A model of the Headturn Preference Procedure: Linking cognitive processes to overt behaviour

Christina Bergmann*†, Lou Boves*, Louis ten Bosch*
*Centre for Language & Speech Technology, Radboud University, Nijmegen, The Netherlands
†International Max Planck Research School for Language Sciences, Radboud University, Nijmegen, The Netherlands

***Abstract*—The study of first language acquisition still strongly relies on behavioural methods to measure underlying linguistic abilities. In the present paper, we closely examine and model one such method, the headturn preference procedure (HPP), which is widely used to measure infant speech segmentation and word recognition abilities Our model takes real speech as input, and only uses basic sensory processing and cognitive capabilities to simulate observable behaviour. We show that the familiarity effect found in many HPP experiments can be simulated without using the phonetic and phonological skills necessary for segmenting test sentences into words. The explicit modelling of the process that converts the result of the cognitive processing of the test sentences into observable behaviour uncovered two issues that can lead to null-results in HPP studies. Our simulations show that caution is needed in making inferences about underlying language skills from behaviour in HPP experiments. The simulations also generated questions that must be addressed in future HPP studies.**

I. Introduction

Experimental research into early first language acquisition has blossomed over the past decades.

Most studies in this line of research use paradigms that rely on behavioural responses. One example of a method to study infant's underlying linguistic abilities using overt behaviour is the Headturn Preference Procedure (HPP) [6]. With this method speech processing skills in pre-verbal infants as young as 4 months can be investigated. The main application of the HPP concerns investigations into infant segmentation and word recognition skills. The work by Jusczyk and Aslin [6] serves as seminal study. The authors showed in HPP experiments that 7.5-month-olds can memorise recurrent acoustic patterns (words spoken in isolation or embedded in short paragraphs) and are subsequently able to recognise these patterns despite the fact that they are now presented in a different context (isolated words are now embedded in short paragraphs or vice versa).

The change between isolated words and sentences introduces variation in the acoustic signal that the infant is able to cope with. Even within speakers pronunciations of the same word vary, more so when these words are spoken in different contexts and in isolation. But the amount of variability that infants tolerate is limited: if the words are mispronounced (e.g., 'cut' instead of 'cup'), infants treat these words as if they were not familiarised [6]. Yet, there is no evidence that infants at 7.5 months have acquired a phonological system that can support the detection of language-specific sound contrasts [7]. This is evidenced by findings that changing speaker identity or voice quality between familiarisation and test leads to infant behaviour that does not distinguish between familiar and unfamiliar words [4], [9].

However, it has to be noted that infant studies only can observe preference and that conclusions about abilities have to be made with caution. As an example serves the work by van Heugten and Johnson [12], who found that infants of the same age range as in previous studies could recognise familiarised words regardless of speaker change. This finding demonstrates that the HPP is still not fully understood and the knowledge about the processes it taps into is limited.

A. The Headturn Preference Procedure

HPP studies are usually split into two parts, where first the infant is presented with a familiarisation stimulus for a predetermined amount of time. Jusczyk and Aslin [6] presented several tokens of two words (out of four, counterbalanced across infants) in the form of alternating lists for 30s per word to ensure that the infants have become sufficiently familiar with each word. In the subsequent test phase sentences containing either a familiarised or a novel word were presented. Using the HPP, Jusczyk and Aslin [6] found that 7.5-month-olds tend to listen longer to sentences that contain familiarised words. This is interpreted as the ability of the infants to segment, store, and compare word tokens, even when they are embedded in continuous speech.

The HPP measures listening times via a behavioural response, the eponymous headturns. To this end, the infant is placed in a booth where lamps are installed to the left and right of the infant's position. An additional centre lamp attracts the infant's attention at the beginning of each trial. When the infant's head is directed towards this centre lamp, one of the side lamps begins to flash. As soon as the infant turns the head towards the flashing side lamp with a sufficient angle, speech is presented from the corresponding side. While the head remains turned towards this lamp, the sound continues to play until a trial is finished. Trials can also end early when the infant's head was turned away from the lamp for at least two consecutive seconds (turning away for a shorter time is measured, but does not end the trial). Thus, the headturn is both an on-line control during the experiment that can end a trial and it is the dependent response. A neutral experimenter monitors the head-direction on-line based on visual inspection of a live video feed.

B. Goals

Our model of HPP experiments has two goals. First, by formalising the procedure and implementing these processes computationally we aim to uncover and evaluate explicit and implicit assumptions that play a role in HPP studies. Second, we aim to explain the underlying cognitive processes that generate the behaviour usually observed in HPP studies.

Linking overt behaviour in HPP experiments to underlying cognitive processes requires a number of strong assumptions. First, it is assumed that the amount of time during which the infants' head is turned toward a flashing light can be taken as a measure of interest based on processing of the acoustic stimuli. Our model is also based on this assumption. A second, perhaps even stronger, assumption is that longer listening times for the familiarised words are evidence that these words are segmented from the sentences and subsequently 'recognised'. Our model does not require this assumption. On the operational side, it is assumed that on-line decisions of an experimenter are unlikely to affect the overall outcome of an experiment, and that differences between experimenters are small and systematic, ensuring comparability across coding protocols and therefore between research groups and studies.

To maximise the explanatory power of the model we abstain from using knowledge or skills that 7.5-month-olds may not yet have acquired. Specifically, we assume that infants do not decode and memorise speech in the form of sequences of phonemes or similar 'abstract' discrete units [7], [9]. Rather, we follow the proposition that there are episodic representations at play [3]. The input to the model consists of real speech. The only meta-level information given is utterance start and end (c.f. [5] for a similar proposal regarding infant speech processing). The model's output allows for analysis of the continuous behaviour over time and of summed listening times, the latter being the basic unit of analysis in HPP studies.

Thus, our model makes it possible to investigate the architecture of the speech processing mechanism, including the memory and the matching procedure. At the same time the model makes the relation between internal processes and overt behaviour explicit. In this paper we focus on a number of technical and computational aspects of the model that we are developing. We will interpret the results of simulation experiments in terms of consequences for the cognitive interpretation of behavioural HPP experiments as well as in terms of the execution of those experiments.

II. The Model

The flow of information in the model is depicted in Fig. 1. Below, all components of the model are explained in detail.

A. Acoustic Preprocessing

Discovering that a novel stimulus is similar to previous experience requires some mechanism to store 'old' information (at least for the duration of the experiment) and to compare stored and new stimuli. This capability must be implemented without taking recourse to phonetic or linguistic knowledge that the infants in the experiments have not yet acquired [7].

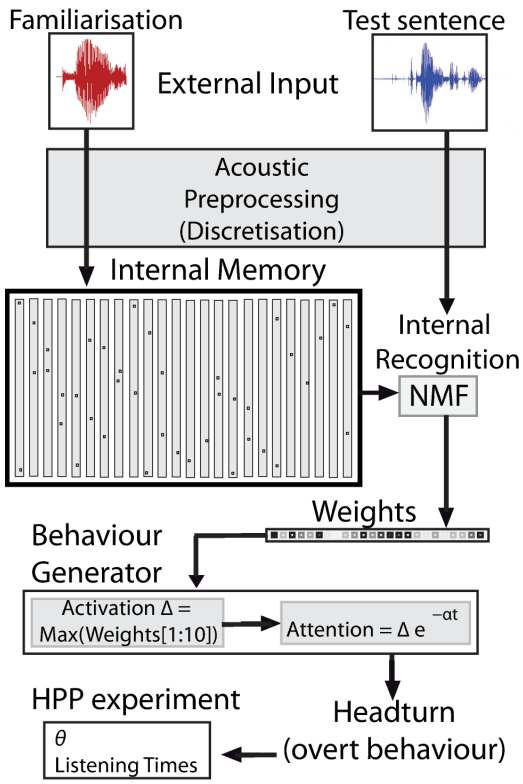

Fig. 1: The Headturn Preference Procedure model, with processing stages and flow of information from top (external input) to bottom (overt behaviour in an experimental setting).

We are following the assumption that the auditory system of 7.5-month-olds is very similar to the adult auditory system [10]. Short-time Mel-frequency spectra and their first and second order time derivatives can thus be seen as a useful analogue to human acoustic processing. We used a window length of 20ms with a frame shift of 10ms. 30 Mel-filter coefficients were transformed into 12 cepstral coefficients and log-energy. We assume that the infant auditory processing system makes it possible to estimate similarities between short-time spectral representations. This justifies the use of k-means clustering to build three code books with 250, 250 and 100 labels for the static cepstral coefficients and their first and second-order time derivatives, respectively, which represent short acoustic events.

Methods for measuring similarity between stimuli usually require a representation of the input as equally long vectors. Therefore, we need to find a way of representing spoken utterances of variable duration in a fixed-length format. For this purpose, we borrow an idea from text processing, where texts are represented as the number of times words from some index occur. This turns arbitrary texts into vectors the length of the index. While it is evident that representing the works of Shakespeare as a list of words and the number of times they occur destroys the artistic value, it is very difficult to find higher-level structural information that improves text processing performance significantly beyond what can be

obtained with such a bag-of-words representation [15].

We use a similar approach to convert arbitrary length utterances into fixed-length known vectors by means of the histogram of acoustic co-occurrences (HAC) [13]. A HAC vector for an utterance is created by counting the number of occurrences of all individual acoustic events and the number of times that these acoustic events co-occur (reminiscent of bigrams in text processing) as a means for covering the most salient aspects of the temporal structure in the speech signals. We used pairs of events that are separated by 2 frames (20ms lag), and by 5 frames (50ms lag). HAC vectors can be built without using language-specific phonetic knowledge.

Despite the general cognitive plausibility of the procedures used to create HAC representations, the practical implementation of the procedure cannot claim neural or cognitive plausibility, and all details are open to discussion [2]. As with text comprehension, for which the order of the words does matter in certain conditions, more mature representations of the speech signal will need to go beyond HAC encoding.

B. Internal Memory

Using the HAC encoding of individual utterances, we created an internal memory that represents what an infant brings to the task in an HPP experiment. In the present model, the internal memory consists of two parts. First, to model that infants have been exposed to the ambient language, we store HAC-coded utterances randomly selected from a corpus of infant-directed speech [1]. The second part contains several HAC-coded tokens of the two words spoken in isolation with which the infants were familiarised. In addition, one vector encoding silence was added which can be interpreted as a non-linguistic noise-filter.

C. Internal Recognition: Non-negative Matrix Factorisation

If we assume some form of episodic memory, we need a cognitively plausible method for matching new stimuli with the contents of that memory. We expect that all stored episodes are activated to some degree and the amount of activation denotes how well each episode matches with the new input. To avoid pair-wise comparisons of the input against all episodes, we chose a machine-learning procedure that considers all content of the memory simultaneously.

Non-negative Matrix Factorisation (NMF) [8] is a computational approach by which a new token is interpreted in terms of stored representations. It is based on the assumption that new input can be 'reconstructed' as a positive weighted sum of previous experience. Interestingly, NMF can be phrased in the same terms as activation and inhibition in neural networks [14], which helps to underpin the claim that NMF does not violate known restrictions on cognitive processing.

The variant of NMF used in the present paper minimises the Kullback-Leibler divergence between the HAC vector of a novel stimulus and the internal memory representing past experience. During the HPP test the model's internal memory does not adapt to new inputs. While this is not entirely plausible, it simplifies interpretation of the following steps.

D. From Discrete Scores to Continuous Behaviour

NMF decoding of an unknown utterance results in a vector with positive weights for all episodes in the memory. The weights are normalised to sum to one to allow comparisons between the decoding of different utterances (this bears no relation to probabilities; any constant would be appropriate). To convert these weights into overt behaviour, we first define a process to obtain a *familiarity score* for each test sentence, which is subsequently turned into a continuous function controlling overt behaviour.

1) Familiarity Score: We devised two processes to obtain a measure of familiarity based on the normalised weights that are returned by NMF. Each process implies slightly different cognitive operations and representations.

The **Max** approach takes the maximum across all episode activations corresponding to the familiarised words as the familiarity score. Thus, the familiarity score is solely determined by the best matching episode, irrespective of the word it corresponds to.

In the **Words** process the stored episodes are first grouped according to the two familiarised words. The measure of familiarity of each word is calculated as the sum of the weights of its corresponding episodes. The output is the familiarity score of the word that is activated most. This approach assumes that the infant realises that two different words are presented during the familiarisation and that all stored episodes for both words are used during test. Thus, the **Words** process postulates an additional internal operation, both during familiarisation and during recognition in the test phase that involves grouping stored information according to the words present in the experiment.

2) Continuous Behaviour: In HPP studies, the headturns of an infant are measured as an overt sign of underlying attention to the speech stimuli. In this section we describe how internal recognition and familiarity can be transformed into observable, attention-driven behaviour. We assume that the familiarity score, based on the weights of episodes, is congruent with attention.

To compute the time during which the infant pays attention to the test stimuli we need to convert the discrete-time familiarity scores into a continuous function. Since we assume that the familiarity score is available immediately after the end of the utterance and we know the duration of all utterances, the discrete familiarity values can be converted to Dirac δ pulses with an amplitude equal to the familiarity score, separated by the duration of the utterances. The sequence of δ pulses is converted into a continuous attention function by applying an exponential decay. This is based on the finding that the exponential function appeared to be a very good choice to model memory effects in delayed retrieval tasks [11].

The exponential decay function is $\textit{familiarity}(t) = e^{-\alpha t}$ in which α is a (positive) parameter specifying the decay rate and t denotes time. This exponentially decaying attention function can be interpreted as the degree of headturn. While the function value is high, we assume that the infant's head

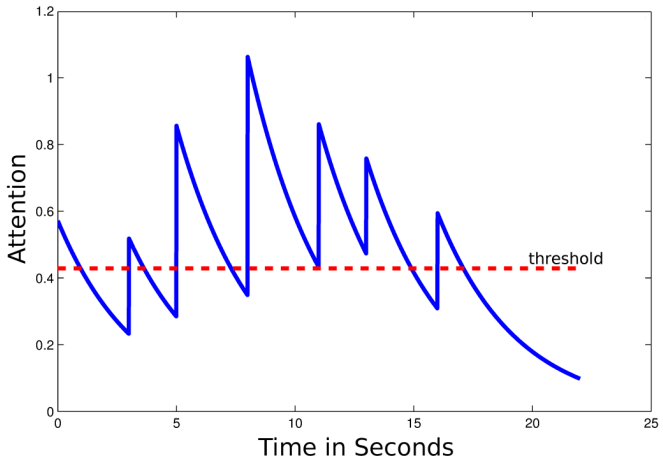

Fig. 2: Exemplar attention function, after applying exponential decay with $\alpha = 0.3$ on a sequence of δ spikes. The attention at $t = 0$ reflects the initial interest at the start of the trial. The horizontal line represents a threshold with $\theta = 0.42$.

is turned towards the flashing light. As the attention value decreases, the head is gradually turned away from the lamp.

E. Modelling Headturns in an Experimental Setting

During the HPP procedure, the experimenter interprets the angle of the head relative to the center and side lamps in terms of discrete states. When the head is turned too far away for more than two consecutive seconds, the experimenter ends the trial. In addition, the time spent with the head turned towards the lamp is measured as the dependent variable in HPP studies. However, in the usual HPP setting it is difficult to exactly measure degree of headturn. There are usually a number of situations in infant experiments where different decisions are possible regarding what can be counted toward the total listening time. While these decisions are often consistent within experimenters, there is little documentation and exchange regarding this topic across different labs.

In the model, the experimenter's decisions are implemented as a Finite State Machine (FSM). The FSM takes the continuous attention function as input and calculates the listening time (for each 'paragraph'). To that end, the FSM uses a threshold θ. If attention values exceed θ, the head is turned in the direction of the flashing light. As soon as the attention level drops below θ, it is assumed that the infant is no longer listening, as indicated by an angle of the headturn that is too far away from the lamp. If attention stays below θ for more than two consecutive seconds, the trial is terminated (analogous to infant HPP).

An additional parameter is used which models the start attention level. It can be conceptualised as the degree of interest in the flashing lamp. At $t = 0$ this value must exceed θ, since a trial only starts when the infant's head is turned towards the lamp. Since we cannot know the absolute value of the initial attention, this parameter is defined in relative terms as $\rho + \theta$, with $\rho \geq 0$.

III. Experiment: Parameter Investigation

We performed a large number of simulations to investigate whether there is a range of values of the parameters ρ, θ and α for which a robust difference in listening times between sentences with familiar or novel words can be established. Note that the amplitude of the δ pulses that represent the familiarity value have a maximum value of 1, which only occurs in the unlikely case that a test sentence would result in a HAC vector that is exactly equal to the HAC vector corresponding to one of the tokens of a familiarised word (see Sec. II-D). The ranges for the free parameters were chosen between 0 and some positive maximum value. For θ, which represents the threshold used by the experimenter to decide whether the degree of headturn towards the side lamp is sufficient, the maximum value was set to 2. The larger θ is, the shorter the time during which the experimenter considers the infant to be listening. The maximum value for α was set to .5. Larger values of α correspond to faster decay of attention and thus headturns that suffice to show attention. ρ can take on values up to .9; this parameter acts as a safety margin above the threshold θ. In addition, we investigated whether the two different methods for generating familiarity scores **Max** and **Words** lead to different results in terms of the values for the three parameters.

A. Material

For each familiarised word stored in memory, we used five different pronunciations of monosyllabic words spoken by a female native speaker of English and recorded in a virtually noise-free environment [1]. The words chosen as familiarisation stimuli were either *frog* and *doll* (words 1, 2) or *duck* and *ball* (words 3, 4). We randomly selected 24 short sentences for each of these words in variable contexts from the same corpus and spoken by the same speaker as test sentences. These test sentences contained all four words and could thus be used in both familiarisation conditions, as novel or as familiar stimuli, respectively.

In the simulations we used an internal memory comprising of 111 slots, 10 containing tokens of the familiarised words, 100 containing sentences spoken by the same female speaker that did not contain one of the four target words, and one containing background noise. The 101 non-target slots were identical in all simulations.

B. Results

*1) **Words** versus **Max**:* As a first step, we ensured that the familiarity score can indeed differentiate between test sentences containing either novel or familiar words. The familiarity scores can be computed independent of the three parameters α, θ and ρ, all of which affect the conversion of the familiarity scores to listening times. Two familiarity scores were computed for each test sentence, one with the tokens of the word pair *frog, doll* in the memory and one with the tokens of *duck, ball*. The scores were obtained with the **Max** and **Word** methods for computing familiarity.

For the method **Max**, which uses the highest value of the familiar word episodes, we found a mean familiarity value $\mu = 0.066$, and standard deviation $\sigma = 0.04$ for the familiar condition, and $\mu = 0.055, \sigma = 0.03$ for the novel words. The familiarity scores are significantly different for familiar and novel input, according to a Mann-Whitney-U test yielding $U = -453.0, p < 0.001$. For the method **Words**, which requires grouping of the episodes of the two familiar words, mean and standard deviation are $\mu = 0.127, \sigma = 0.05$ for the familiar words and $\mu = 0.09, \sigma = 0.04$ for the novel words. These values are significantly different with $U = -112.0, p < 0.001$.

We did not further investigate possible differences according to the specific words used, since the overall discrimination ability between familiar and novel words was sufficient for the aims and purposes of the present paper.

2) Behaviour-generating parameters: Since there is no upfront difference between the **Max** and **Word** approaches, we ran simulations with both to investigate the impact of the three parameters. We did not find interesting differences between the two approaches. Therefore, we will drop this distinction in the analysis of the effect of the parameters α, θ and ρ.

As a means to compare how the model fares when replicating HPP data, we use differences in listening times to test passages containing either novel or familiar words. Longer listening times for sentences containing familiarised words indicate that the model behaviour reflects the familiarity preference found in [6].

To obtain a reliable measure of model performance, we generated 30 test passages for each of the four words. To this end, we randomly chose sets of 6 sentences out of the 24 available sentences per test word. The differences obtained with the 30 test passages were averaged and form the basis for the analysis of the impact of the three parameters α, θ and ρ in the process that converts the familiarity scores into overt behaviour.

In Fig. 3 an example of the model performance for different values for α and θ for a fixed value $\rho = 0.4$ is depicted. The black squares represent the average difference in listening time between the familiarised and novel test conditions. The size of these squares indicates the size of the difference: the larger a square, the greater the listening time advantage for paragraphs containing familiar words.

From Fig. 3 it can be seen that the model yields positive differences in listening times across a wide range of parameter settings, but that the differences become smaller as the parameter values are more extreme. It can be seen that at high values of α, which correspond to short attention spans, differences in familiarity scores only become apparent if a very lenient criterion for head turn direction is used (low values of θ). If the criterion for headturn direction is very strict (high values of θ) even moderately long attention spans are no longer enough to bring about the difference in familiarity scores in overt behaviour.

Increasing the value of ρ leads to a wider range of values of α and θ for which positive differences in listening time are obtained (not depicted). When the initial value of the familiarity function $(\theta + \rho)$ increases, steeper decays can be tolerated before the value of the function drops below θ. As an example, for $\rho = 0$ we obtain positive differences in listening time for all $0 \leq \theta \leq .2$ and $0 \leq \alpha \leq .04$; for $\rho = .9$ the corresponding values increase to $0 \leq \theta \leq .6$ and $0 \leq \alpha \leq .2$.

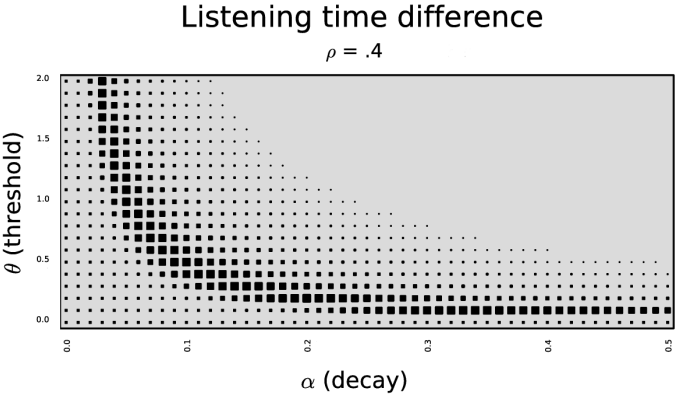

Fig. 3: Listening time difference to familiar versus novel stimuli across different parameters for attention α and headturn threshold θ. The initial attention is $.4 + \theta$.

IV. Discussion

In this paper we present a computational model that can simulate the outcomes of experiments that use the Headturn Preference Procedure (HPP) to investigate language skills in infants. Importantly, the model makes no assumptions about phonetic and phonological skills. The fact that we can robustly simulate the results of HPP experiments enhances the credibility of the HPP approach. At the same time it calls into question the cognitive interpretations in terms of word segmentation skills that are attached to the results of experiments using the HPP. After all, positive differences in listening times can be obtained within the model using only very general perception and matching skills. This need for caution is emphasised by the fact that the target behaviour can be simulated without assuming that repeated tokens of the same word are clustered to a unique representation of that word. Furthermore, we use different conditions between familiarisation and test, where words familiarised in isolation have to be 'recognised' when they are embedded in running speech. The model overcomes this obstacle despite the fact that no explicit segmentation procedure is implemented. Two assumptions are made within the model. First, auditory stimuli can be stored in an episodic memory, where they are encoded as a histogram of acoustic events and their co-occurrences. Second, we assume a procedure for matching incoming new to stored stimuli.

Our model includes an explicit account of the process that converts the matching score to overt behaviour. Thus, we uncovered two factors that can lead to null-results, despite infants' ability to treat familiar and unfamiliar stimuli differently. First, if infants are easily distracted leading to a short attention span, modelled as decay with a large value of α, the difference between internal processing of familiar and novel

stimuli can become invisible in the overt behaviour. Second, our simulations drew attention to a possible experimenter effect: if the experimenter is too critical in scoring the angle of the headturn, (which in our simulations corresponds to a high value of θ), possible differences in the internal processing of familiar and novel stimuli can also become invisible in the overt behaviour.

A. Future Work

The HPP model presented in this paper offers many opportunities for future investigation of specific aspects of the procedure. The results reported here for one female speaker of English need to be repeated with more speakers and more languages. We are confident that this will confirm our findings. As a next step, we envision a replication of experiments with mispronunciations [6]. We plan to also replicate experiments where speakers change between familiarisation and test [4], [12]. These experiments might yield smaller ranges of the parameters α, θ and ρ within which behavioural results can be reproduced and will help unite seemingly conflicting results across studies [4], [12].

The present model was designed to only produce familiarity effects. However, there are a number of conceivable ways to model the novelty preference that has been found in some HPP studies [12]. The model will be able to shed light on the factors that give rise to either a familiarity or a novelty preference within one framework.

Another extension concerns the possibility that the infants' internal representations change during an experiment. This can be implemented in the model with additional processing and learning steps that were omitted for simplicity in the present paper.

Last but not least, our model makes specific predictions that have to be tested in behavioural experiments. First, we will examine how the individual experimenter factor, modelled as θ, influences outcome. Second, we need to devise independent methods for estimating the infant's attention span, the parameter α, and its effect on behaviour in HPP experiments.

V. Conclusion

We presented an end-to-end model that successfully simulates behaviours observed in experiments that use the HPP to investigate language skills of pre-verbal infants. End-to-end means that the input of the model is real speech, and the output can be interpreted as observable behaviour, the headturn angle.

The HPP model demonstrates that the familiarity preference, a behavioural pattern that is usually interpreted as an indication that at 7.5 months infants are able to segment words from continuous speech, can be simulated without assuming such language skills. Our model shows that to exhibit this behaviour it suffices that the model (or the infant) is able to form uninterpreted episodic representations of spoken words and to match new stimuli with stored representations of previously heard stimuli.

Next to providing an explicit account of auditory processing and matching procedures, we also examine the processes that convert the result of the match into observable behaviour (headturn) and the experimenter's scoring of this behaviour. We identify two issues that can lead to null-results despite the fact that infants process familiarised stimuli in a different way than new ones. The first issue concerns the attention span of the infants: if that span is very short, potential differences between the processing of the two types of stimuli will not yield observably different behaviours. The second issue concerns the potential measurement bias introduced by the experimenter. This factor has implications for comparisons between research groups and reproducibility of experiments. The effects of these issues must be investigated in future HPP experiments.

Future model research will simulate headturn behaviour in a wider range of experiments. In addition, we plan to extend the model in such a way that it can simulate a novelty effect, along with the familiarity effect, both of which have been observed in published HPP studies [4], [9], [12].

Acknowledgements

The research of Christina Bergmann is supported by grant no. 360-70-350 from the Dutch Science Organisation NWO.

References

[1] T. Altosaar, L. ten Bosch, G. Aimetti, C. Koniaris, K. Demuynck, & H. van den Heuvel "A Speech Corpus for Modeling Language Acquisition: CAREGIVER", *Proceedings LREC'10*, 2010.

[2] J. Driesen, *Discovering Words in Speech using Matrix Factorization*, PhD thesis, KU Leuven, Belgium, 2012.

[3] S.D. Goldinger, "Echoes of echoes? An episodic theory of lexical access", *Psychological review*, 105(2): 251–279, 1998.

[4] D.M. Houston & P.W. Jusczyk, "The role of talker-specific information in word segmentation by infants", *Journal of Experimental Psychology: Human Perception and Performance*, 26(5): 1570–1582, 2000.

[5] K. Hirsh-Pasek, D.G. Kemler Nelson, P.W. Jusczyk, K.W. Cassidy, B. Druss, & L. Kennedy, "Clauses are perceptual units for young infants", *Cognition*, 26(3): 269–286, 1987.

[6] P.W. Jusczyk & R.N. Aslin, "Infants' Detection of the Sound Patterns of Words in Fluent Speech", *Cognitive Psychology*, 29(1): 1–23, 1995.

[7] P.K. Kuhl, "Early language acquisition: cracking the speech code", *Nature reviews neuroscience*, 5(11): 831–843, 2004.

[8] D.D. Lee & H.S. Seung, "Learning the parts of objects by non-negative matrix factorization", *Nature*, 401(6755): 788–791, 1999.

[9] R.S. Newman, "The level of detail in infants' word learning", *Current directions in Psychological Science*, 17(3): 229–232, 2008.

[10] J.R. Saffran, J. Werker, & L. Werner, "The infant's auditory world: Hearing, speech, and the beginnings of language", In: R. Siegler and D. Kuhn (Eds.), *Handbook of Child Development*. New York: Wiley, 58–108, 2006.

[11] J. Sargisson & K.G. White, "On the form of the forgetting function: The effects of arithmetic and logarithmic distributions of delays", *Journal of the Experimental Analysis of Behavior*, 80: 295–309, 2003.

[12] M. van Heugten & E.K. Johnson, "Infants Exposed to Fluent Natural Speech Succeed at Cross-Gender Word Recognition", *Journal of Speech, Language, and Hearing Research*, 55: 554–560, 2012.

[13] H. Van hamme, "HAC-models: a novel approach to continuous speech recognition", *Proceedings Interspeech 2008*, 2554–2557, 2008.

[14] H. Van hamme, "On the relation between perceptrons and non-negative matrix factorization", *Signal Processing with Adaptive Sparse Structured Representations Workshop*, 2011.

[15] S. Verberne, L. Boves, N.H.J. Oostdijk, & P.A.J.M. Coppen, "What is not in the Bag of Words for Why-QA?", *Computational Linguistics*, 36(2): 229–245, 2010.

[16] J. Zevin, "Word Recognition", In: Larry R. Squire, *Encyclopedia of Neuroscience*, Academic Press, Oxford, 517–522, 2009.