurring in isolation (Ambridge & Lieven, 2011; Rowland, 2014; Gobet, 2015).

Of these cue types, we are particularly interested in forward transitional probabilities; predictions of syllable occurrence in a word based on the final syllable of the preceding word. These probabilities provide useful information for segmenting the speech stream, even if they are learnt approximately. The importance of forward transitional probabilities for word segmentation was demonstrated in a classic experiment by Saffran, Aslin and Newport (1996).[1] The paper proposes that infants use mechanisms which allow them to compute statistical properties of their language rapidly and such mechanisms are demonstrably active in 8 month-old infants.

SAN studied two groups of 24 8-month-old infants in two experiments. Each experiment was composed of a *learning* and *test* phase. During learning, each participant was played a continuous stream of four nonsense words, composed of three syllables (e.g. "bidaku") and randomly assorted so that no word occurred in succession in the stream, for 2 minutes. Words were uttered at a rate of 270 syllables per minute and the stream was designed so no information about prosody or cues to word boundaries existed, except for the transitional probabilities between words (unknown to the participants). Each group of infants were assigned to one of two experiment conditions, A or B, which varied the words used in the learning stream. During testing, a blinking light appeared on the wall ahead of the infant to direct their attention. Following fixation on this, a blinking light randomly appeared above a speaker on one of two side walls. Following fixation on this light, a test word from a set of 4 was played until the participant either turned their head 30° from the speaker or until the participant had fixated on the speaker for 15s (the "headturn preference procedure" (Jusczyk & Aslin, 1995)). Two words in the set were present in the learning stream (*familiar* words), the other two words (*novel* words) were either not present in the learning stream (experiment 1) or were *part-words*, i.e. composed of the last syllable of a familiar word and the first two syllables of another familiar word.

SAN's results demonstrated that infants listened to familiar words for less time than novel words in both experiments: 7.97s (SE = 0.41) vs. 8.85s (SE = 0.45) in experiment 1 ($P < 0.04$) and 6.77s (SE = 0.44) vs. 7.60 (SE = 0.42) in experiment 2 ($P < 0.03$). Thus, after an exposure of only 2 minutes, infants were able to learn enough information about the order of syllables to discriminate between familiar and novel words. Experiment 2 directly tested the hypothesis that infants can learn forward transitional probabilities; SAN took their results as evidence that they had.

Several computational models have simulated aspects of

[1] "SAN" hereafter.

SAN's study. These include simple recurrent networks (Elman, 1991; French, Addyman, & Mareschal, 2011), connectionist autoassociators (French et al., 2011), Kohonen networks (Anderson, 1999), and PARSER (Perruchet & Vinter, 1998), a symbolic model. In these models, transition probabilities are approximated by learning mechanisms based on connectionist algorithms or the creation of chunks. However, to our knowledge, no model has been used to simulate the ability of infants in SAN's study to discriminate between words, non-words and part-words and to replicate the times recorded by SAN when the headturn preference procedure is used to assess familiarity with words.

In this paper we aim to account for SAN's results using chunking mechanisms that indirectly implement the learning of forward transitional probabilities. Most importantly, we aim to simulate SAN's exact timing data, as opposed to simply demonstrating notable discrimination between the words used by SAN. Doing this adds confidence that a model captures the actual cognitive processes used by the infants rather than an abstract measure of their discrimination ability.

Participant Modelling

Each participant is simulated using an instance of the computational CHREST (**C**hunking **H**ierarchy and **R**etrieval **ST**ructures) model; a symbolic cognitive architecture (Gobet & Lane, 2010). In this study, CHREST is augmented with a phonological loop (Baddeley & Hitch, 1974) that is subject to time-sensitive trace decay (Baddeley, 2007).[2] Since CHREST's cognitive operations are also time-sensitive, we show how these structures can interact to segment speech and attempt to replicate the times recorded in SAN's study.

CHREST

CHREST has been used to account for data in several domains characterized by the acquisition of implicit learning. These include the acquisition of expertise in board games (de Groot & Gobet, 1996), implicit learning tasks (Lane & Gobet, 2012), children's acquisition of vocabulary (Jones, Gobet, & Pine, 2000) and syntax (Freudenthal, Pine, & Gobet, 2009).

A CHREST model consists of three major components: long-term memory (LTM), short- term memory (STM) and an input interface (the phonological loop, in this case). CHREST uses *patterns*, ϕ, generated from the environment using its input interface to create, modify and retrieve LTM nodes, i.e. to learn and recognise. These nodes are organised in a discrimination network: nodes are first organised according to their modality (action, visual or verbal) then according to the information that differentiates them from other nodes. STM is a fixed-size, first-in, first-out list structure that stores retrieved LTM nodes (whereupon they are called "chunks") to facilitate recognition. Note that there exists a STM structure for each modality.

[2] A simplified phonological loop has been used in earlier CHREST models, see (Jones, Gobet, & Pine, 2005; Lane & Gobet, 2012) and references therein.

The version of CHREST used in this paper simply learns and recognises verbal information when requested. To learn, content from the phonological loop is added to a ϕ in the order it was inserted into the phonological loop (oldest information added first) and ϕ is then sorted through the model's LTM. The LTM node reached, θ, is simultaneously placed into STM and its *image* (the pattern contained in θ) is analysed to determine if ϕ contains any new information. If there is new information in ϕ, CHREST will attempt to learn it either either by *discriminating* or *familiarising*. When recognising, CHREST simply checks if ϕ is present in a verbal STM chunk retrieved after sorting ϕ through LTM.

Discrimination increases the number of nodes in a CHREST model's LTM and occurs either when θ is a modality root node (ϕ is not recognised at all), or θ is not a modality root node but its image either mismatches ϕ or is *finished*, i.e. no new information can be added to it. In the first case, a new node is created and connected to the relevant modality root node (see Figure 1(b)); the new node is technically called a *primitive*. In the second case, a new node is created and connected to θ (see Figure 1(c)). The connection created is a *test-link* that contains the first unit of information that is present in ϕ but not θ's image.

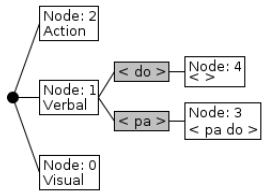
Familiarisation (see Figure 1(d)) adds new information to images of existing LTM nodes and occurs when ϕ contains a sub-pattern, ρ, that exists as a primitive; ρ is not present in θ's image; information preceding ρ in θ's image and ϕ are the same; θ's image is not finished.

Learning consumes a CHREST model's cognitive resource for a specified period of time. Consider a CHREST model, *c* that takes 10000ms to discriminate, 10ms to traverse a link in LTM and 50ms to update its STM. If *c*'s LTM is in the state depicted in Figure 1(a) and it is asked to learn the verbal pattern < `pa go` > at time t (resulting in the production of the LTM state seen in Figure 1(b)), *c* will take 20ms to sort < `pa go` > through LTM (10ms to sort its modality and 10ms to traverse the link to node 3). When node 3 is retrieved, it will be placed into verbal STM at time $t + 20\text{ms} + 50\text{ms}$ and node 5 will be created at time $t + 20\text{ms} + 10000\text{ms}$. Any requests to learn another pattern between t and $t + 20\text{ms} + 10000\text{ms}$ will be blocked.

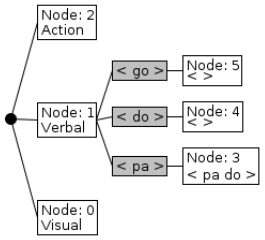
Phonological Loop

The phonological loop implemented adheres closely to the architecture proposed by Baddeley and Hitch (1974). However, we only implement the *phonological store* and omit an *articulatory rehearsal mechanism* since sub-vocal rehearsal of phonological loop content is not believed to emerge until around 7 years of age (Baddeley, Gathercole, & Papagno, 1998; Cowan & Kail, 1996; Gathercole & Hitch, 1993).

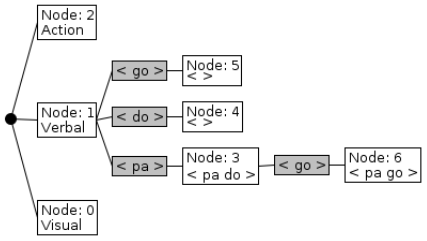
The phonological store is a first-in, first-out list structure of arbitrary length whose spaces can be occupied by, at most, one syllable (the smallest unit of phonological input specified by SAN). The debate regarding how phonological store



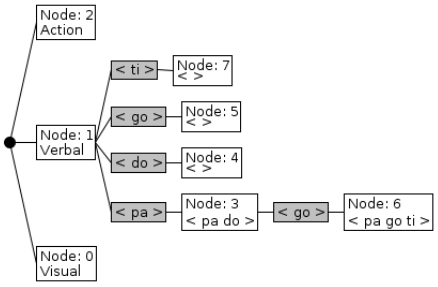
(a) Before pattern presentation.



(b) After first presentation of pattern (go unrecognised; discrimination).



(c) After second presentation of pattern (mismatch on chunk retrieved, < pa do >; discrimination).



(d) After fourth presentation of pattern (< ti > learned as primitive during third presentation but not present in chunk retrieved after fourth presentation, < pa go >; familiarisation).

Figure 1: LTM states when learning verbal pattern < pa go ti >. Tests on links are indicated by grey rectangles.

capacity is limited is noted.[3] Since this debate is unresolved and Baddeley appears to favour the trace decay explanation of phonological store capacity limitation (Baddeley, 2007), we implement this mechanism. Consequently, when syllables are added to a phonological store, they will only be retained for the trace decay value specified before they are removed.

We set the default trace decay value to 600ms according to data reported by Glass, Sachse, and Suchodoletz (2008) who estimate the duration of sensory auditory memory to be between 1 and 2 seconds in 2- and 3-year-old children; extrapolating infant trace decay to 600ms seems reasonable.

[3]The phonological store's capacity constraint may be caused by a bound on the number of items that it can store at any time or an artefact of content trace decay. See Baddeley (2007) for a review.

Experiment Simulations

A complete simulation run consists of repeating SAN's experiment battery (experiments 1 and 2), with 24 simulated participants in each experiment, 50 times. This set-up ensures that the sample sizes in our simulations are equal to the sample size of SAN, making statistical comparisons between the simulated and human data more valid.

Each experiment consisted of a 2 minute learning phase followed by a test phase. There are some notable assumptions we made when considering how to implement the experiments computationally that should be noted:

- SAN do not indicate a participant's prior familiarity with the syllables used so every simulated participant is initialised with a LTM containing nodes for each syllable used in their experiment. The alternative is to initialise simulated participants with empty LTMs; this is implausible since SAN's participants must have learned some of the syllables used in the study before they took part, since the syllables are taken from the infant's native language.

- The time taken by infants in SAN's study to initially fixate on the lights in the test phase is not specified. So, before the presentation of each word in the simulated testing phase, the simulated participant's S/LTM operations were allowed to complete and both their phonological store and STM were cleared.

During the learning phase, a learning string (randomly generated for each participant according to SAN's conditions) was used to populate the phonological store of each simulated participant. SAN report that 270 syllables are uttered every minute and a learning phase lasts for 120000ms. So, if every syllable took an equal amount of time to utter, a syllable should be uttered every 222ms: $120000 \div (270 \times 2) = 222$. Since every word is composed of 3 syllables, a word should be uttered every 666ms ($222 \times 3 = 666$). Consequently, according to our best estimate, a syllable is placed into a simulated participant's phonological store every 222ms and simulated participants constantly tried to learn the contents of their phonological store.[4]

During the test phase, the test words used by SAN in their study were presented to each simulated participant. Again, their phonological store was populated according to the timings specified in the previous paragraph. However, like SAN's study, after a whole word had been uttered, a 500ms break occurred before the next test word was presented. Simulated participants attempted to recognise the test word constantly; if recognition failed then the participant would attempt to learn the contents of its phonological store.

To ensure that the only variability in recognition times would be produced by the ordering of words in the learning phase (rather than the order of learning in the testing phase

[4]If a phonological store was empty, no attempts to learn would occur thus the simulated participant's attentional and cognitive resources were not consumed.

too), the order of test word presentation for each simulated participant was not varied over the entire simulation. We assume this consideration was also afforded by SAN; however, there is no information regarding this.

We approximated the headturn preference procedure as follows: as in SAN's study, test word presentation occurred continuously. However, unlike SAN's study, presentation ended if the simulated participant either recognised the test word or the test word was not recognised after 15 seconds from first presentation; the presentation time length was then recorded. This strategy allows for direct testing of the headturn preference procedure's proposal that novel (unrecognised) stimuli causes perceptual fixation. It also allows us to investigate whether there is evidence to posit that the participant's fixation on a speaker in SAN's study is an indicator of the participant's engagement in discrimination or familiarisation. Since learning in our simulations is analogous to a participant fixating on a speaker in SAN's study, if the presentation times recorded in our simulations are a good fit to the looking times recorded in SAN's study, this hypothesis could be accepted.

We first report results obtained with default CHREST values for discrimination (10000ms) and familiarisation (2000ms); trace decay in the phonological store is set to 600ms (as justified earlier). We then report the results of a grid search aimed at finding the optimal values for these three parameters. In 27 simulations, each repeated 50 times, simulated participants were set-up with a unique combination of values for the three parameters mentioned:

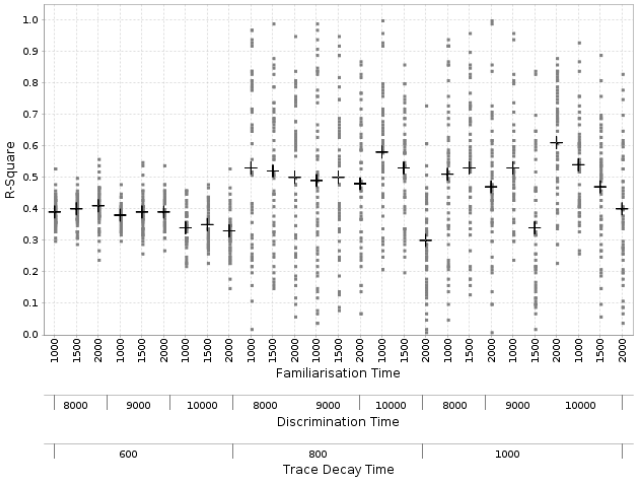
- Phonological store trace decay: 600ms, 800ms, 1000ms
- Discrimination: 8000ms, 9000ms, 10000ms
- Familiarisation: 1000ms, 1500ms, 2000ms

The mean familiar word presentation times and mean novel word presentation times for each simulated participant type repeat were used together with SAN's mean listening times for familiar and novel words in both experiments to calculate r^2 and root mean square error (RMSE) values. This data was then used to answer the following questions:

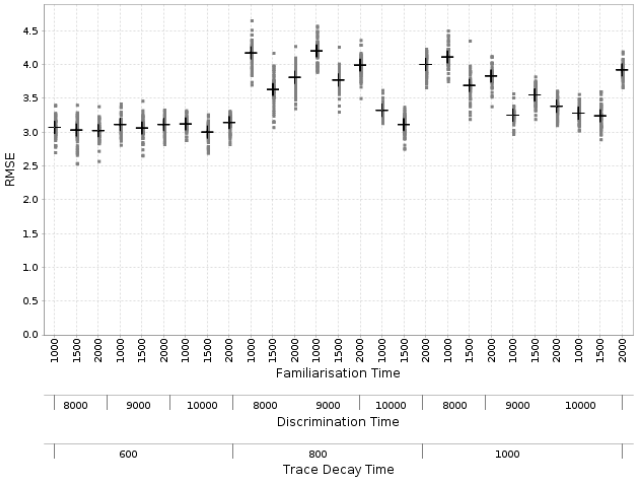
1. Do simulated participants with default times for trace decay, discrimination and familiarisation produce presentation times that are a good fit to SAN's data?
2. Do a different combination of values for these parameters offer a better fit to SAN's data?

Results

Figure 2 displays the r^2 and RMSE values calculated for each repeat of the 27 distinct combinations of the trace decay, discrimination time and familiarisation time parameters along with their means. Note that, whilst higher r^2 values indicate a better fit of the simulation data to the human data, higher RMSE values indicate a worse fit.



(a) r^2 values.

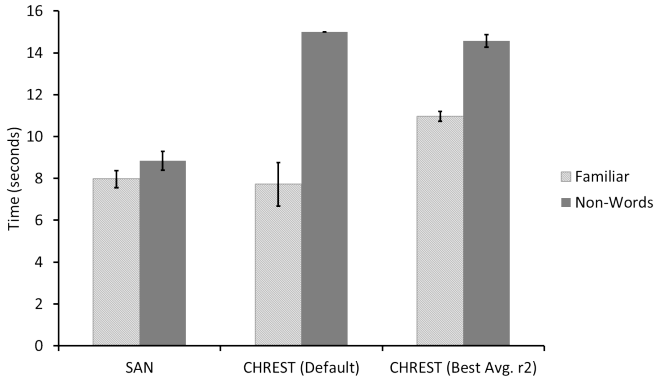


(b) Root mean squared error (RMSE) values.

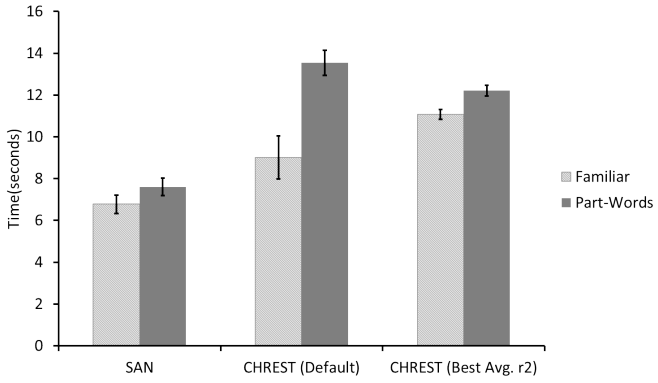
Figure 2: Model fit data for each simulated participant type's repeat experiment battery (mean values denoted by '+').

Default Parameter Model Fit

CHREST clearly discriminates between familiar words and non-words (experiment 1) and familiar words and part-words (experiment 2), as shown in Figure 3. Since 50 replications were carried out, we used Fisher's method to aggregate the p values achieved by the model with default parameter values. The resulting chi squares were highly significant: $\chi^2(100) = 1,792.8$, $p < 4.09 \times 10^{-308}$, for experiment 1, and $\chi^2(100) = 681.6$, $p < 2.35 \times 10^{-87}$, for experiment 2. However, the fit with human data was poor: r^2 = 0.33, RMSE = 3.14. The problem is that CHREST magnifies the effect with novel stimuli, compared to the human data; infants appear to become bored much more quickly than the model. Interestingly, French et al. (2011) report the same effect with the SRN model after measuring the proportion of familiar words better recognised than novel words.



(a) Experiment 1



(b) Experiment 2

Figure 3: Human listening times vs. CHREST presentation times.

Better Model Fits

When trace decay = 600ms, average r^2 values are poor and no clear pattern of r^2 improvement or deterioration emerges as trace decay values increase. The same is true when one considers how r^2 values change as discrimination and familiarisation times increase. The best r^2 on average (0.61) was obtained when trace decay = 1000ms, discrimination = 9000ms and familiarisation = 2000ms.

Conversely, with regard to RMSE values, the best fits are obtained when trace decay = 600ms. Whilst RMSE values generally deteriorate as trace decay is incremented to 800ms and 1000ms, pockets of high RMSE values are still observed. The best RMSE value on average (3.00 seconds) was observed when trace decay = 600ms, discrimination = 10000ms and familiarisation = 1500ms.

Figure 3 illustrates how the average presentation times for the simulated participants that achieve the best r^2 values on average compare to the human data. It can be seen that, whilst the general trend is well captured, the absolute values obtained by the model are still incongruent with SAN's data.

Conclusions

In this paper, we have used CHREST to simulate the listening times displayed by the infants in SAN's highly influential experiment. The key assumptions were that syllables were maintained in a phonological store and that listening time was directly mapped to the time taken to learn and recognize sequences of these syllables. Based on the previous literature, we speculated that the likely value for the trace decay time of the phonological store would be 600ms.

Simulated participants with default values for the trace decay, CHREST discrimination time and CHREST familiarisation time parameters succeeded in capturing the result that infants can significantly discriminate between familiar words and novel words (both non-words and part-words). In fact, the times achieved by simulated participants showed that differentiation between the two types of words was sharper than that observed with infants. Conversely, the sensitivity analysis provided inconsistent results: simulated participants whose trace decay is set to 600ms achieve both the best RMSE on average along with one of the worst r^2 on average. The best r^2 on average was obtained with simulated participants whose trace decay = 1000ms, discrimination = 9000ms and familiarisation time = 2000ms.

Interestingly, a trace decay time of 600ms means that a simulated participant is able to retain only three syllables or, a word using SAN's set-up, in its phonological store at any time. At the beginning of the learning and testing phase, a simulated participant will "hear" the first syllable of the first word at 222ms. This syllable will therefore decay at time 822ms, 156ms after the third syllable is heard at 666ms but 66ms before the fourth syllable (the first syllable of a new word) is heard. This buttresses statements presented by Gathercole and Adams (1993): "...in a 3-year longitudinal study of children, we found that 10% of children aged between 2 years 10 months and 3 years 1 month could already achieve a digit span of four, whereas 36% of the same cohort did not reach this level until 2 years later". If we accept that phonological store capacity is mutable and that the digit span task is a good indicator of phonological store capacity, the data obtained in this paper would lend credence to the proposal that the trace decay time of the phonological store is around 600ms for very young infants.

However, we acknowledge that the model's fit overall was low. The low fit with regard to r^2 measurements is in part explained by the fact that SAN only published 4 data points making goodness of fit comparisons difficult. In addition, it is generally accepted by developmental researchers that the headturn preference procedure is inherently noisy (Bergmann, ten Bosch, Fikkert, & Boves, 2013).

The high RMSEs observed may be explained by one of our design choices. We approximated the headturn preference procedure by assuming that presentation of the test word continued until CHREST had fully recognised it. The fact that the simulated times are too long suggests that this condition might be too strong. Another approach would be to assume that, rather than needing to have fully familiarized the LTM node image with a word in order for the word to be "recognised", it is sufficient to recognise *most* of the word

(2 syllables); this would certainly reduce the times recorded. Whether this change would produce times that fit SAN's data better can only be established by running more simulations.

Word segmentation is an essential aspect of (implicit) language acquisition, and this study has illuminated how time-limited cognitive processes, together with chunking mechanisms, mediate statistical learning, showing that transitional probability learning is not necessary. The paper's novelty was to simulate how listening times, as opposed to just the degree of recognition between familiar and novel nonsense words constructed using considerations of forward transition probabilities, differed as a function of key timing parameters associated with the phonological store and cognitive processes.

While the results of the simulations were encouraging, there is room for improvement. Future research should apply the model to more experimental results on word segmentation obtained with the headturn preference procedure, so that goodness of fit can be calculated with more data points resulting in more reliable conclusions. In particular, more accurate simulation of "head-turning" will hopefully provide data whose accuracy is much improved. This could be achieved using CHREST's simulated eye movements. We also intend to establish if a trace decay value of 600ms for the phonological store generalises to similar word segmentation studies of infants around the age of 8-months old. This is perhaps the most interesting and exciting result obtained in this paper.

References

Ambridge, B., & Lieven, E. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press.

Anderson, B. (1999). Kohonen neural networks and language. *Brain and Language*, *70*(1), 86–94.

Baddeley, A. (2007). Working memory, thought, and action. In (pp. 35–62). Oxford University Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), (Vol. 8, pp. 47–89). Academic Press.

Bergmann, C., ten Bosch, L., Fikkert, P., & Boves, L. (2013). A computational model to investigate assumptions in the headturn preference procedure. *Frontiers in Psychology*, *4*(676).

Cowan, N., & Kail, R. (1996). Models of short-term memory. In S. E. Gathercole (Ed.), (pp. 29–50). Erlbaum Associates, Ltd.

de Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess: Heuristics of the professional eye*. Assen: Van Gorcum.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*, 195–225.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614–636.

Freudenthal, D., Pine, J. M., & Gobet, F. (2009). Simulating the referential properties of Dutch, German and English root infinitives in MOSAIC. *Language Learning and Development*, *15*, 1–29.

Gathercole, S. E., & Adams, A.-M. (1993). Phonological working memory in very young children. *Developmental Psychology*, *29*(4), 770–778.

Gathercole, S. E., & Hitch, G. J. (1993). Developmental changes in short-term memory: A revised working memory perspective. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 189–210). Erlbaum.

Glass, E., Sachse, S., & Suchodoletz, W. (2008). Development of auditory sensory memory from 2 to 6 years: an MMN study. *Journal of Neural Transmission*, *115*(8), 1221–1229.

Gobet, F. (2015). Vocabulary acquisition. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 226–231). Oxford: Elsevier.

Gobet, F., & Lane, P. C. R. (2010). The CHREST architecture of cognition: The role of perception in general intelligence. In E. Baum, M. Hutter, & E. Kitzelmann (Eds.), *Proceedings of the 3rd conference on artificial general intelligence* (Vol. 10, pp. 7–12).

Jones, G. A., Gobet, F., & Pine, J. M. (2000). A process model of children's early verb use. In L. R. Gleitman & A. J. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Meeting of the Cognitive Science Society* (pp. 723–728). Mahwah, NJ: Lawrence Erlbaum.

Jones, G. A., Gobet, F., & Pine, J. M. (2005). Modelling vocabulary acquisition: An explanation of the link between the phonological loop and long-term memory. *Journal of Artificial Intelligence and Simulation of Behaviour*, *1*, 509–522.

Jusczyk, P., & Aslin, R. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23.

Lane, P. C. R., & Gobet, F. (2012). CHREST models of implicit learning and board game interpetation. In J. Bach, B. Goertzel, & M. Ikle (Eds.), *Proceedings of the 5th conference on artificial general intelligence* (Vol. LNAI 7716, pp. 148–157). Berlin, Heidelberg: Springer-Verlag.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246–263.

Reber, A. S. (1993). *Implicit learning and tacit knowledge*. Oxford University Press.

Rowland, C. (2014). *Understanding child language acquisition*. London: Routedge.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.