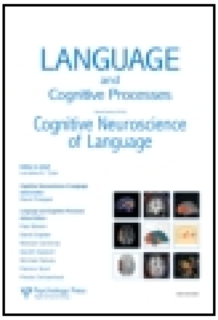

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp20

Computational modelling of phonological acquisition: Simulating error patterns in nonword repetition tasks

Marco Tamburelli [a], Gary Jones [a], Fernand Gobet [b] & Julian M. Pine [c]

[a] Division of Psychology, Nottingham Trent University, Nottingham, UK

[b] Brunel University, Uxbridge, UK

[c] University of Liverpool, Liverpool, UK

Available online: 06 Jan 2012

LANGUAGE AND COGNITIVE PROCESSES, 2012, iFirst, 1–46

Psychology Press
Taylor & Francis Group

Computational modelling of phonological acquisition: Simulating error patterns in nonword repetition tasks

Marco Tamburelli[1], Gary Jones[1], Fernand Gobet[2], and Julian M. Pine[3]

[1]Division of Psychology, Nottingham Trent University, Nottingham, UK
[2]Brunel University, Uxbridge, UK
[3]University of Liverpool, Liverpool, UK

Nonword repetition tasks (NWRTs) are employed widely in various studies on language development and are often relied upon as diagnostic tools. However, the mechanisms that underlie children's performance in NWRTs are very little understood. In this paper we present NWRT data from typically developing 5- to 6-year-olds (5:4–6:8) and examine the pattern of their phonological errors within the syllabic domain. We show that the children display a strong tendency for errors at the syllable onset, with fewer errors in coda position. We then show how the same pattern can be simulated by a computer model, thus shedding some light on the cognitive mechanisms that underlie specific error patterns as well as general phonological development.

Keywords: Nonword repetition; Computational modelling; Syllabic position; Onset; Coda.

Correspondence should be addressed to Marco Tamburelli, School of Linguistics and English Language, Bangor University, Bangor, Gwynedd, LL57 2DG, UK. E-mail: m.tamburelli@bangor.ac.uk

The authors would like to thank the Leverhulme Trust who funded this research in the form of a Research Grant (ref F/01 374/G) to the second author. We would also like to thank Sarah Watson for collecting and transcribing the data, Gabrielle Le Geyt for recording the nonword stimuli, Hannah Witherstone for helping with inter-rater reliability for the children's nonwords, and Chloe Marshall and an anonymous reviewer for their valuable comments and suggestions on earlier versions of this paper. Our thanks are also due to the schools and children who participated in the study.

http://www.psypress.com/lcp http://dx.doi.org/10.1080/01690965.2011.583510

INTRODUCTION

The error patterns that children exhibit in their speech production are of the utmost importance to our understanding of language development. The study of error patterns allows researchers to develop a clearer picture of what constitutes typical language development, a condition which is not only an empirical necessity for the development of sound acquisitional theories but also the *sine qua non* for achieving a more refined definition as well as a deeper understanding of atypical development.

A central part of language development is the acquisition of the phonotactic rules of one's native language, whose importance is at least twofold. Firstly, one may not master a language without learning what phonological sequences are possible in that language as well as what positional restrictions may be placed on those sequences. Secondly, and perhaps more importantly, the ability to generalise from individual phonetic strings to larger, more general rules is arguably the single most important skill underlying successful language development. Hence, the understanding of the patterns involved in phonological errors is essential to our understanding of language development as a whole.

Nonword repetition tasks (NWRTs) offer the possibility of collecting large amounts of phonological errors by using strings that conform to the phonotactics of the language at issue whilst ensuring that the children have no previous knowledge of the specific strings.

NWRTs have been argued to be a powerful diagnostic for Specific Language Impairment (SLI) (Bishop, North, & Donlan 1996; Conti-Ramsden & Hesketh 2003; Dollaghan & Campbell, 1998; *inter alia*). However, the phonological mechanisms that underlie the performance of both children with SLI and typically developing children in NWRTs are poorly understood. The need for a deeper understanding of these mechanisms is particularly pressing when one considers that phonology has been argued to be at the core of morpho-syntactic and lexical impairments (e.g., Chiat, 2001; Joanisse & Seidenberg, 1998).

One question that has received little attention is what part—if any—syllabic positions[1] might play in children's performance in NWRTs. In this paper we analyse the patterns of phonological errors in typically developing children, aged 5:4 to 6:8 years, and compare them to the patterns emerging from a newly updated version of the EPAM-VOC computational model (EPAM-VOC II), based on the EPAM architecture (Feigenbaum & Simon, 1984; Gobet & Simon, 2000). The aim is to shed new light on the mechanisms that underlie the children's performance, particularly with

[1] Throughout the paper, the term "syllabic position" refers to the position of phonemes within syllables.

respect to three widely recognised syllabic positions: onset, nucleus, and coda (Giegerich, 1992; Jones, 1997; Harris, 1994; Laver, 1994; *inter alia*).

The paper is organised as follows. First, we discuss relevant literature concerning NWRTs and the analysis of errors in relation to syllabic structures. Second, we describe the EPAM-VOC model, explaining our latest additions. Third, we present newly collected data on two separate NWR experiments. Fourth, we show that the model can account for the children's performance at the general level—by successfully simulating performance effects across nonword types—as well as at a more specific level, by matching the specific type of errors that children make *within* nonword types, and in particular at the level of the syllable. A general discussion of the findings of our experiment and our simulations will follow.

Nonword repetition tasks and the study of phonological development

Whilst it is fairly uncontroversial that children's phonological knowledge plays an important part in their performance on NWRTs (e.g., Gathercole, Willis, Emslie, & Baddeley, 1991; Hulme, Maughan, & Brown, 1991), the potentially important role of positional information in NWRT performance has received very little attention. Although NWRTs have become a widely used tool in psycholinguistics, much of the research has focused on factors such as word length or consonantal combinations (e.g., Botting & Conti-Ramsden, 2001; Gathercole & Baddeley, 1989; Gathercole, Willis, Baddeley, & Emslie, 1994), without distinguishing between different syllabic positions (see for example Gimson, 1980; Roach 2000; among many others). Although often recognising syllables as sub-lexical entities, the research has yet to examine the relationship between syllabic and phonemic information, or the role of phonemes as syllabic components. Many researchers have relied on the syllable solely as a means of measuring nonword length without exploring its internal structure, thus overlooking the potential importance of analysing repetition errors as a function of their syllabic position. However, in domains outside of NWRTs, syllabic structures have been examined in depth for the past 30 years. Indeed, the recent psycholinguistic literature on children's phonological performance could be divided into two strands. On the one hand, there is research that investigates children's performance on NWRTs but ignores the potential involvement of syllabic structure in explaining such performance. On the other hand, there is research that investigates children's errors from a syllabic perspective but does not employ NWRTs, thus providing no direct information on the extent to which syllabic positions may influence children's performance on these tasks. We will consider these two research strands in turn.

Error analysis in NWRTs

Children's performance on NWRTs has received a lot of attention in the psycholinguistic literature, and NWRTs have been used to investigate linguistic development in a variety of languages (for example Gathercole & Baddeley, 1989 for English; Radeborg, Barthelom, Sjöberg, & Sahlén, 2006, for Swedish; Ho & Lai, 1999, for Cantonese). Some of these studies have not been concerned with phonological development, however, and concentrated mainly on lexical issues. For example, Munson, Edwards and Beckman (2005) investigated the role that lexicality plays in aiding children's recall of nonwords in NWRTs and argued that error patterns could be at least partly explained in terms of familiarity (see also Dollaghan, Biber, & Campbell, 1995; Gathercole, 1995).

However, those researchers that have worked on the phonological issues associated with NWRT performance have often provided error analyses that involved separating nonwords into phonemes and syllables without considering the relationship that holds between the elements within these two phonological entities. This has been, for example, the approach taken by researchers investigating the role of memory in NWRTs and its interaction with long-term phonological knowledge (e.g., Botting & Conti-Ramsden, 2001; Gathercole & Baddeley, 1989; Gathercole et al., 1994). In this work, performance on NWRTs has been analysed chiefly in terms of nonword length, together with considerations regarding phonetic membership (i.e., consonantal vs. vocalic) of the component phonemes.

A similar method of analysis has been employed by researchers investigating the role of sub-phonemic features. For example, Edwards and Lahey (1996) compared error rates for stressed and unstressed syllables, and considered whether specific manners of articulation might have a particularly negative effect on children's performance (see also Snowling, 1981). Although the inclusion of prosodic and sub-phonemic information is doubtlessly a step forward, this work still treats phonemes as stand-alone elements rather than as components of larger phonological units.

A more complex experimental paradigm has been developed by Bowey (1996, 2001) and Metsala and colleagues (Metsala, 1999; Metsala & Walley, 1998; Walley, Metsala, & Garlock, 2003), whose work on NWRTs also included gathering information on children's phonological awareness, as measured by a task of phonemic discrimination. For example, in Bowey (2001) children were presented with a number of spoken words and were then asked to pick out those that ended in a specific sound. For the sound [f], for instance, children were asked "Which of these ends with a [experimenter makes [f] sound], *mop* or *leaf*?" (2001, p. 450). Researchers have claimed that this paradigm provides evidence in favour of the view that performance in NWRTs is affected by the ability to perform a thorough segmentation of the input during phonological processing.[2] Although we do not necessarily disagree with this view, we believe that it is unwise to draw conclusions regarding the role of phonological segmentation without first providing an analysis of children's errors that takes into account *all* the various syllabic constituents, as these have long been considered crucial components in the segmentation of phonological units (see for example Blevins, 1995; Goldsmith, 1992; Onnis, Monaghan, Richmond, & Chater, 2005; among many others). Unfortunately, however, research on NWRTs has generally shied away from considering these components.

Analyses of syllabic constituents

Naturally, there have been exceptions to this trend. For example, Kehoe and Stoel-Gammon (2001) have investigated heavy syllables in early development, a task that involved examining the relationship between syllable nuclei and syllable codas. They reported that the two phonological features that make up heavy syllables (i.e., long vowels and VC sequences) develop hand in hand, with neither preceding the other. They also provided further evidence for a developmental asymmetry concerning the late emergence of voiced obstruents as compared to voiceless sonorants in coda position (see also Stoel-Gammon, 1985; Bernhardt & Stemberger, 1998). However, as this study was not concerned with pre-vocalic elements, it did not provide information about syllable onsets, thus falling short of providing an inclusive account of the role of syllabic membership in children's development.

A more comprehensive study was that of Kirk and Demuth (2005) who examined the interaction between prosody and morphology (also Demuth 2001), including the potential role of syllabic structure in explaining asymmetries in the acquisition of word-initial versus word-final clusters. Specifically, they investigated children's accuracy in repeating consonant clusters through a picture identification task and reported that their participants were more accurate in repeating clusters word-finally than word-initially.[3] They suggested that these results must be due to articulatory constraints, as they cannot be explained either in structural terms or in terms of frequency. In particular, their decision to treat consonant clusters as structured combinations instead of simple linear sequences allowed them to assess the empirical value of different theoretical frameworks relating to

[2] Metsala and Walley (1998) and Walley, Metsala & Garlock (2003) have incorporated this as part of their Lexical Restructuring Model.

[3] Kirk and Demuth's study included comparison of two groups of clusters. Word initially: /s/ + stop and /s/ + nasal; word-finally: stop + /s/ and nasal + /z/. Only these consonants were considered since, as far as English is concerned, they are the only consonants that can combine to form both word-initial and word-final sequences.

various levels of phonological performance, both at the processing and at the articulatory level.

In this paper, we intend to build on this approach, and therefore we will be viewing phonemes not just as phonological units, but also as syllabic components whose properties are at least partly dependent on their position within larger phonological units. As Kirk and Demuth (2005) have demonstrated, this approach can provide a more far-reaching analysis of children's errors, allowing for the detection of important patterns that may otherwise go unnoticed. Moreover, we will make use of these analytical tools to investigate performance in NWRTs, an experimental paradigm that has been argued to tap more directly into the phonological as well as the lexical components of children's linguistic development.

Syllabic analysis and NWRTs

A notable exception to the research strands outlined above is the work of Marshall and colleagues (Marshall, Ebbels, Harris, & van der Lely, 2002; Marshall & van der Lely, 2009), who have investigated the potential role that syllabic information, and its interaction with stress patterns, might play in the development of spoken and written impairments (SLI & dyslexia). For example, Marshall and van der Lely (2009) have used NWRTs to investigate children's performance on word-initial versus word-medial clusters, effectively building on the work of Kirk and Demuth (2005). They reported that the clinical groups showed an asymmetry in cluster production, with word-medial clusters attracting more errors than word-initial ones. Moreover, their findings also suggest that positional information is closely linked with language impairments, as no asymmetry was found for the typically developing children, a fact that highlights once again the importance of positional analyses.

Whilst this study provided important insight into the interaction between position and phonological complexity, it was restricted to one syllabic component, namely the onset. Hence, we will expand our error analysis to include the two other syllabic constituents, namely nucleus and coda. Our aims are threefold: firstly, to study whether the syllabic position in which a phoneme occurs has an effect on children's errors; secondly, to investigate whether phonotactic probabilities of phonemic sequences interact with syllabic positions in creating error patterns; and finally, to go beyond empirical generalisations by providing a computational simulation of children's performance. As a computational model is first and foremost a detailed set of theoretical assumptions, we aim to provide well-specified theoretical explanations of the phenomena observed through NWRTs. These will include well-established phenomena, such as word-length effects (e.g., Gathercole & Baddeley, 1989; Stokes, Wong, Fletcher, & Leonard, 2006), as well as newly discovered phenomena that relate to the positional role of phonemes within the syllabic template.

We believe that the development of a computational model of children's performance is essential for furthering our understanding of the mechanisms involved in the processing of nonwords. In particular, a computational model can help us understand which specific patterns of errors can be generalised from the input data, and which may result from the specific heuristics that children might apply when faced with NWRTs. The immediate advantage of a computational model is that it allows direct and detailed exploration of the consequences that follow from our theoretical assumptions (see also Fum, Del Missier, & Stocco, 2007). Thus, given that even small changes in our fundamental assumptions can have important—and often unanticipated—knock-on effects, a computational model can offer invaluable help in investigating a system as complex as language.

A MODEL OF CHILDREN'S VOCABULARY LEARNING: EPAM-VOC

Introduction

EPAM-VOC (Jones, Gobet, & Pine, 2007, 2008) is a model of vocabulary acquisition based on the EPAM modelling architecture (Feigenbaum & Simon, 1984; Gobet & Simon, 2000). It takes in phonemic information from naturalistic speech input and organises it into a hierarchy of phoneme sequences (not unlike the learning mechanism suggested by Ellis, 1996) which it then stores as long-term phonological knowledge. The hierarchy is represented via a tree structure with a null root-node under which are a number of progressively longer sequences connected by "links".

Previous research (Jones, Gobet, & Pine, 2007) has shown that the EPAM-VOC model can approximate performance on NWRTs from different age groups (2–3 and 4–5 year olds) and across nonword-types. Although this work represented an important step in the modelling of NWR performance, it did not explore error patterns *within* nonword-types, and thus did not address children's treatment of syllabic structures. In this paper, we present a more fine-grained analysis of children's as well as model data, particularly in relation to error-types as dictated by standard syllabic structure. In describing version II of the EPAM-VOC model, any departure from the previous version of EPAM-VOC will be highlighted, together with the reason for the alteration. Also, whilst the current version of the model presents a number of changes and improvements, it nevertheless remains a version of EPAM-VOC since it accounts for all the data that were accounted for by previous versions of the same model.

Representing long-term knowledge

EPAM-VOC encodes phonological knowledge as a hierarchy of phonemic sequences, or "chunks". Top-level chunks contain only individual phonemes, while lower-level chunks involve phonemic sequences, provided that the model has been exposed to a reasonable amount of input. Chunks get progressively larger as one proceeds further down the hierarchy.

Knowledge of the phonemic inventory is programmed into the model prior to the beginning of the learning procedure. Hence, each top-level chunk corresponds to an English phoneme, for a total of 39 single-phoneme chunks. The decision to programme the phonemic inventory into the model in this way follows our aim to compare the model's performance with that of children between 5 and 6 years old, by which age the constituent phonemes of their native language are generally assumed to be established (e.g., Bailey & Plunkett, 2002). Figure 1 provides an illustration of the hierarchical structure of EPAM-VOC[4]:

EPAM-VOC relies on its network of long-term knowledge in order to parse phonemic sequences

To illustrate how this works, let us take the network in Figure 1 as an example, together with the hypothetical input /sʌm/. On exposure to this input, the model checks whether it already knows the phoneme sequence by attempting to traverse the network in a top-down fashion. Beginning at the top node, it searches for a link that matches the first phoneme in the input.

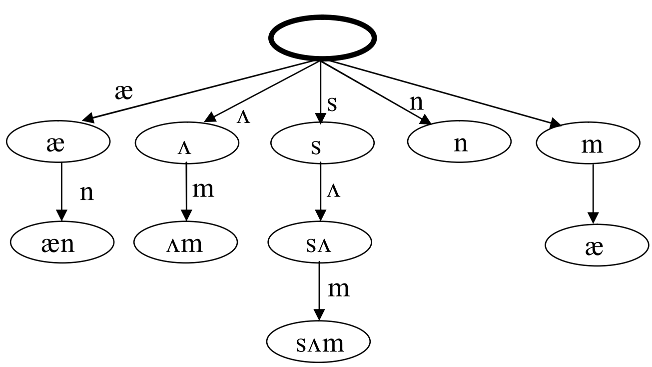

Figure 1. An illustration of the EPAM-VOC architecture. Nodes are represented by ellipses and links by arrows.

[4] For the purposes of this example, only five individual phonemes are illustrated below the top node. However, as mentioned above, the model is programmed to know the whole phonemic inventory of English prior to the beginning of the learning procedure.

If the search is successful, the chunk associated with the newly traversed link becomes the current chunk, and parsing will carry on in a left-to-right fashion.

Starting at the top node, EPAM-VOC will first match the phoneme /s/ (Figure 1, third link from the left), which will then become the current node. The model will then try to find a match for the next phoneme in the input, namely /ʌ/. As the first level of the hierarchy has already been traversed, the search will only consider links that are below the current node, which in this case is the node containing the phoneme /s/.

In the case of Figure 1, an /ʌ/ link can be found below the current node, providing the sequence /sʌ/ as the new chunk. The final phoneme is then examined and the /m/ link traversed, with /sʌm/ becoming the current chunk. As the input contains no further phonemes, the current chunk is returned, indicating that the phoneme sequence /sʌm/ is part of the model's long-term knowledge.

Although some input may be known to the model as a whole sequence (i.e., as an entire chunk, as for /sʌm/ in the example above), the vast majority of input will require several chunks in order to be represented. When a whole match cannot be achieved, the model will attempt to match the remainder of the input by applying its parsing procedure alternately to the right and left edges of the input string. This change in the parsing mechanism from the previous version of EPAM-VOC is motivated by evidence from current psychological research, particularly in relation to the role of primacy and recency effects, as reported in the literature on serial recall (e.g., Hulme, et al., 1997) as well as in the NWRT literature (e.g., Gupta, 2005).

Within the EPAM-VOC architecture, there are at least two approaches one could adopt in order to implement these effects. One of these would be to assign some "special" value to the chunks at the beginning and at the end of a sequence, in order to mark them more prominent. However, this implementation would necessarily involve the assignment of an arbitrary value in order to render certain chunks more easily accessible than others.

An alternative option, and the one we follow here, is to implement a mechanism that begins parsing the input from the left-edge (i.e., in the order in which phonological strings are heard), but then moves to the right edge of the input string, thus shifting its computational focus to that part of the phonological string which is more active in memory. This method has the advantage that it gives special status to word-initial sequences (by parsing them first) as well as word-final ones (by parsing them before any preceding subparts) without introducing any arbitrarily chosen computational weight. This change in the parsing strategy does not affect the fundamental architecture of the model, as the parse still proceeds in three basic steps:

(i) parsing a subpart of the input, (ii) returning the largest possible chunk, and (iii) resuming the parse for the remainder of the input. The only difference being that step (iii) begins elsewhere, namely at the opposite edge of the input string. This sequence of events necessarily assumes the involvement of a buffer where the remainder of the input string is held whilst each subpart is being parsed. This buffer, which is not unlike the episodic buffer discussed by Baddeley (2000a, 2000b), offers a parsimonious method of integrating primacy and recency effects as emergent properties of the parsing architecture rather than being programmed in *ad hoc* through a weighting procedure.

In order to illustrate how this works, let us take as a working example the input utterance "Some man" (phonemic representation: /sʌmmæn/), together with the network shown in Figure 1. Beginning at the left-edge, EPAM-VOC II can traverse the hierarchy in a left-to-right fashion as far as the /sʌm/ chunk, at which point traversal ends. This information is then removed from the input and traversal begins again using /mæn/ as input, but this time processing will involve the right-edge, while still proceeding in a left-to-right fashion. Thus, as the full input /mæn/ cannot be matched, EPAM-VOC II will attempt to traverse the sequence /æn/, as opposed to the sequence /mæ/. As this sequence can be traversed successfully (i.e., the network in Figure 1 contains the sequence /æn/), the phoneme /m/ is left as the remainder of the input, which will then be represented as a single phoneme chunk. Note that—given the network in Figure 1—processing the sub-string /mæn/ from the left-edge would have resulted in a different distribution of chunks, leaving the word-final phoneme /n/ (instead of the word-medial /m/) as the one-phoneme chunk. This is because the traversal would have proceeded to match the left-edge sequence /mæ/ instead of the right-edge /æn/. As individual phonemes are more likely to attract errors than phonemes that belong to larger chunks (see section on articulating an input sequence), processing the sub-string /mæn/ from the left-edge would have penalised the phonological material that appears at the end of a string, contrary to what research on recency effects would suggest. The edge at which processing begins is alternated at each traversal. First, the input string is processed left-to-right, and the resulting string is removed from the beginning of the input string. The input is then processed from the end of the input string in a right-to-left fashion, with the resulting string being removed from the input, whereupon left-to-right processing is resumed.

During the parse, the input is also being encoded into working memory. However, since working memory is capacity limited (e.g., Baddeley & Hitch, 1974; Cowan, 1997), the input needs to be restricted in some way such that an input sequence that is parsed into many chunks (or even one chunk with many phonemes within it) is limited in terms of its representation. We accomplish this via an interaction between working memory and long-term knowledge, the details of which are explained below.

Incorporating working memory

EPAM-VOC II simulates the relationship between long-term knowledge (henceforth LTK) and working memory (henceforth WM) through a system of pointers. When parsing an input, a pointer is created to each chunk returned at the end of a traversal, and the same pointer is then placed in WM. The serial order of the input is maintained by noting whether parsing began from the left-edge or the right-edge. For example, the initial parse of /sʌmmæn̩/ will place the pointer to the /sʌm̩/ chunk at the beginning of WM, the following parse will place the pointer to /æn̩/ at the end of WM, and the final parse of /m̩/ places its pointer between the two. In essence, pointers represent a procedure whereby a phonemic sequence is encoded in WM, and they are relied on when the chunks need to be retrieved. That is, any subsequent accessing of phonemic information from WM relies upon a chunk being accessed from its associated pointer placed in WM.

As there is evidence to suggest that WM is time-limited (e.g., Baddeley & Hitch, 1974; Gathercole & Baddeley, 1989), we limit the amount of auditory input that can be processed in WM to a duration of 2,000 ms (following Baddeley, Thompson, & Buchanan, 1975). This limitation interacts with the accessing times necessary to retrieve a phonological chunk from LTK during the parsing procedure which has been estimated by Zhang and Simon (1985) to be 400 ms for the chunk itself plus 30 ms for each phoneme within it, excluding the first phoneme. Thus, a three-phoneme string (e.g., /sʌm/) would total an accessing time of 460 ms (400 ms for the chunk, plus 30 ms each for the phonemes /ʌ/ and /m/).

As WM capacity has been argued to be dependent on the intensity of spreading activation across items rather than on the number of items involved (see for example Cowan, 1997; Cantor & Engle, 1993), we implement WM limitations as a restriction on the amount of time available to access long-term information. According to this implementation, encoding can be performed accurately only as long as the total time necessary to access the relevant input from LTK does not exceed 2,000 ms, with no principled restriction as to the amount of pointers that can be placed in WM. This differs from the previous version of EPAM-VOC that cut-off the remainder of an input once the time limit of 2,000 ms was reached; for a large input, therefore, the previous version would build a WM representation only for the initial part, whereas the new version represents the whole input but places a probability on correctly accessing each part of it.

In cases where the sum total of the accessing time exceeds the 2,000 ms limit, there will be a probability of < 1.00 that a pointer placed in WM can subsequently access its associated chunk. Precisely, the probability will be equivalent to the time limit (i.e., 2,000 ms) divided by the total time needed to access all the chunks from the input. For example, if the input were represented by five chunks whose total access time is equal to 2,300 ms, then the probability of subsequently accessing a chunk from its associated pointer in WM would be 2,000/2,300 = 0.87.

This probability will then have an effect whenever the model requires access to a chunk from its associated pointer. The pointer itself activates the relevant chunks that were accessed during the initial traversal process (e.g., the chunks /s/, /sʌ/ and /sʌm/ for the input /sʌm/ and the network in Figure 1). Accessing a chunk from its pointer therefore involves the traversal of these activated chunks to arrive at /sʌm/. However, when the probability of accessing a chunk falls below 1.00, the model may take an incorrect link during traversal and end up on the wrong path. In such instances, the chunk that is returned may be shorter or longer than the original, or it may differ from the original in one or more of its constituent phonemes. These eventualities simulate three well-known effects in children's phonological development, namely phoneme elision, phoneme epenthesis, and phoneme substitution.

Sequences that are incorrectly retrieved will then have an effect on the learning as well as the articulation procedures, as the functioning of both of these is dependent on accessing chunks from the pointers that are encoded in WM. It is to the learning procedure that we now turn (the articulation procedure will be addressed in the next two sections).

Learning long-term phonological knowledge

The procedure through which EPAM-VOC II achieves long-term phonological knowledge remains unaltered from that of the previous version. When an input sequence cannot be matched with a single chunk, EPAM-VOC will proceed to acquire new knowledge by expanding its network to include some, or all, of the sequence that it failed to match fully (i.e., that it could match only as a series of chunks). Recall that, after encoding a sequence, EPAM-VOC relies on a series of pointers in order to access it, and one case in which this sequence needs to be accessed is in order for EPAM-VOC to learn new information. Learning occurs after the input has been parsed and the input information has been placed as pointers into WM. Pointers are processed in a left-to-right and pair-wise fashion, with EPAM-VOC attempting to access the chunks associated with each pair of pointers in order to create a new chunk that joins the information together.

Let us once again take as an example the input utterance /sʌm/. Prior to any learning, any input can only be encoded one phoneme at a time. In the case of /sʌm/, this would result in three pointers being placed in WM, for the phonemes /s/, /ʌ/, and /m/ respectively. Since the total access time for these chunks is 1,200 ms (3×400 ms), all chunks are accessed successfully. The next step will constitute the learning procedure proper, as EPAM-VOC links the /ʌ/ phoneme to the /s/ phoneme by placing the former under the latter in the hierarchy, effectively learning the phoneme sequence /sʌ/. The same process is then repeated for the next pair of pointers, which link the phonemes /ʌ/ and /m/, obtaining the sequence /ʌm/. As there are no further pairs of pointers in this case, the learning procedure ends (see Figure 2).

Consider now the resulting network receiving the same input again. This time, the input can be encoded as only two chunks, namely /sʌ/ and /m/, with the consequence that only two pointers will be placed in WM. As both chunks can be accessed accurately (access time is 430 ms + 400 ms = 830 ms), EPAM-VOC can proceed to add a /m/ link below the /sʌ/ chunk, thus adding the phoneme sequence /sʌm/ as part of its LTK. Hence, after two presentations of the sequence /sʌm/ the model can be said to have learnt the word "some". This last step is represented as the move from the network in Figure 2 to the one in Figure 3, with Figure 3 showing the resulting network. One may consider learning in EPAM-VOC to be quite rapid; slowing down the learning procedure has been successful for other variants of EPAM models (e.g., Croker, Pine, & Gobet, 2003; Freudenthal, Pine, & Gobet, 2006). However, since EPAM-VOC is only presented with a small subset of the natural language a child may hear, it seems sensible to have learning proceed in the manner described. Slowing down learning will likely yield the same set of results, but over a longer period of time.

In previous versions of EPAM-VOC, repetition performance was simulated by a retrieval procedure. When confronted with an input, EPAM-VOC would attempt to retrieve one or more chunks from LTM that would allow it to match that input. The model performance was then judged based on how successful this retrieval procedure was. While this was a reasonable compromise for an earlier stage of the model's development, the resulting architecture did not involve production of an actual output, and thus

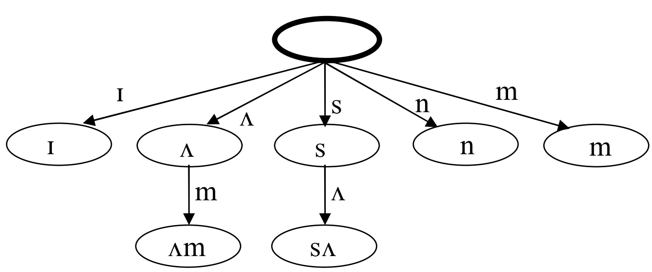

Figure 2. EPAM-VOC II after having been presented with the input /sʌm/ for the first time.

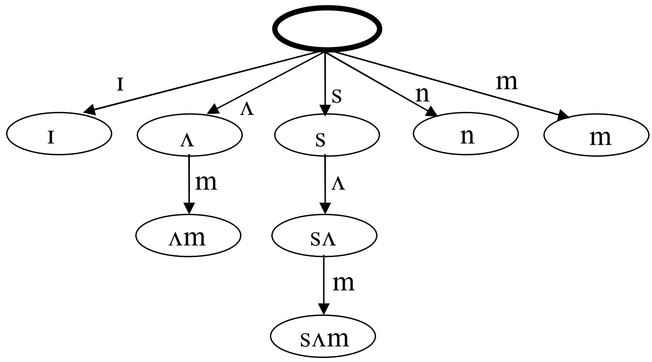

Figure 3. EPAM-VOC after having been presented with the input /sʌm/ twice and having acquired LTK of this sequence.

abstracted away from the articulation procedure that lies between the retrieval of a phonological representation from WM and the actual production of an utterance.

As a further development of the model it therefore seemed important to also include as part of the modelling architecture a procedure that aims at simulating this articulation stage.

Articulating an input sequence (performing the nonword repetition test)

Just as we have seen for real words (see discussions above on long term knowledge and on working memory), when the model is presented with the phonemic representation of a nonword it will proceed to parse it by associating it with as few chunks as possible from the long-term hierarchy, and to subsequently encode the parse by linking these chunk(s) to a number of pointers in WM. Articulation of a sequence involves accessing the relevant chunk information from the pointers in WM, and then articulating each of the phonemes contained in the resulting sequence. As we saw in our discussion on working memory, the former process might involve an error in accessing the relevant chunks, which will then be returned on articulation. However, the updated version of EPAM-VOC allows errors to also occur in the latter stage (i.e., the actual production of the phoneme sequence).

When retrieving a chunk, EPAM-VOC II calculates a probability of correctly articulating the information contained within it based on the processing weight of the chunk. The processing weight of a chunk is calculated based on the frequency of its subsets, as these represent all the chunks that need to be accessed during the traversal phase. As the frequency of single phonemes is on average much larger than any possible sequence involving that phoneme, we multiply the frequency of phoneme sequences by multiples of five, depending on their length.[5] The aim of this procedure is to increase the impact that longer sequences might have, particularly when compared to single phonemes. This decision is based on the assumption that having been exposed to a phoneme sequence is developmentally more advantageous than having been exposed to its component phonemes in separate phonological contexts,[6] as only the former type of exposure may provide the phonological knowledge necessary in order to successfully articulate phonemes in context. Essentially, this follows from a view of phonological knowledge that dates back at least to Jakobson (1941)/1968), and which finds overwhelming support throughout the developmental literature, particularly in relation to the study of Voice Onset Time (see for example Catts & Kamhi, 1984; Johnson & Wilson, 2002; Macken & Barton, 1980) but also in relation to general syllabic development (e.g., Demuth, Culbertson, & Alter, 2006; Goad & Brannen, 2003, *inter alia*). Furthermore, the motivation for basing correct articulation (in part) on frequency is supported by good correlations between the frequency with which phonemes and consonant clusters are used by the children in the Manchester corpus (see simulations reported below) and the age of acquisition of the phonemes and clusters (from Smit, Hand, Freilinger, Bernthal, & Bird, 1990) [$r(48) = -0.51$, $p < .001$].

When an input sequence requires more than one chunk to represent it, then it is clear that the model does not have knowledge of this particular sequence of phonemes as a whole. We therefore divide the calculated frequency of a chunk by the number of chunks required to represent the input sequence, to account for the fact that articulation will be more demanding as the number of chunks required to represent the input increases.

After having obtained the weight of a chunk, a probability for articulation error is calculated by dividing the common logarithm of this weight by 4. In essence, this sets a threshold of 10,000 (as $log\ 10{,}000 = 4$) above which articulation will be free from errors. Logarithms rather than raw counts were used because different phonemes—and phonemic sequences—display massive ranges of variation, a fact that would have compromised the reliability of our comparisons. When an articulation error does occur, a consonant is

[5] The frequency of a bi-phone sequence is multiplied by 5, that of a tri-phone sequence is multiplied by 25, while that of a quadri-phone sequence is multiplied by 125 ($5 \times 5 \times 5$), and so on for longer sequences.

[6] And, in turn, having had the opportunity to *practise* articulation of a phoneme sequence is more advantageous than having practised its component phonemes in separate phonological contexts.

randomly selected within the chunk and substituted for another consonant, also selected at random, from an inventory of English consonants.[7] At this stage of development, we have not yet included a sub-phonemic theory that could provide information on the structures of—and thus on the potential differences between—consonantal and vocalic elements. Consequently, we concentrate on modelling performance solely within the consonantal domain, directly inhibiting vowel errors during the production phase.[8]

As the consonants involved in this procedure are selected at random, the resulting errors do not reflect the types of consonant substitutions observed in children. Such a procedure would require the model to have knowledge of sub-phonemic components, and it is therefore not achievable at this stage of model development. Thus, the focus of this paper will be entirely on the modelling of errors at the positional rather than the melodic level.

THE EXPERIMENTS

Introduction

EPAM-VOC allows us to examine error patterns in NWR performance both across and within nonword types. However, well-established NWR tests contain a high rate of lexical components. For example, the nonwords used by Gathercole and Baddeley (1989) contain lexical (e.g., ***pen**nel*, ***thick**ery*, *dope**late***) and morphological components (e.g., *slad**ing***, *penneri**ful**,* *fenner**iser***). Also, as the nonwords were not controlled for phoneme-pair frequency, it is impossible to deduce whether—or to what extent—error patterns may be influenced by frequency of exposure to certain phonological strings. We have therefore collected additional data in order to assess children's NWR performance across nonword types. The tests included nonword sets involving lexical and morphological components (Gathercole and Baddeley, 1989), as well as a newly devised set that manipulates frequency as the main variable and excludes lexical or morphological components.

Participants

Twenty-five children took part in this research, 12 males and 13 females. They were between 5:4 and 6:8 years of age ($M = 6.1$). The children were recruited from four schools in and around the city of Nottingham, and they were all monolingual native English speakers. Children had to satisfy two

[7] For the purposes of this operation, the consonant that is being substituted is temporarily removed from the inventory.

[8] However, vowel articulation errors can still occur in cases where the single phoneme contained within a chunk is a vowel.

criteria in order to be included in the sample. First, all children had to achieve a minimum score of 85 on a test of non-verbal intelligence. The non-verbal test used was the Wechsler Pre-school and Primary Scale of Intelligence 3 UK test (Wechsler, 2002). Second, in order to ensure that their vocabulary was appropriate to their age, children had to achieve a score of 85 or higher on the British Picture Vocabulary Scale 2 (Dunn, Dunn, Whetton, & Burley, 1997).

Materials

Nonwords set 1: Gathercole, Willis, Baddeley and Emslie (1994)

The children were presented with two separate nonword repetition tests. The first test was a modified version of the nonword repetition test developed by Gathercole, Willis, Baddeley and Emslie (1994), also known as the CNRep. This test originally involved 40 nonwords consisting of two-, three-, four-, and five-syllable items. However, we omitted the five-syllable items following a pilot which showed that younger children had great difficulty repeating nonwords of this length. Thus, the children were presented with three lengths of nonword, each containing 10 items, for a total of 30 nonwords. Nonwords at each length consisted of five items containing only single consonants (i.e., they had a CV.CV or a CV.CVC structure, e.g., bannow, pennel) and a further five containing one or more consonantal sequences. Such sequences were either true clusters (i.e., CCV structures, e.g., **sl**ading) or coda-onset sequences (i.e., VC.C structures, e.g., pri**nd**el).

Presentation of the 30 nonwords was split into two 15 nonword lists, and we kept to the fixed random order that was provided on the original audio cassette (Gathercole & Baddeley, 1996), though the track was transferred to MP3 format for ease of administration. Each nonword list began with the standard instructions provided on the audio-cassette (see next section), though the practice items were removed. This is due to the fact that each set of nonwords was presented in a counter-balanced order—hence no particular set of nonwords could be assigned to be the "first heard". Therefore, no training sets were given for any of the nonword tests.

Nonwords set 2: Newly constructed nonwords

The second set of nonwords was created by the first two authors, consisting of two sets of 3-syllable nonwords, each containing eight nonwords. All nonwords had primary stress on the initial syllable and secondary stress on the final syllable. Unlike the test discussed above, these nonwords were aimed at testing the impact that phonotactic probabilities may have on children's performance in NWRTs. Thus, the two groups were divided into low-frequency (LF) and very low-frequency (VLF) items, based on the number of low-frequency biphones (i.e., phoneme pairs) that they contained. LF nonwords contained a maximum of three low-frequency biphone sequences while VLF nonwords contained a minimum of 4. The nonwords were all constructed to be of relatively low frequency because of the difficulty in constructing high-frequency nonwords that maintained a similar phoneme set to those used in the CNRep.

The two groups (LF and VLF) were matched for length and the number of phonemes. For each group, two nonwords were composed of seven phonemes, four nonwords were composed of eight phonemes and two nonwords were composed of nine phonemes. They were also matched for vowel quantity (i.e., the number of short vowels and long vowels/diphthongs that they contained), and for consonantal markedness (i.e., the number of consonantal phonemes that have been documented as appearing relatively late in acquisition; see Clark, 2003; McLeod, Doorn, & Reed, 2001). The mean spoken duration was 0.91 s for the LF set and 0.92 s for the VLF set. The spoken duration across conditions was equivalent to [t (7) = 0.15, p = .89].

Biphone frequencies were calculated based on their occurrence in the Children's Printed Word Database (http://www.essex.ac.uk/psychology/cpwd/), a database of word frequencies for 5- to 9-year-old children. Probabilities were then aggregated on a biphone basis. For example, the word /trɪp/ has a frequency of 116 and consists of four phonemes. Therefore, the biphones /tr/, /rɪ/, and /ɪp/ had 116 added to their respective biphone frequencies. This method is very similar to the one used in a series of works by Luce and colleagues (e.g., Jusczyk, Luce, & Charles-Luce, 1994; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997; Vitevitch & Luce, 1998) and has been found to be a good predictor of word-likeness ratings in adults (see Frisch, Large, & Pisoni, 2000).

Finally, care was taken to use only constituent syllables that did not exist as lexical items in the children's printed word database and to employ biphones that appeared in the children's printed word database, hence none of the biphones used had a frequency of zero.

The nonwords were recorded onto a Sony ICD-MX20 digital voice dictaphone by a researcher unrelated to the project who was native to the Nottingham area, so that the dialectal features would be familiar to the children. All nonwords were converted to MP3 format using Sony Digital Voice Editor, version 3.1 (available at http://esupport.sony.com/). The nonwords were then arranged into the required order using Audacity (http://www.audacity.sourceforge.net/). Two sound files were produced: one with the nonwords in a randomised order and a second one that was the reverse of this order. Each spoken nonword was succeeded by a 4.0 s pause (as per the CNRep nonwords). The nonword lists were then turned back into

MSV files so that they were compatible with the voice recorder. No practice nonwords were provided on any of the recordings because the order of presentation of the nonword tests was counterbalanced across all tests. The children heard each nonword only once. The following sentences, based on Gathercole and Baddeley (1996), were provided as instructions to the children: "*Hello, in a few seconds you will hear a funny made up word. Please say the word aloud yourself as soon as you hear it*".

Standardised tests

Three standardised tests were used to ensure that the children's IQ and vocabulary were appropriate for their age, and that they could articulate the sounds within the nonwords. The core non-verbal tests (block design, matrix reasoning, and picture concepts) of the Wechsler Preschool and Primary Scale of Intelligence 3 UK (WPPSI-3 UK, Wechsler, 2002) assessed non-verbal IQ. The British Picture Vocabulary Scale 2 (BPVS-2, Dunn, Dunn, Whetton, & Burley, 1997) assessed vocabulary level. Appropriate levels for age were scores of 85 or higher on the standardised scores for IQ and vocabulary. Ability to correctly articulate phonemes was verified using the Diagnostic Evaluation of Articulation and Phonology (Dodd, Hua, Crosbie, Holm, & Ozanne, 2002).

Experimental design

For nonwords set 1 (CNRep) there were two independent variables: Nonword Length and Consonant type. The first independent variable had three levels: two-, three-, and four-syllables; and the second had two levels: single and clustered. Both were repeated measures variables.

For nonwords set 2 there was only one independent variable—nonword frequency—which had two levels: low frequency and very low frequency. This was also a repeated measures variable.

For both sets of nonwords, the error position (onset, nucleus, or coda) was also an independent variable, since we measured as our dependent variable the number of errors made in onset position, nucleus position, and coda position.

Procedure

The children were assessed on a one-to-one basis in their school, in a quiet room away from their classroom. Up to four sessions were used for each child to administer the standardised tests and the nonword tests. Each session lasted a maximum of 15 minutes. Administration of the WPPSI-3 UK, the BPVS-2, and the DEAP, together with the nonword tests, were interspersed across sessions. The specific nonword repetition tests (the two that formed the CNRep and the two that formed the new nonword test) that each child performed within sessions was randomised. Nonwords were played from a Sony ICD-MX20 digital voice dictaphone (Memory Stick Digital Recorder) through Creative TravelDock 900 Portable speakers. All children's repetitions were recorded onto a Sony ICD-MX20 digital voice dictaphone.

Transcription

Nonwords were broadly transcribed from the recordings, and a random sample of 15% was then transcribed by two other researchers.[9] Inter-rater reliability was 87.3% (range: 85–92.3%). The repetition errors were then divided according to the syllabic position in which they occurred: onset, nucleus, or coda. As there are different approaches to the question of syllabification, we analysed each set of errors according to the two principles that are most commonly employed in the literature on English syllables (see appendixes 1 and 2). The first of these is traditionally known as the "Maximal Onset Principle", which dates back at least to the work of Pulgram (1970) and is often relied upon in current phonological theory (see for example the volume edited by Bernhardt & Stemberger, 1998; as well as much of the literature on Optimality Theory, Prince & Smolensky, 1993/2004). According to this principle, word-medial consonants are syllabified as onsets, unless they create a consonantal sequence which is impossible for an English word-initial onset.

However, it has been argued that in the case of intervocalic consonants syllabification should also take into consideration the phonotactic regularities that apply to vocalic elements (e.g., Jones, 1997; Roach, 2000). For example, according to Maximal Onset, the word "platter" would be syllabified as /plæ.tə/. However, when looking at what constitutes a possible word-final sequence in English, we find that the following vowels never occur in a stressed position without a coda: /ɪ e æ ʌ ɒ ʊ/. Some authors have therefore adhered to the standard proposed by the English Pronouncing Dictionary (Jones, 1997) and allow for the Maximal Onset Principle to be outranked, thus syllabifying "platter" as /plæt.ə/.

Consequently, in our analysis of children's errors we also consider this syllabification pattern, which we label "Closed Syllabification", in reference to the fact that it bans certain vowels from ending as an open syllable.

[9] These were the first author and a second researcher not involved in this project but experienced in coding nonword repetitions.

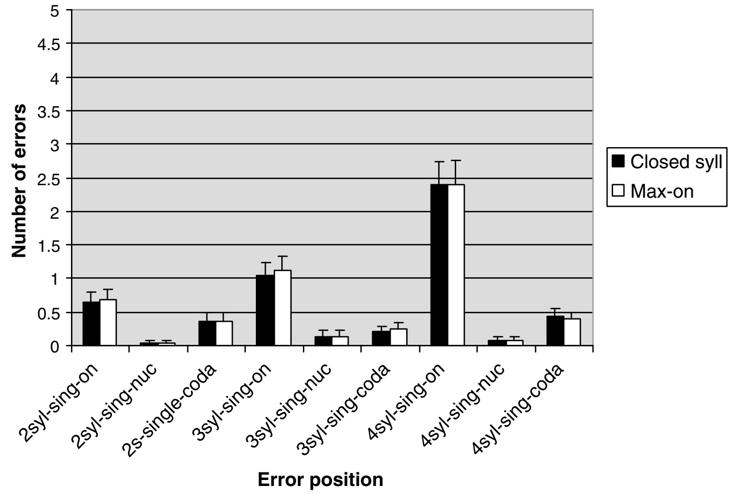

Figure 4. Onset, nucleus, and coda errors for the children for the CNRep single consonant nonwords, for both syllabification methods.

Simulating children's performance

The same sets of nonwords were run through the EPAM-VOC II model after it had been exposed to naturalistic input. The simulations used 12 sets of maternal utterances from the Manchester corpus (Theakston, Lieven, Pine, & Rowland, 2001) on the CHILDES database (MacWhinney 2000). They had been uttered by mothers while interacting with 2- to 3-year-old children. The average number of utterances was 25,519 (range 17,474–33,452).

The amount of input in these sets of utterances governed the amount of input that was given to the model for each "mother". That is, 12 sets of input were used, one for each mother. However, since we were simulating 5-year-old children, our input also included parental utterances aimed at 4- to 5-year-old children as well as stories from children's books.[10] The shift from input based on 2- to 3-year-olds and that based on 4-to 5-year-olds was accomplished by increasing the amount of the 4- to 5-year-old input as learning progressed (and subsequently decreasing the amount of 2-to 3-year-old input). For example, for any one mother, the initial input consisted of the utterances spoken to her 2- to 3-year-old child. After 25% of the utterances had been seen (in line with Jones et al., 2007, 2008), increasing amounts of the 4-to 5-year-old input were introduced at the expense of some of the 2- to 3-year-old input (such that in the final stages of learning, 80% of the input was based on the 4- to 5-year-old stimuli). Each line of 4- to 5-year-old input

[10]These were added to represent the increase in input during the first year of schooling.

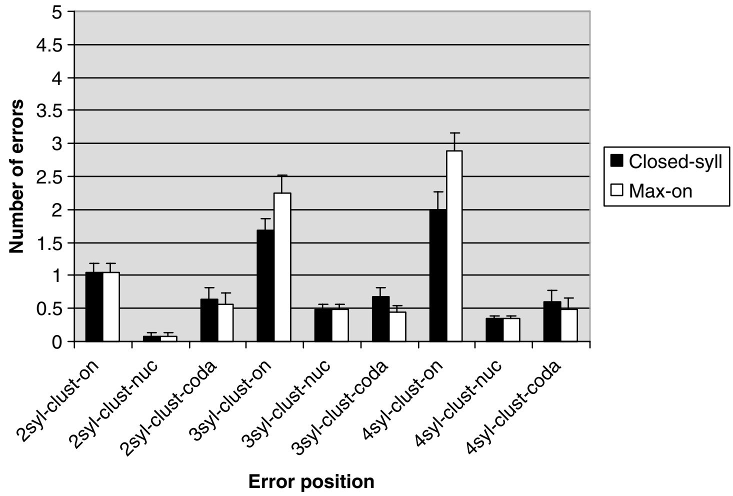

Figure 5. Onset, nucleus, and coda errors for the children for the CNRep clustered-consonant nonwords, for both syllabification methods.

was randomly selected without replacement from the full set of 4- to 5-year-old input corpora. This means a high likelihood that for every simulation of the child data, a different set of "older" input is used.

The parent–child utterances to 4- to 5-year-old children were from the English-speaking files of the Watkins' corpus, available from the CHILDES database (MacWhinney 2000). The written input was taken from books that would normally be read by or to 5-year-old children. Examples of the stories used are "The Ugly Duckling" (Hans Christian Andersen) and "Snow White" (The Brothers Grimm).

All utterances were converted into their phonemic representation using the CMU Lexicon database (available at http://www.speech.cs.cmu.edu/cgi-bin/cmudict), which contains the phonemic representation of over 120,000 and allows automatic conversion of utterances into phoneme sequences. The input did not distinguish word boundaries, hence no word segmentation had been performed on the input that was fed to the model.

A total of 120 simulations were carried out (10 for each of the mothers). This was because, with the introduction of random utterances from the 5-year-old input sets, we wished to ensure that the output constituted a reliable representation of the model's performance and was not simply due to a potentially biased sample from the input. In addition, since the retrieval and articulation procedures have probabilistic elements to them (cf. sections on working memory and on articulating input sequences), each nonword test was carried out 10 times within each simulation. This resulted in 1,200 nonword test simulations. All NWR results for the model are based on the

TABLE 1
Nonword repetition accuracy (percentage accuracy, standard deviation in parentheses) for both nonword sets, for the children and the model

	Nonword set 1 (CNRep nonwords)					
	Single consonant			*Clustered consonant*		
	2-Syllable	*3-Syllable*	*4-Syllable*	*2-Syllable*	*3-Syllable*	*4-Syllable*
Children	79.20 (15.79)	73.60 (19.77)	48.00 (23.09)	67.20 (19.90)	48.80 (22.42)	36.00 (25.82)
Model	78.15 (17.02)	62.38 (21.52)	41.72 (21.83)	65.25 (22.22)	51.05 (22.01)	34.60 (20.86)
	Nonword set 2 (new nonwords)					
	LF			VLF		
Children	42.50 (16.93)			43.50 (18.44)		
Model	48.15 (17.21)			47.13 (16.70)		

TABLE 2
ANOVA results for children's performance on the CNRep nonwords, for both syllabification methods

	Maximal Onset Principle			
	F	*df*	*p*	η_p^2
Nonword type	20.10	1,24	<.001	0.46
Nonword length	14.76	2,48	<.001	0.41
Error position	108.77	2,48	<.001	0.82
Nonword type × Nonword length	2.42	2,48	.100	0.09
Nonword type × Error position	5.54	2,48	.007	0.19
Nonword length × Error position	20.39	4,96	<.001	0.46
	Closed syllabification			
Nonword type	11.50	1,24	.002	0.32
Nonword length	14.40	2,48	<.001	0.38
Error position	88.28	2,48	<.001	0.79
Nonword type × Nonword length	3.16	2,48	.051	0.12
Nonword type × Error position	0.19	2,48	.825	0.01
Nonword length × Error position	10.70	4,96	<.001	0.31

average results for each set of mother input—resulting in 12 averaged datasets. For example, for "Anne", the 10 runs of the simulation that was based on Anne's Mother's utterances and the 10 nonword repetition tests for each run of the simulation (i.e., 100 nonword repetitions for each nonword) were averaged. This was done in order to achieve a comparable number of simulations to the number of children who participated in the nonword study.

The model's repetition errors were also divided according to the syllabic position in which they occurred. Note that syllabification was used only as an analytic tool in order to examine the error positions in the children's and model's performance, and was not part of the mechanisms through which the model achieves its output. In fact, as explained above, the chunking procedure employed by EPAM-VOC relies solely on frequency considerations and has no built-in definition of what a "syllable" consists of.

Results

Table 1 shows the nonword repetition accuracy data for both nonword sets, for the children and the model. As can be seen from the table, the model shows a good fit to the child data (within ±10%) for seven of the eight datapoints.

Nonwords set 1: Children's results

Figures 4 and 5 show the raw number of onset, nucleus, and coda errors for the CNRep nonwords, for both syllabification methods. Table 2 shows the results of the ANOVA analyses for both syllabification methods. As the figures clearly illustrate, exactly the same effects are seen irrespective of how nonwords are syllabified. Table 2 also illustrates that a very similar set of statistical effects are seen for both syllabification methods (all three main effects and two of the three pairwise interactions are closely matched for significance)—any further statistics in this section are therefore based on the closed syllabification data. A 2 (nonword type: single or clustered) $\times$ 3 (nonword length: 2, 3, or 4 syllables) $\times$ 3 (error position: onset, nucleus, or coda) within subjects ANOVA was performed on the children's data. All statistics are shown in Table 2. There was a significant effect of nonword type, with nonwords containing consonantal sequences attracting more errors than their single consonant counterparts. There was also a significant effect of nonword length, with longer nonwords attracting more errors than shorter ones. *Post hoc* Bonferroni tests indicated that there were significantly more errors made for 4-syllable nonwords than both 2- and 3-syllable nonwords ($p < .001$ and $p = .032$ respectively) and more errors made for 3-syllable nonwords than 2-syllable ones ($p = .033$).

Moreover, there was an effect of error position, with syllable onsets consistently attracting more errors than syllable codas, and syllable nuclei being the least error-prone elements. *Post hoc* Bonferroni tests indicated that there were more onset errors than coda ($p < .001$) and nucleus ($p < .001$) errors, and significantly more coda errors than nucleus errors ($p = .004$). There were no interactions between nonword type and nonword length or nonword type and error position. However, there was a significant interaction between nonword length and error position, indicating that as nonword length increased, the number of onset errors increased accordingly, yet the number of nucleus and coda errors remained stable.

Note that the increase in onset errors as nonword length increases cannot simply be due to the fact that opportunities for making onset errors increase with nonword length. Although longer words do have a higher ratio of onsets

TABLE 3
Ratio of onsets to codas for nonwords in the CNRep, expressed as a percentage (the number of onsets divided by [the number of onsets + the number of codas])

Syllabic length	*Onsets N*	*Codas N*	*Ratio (% onsets)*
2syll	18	19	48.6
3syll	30	20	60
4syll	40	24	62.5

to coda consonants, the difference is not sufficiently high to explain the pattern of error, thus excluding the possibility that the increase in onset errors is simply a side-effect of a high number of onsets present in the stimuli. Table 3 shows the ratio of onsets to codas in the nonwords.

Based on the raw number of onsets and codas that each nonword type contains, it can be seen that the ratio of onset:coda increases with nonword length. However, in the children the ratio of onset error:coda errors increases over and above what would normally be expected based on the raw numbers of onsets and codas. This is evidenced by calculating the onset error:coda error ratio for each child at each nonword length, and performing one-sampled t-tests to compare this ratio with what would be expected by chance. At all lengths of nonword, there were significantly more onset errors than would be expected if onset errors were simply a byproduct of the number of onsets versus the number of codas contained in each nonword [$t(24) = 4.24$, $p < .001$ for 2-syllable; $t(23) = 6.46$, $p < .001$ for 3-syllable; $t(23) = 5.11$, $p < .001$ for 4-syllable].

It is in principle possible that the difference in error position we have observed can be explained without having to appeal to the concept of "syllable onset". Since syllable onsets include word-initial phonemes, it might be the case that our analysis in terms of syllable position masks what is actually a tendency for word-initial errors. If this is the case, the phenomenon at issue could be explained in a linear fashion (i.e., by appealing to the concept of "word-initial phoneme"), without the need for a syllabic analysis. In order to exclude this possibility, we compared word-initial onsets to word-medial onsets and, for consistency, word-medial codas to word-final codas. Word-initial and word-medial onset errors were totalled across all nonword lengths. For single consonant nonwords, there were significantly less word-initial onset errors than word-medial onset errors [word-initial $M = 1.92$, $SD = 1.21$; word-medial $M = 4.00$ $SD = 1.79$; $t(23) = -5.22$, $p < .001$]. The same was true for the clustered consonant nonwords [word-initial $M = 1.71$, $SD = 1.46$; word-medial $M = 4.96$ $SD = 2.80$; $t(23) = -4.70$, $p < .001$]. The coda data are as follows. For single consonant nonwords, word-final codas showed significantly more errors than word-medial codas [word-medial $M = 0.00$, $SD = 0.00$; word-final $M = 1.12$, $SD = 0.92$; $t(23) = -6.23$, $p < .001$]. The same was true of the clustered consonant nonwords [word-medial $M = 0.96$, $SD = 0.91$; word-final $M = 2.50$, $SD = 1.41$; $t(23) = -4.21$, $p < .001$]. Clearly, our onset bias is not caused by a large amount of word-initial onset errors. However, it would seem that for codas, there is a tendency for an error to occur word-finally more than word-medially.

At first, this might look at odds with the primacy/recency advantage described earlier, under which word-final codas would be expected to be less error-prone than word-medial codas. However, position is only one of the contributing factors affecting target accuracy, with frequency being the other, potentially more powerful, factor. While we know from primacy and recency effects that word-initial and word-final positions have the potential to hold a repetition accuracy advantage over word-medial positions, it is also the case that more frequent phoneme sequences hold a repetition advantage over lees frequent sequences[11] (e.g., Munson, Kurtz & Windsor, 2005). In the case of clustered consonants nonwords, the role of frequency in different coda positions is rather straightforward: word-medial codas have much stronger phonotactic restrictions—and therefore have a higher positional frequency—than word-final codas (e.g., Selkirk, 1982). The fact that word-final codas are more prone to error in our results shows how the role of recency is modulated by that of frequency and—in the case of codas—frequency has the upper hand. For single-consonant nonwords, the situation must be viewed on a case-by-case basis (at least for longer nonwords), since no general phonotactic restrictions apply. For shorter nonwords, on the other hand, the fact that the parsing procedure begins in a left-to-right fashion gives an inherent disadvantage to word-final segments, since they are more likely to end up in a single-segment chunk. As we have seen in our discussion on long term knowledge, for longer nonwords the situation is more complex. What emerges from this discussion is exactly what we think is a major advantage of computational modelling: it allows us to observe what happens when different contributing factors combine. The potential conflict between primacy/recency on the one hand and frequency on the other is a case in point.

Nonwords set 1: Model's results

Figures 6 and 7 show the raw number of onset, nucleus, and coda errors for the CNRep nonwords, for both the children and the model. A 2 (nonword type: single or clustered) $\times$ 3 (nonword length: 2, 3, or 4 syllables) $\times$ 3 (error position: onset, nucleus, or coda) within subjects ANOVA was performed on the model's data. There was a significant effect of nonword type [$F(1, 11) = 29.46$, $p < .001$, $\eta_p^2 = 0.73$], with nonwords containing consonantal sequences attracting more errors than their single-consonant counterparts. There was also a significant effect of nonword

[11] In fact, we also analysed primacy/recency effects for all three syllable nonwords used in the nonword tests presented here (three syllable nonwords were used because both nonword sets contain this length of stimuli and this length also allows for the examination of primacy and recency). No primacy and recency effects were observed ($F(2,50) = 2.52$, $p = .091$, $\eta_p^2 = .09$). As discussed above, this is most likely because the nonwords were not designed in order to examine primacy and recency effects, which are only likely to emerge when other contributing factors are controlled for.

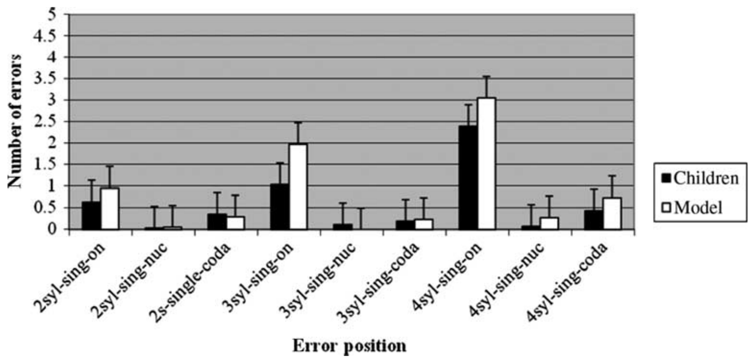

Figure 6. Onset, nucleus, and coda errors for the children and model for the CNRep single consonant nonwords.

length [$F(2, 22) = 144.41$, $p < .001$, $\eta_p^2 = 0.93$], with longer nonwords attracting more errors than shorter ones. *Post hoc* Bonferroni tests indicated that there were significantly more errors made for 4-syllable nonwords than both 2- and 3-syllable nonwords, and more errors made for 3-syllable nonwords than 2-syllable ones ($p < .001$ in all cases).

There was also an effect of error position [$F(2, 22) = 717.16$, $p < .001$, $\eta_p^2 = 0.99$], with syllable onsets consistently attracting more errors than syllable codas. *Post hoc* Bonferroni tests indicated that there were more onset errors than coda and nucleus errors, and significantly more coda errors than nucleus errors ($p < .001$ in all cases).[12] There was no interaction between nonword type and nonword length [$F(2, 22) = 2.02$, $p = .156$, $\eta_p^2 = 0.16$]. However, there was a significant interaction between nonword type and error position [$F(2, 22) = 38.74$, $p < .001$, $\eta_p^2 = 0.78$] indicating that coda errors increased for the clustered consonant nonwords, whereas onset and nucleus errors remained stable across single and clustered consonant nonwords. There was also a significant interaction between nonword length and error position [$F(4, 44) = 62.76$, $p < .001$, $\eta_p^2 = 0.85$], indicating that as nonword length increased, the number of onset errors significantly increased while the number of coda errors increased only marginally, with the number of nucleus errors remaining stable.

As with the children's data, the ratio of onset errors to coda errors was over and above what would be expected on the basis of the raw frequency of onsets and codas. This was the case for both 2-syllable [$t(11) = 6.43$,

[12]As far as the model is concerned, vowel errors will not be discussed as they were deliberately inhibited (see section on articulating an input sequence). We leave the question of simulating consonantal vs. vocalic errors for further research.

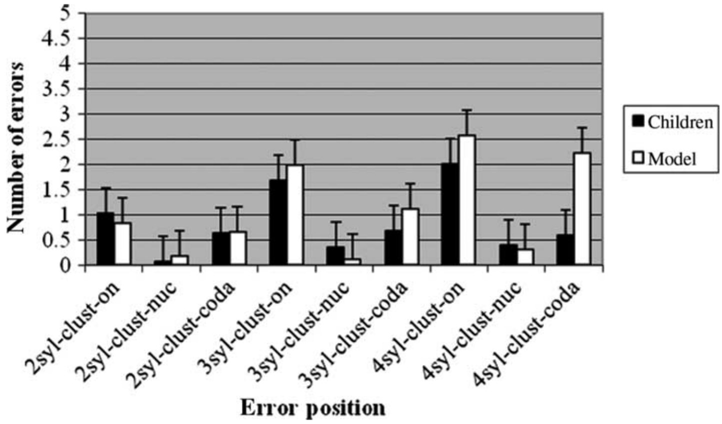

Figure 7. Onset, nucleus, and coda errors for the children and model for the CNRep clustered-consonant nonwords.

$p < .001$] and 3-syllable nonwords [$t(11) = 10.69$, $p < .001$], though not for 4-syllable nonwords [$t(11) = 0.57$, $p > .05$].

Nonwords set 2: Children's results

Figure 8 shows the raw frequency of onset and coda errors for nonword set 2, for both syllabification methods; and Table 4 shows the statistical analysis for both syllabification methods, based on a 2 (nonword type: low

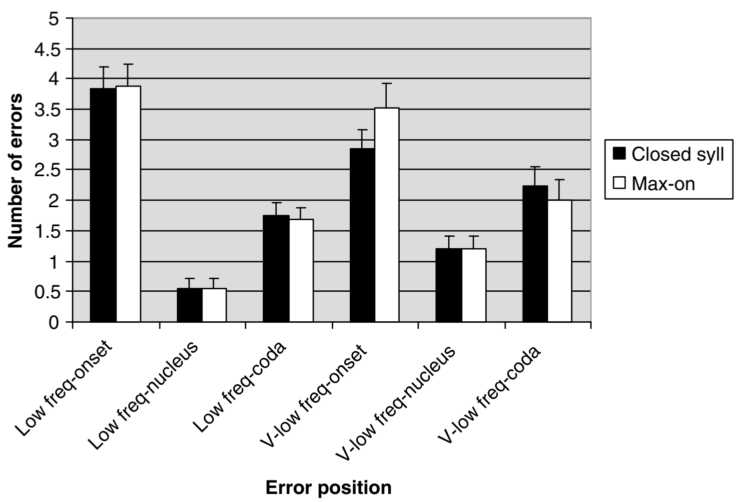

Figure 8. Onset, nucleus, and coda errors for the children for the new set of nonwords, for both syllabification methods.

TABLE 4
ANOVA results for children's performance on the new set of nonwords, for both syllabification methods

	F	*Df*	*p*	η_p^2
	Maximal Onset Principle			
Nonword type	0.66	1,24	.425	0.03
Error position	56.78	2,48	<.001	0.70
Nonword type × Error position	3.39	2,48	.042	0.12
	Closed syllabification			
Nonword type	0.00	1,24	.962	0.00
Error position	58.00	2,48	<.001	0.71
Nonword type × Error position	8.25	2,48	.001	0.26

frequency or very low frequency) × 3 (error position: onset, nucleus, or coda) within subjects ANOVA. Exactly the same effects are seen irrespective of how nonwords are syllabified, and any subsequent tests in this section are therefore based on the closed syllabification data. There was no effect of nonword type but there was an effect of error position. *Post hoc* Bonferroni tests indicated that there were significantly more onset errors than coda ($p < .001$) and nucleus ($p < .001$) errors, and significantly more coda errors than nucleus errors ($p < .001$). There was also a significant interaction between nonword type and error position indicating that the difference between the number of onset and coda errors was greater in the low frequency than the very low frequency nonwords.

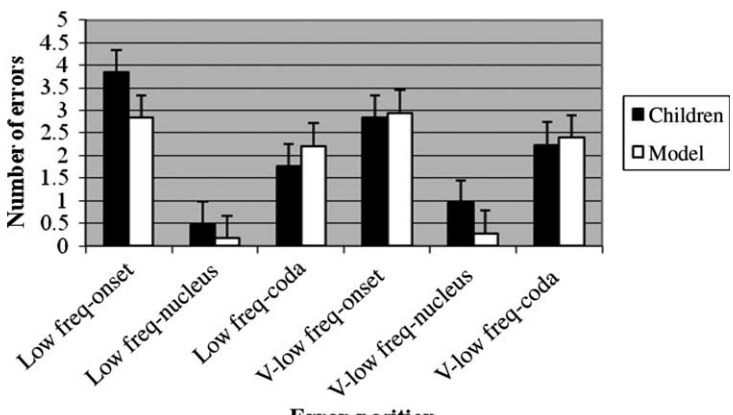

Figure 9. Onset, nucleus, and coda errors for the children and model for the new set of nonwords.

Nonwords set 2: Model's results

Figure 9 shows the raw frequency of onset and coda errors for nonword set 2, for both the children and the model. A 2 (nonword type: low frequency or very low frequency) × 3 (error position: onset, nucleus, or coda) within subjects ANOVA was performed on the model's data. There was no effect of nonword type [$F(1, 11) = 2.15$, $p = .171$, $\eta_p^2 = 0.16$] but there was an effect of error position [$F(2, 22) = 377.00$, $p < .001$, $\eta_p^2 = 0.97$]. *Post hoc* Bonferroni tests indicated that there were significantly more onset errors than coda ($p < .001$) and nucleus ($p < .001$) errors, and significantly more coda errors than nucleus errors ($p < .001$). However, there was no significant interaction between nonword type and error position [$F(2, 22) = 0.05$, $p = .949$, $\eta_p^2 = 0.01$], indicating that—unlike the children—the model kept a constant relationship between onset errors and coda errors across the two nonword groups.

DISCUSSION

Nonwords set 1: CNRep

Our results for the CNRep are consistent with those reported by Gathercole and Baddeley (1989), who also observed effects of nonword type (i.e., single vs. clustered consonants) as well as nonword length. Errors increase proportionally to nonword length, and nonwords containing consonant sequences attracted more errors than their single-consonant counterparts.

However, our analysis also revealed an effect for error position. Errors occurred most often of all in the onset position, followed by errors in coda position, and with errors in the nuclear position trailing behind. The strength of the effect was dependent on the syllabification method used in the description of the results. When applying the maximal Onset method, the effect was observed only for the clustered-consonant nonwords.[13] As far as the closed syllabification method is concerned, however, the pattern is consistent across nonword types as well as nonword lengths. As we have seen (Table 3), this cannot be explained in terms of the ratio of onsets to codas in the stimuli, thus appearing to be a true bias in the children's performance.

[13] A reviewer points out that, under onset maximisation, it might be the sequence stop + liquid that is responsible for the high rate in onset errors rather than the onset position itself. Although the type of sequence may well play a part in the distribution of errors, this possibility is hard to evaluate since the manner in which the CNRep was constructed does not allow to control for differences in melodic sequences. There is, however, at least one set of nonwords in which the type of onset sequence does not account for the observed onset effect. For the 3syllable nonwords there are three stop + liquid sequences word-initially and none word-medially. Nevertheless, our data show a tendency for w-medial onset errors (34 vs 22 w-initial).

We are not aware of any NWRTs study which has identified the syllable onset as particularly prone to repetition errors, though onsets have been found to be particularly error-prone in relation to slips of the tongue (see for example Berg, 1991), and to be unlike other syllabic components in affecting naming latency (Santiago, MacKay, Palma, & Rho, 2000).[14]

Another interesting finding regards the interaction between error position and length effects under close-syllabification. A comparison of error positions revealed that the increase in the rate of errors that we observe across nonword lengths is not due to a general increase of errors in all positions, but almost entirely to a large increase in onset errors.

Nonwords set 2: Biphone frequency

As we have seen, no effect of nonword type was found for nonword set 2. *Prima facie*, this might be taken to indicate that frequency is not a reliable predictor for NWRT errors or, alternatively, that the two groups of nonwords (i.e., Low Frequency and Very Low Frequency)—being both of relatively low frequency—are not sufficiently different for a frequency effect to emerge. However, this conclusion would be too simplistic in view of the results that emerged from the analysis of error position. In fact, for the Low Frequency nonwords errors occurred most often of all in the onset position, followed by errors in coda position, and with errors in the nuclear position trailing behind. This time the pattern is consistent regardless of the syllabification method employed. This effect, however, was much smaller for the Very Low Frequency nonwords, a finding which highlights once again how a syllabic analysis can provide vital clues to the understanding of error patterns in NWRTs. In this case, such an analysis has enabled us to discover a crucial difference in children's performance which would otherwise have gone unnoticed. Although children's overall performance does not seem to be affected by the difference between low and a very low biphone frequency, their pattern of errors within the syllable is somewhat different. Whilst low frequency nonwords behave similarly to nonwords whose subparts are lexically and morphologically more familiar (i.e., akin to those from the CNRep, at least as far as closed syllabification is concerned), nonwords that consist only of very low frequency biphones have a less distinct difference between the number of onset and coda errors they attract.

[14] Treiman & Danis (1988) reported a tendency for codas to attract more errors than onsets in an experiment that tested subjects' ability to repeat lists of nonwords. However, this was not a standard NWRT, as it was concerned with lists rather than individual nonwords, and it presumably tapped on a slightly different set of abilities particularly in relation to the interaction between short-term memory and phonological performance.

In sum, our results are in line with previous research, as we observed effects of length and nonword type for the CNRep. However, we also observed a tendency towards an onset bias for the newly devised nowords. This tendency was particularly strong for the Low Frequency nonwords, and less so for the Very Lowe Frequency set. We believe that this highlights the importance of moving beyond linear approaches to nonword analysis and towards more detailed analyses which take into consideration the syllabic structure of the nonwords and their role in determining phonological complexity, and thus nonword test design. For the CNRep, when applying closed syllabification we observed that the onset bias increases as the nonword length increases, so much so that virtually the whole length effect could be explained in terms of onset errors. However, the potential presence of an onset bias in the CNRep results remains inconclusive, as it is dependent on the syllabification method used.

A comparison with the model's performance

Nonwords set 1

For the CNRep, the model produced repetition results that matched those of the children very closely. First, the model's performance produced effects of nonword type (single vs. clustered consonants) as well as nonword length. Second, the model made a higher number of errors in onset than in coda positions,[15] showing that the EPAM-VOC architecture does not simply match children's overall performance, but can actually pick up on the more fine-grained distinctions that emerge from a within-syllable analysis. Just as we saw for the children, this pattern is consistent across nonword types as well as nonword lengths. Moreover, the children's tendency to increase onset errors above all other error positions was also successfully simulated. Although this tendency was not as strong as the one we observed for the children (see Figures 6 and 7), it shows that the children's performance is at least in part generalisable from the input data.

In sum, although it is based on a fairly simple learning mechanism,[16] EPAM-VOC II can produce repetition results that are comparable to those of the children at three levels: (i) overall effects (of length and nonword type), (ii) effect for error position (onsets over codas), and (iii) a tendency to increase onset errors as nonword length increases. All three points are

[15] Vowel errors are not discussed as they were deliberately inhibited due to the fact that at this stage we are primarily concerned with the modelling of consonantal errors. We leave the question of consonantal vs. vocalic errors for further research.

[16] In fact, this simplicity could be viewed as a further strength of the EPAM-VOC model, as it has been argued that simpler models are preferable to more complex ones, as the latter are less readily falsifiable, and thus have less explanatory power (Fum et al. 2007, Myung, 2000).

matched systematically, and particularly good matches are given for points (i) and (ii).

Nonwords set 2: Biphone frequency

For nonword set 2, the model also matched the children's performance in that it did not show an effect of nonword-type but did reveal an effect of error position, with more onset than coda errors. However, unlike the children, no interaction between nonword-type and error position emerged from the model analysis, as the ratio of onset to coda errors was virtually identical across the two nonword sets.

A comparison with other modelling architectures

Alternative models of NWR performance have been discussed in the literature. Although they have all contributed to our understanding of repetition performance and the interaction between phonological knowledge and WM, none of these models can produce simulations that match children's performance on all the three levels reported above.

Hartley and Houghton (1996) outline a connectionist model that encodes syllabic as well as phonemic information in the form of built-in structural templates. The network is then presented with nonword stimuli during a training phase and is subsequently tested on the same nonwords during the recall phase. By relying on a system of weight decay and node competition, this model is able to incorporate length effects (i.e., it recalls longer nonwords less accurately than shorter nonwords) as well as phoneme substitutions.

Gupta and Tisdale (2009) adapted a "simple recurrent network" model developed by Botvinick and Plaut (2006). The model uses activation patterns to represent well-formed phonemic sequences as part of syllabic slots and input is presented to it one syllable at a time. The model simulates overall accuracy patterns as well as length effects and error types (i.e., substitution, insertion, or deletion) in NWR.

In terms of the three levels discussed previously, both of these models can simulate length effects and error types, though they cannot account for effects of nonword type or error position. Moreover, both models have a large amount of built-in information, even in domains where information is arguably available from—and therefore extractable from—the input data. For example, both models have built-in knowledge of phonotactic regularites which—almost by definition—are patterns of regularity in the input data. Furthermore, the model outlined by Gupta and Tisdale (2009) also has knowledge of what constitutes a possible syllable.

A third model, presented in Sibley, Kello, Plaut, & Elman (2008), relies on a connectionist architecture that takes in phoneme sequences, encodes them as a fixed-width patterns, and then uses the encoded pattern to produce the input sequence as output. After having been exposed to a body of input sequences the model is able to simulate word-likeness effects, as it outputs phonotactically legal nonwords more accurately than ill-formed nonwords. Just as the two previous models, however, this too falls short of the three levels simulated by EPAM-VOC II, as it only matches empirical data on overall accuracy.

A model developed in order to simulate more specific positional effects (rather than overall effects) is OSCAR, as described by Vousden, Brown and Harley (2000). This is a dynamic oscillator-based model that simulates the kind of ordering effects that make up naturalistic speech errors, also known as "slips of the tongue". Importantly, it simulates the procedure by which slips of the tongue affect syllable onsets more often that syllable codas. However, this model was not developed to simulate NWR performance, and therefore it does not simulate some of the fundamental patterns of NWRT, such as length effects or effects of word-type.

In sum, EPAM-VOC II seems to be the only model that can match empirical data on all three levels outlined previously, as well as provide a detailed explanation of how long-term phonological knowledge interacts with working memory in encoding, retrieving, and subsequently articulating phonological information.

Simulating the children's performance

Length effects

As already shown in earlier versions (Jones et al. 2007), length effects are the result of the mechanisms that make up the WM store of EPAM-VOC (see discussion on WM above). As longer words are more likely to exceed the 2,000 ms limit, they are more prone to encoding errors, which then emerge at the articulation stage. Moreover, longer nonwords will be divided into a higher number of chunks, which also places a heavier burden on the articulatory stage, raising the potential for error even further. Note that although the manner in which WM is implemented has changed, this does not affect the explanation presented in Jones et al. (2007), since such explanation is based on restrictions imposed by the 2,000 ms time limit, a feature that is still operative in the current version of the model.

Single versus clustered consonants

The fact that performance is better on single-consonant than on clustered-consonant nonwords can be explained in terms of the length effect discussed previously, as clustered consonant nonwords contain on average more

TABLE 5
Average phonemic length of nonwords in the CNRep

Nonword type	*Syllables*	*Average phonemic length*
Single consonant	2	4.6
	3	6.6
	4	8.8
Clustered consonants	2	6.8
	3	9.2
	4	12.0

phonemes than their single-consonant counterparts, as shown in Table 5 (see also Jones et al., 2007).

Due to being relatively longer, clustered nonwords are also more likely to exceed the time limit placed on WM, thus attracting further errors. For the same reason, they are also more likely to be divided into a higher number of chunks compared to single consonant nonwords, thus resulting in poorer performance.

If this explanation is on the right track, it will be the case that children make more errors on clustered nonwords because—on average—they are harder to store in WM *and* more difficult to access at the articulatory stage.

Onsets versus codas

The onset bias is achieved through a combination of factors which we will consider in turn. Firstly, it must be noted that—at the stages of learning considered in this article—the model tends to divide nonwords into chunks of two or three phonemes.[17] For example, for the CNRep nonwords, the first four phonemes of single consonant nonwords are chunked either as [C V] [C V] or [C V C] [V. . .].[18] In the first case, the second consonant is almost always followed by a schwa, due to the prosodic structure of the nonwords, whose weak syllable is almost invariably in second position. Given that schwa is by far the most frequent vowel in the model's inventory,[19] its presence will lower the chance of error for any consonant that happens to occur in the same chunk. Thus, any consonant that is chunked up together with a schwa will be less prone to error compared to a word-initial consonant, which is almost always chunked with a vowel different from schwa, and whose frequency is necessarily lower. Although this tendency is present regardless of the

[17] This might be different if the model received more input, for example if it was to simulate older children.

[18] The [. . .] indicate that the second vowel may either end up on its own or be chunked up with another phoneme.

[19] Its frequency is 269,780, which is twice as much as the second most frequent vowel.

syllabification method used, it is stronger when applying closed syllabification since, by requiring that a short stressed vowel have a coda, and given that the first vowel is often a short vowel, the second C in our [C V] [C V] sequence will often be a coda. Consequently, any advantage in favour of the second consonant is effectively an advantage in favour of coda position. A clear example of this pattern can be seen when we compare the 2-syllable nonwords "pennel" and "bannow" from the CNRep, for which the children show an onset and a coda bias respectively. For "pennel", which shows a strong onset bias, the coda is consigned to the same chunk as the schwa, hence reducing its opportunity for error since chunks involving the schwa tend to be of a high frequency. However, for "bannow", which does not contain a schwa, the coda is consigned to a relatively low frequency vowel, thus attracting more errors (coda consonants are in bold) (Table 6).[20]

Thus, the chunking pattern that results from the current stage of learning, together with the prosodic structure of the nonwords, work towards building a general bias in favour of onset errors. Importantly, this bias emerges entirely from the modelling architecture without the need for any built-in system of syllabification.

Secondly, there are also some phonotactic pressures which must be considered, as these too work in favour of an onset bias. It is a well-known cross-linguistic fact that languages have restrictions on what can appear in specific syllabic position, and English is no exception. Of particular interest to our discussion is that there exists only one phoneme which cannot appear in onset position, namely /ŋ/. On the other hand, there are four phonemes which can never fill a coda position; these are /h/, /j/, /w/, and /r/.

It follows that, all else being equal, an onset position is more likely to host a low-frequency phoneme, since 24/25 consonants (i.e., 96% of consonants) are under distributional restrictions that limit them to the onset position, while only 21/25 consonants (i.e., 84% of consonants) are restricted to coda position. A look at the average frequency of these consonants within EPAM-VOC shows that their distributional limitations tend to translate into lower

TABLE 6
Most frequent chunking pattern of syllabic components for nonwords *pennel* and *bannow* (CNRep)

Syllabification	*Chunking*	*Error bias*
pen.əl	[pe] [**nəl**]	Onset (/p/)
bæn.əʊ	[bæ] [**nəʊ**]	Coda (/n/)

[20] Only the most recurrent chunking patterns are given here.

frequencies. The average frequency for a consonant is just below 80,000 (77,779.08) with /t/ and /ʒ/ at the highest and lowest end of the spectrum, scoring 228,472 and 648 respectively. If we compare this to the frequency of those consonants that suffer from distributional restrictions, we find that they tend to be in the mid and lower half of the spectrum, apart from /r/: /j/ (20,599); /ŋ/ (33,223); /h/ (63,638); /w/ (80,669); and /r/ (126,764). Thus, phonotactic restrictions lead to an imbalance that further contributes towards an onset bias in the model's performance. This works together with the chunking mechanism and the prosodic structure of the nonwords.

These factors lead to an onset bias in the model's performance, and subsequently to the successful simulation of children's performance, showing that the syllabic positions of consonantal errors can be accounted for by a mechanism that relies solely on frequency information and does not "know" about syllables.

Notably, this explanation makes a precise prediction with regard to developmental factors: as the model extends its LTK, it will gain the ability to divide nonwords into progressively larger chunks, containing four or perhaps even five phonemes. Thus, the schwa will no longer be contributing its high frequency solely in favour of the coda consonant, as the word-initial CV sequence may also end up in the same chunk, which—for single consonant nonwords—would take the form [C V C V]. In these cases, the high frequency of the schwa no longer brings an advantage to one specific consonant, as the probability of errors is left entirely to the frequency of each individual consonant. Thus, the prediction that follows is that—for older children—the onset bias will be much less pronounced than for younger children. At the moment, however, whether or not this is the case remains an empirical question, since—to the best of our knowledge—no other research has been carried out on the interaction between onset and coda errors in NWRTs.

Onset errors increase as length effect

As shown in our results section, longer nonwords attract a much higher number of onset errors than shorter ones, for both children and model. Interestingly, this increase in onset errors is not entirely explicable in terms of the relative increase in the number of onset consonants. Although longer nonwords tend to include a higher ratio of onsets, the increase in onset errors significantly exceeds this factor (see results section), thus calling for an explanation.

The explanation that emerges from the EPAM-VOC II performance has two aspects to it, both of which are in relation to frequency considerations. As discussed in the section on articulating an input sequence, the model's articulatory performance is dependent on the chunking procedure that takes place at the parsing stage the lower the weight of a chunk, the higher the chance that it will attract an error. As discussed above (section on onsets versus codas), coda consonants tend to be chunked up with the schwa, thus becoming part of relatively high frequency chunks. When we look across different nonword lengths, it becomes apparent that this is part of a more general tendency that involves coda consonants being regularly chunked up with phonemes at the high end of the frequency spectrum, the schwa being but the maximum exemplar.

To illustrate this, we have taken a sample of the most common chunking patterns for each nonword and divided each chunk according to whether it contained an onset or a coda.[21] These chunks were then subsequently divided according to the weight range that they fell into. Weight was calculated following the procedure explained in the section on articulating an input sequence. The weight range was divided into four equal parts with 10,000 set as the maximum as this is the frequency beyond which no error is made. The descriptive statistics are given in Table 7.

As Table 7 shows, for the 2-syllable nonwords, onsets and codas seem to be almost evenly distributed, though a slight accuracy bias is evident in favour of codas, 20% of which end up in the mid-to-high frequency range. However, as nonword length increases, the distribution of onset and coda chunks in the lower frequency ranges diverges: many more onsets than codas are low frequency. Hence we would expect to see a sharp increase in the onset errors relative to coda errors as nonword length increases. This is in fact what is seen in both the child and model data.

In sum, the error bias that sees onsets being targeted more often than codas is due to a distributional effect emerging from the interaction between the chunking mechanism employed by EPAM-VOC *and* the phonemic make-up of the experimental stimuli. On the one hand, the combination of these two factors creates a general bias whereby chunks containing onset

TABLE 7
Low-frequency chunks in EPAM-VOC II containing either onsets or codas, for nonwords in the CNRep

Frequency range	*2-syllable chunks*		*3-syllable chunks*		*4-syllable chunks*	
	Onset (%)	*Coda (%)*	*Onset (%)*	*Coda (%)*	*Onset (%)*	*Coda (%)*
0–2,500	25.00	20.00	66.67	14.29	84.21	60.00
2,500–5,000	66.67	60.00	33.33	71.43	15.79	40.00
5,000–7,500	8.33	20.00	0.00	14.29	0.00	0.00
7,500–10,000	0.00	0.00	0.00	0.00	0.00	0.00

[21] Chunks that contained both elements were excluded from the count.

consonants tend to be relatively lower in frequency, as reported above. On the other hand, the phonemic structure of the experimental stimuli provided by the CNRep is such that the bias is more marked in longer nonwords than in shorter ones, thus resulting in a sharp increase in onset errors.

Summary

We have presented a computational model of phonological learning, EPAM-VOC, that is able to simulate not only the basic patterns of error across two nonword repetition tests, but also the pattern of errors within specific syllabic positions. We have shown that, for the nonwords used in the two tests outlined, there is a clear onset bias for errors in the children. This effect is also simulated in EPAM-VOC. Furthermore, the model also presents an explanation for a bias in onset errors: onsets tend to be contained in chunks that are of a low frequency, whereas codas tend to be contained in chunks that are of a relatively higher frequency. Of course, the onset bias may well be influenced by the particular nonword sets used, and it will therefore be interesting to see if EPAM-VOC is also able to simulate children's repetitions that show a coda bias for errors. This is one of the next steps for the model.

While EPAM-VOC has provided good simulations of the error data in the studies described, the model is also still some way from providing a full account of phonological learning. For example, errors themselves are not based on phoneme similarity or lexical influences. This is the next stage of the model's development, whereby an account of the actual types of errors that occur in the children's data can be examined.

In summary, we have provided a detailed account of the onset and coda errors that are seen in NWRTs, illustrating an onset bias in children's errors, and how this bias increases with nonword length. Both of these phenomena are also simulated in our model of phonological acquisition. EPAM-VOC provides a good account of nonword repetition that illustrates how reasonable assumptions concerning working memory, long-term learning, and the interaction between the two can simulate a detailed set of children's error data without the need for any built-in knowledge about syllables or their components.

Manuscript received 8 December 2009
Revised manuscript received 15 April 2011
First published online 6 January 2012

REFERENCES

Baddeley, A. D. (2000a). The magic number and the episodic buffer. *Behavioral and Brain Sciences, 24*, 117–118.

Baddeley, A. D. (2000b). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47–90). New York, NY: Academic Press.

Baddeley, A. D., Thompson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour, 14*, 575–589.

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development, 17*, 1265–1282.

Berg, T. (1991). Phonological processing in a syllable-timed language with pre-final stress: Evidence from Spanish speech error data. *Language and Cognitive Processes, 6*, 265–301.

Bernhardt, B. H., & Stemberger, J. P. (1998). *Handbook of phonological development from the perspective of constraint-based non-linear phonology.* San Diego, CA: Academic Press.

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry, 37*, 391–403.

Blevins, J. (1995). The syllable in phonological theory. In John A. Goldsmith (Ed.) *The Handbook of Phonological Theory* (pp. 206–244). Cambridge, MA and Oxford: Blackwell.

Botting, N., & Conti-Ramsden, G. (2001). Nonword repetition and language development in children with language impairments. *International Journal of Language and Communication Disorders, 36*, 421–432.

Botvinick, M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113*, 201–233.

Bowey, J. A. (1996). On the association between phonological memory and receptive vocabulary in five-year-olds. *Journal of Experimental Child Psychology, 63*, 44–78.

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics, 22*, 441–469.

Cantor, J., & Engle, R. W. (1993). Working-memory capacity as long-term memory activation: An individual-differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1101–1114.

Catts, H. W., & Kamhi, A. G. (1984). Simplification of /s/ + stop consonant clusters: a developmental perspective. *Journal of Speech and Hearing Research, 27*, 556–561.

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions and evidence. *Language and Cognitive Processes, 16*, 113–142.

Clark, E. (2003). *First language acquisition*. Cambridge: Cambridge University Press.

Conti-Ramsden, G., & Hesketh, A. (2003). Risk markers for SLI: A study of young language-learning children. *International Journal of Language and Communication Disorders, 38*, 251–263.

Cowan, N. (1997). The development of working memory. In Cowan N (Ed.), *The development of memory in childhood* (pp. 163–200). Hove, UK: Psychology Press.

Croker, S., Pine, J. M., & Gobet, F. (2003). Modelling children's negation errors using probabilistic learning in MOSAIC. In F. Detje, D. Dörner & H. Schaub (Eds.), *Proceedings of the Fifth International Conference on Cognitive Modeling* (pp. 69–74). Bamberg: Universitäts-Verlag.

Demuth, K. (2001). A prosodic approach to filler syllables. *Journal of Child Language, 28*, 246–249.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech, 49*, 137–174.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic evaluation of articulation and phonology.* London: Psychological Corporation.

Dollaghan, C. A., Biber, M. E., & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics, 16*, 211–222.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1136–1146.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale* (2nd ed.). Windsor, UK:NFER-Nelson.

Edwards, J., & Lahey, M. (1996). Auditory lexical decisions of children with specific language impairment. *Journal of Speech and Hearing Research, 39*, 1263–1273.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition, 18*, 91–126.

Freudenthal, D., Pine, J., & Gobet, F. (2006). Modelling the development of children's use of optional infinitives in English and Dutch using MOSAIC. *Cognitive Science, 30*, 277–310.

Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science, 8*, 305–336.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language, 42*, 481–496.

Fum, D., Del Missier, F., & Stocco, A. (2007). The cognitive modelling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research, 8*, 135–142.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*, 83–94.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200–213.

Gathercole, S. E., & Baddeley, A. D. (1996). *The Children's Test of Nonword Repetition*. London, UK: The Psychological Corporation.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory, 2*, 103–127.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1991). The influence of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12*, 349–367.

Giegerich, H. J. (1992). *English phonology: An introduction*. Cambridge: Cambridge University Press.

Gimson, A. C. (1980). *An introduction to the pronunciation of English* (3rd ed.). London: Edward Arnold.

Goad, H., & Brannen, K. (2003). Phonetic evidence for phonological structure in syllabification. In Joost van de Weijer, Vincent J. van Heuven & Harry van der Hulst (Eds.), *The phonological spectrum* (pp. 3–30). Amsterdam: John Benjamins.

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science, 24,* 651–682.

Goldsmith, J. (1992). Local Modeling in Phonology. In Steven Davis (Ed.), *Connectionism: Theory and Practice* (pp. 229–246). Oxford: Oxford University Press.

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory, 13*, 318–324.

Gupta P., & Tisdale J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language, 61,* 481–502.

Harris, J. (1994). *English sound structure*. Oxford: Blackwell.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35*, 1–31.

Ho, S.-H., & Lai, N.-C. (1999). Naming-speed deficits and phonological memory deficits in Chinese developmental dyslexia. *Learning and Individual Differences, 11*, 173–186.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 385–401.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, M., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology, 23*, 1217–1232.

Jakobson, R. (1941/1968). *Child language, aphasia, and phonological universals* (A.R. Keiler, Trans.). The Hague: Mouton.

Joanisse, M. F., & Seidenberg, M. S. (1998). Specific Language Impairment in children: An impairment in grammar or processing? *Trends in Cognitive Sciences, 2*, 240–246.

Johnson, C. E., & Wilson, I. L. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *International Journal of Bilingualism, 6*, 271–289.

Jones, D. (1997). In P. Roach & J. Hartman (Eds.), *English Pronouncing Dictionary* (15th ed.). Cambridge: Cambridge University Press.

Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Developmental Science, 10,* 853–873.

Jones, G., Gobet, F., & Pine, J. M. (2008). Computer simulations of developmental change: The contributions of working memory capacity and long-term knowledge. *Cognitive Science, 32*, 1148–1176.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*, 630–645.

Kehoe, M., & Stoel-Gammon, C. (2001). Development of syllable structure in English-speaking children with particular reference to rhymes. *Journal of Child Language, 28*, 393–432.

Kirk, C., & Demuth, K. (2005). Asymmetries in the acquisition of word-initial and word-final consonant clusters. *Journal of Child Language, 32*, 709–734.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language, 7*, 41–74.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Marshall, C. R., Ebbels, S., Harris, J., & van der Lely, H. K. J. (2002). Investigating the impact of prosodic complexity on the speech of children with specific language impairment. *University College London Working Papers in Linguistics, 14*, 43–68.

Marshall, C. R., & van der Lely, H. K. J. (2009). Effects of word position and stress on onset cluster production: Evidence from typical development, SLI and dyslexia. *Language, 85*, 39–57.

McLeod, S., van Doorn, J, & Reed, V. A. (2001). Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology, 10*, 99–110.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*, 3–19.

Metsala J. L., &. Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Hillsdale: Erlbaum.

Munson, B., Edwards, J., & Beckman, M. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research, 48*, 61–78.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*, 1033–1047.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory & Language, 53*, 225–237.

Prince, A. & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in generative grammar* (Technical Report). Boulder: Rutgers University and University of Colorado. Blackwell (Rev. ed. 2004).

Pulgram, E. (1970). *Syllable, word, nexus, cursus*. The Netherlands: The Hague, Mouton.

Radeborg, K., Barthelom, E., Sjöberg, M., & Sahlén, B. (2006). A Swedish nonword repetition test for preschool children, *Scandinavian Journal of Psychology, 47*, 187–192.

Roach, P. (2000). *English phonetics and phonology: A practical course* (3rd ed.). Cambridge: Cambridge University Press.

Santiago, J., MacKay, D. G., Palma, A., & Rho C. (2000). Sequential activation processes in producing words and syllables: Evidence from picture naming. *Language and Cognitive Processes, 15*, 4–44.

Selkirk, E. (1982). Syllables. In H. van der Hulst & N. Smith (Eds.), *The Structure of Phonological Representations* (vol.2, pp. 337–384). Dordrecht: Foris.

Sibley, D. E., Kello, C. T., Plaut, D., & Elman, J. L. (2008). Large-scale modelling of wordform learning and representation. *Cognitive Science, 32*, 741–754.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*, 779–798.

Snowling, M. J. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research, 43*, 219–234.

Stoel-Gammon, C. (1985). Phonetic inventories, 15–24 months: A longitudinal study. *Journal of Speech and Hearing Research*, *28*, 505–512.

Stokes, S. F., Wong, A. M-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of Cantonese. *Journal of Speech, Language, and Hearing Research, 49*, 219–236.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of Verb-Argument structure: An alternative account. *Journal of Child Language, 28*, 127–152.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*, 325–329.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*, 47–62.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology 41*, 101–175.

Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: its role in the development of phoneme awareness and early reading ability, *Reading and Writing, 16*, 5–20.

Wechsler, D. (2002). *The wechsler preschool and primary scale of intelligence*. London: Psychological Corporation.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypothesis. *Memory and Cognition, 13*, 193–201.

APPENDIX 1
Syllabification according to Maximal Onset[22]

CNRep	*2-syllables*	*3-syllables*	*4-syllables*
Single consonants	/pe.nəl/	/dɒ.pə.leɪt/	/wʊ.gə.læ.mɪk/
	/bæ.ləp/	/bæ.nɪ.fə/	/fe.nə.raɪ.zə/
	/ruː.bɪd/	/bæ.rɪ.zən/	/kə.mɪː.sə.teɪt/
	/dɪ.lə/	/kɒ.mə.riːn/	/lɒ.də.neɪ.pɪʃ/
	/bæ.nəʊ/	/θɪ.kə.riː/	/pə.ne.rɪ.fʊl/

CNRep	*2-syllables*	*3-syllables*	*4-syllables*
Clustered consonants	/hæm.pənt/	/glɪ.stə.rɪŋg/	/kən.træm.pə.nɪst/
	/glɪ.stəʊ/	/fre.skə.vent/	/pə.plɪ.stə.rɒŋk/
	/slæ.dɪŋg/	/trʌm.pə.tiːn/	/blɒn.tə.steɪ.pɪŋg/
	/tæ.flɪst/	/bræ.stə.rə/	/stɒ.pə.græ.tɪk/
	/prɪn.dəl/	/skɪ.tɪ.kʌlt/	/em.plɪ.fɔː.vənt/

CNRep	*Low-frequency*	*V-Low-frequency*
New set	/pɒn.tə.kʊː/	/ɒ.stə.vɪs/
	/bɪp.fə.næk/	/tʌl.bə.rɪk/
	/prɒ.nə.diː/	/zɪm.bə.læt/
	/læt.mə.nɒz/	/jɒ.glə.mɪn/
	/mɒk.sə.fɪkt/	/trəʊ.bək.nɪs/
	/bɪ.stə.lɔːθ/	/æp.kə.sʌg/
	/spem.pə.teɪf/	/weg.nə.tɛːk/
	/wɪ.fəl.dɒp/	/væl.bə.rɪst/

[22] Please note that, unlike most dialects, the English spoken in Nottingham and the Midlands still realises velar nasal-plosive sequences word-finally, hence the /ŋg/ transcription in examples such as /glɪsterɪŋg/, where most English speakers would produce /ŋ/.

APPENDIX 2
Syllabification according to closed syllabification

CNRep	*2-syllables*	*3-syllables*	*4-syllables*
Single consonants	/pen.əl/	/dɒp.ə.leɪt/	/wʊg.ə.læm.ɪk/
	/bæl.əp/	/bæn.ɪ.fə/	/fen.ə.raɪ.zə/
	/ruː.bɪd/	/bæ.rɪ.zən/	/kə.mɪː.sə.teɪt/
	/dɪl.ə/	/kɒm.ə.riːn/	/lɒd.ə.neɪ.pɪʃ/
	/bæn.əʊ/	/θɪk.ə.riː/	/pə.ne.rɪ.fʊl/

CNRep	*2-syllables*	*3-syllables*	*4-syllables*
Clustered consonants	/hæm.pənt/	/glɪs.tə.rɪŋg/	/kən.træm.pə.nɪst/
	/glɪs.təʊ/	/fres.kə.vent/	/pə.plɪs.tə.rɒŋk/
	/slæd.ɪŋg/	/trʌm.pə.tiːn/	/blɒn.tə.steɪ.pɪŋg/
	/tæf.lɪst/	/bræs.tə.rə/	/stɒp.ə.græt.ɪk/
	/prɪn.dəl/	/skɪt.ɪ.kʌlt/	/em.plɪ.fɔː.vənt/

CNRep	*Low-frequency*	*V-Low-frequency*
New set	/pɒn.tə.kʊː/	/ɒs.tə.vɪs/
	/bɪp.fə.næk/	/tʌl.bə.rɪk/
	/prɒn.ə.diː/	/zɪm.bə.læt/
	/læt.mə.nɒz/	/jɒg.lə.mɪn/
	/mɒk.sə.fɪkt/	/trəʊ.bək.nɪs/
	/bɪs.tə.lɔːθ/	/æp.kə.sʌg/
	/spem.pə.teɪf/	/weg.nə.tɛːk/
	/wɪf.əl.dɒp/	/væl.bə.rɪst/