THE STATUS OF VARIABLE PHONETIC OUTPUT IN EARLY SPEECH

Elinor Payne[a], *Eftychia Eftychiou*[a], *Brechtje Post*[b], *Lluïsa Astruc*[b,c], *Pilar Prieto*[d,e] & *Maria Vanrell*[d]

[a]Oxford University, UK; [b]Cambridge University, UK; [c]Open University, UK; [d]Universitat Pompeu Fabra, Spain; [e]ICREA

elinor.payne@phon.ox.ac.uk; eftychia.eftychiou@phon.ox.ac.uk; bmbp2@cam.ac.uk; mla28@cam.ac.uk; pilar.prieto@upf.edu; MariaDelMar.Vanrell@uab.cat

ABSTRACT

This paper presents an analysis of variable outputs in the speech of English-, Spanish- and Catalan-speaking 2 year-olds, to evaluate the evidence for multiple targets. Temporal analysis shows syllable duration is determined by actual segmental composition, compatible with either multiple targets or a process of categorical deletion. However, spectral analysis suggests deletions may be 'gradient', with the possibility of a single target subject to variable implementation. This mismatch between (pervasively categorical) temporal and (optionally gradient) articulatory behaviour suggests variable output is not due to a multiple-entry lexicon as such, but to a shifting association between articulation and timing tiers, possibly as a result of competing structural templates.

Keywords: development, lexical representation, variable output, duration, articulation

1. INTRODUCTION

As is well-documented, young children often reduce words in regular ways to fit their production capacities [3, 5, 6], e.g. reducing consonant clusters (producing /snəʊ/ as [nəʊ]). Furthermore, even once phonetic mastery is acquired, children often show 'inertia' [3] in up-dating these forms, and go through a period of producing variable outputs, suggesting they retain earlier structures for some time [7]. [3] assert that in acquiring the adult system, each child "must construct his/her own version [..] for word production, for there is no way to "internalize" the adult system without this". The questions arise: what is the status of these variable outputs and how is the child's transitional system structured? This study hopes to shed light on these questions by examining cross-linguistic speech productions during the period of variable outputs. In particular, it examines the extent to which variable outputs can be described as categorically different (in temporal and articulatory terms), and hence potentially correspond to multiple targets in the developing lexicon. E.g. when a child produces [pa] for target /pla/, are there temporal or spectral traces of the 'missing' segment, or could [pa] plausibly correspond to a competing /pa/ target?

2. DATA

The data came from a corpus of recordings of semi-structured dialogues between children and their mothers, elicited through an interactive computer game based on short animated clips. These showed simple, everyday scenes easily describable in words highly familiar to the children. E.g. one showed a girl blowing soap bubbles, another showed a boy playing a trumpet. The mother then asked her child to describe what was happening in each clip, then praise the child for getting it right, and repeat what the child had said. A typical dialogue went thus:

(1) Mother: "*What's the little girl doing?"*
Child: *"(She's) blowing bubbles!"*
Mother: "*Yes! She's blowing bubbles!*"

There were 23 utterances per dialogue. The corpus contains speech data from children aged 2-, 4- and 6-years (+/- 2 months), for English, Spanish and Catalan. There were 3 child-mother pairs for each age and language (27 dialogues in total). The speech files were segmented into syllables in PRAAT (see [4]), and labelled for syllable type, and for prosodic information (to control for prosodically conditioned duration variation):

- Realized structure: [CV], [CVC], [CCV], etc;
- Presumed target structure: /CV/, /CVC/, /CCV/;
- Stress: unstressed; stressed; nuclear accented;
- Proximity to boundary: initial, medial, final.

2.1. Selection of data

Syllable information (structure type, prosody, duration) was extracted by script, and a subset of

the data analysed. Since deletions were very sporadic or non-existent in the speech of 4 and 6 year-olds, only the speech data from 2 year-olds (9 dialogues in total) were analysed. As is well known, child outputs can present various kinds of mismatch with adult targets, including epenthesis and substitution [5]. Since the aim was to investigate possible traces of target structures in cases of apparent deletion, we only selected syllables where apparent deletion of one or more segments had occurred, and those realized with the target syllable structure as a control. The focus was on syllable structure rather than specific segment type: e.g. /kat/ realized as [tat] was deemed on target; /kat/ realized as [ka] was not.

Thus in the first instance the main kind of evidence under investigation was durational: any trace of extra duration in a realisation like [ka] would suggest the existence (and persistence) of abstract schematic timing properties from a single lexical representation /kat/ that is varyingly implemented. The investigation was done through two types of comparison: i) *different realisations of the same target structure*; ii) *surface-identical realizations of different target structures*. If abstract timing properties from a single target exist, and persist even during 'deletion' of segmental material, this should manifest itself in a difference in duration in i), and no difference in ii). Although the corpus contains several different syllable structure types (both target and realized), only a sub-set occur with sufficient frequency for statistical analysis (summarised below). Coda deletions were more common in unstressed syllables, while onset reductions/deletions were more common in stressed syllables.

2.1.1. Different realizations of same target

The only target syllable structure types occurrng in sufficient numbers for analysis were / CV / and / CVC /. The incidence of these as a function of their realised syllable structure is given in Table 1. In general, onset deletions were spread relatively evenly across the three languages, while coda deletions were much more common in Spanish.

Table 1: Number of tokens analysed for different realizations of same target.

Realisation / target	[V]	[CV]	[VC]	[CVC]
/ CV /	35	329	n/a	n/a
/ CVC /	5	39	15	141

2.1.2. Same realization of different targets

The only realised syllable structure types occurring with sufficient frequency were [V], [CV] and [CVC]. The incidence of these as a function of their presumed target is given in Table 2. Again onset deletions were spread relatively evenly across the three languages, while coda deletions were much more common in Spanish (marked*).

Table 2: Numbers of tokens analysed for same realization of different target.

Realisation / target	[V]	[CV]	[CVC]
/ V /	42	-	-
/ CV /	15	329	-
/ (C)CCV /	-	27	-
/ CVC /	5*	39*	141
/ CCVC /	-	15*	27

2.1.3. Statistical analysis

A series of univariate anovas was carried out, one for each target or realization type, with syllable duration as the dependent variable, and 4 independent variables: Language, Stress, Phrase Boundary (PB), Deletion Type (DT). In each case, there were 3 levels of Language (English, Spanish, Catalan); 3 levels of Stress (unstressed, stressed, nuclear accented); and 3 levels of PB (initial, medial, final).

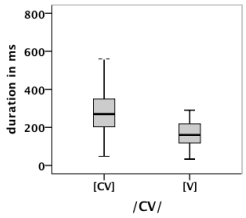
3. RESULTS OF DURATIONAL ANALYSIS

3.1. Different realisations of same target

3.1.1. /CV/ Target

DT had a main effect on duration ((333, 1) F(5.149, $p<0.05$): /CV/ syllables that had undergone onset deletion > [V] were significantly shorter than those retaining their onset (see Fig. 1). There were also main effects of Language (333, 2) F(3.185), $p<0.05$; Stress (333, 2) F(4.014), $p<0.05$; and PB (333, 2) F(8.738), $p<0.005$; and a significant interaction between Lang. and PB (333, 4), F(7.487), $p<0.005$.

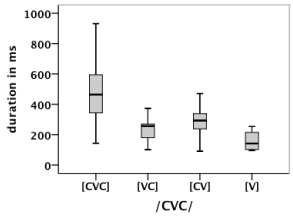
Figure 1: Mean duration of target /CV/ realized as [CV] & [V] across 3 languages.



3.1.2. /CVC/ Target

DT had a main effect on duration ((158, 3), F(5.235), $p<0.05$). /CVC/ syllables that had undergone apparent deletion > [VC], [CV] or [V] were significantly shorter (see Fig. 2). There was also a main effect of PB (158, 2), F(13.395), $p<0.005$.

Figure 2: Mean duration of target /CVC/ realized as [CVC], [VC], [CV] and [V].



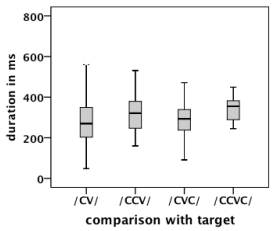
These two sets of results show duration is determined by actual segmental composition, with no evidence of abstract temporal properties being projected onto phonetic interpretation.

3.2. 'Same' realizations of different targets

3.2.1. Realisations of [CV]

Main effects on duration were found for Stress (366, 2) F(5.301) p=0.5; and PB (366, 2), F(29.852), p<0.005, with a significant interaction between Language and PB: (366, 2) F(5.887) p<0.005. However, there was no main effect of DT, suggesting no difference between realizations of [CV] as a function of their 'target type' (see Fig. 3).

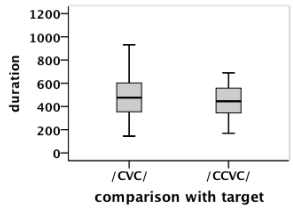
Figure 3: Mean duration of [CV] for different targets, across 3 languages.



3.2.2. Realisations of [CVC]

Main effects were found for Lang (140, 2), F(3.905), p<0.05, and PB: (140, 2), F(5.729), p<0.05. As with [CV], no main effect was found for DT (Fig 4).

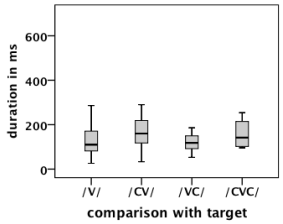
Figure 4: Mean duration of [CVC] for different targets, across 3 languages.



3.2.3. Realisations of [V]

Main effects were found for Language: (49, 2) F(7.369) p<0.005; Stress: (49, 2) F(15.182) p=0.005; and PB: (49, 2), F(26.703), p<0.005. Once more, no main effect was found for DT (see Fig. 5).

Figure 5: Mean duration of [V] for different targets, across 3 languages.



3.3. Summary of durational findings

Firstly, the durational findings reinforce/extend a number of findings from previous studies. It is known, e.g. [1] that very young children are perceptually sensitive to the durational cures of lexical stress and prosodic phrasing. Here we present evidence that by 2 years children are already using these cues in their own first connected speech: syllable duration was significantly longer for all three languages when close to a phrase boundary, and when stressed or accented. Also, the higher frequency of coda 'deletions' in Spanish confirms previous findings that coda acquisition occurs later in languages in which codas are not very frequent, e.g. Spanish, than in those in which they are more common, e.g. Catalan and English [2].

Secondly, and most importantly for the present study, the durational findings reveal that, prosodically conditioned variability aside, the duration of syllables in early child productions is essentially determined by surface segmental composition, and not by any abstract schematic properties of the presumed target. In other words, when 2 year-olds produce e.g. [CV], the resulting duration is determined by this structure, and not influenced by the target structure (which may be different, e.g. /CVC/). This would indicate either the existence of multiple targets, or a single target undergoing a process of categorical deletion.

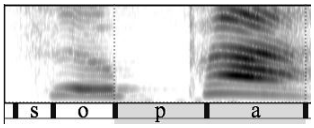
4. QUALITATIVE ANALYSIS

Close spectrographic inspection was made of tokens auditorily classed as having undergone deletion. Greater attention was given to cluster reductions, where effects could be manifested on remaining consonants and thus potentially be more overt. For space reasons, only a summary is presented here.

The findings were varied and point to a possibility for both categorical and gradient deletions. A possible example of 'clean' (categorical) deletions is given in Fig. 6, which shows a rendition of Spanish /soplar/ as [sopa] (/CCVC/ > [CV]). Though subtle traces of target

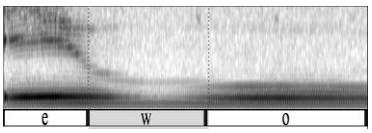
segments may remain, there is no *overt* evidence of either the lateral or rhotic, and so an interpretation of categorical deletion is at least plausible.

Figure 6: Spectrogram of Spanish 2-year-old saying [sopa] for /soplar/.



However, in several cases, a change of place and/or manner in the remaining segment is observable, suggesting a more gradient process of deletion. Fig. 7 shows Spanish target syllable /glo/ articulated as [wo]: here we see a coalescence of properties from both segments, with the resulting segment combining the sonorant nature of the target liquid, and the tongue back raising for the target velar stop. Arguably such deletions cannot be modeled as categorical, at least not on the articulatory tier.

Figure 7: Spectrogram of Spanish 2-year-old saying [wo] for /glo/.



5. DISCUSSION

The temporal findings suggest a categorical difference between realizations of e.g. [CV] and [CVC] for /CVC/, and this is compatible with either a multiple representation approach (with e.g. both /CV/ and /CVC/ versions), or a single representation subject to an optional but categorical process of deletion. However, close spectrographic inspection reveals, at least in some cases, the possibility of a gradient 'deletion' (perhaps better to be modeled as gestural overlap resulting in a blended articulation).

The apparent mismatch, in some tokens, between temporal and articulatory evidence is difficult to explain. One explanation might be that children are conforming to a structural template [7], which can be achieved either through 'blunt', segmental deletion (cf. where timing and articulation match), or through overlapping gestures (cf. where articulatory deletion is non-categorical). However, the pervasiveness of this template is questionable since children are also producing [CVC] at this stage. What we may be seeing is the co-existence of *multiple templates* (the selection between which determines temporal properties), but with a *single representation* with respect to segmental material (which may or may not be fully implemented). Hence a child can produce i) [glo]; ii) [go] or [lo]; iii) [wo], all for target /glo/ combined with either the template CVC (to produce i) or CV (to produce ii and iii).

6. CONCLUSION AND NEXT STEPS

This paper has presented an analysis of variable outputs in early speech and the evidence for multiple targets. An apparent mismatch in some cases between (pervasively categorical) temporal and (optionally gradient) articulatory findings suggests variable output is not due to a complex, multiple-entry lexicon, but to a complex and shifting association between articulation and timing tiers, possibly in conjunction with multiple structural templates. This is compatible with the hypothesis of 'inertia' in the retention of previous structures even when phonetic mastery has been acquired [3]. On this interpretation, there would be no need to posit explicit processes of deletion to explain variable output; instead, variable output would result from unstable associations between competing structural templates and phonetically rich representations. In time these associations settle (in favour of the template best matching the representation), and outputs cease to vary in syllable structure.

This raises questions about the substance of early lexical representations and the role of structural templates. These questions will be pursued in a follow-up perceptual study examining how children *respond* to variable *inputs*.

7. REFERENCES

[1] Jusczyk, P.W., Luce, P.A., Charles-Luce, J. 1994. Infants' sensitivity to phonotactic patterns in the native language. *J. of Mem. and Lang*. 630-645.

[2] Lleó, C. 2003. Prosodic licensing of codas in the acquisition of Spanish. *Probus* 15, 257-281.

[3] Menn, L., Stoel-Gammon, C. 2004. Phonological development. In: Fletcher, P., MacWhinney, B. (eds.), *The Handbook of Child Language*. Oxford: Blackwells.

[4] Payne, E., Post, B., Astruc, L., Prieto, P., Vanrell, M.M. Forthcoming. Measuring child rhythm. *Lang. and Speech.*

[5] Vihman, M. 1992. Early syllables and the construction of phonology. In Ferguson, C.A., Menn, L., Stoel-Gammon, C. (eds.), *Phonological Development: Models, Research, Implications.* Timonium, MD: York Press.

[6] Vihman, M, 1993. The construction of a phonological system. In de Boysson-Bardies, B., de Schonen, S., Jusczyk, P., MacNeilage, P., Morton, J. (eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. Dordrecht: Kluwer Academic.

[7] Vihman, M. 1993. Variable paths to early word production. *Journal of Phonetics* 21, 61-82.