The dynamical landscape: phonological acquisition and the phonology–phonetics link*

Beata Łukaszewicz (ORCID)
University of Warsaw

During acquisition children internalise adult-based phonological patterns and alternately adopt and discard child-specific patterns reflecting their unskilled production. The child-specific patterns are often assumed to be low-level phonetic effects, and so, in a classical modular feedforward grammar, they should not interfere with the higher-level adult-based phonology. This paper reports an interaction in which the application of a categorical adult-based process (Voice Assimilation) is conditioned by a gradient child-specific process (fricative devoicing). Acoustic analyses of longitudinal data from a Polish-speaking child reveal variable Voice Assimilation effects in target voiced fricative–stop/stop–fricative clusters (voicing and devoicing), correlated with the extent of voicing in fricatives in non-assimilatory contexts. I analyse this phonology–phonetics trade-off by appealing to symbol-like dynamical representations, expressed in the language of non-linear mathematics. Such representations offer a non-derivational link between the qualitative and quantitative aspects of speech. Variability ensues as a natural consequence of grammar change.

1 Introduction

In the process of acquisition, children internalise phonological patterns on the basis of perceived regularities in the adult input. Also, and to some

extent in parallel, they alternately adopt and discard child-specific patterns, reflecting their immature motor skills.[1]

While the acquisition of adult-based phonological patterns has been assumed to be within the realm of phonological competence, child-specific patterns have been viewed as performance-related (cognitively controlled or biomechanical) effects, situated within or outside of the scope of phonological computation depending on the approach (see discussions in Smith 1973, 2010, Kiparsky & Menn 1977, Hayes 2004, Hale & Reiss 2008). The phonological *vs.* phonetic status of these patterns has been a protracted controversy. It suffices to note that, at its extremes, the generativist stance has produced two contradictory views on the status of such patterns: at one end they have been considered to be by-products of the unmarked initial stage, the Universal Grammar (Smolensky 1996); at the opposite end as performance effects of no linguistic significance (Hale & Reiss 2008). Under the latter view, child-specific production patterns are regarded as merely physiological acts, not having phonetic representations – analogously to Sapir's (1925) example of the 'candle-blowing sound', which is superficially similar to a voiceless [ʍ], but is not a speech sound (Hale & Reiss 2008: 66–67).

More commonly, child-specific patterns are viewed as emergent grammaticalised solutions to phonetic problems (e.g. Kiparsky & Menn 1977, Menn 1978, Hayes 2004). A cophonological organisation between the adult-based comprehension grammar and the child-specific production grammar has been assumed by some writers (e.g. Kiparsky & Menn 1977, Hayes 2004). The two-component organisation, schematised in Fig. 1, is analogous to standard modular 'feed-forward' models of phonetic implementation developed for adult speech (e.g. Keating 1988; see also discussion in Pierrehumbert 2002). This presupposes a strict division of labour and ordering between the two components. In the feedforward model, the adult-based phonological component within the child's system acts independently of child-specific production patterns; child-specific patterns target the output of the adult-based component, i.e. the adult surface form. In Fig. 1, and throughout this paper, target forms (i.e. adult surface representations) are given in / / and the child's pronunciations in [ ]. Adult underlying representations, where relevant, appear in // //.

---

[1] The term 'child-specific' refers to a pattern that has no parallel in the ambient language, and as such needs to be discarded in the course of development. The focus here is on child-specific patterns typical of phonological organisation of the third year of life. Overall, beyond the first 18 months of age, the child speech output is likely to exhibit segment-like organisation resembling adult speech (cf. the 'phoneme plane' of Werker & Curtin 2005), making it possible to draw correspondences between child and adult forms in terms of phone-sized units; it is also the period when children begin to concatenate words into larger utterances; see the discussion in MacNeilage (1997: 326ff).

A	‖ underlying representations ‖ hypothesised by the child

A → B
adult-based

$\downarrow$

B	/ phonetic representations / perceived by the child ($\approx$ adult surface forms)

B → C
child-specific

$\downarrow$

C	[child's pronunciations]
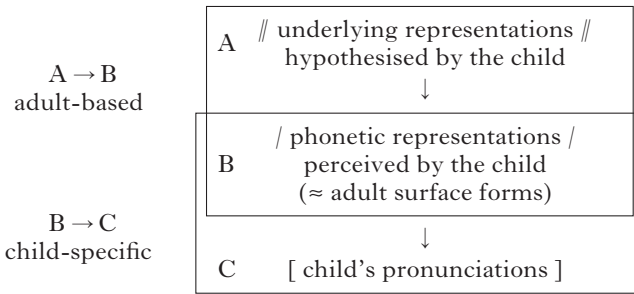
Figure 1
A feedforward organisation of the phonological system
in the child (cf. Kiparsky & Menn 1977: 61).

The constraint-based framework of Optimality Theory (OT; Prince & Smolensky 1993, McCarthy & Prince 1995) allows for integrating adult-based generalisations and child-specific effects within the child's production grammar. The child-specific component (B → C in Fig. 1), which arises from the child's proprioception and auditory perception of his or her own vocal behaviour as compared with the perceived vocal behaviour of adults, is then hypothesised to operate in terms of the same phonological constraints as his or her comprehension (using the A → B system) (Hayes 2004: 196). This is in accordance with the findings that child-specific patterns in many respects resemble adult phonological phenomena; for example, they can be interrelated, and form 'conspiracies' (e.g. Smith 1973, Menn 1978, Pater & Barlow 2003, Łukaszewicz 2007).

The separation of the comprehension grammar (A → B) in Fig. 1 is corroborated by the well-documented developmental perception–production gap. The major transition from a universal categorical-like perception, which is present at birth (Eimas *et al*. 1971), towards the language-specific perceptual ability – sensitivity to adult contrasts (Werker & Tees 1984, Kuhl 1991, Pegg & Werker 1997) and phonotactics (Jusczyk *et al*. 1994) – emerges in the second half of the first year of life.[2] Production accuracy develops much later; mastering certain adult distinctions and phonotactics in production extends well into pre-school years (e.g. MacNeilage 1997). The early acquisition of adult-based phonotactics, contrasts and allophony has been modelled in terms of formal and statistical algorithms (e.g. Tesar & Smolensky 1993, Boersma & Hayes 2001, Hayes 2004, Peperkamp *et al*. 2006, Hayes & White 2013) solely on the

---

[2] This early sensitivity does not automatically entail full lexical contrast specification at later stages. In 14-month-olds, the capacity of contrast discrimination can be seen in non-lexical tasks, but not in word-learning tasks (Stager & Werker 1997, Pater *et al*. 2004), which suggests that pairing sound with meaning is taxing, and not all distinctions are encoded in early lexical representations. The ability to detect contrastive changes in word forms has been reported for 30-month-olds (Tamási *et al*. 2019); at this stage, children's representations can be assumed to be specified for information corresponding to subphonemic features.

basis of the adult (or adult-like) input (B), without reference to the development of the child's production skills and the child-specific component (B → C). From this perspective, the role of child-specific patterns is purely implementational.

One particularly noteworthy aspect of child-specific patterns, evident from longitudinal studies (e.g. Smith 1973, 2010, Menn 1976, Becker & Tessier 2011), is that they exhibit only temporary stability. Enhanced variation and, as reported in this paper, gradience along the continuous phonetic dimension are found in the periods when these patterns dissolve. Grammaticalising these patterns in terms of OT constraints obscures the intuition that their gradual disappearance depends largely on the gradual development of the child's production skills. In this paper, I set out to tackle the problem of the interaction between the child's unskilled performance and adult-based phonology in terms of a different formal language, one which naturally links the continuous and discrete aspects of speech, and provides tools for describing mechanisms of change – the non-linear mathematics of the dynamical landscape.

1.1 The contribution of this paper

This paper seeks to deepen our understanding of how child-specific patterns relate to adult phonology by considering the problem in the context of a broader issue, not specific to acquisition but pertaining to human speech in general: the phonological–phonetic (or cognitive–physical) duality of speech structure and the nature of the phonology–phonetics link (for various views on this widely debated topic see Keating 1988, Browman & Goldstein 1989, 1995, Hayes 1999, Steriade 1999, Pierrehumbert 2001, 2002, Gafos 2006). To this end, it investigates the interaction in Polish between child-specific fricative devoicing, exhibiting enhanced gradience along the continuous phonetic dimension, and adult-based Voice Assimilation, which patently reflects categorical phonological organisation. The analysis is conducted on the basis of longitudinal acoustic data from a Polish-speaking child, Jula (/jula/; henceforth J), aged 2;3–2;8. Fricative devoicing is one of the substitution processes reported in child speech worldwide. Voice Assimilation produces uniformly voiceless or voiced obstruent clusters in adult Polish. An important characteristic of the latter process is that it is not only categorical, but also transparent at the utterance level. The phonotactics that underlie this process are learnable at the earliest stages, long before access to the morphological structure of words and the underlying representations of morphemes are available to the child (for details see §2). We can expect that Polish-speaking children use this process effectively in their production as soon as they develop the ability to produce obstruent clusters and contrastive voicing.

The child's data in (1a) reveal stable and target-appropriate application of Voice Assimilation in target voiceless clusters. However, target voiced fricative–stop and stop–fricative clusters are realised variably as voiced or voiceless, as in (b).

(1) *Voice Assimilation effects in J's speech*

   a. *Target voiceless clusters* (*stable*)

pieska	/pʲɛska/	[pʲɛɕka]	'dog (DIM.GEN.SG)'
jest kotek	/jɛs(t) kɔtɛk/	[jɛɕ kɔtɛk]	'(this) is a cat (DIM)'

   b. *Target voiced clusters* (*variable*)

zdejmij	/zdɛjmij/	[ɕtɛjmij]	'take off (IMP)'
jest głowa	/jɛz(d) gwɔwa/	[jɛɕ kwɔwa ~ jɛz̮ gwɔwa]	'(this) is the head'

For example, in *zdejmij* /zdɛjmij/ [ɕtɛjmij] *[ɕdɛjmij] 'take off (IMP)', the devoicing of the fricative (/z/ → [ɕ]) is accompanied by devoicing of the following prevocalic voiced stop (/d/ → [t]), despite Voice Assimilation being a regressive process in adult Polish. The devoicing of the stop must be interpreted as assimilatory because, unlike fricatives, target voiced stops are produced correctly as voiced in contrastive presonorant non-assimilatory contexts; e.g. *dużo* /duʐɔ/ [duɕɔ] 'a lot' (2;2.30), as in (2). The diacritic in the transcription of the fricative in [duʐ̥ɔ] (2;4.24) in (2) is shorthand for an intermediate stage of the gradient fricative devoicing pattern: it reflects the intermediate median value for the voicing parameter in the five tokens of the word at this stage in comparison with the tokens of the same word recorded at the other two stages.

(2) *Gradient fricative devoicing* vs. *stable voicing in stops*

dużo	/duʐɔ/	[duɕɔ]	(2;2.30)	'a lot'
		[duʐ̥ɔ]	(2;4.24)	
		[duʐ̥ɔ]	(2;7.19)	

The reported influence of a phonetic factor (lack of motor skills) on the application of a phonological process is in conflict with the classical modular feedforward grammar, which is founded on the temporal precedence metaphor of 'later' phonetic and 'earlier' phonological rules. Voice Assimilation acts as a 'late repair', which is unexpected in the modular feedforward approach. As revealed by acoustic analyses, the extent of this effect varies over time, and is correlated with the growth of voicing control in fricatives along the continuous phonetic dimension.

   The interaction between phonological and phonetic factors in acquisition is here analysed within the dynamical systems approach, using formal tools of non-linear mathematics proposed in earlier work for the analysis of incomplete neutralisation in adult speech (Gafos 2006, Gafos & Benus 2006). In a nutshell, the dynamical systems approach offers a formal language for integrating the discrete and continuous aspects of speech, while keeping the distinction between them valid. The integration is done in terms of dynamical linkage, not in terms of translating one type of structure (phonological symbols) into another (continuous phonetic parameters). Moreover, enhanced variability and gradience are predicted to be the necessary consequences of grammar change. Building upon this

approach, I propose that child-specific patterns arise from performance-related effects (lack of phonetic skills), and can be formalised as consequences of scaling a dynamical system's control parameters. The growth of the phonetic capacity, reflected in the continuous shift in the control parameter, brings about qualitative differences in the dynamical landscape, i.e. the phonological organisation.

The remainder of the paper is organised as follows. §2 presents basic facts about child-specific fricative devoicing and Voice Assimilation in adult Polish, as described in the literature. The effect of Voice Assimilation is further illustrated with acoustic measurement data from adult Polish (among others, the speech of J's caregiver). §2.3 schematises the hypotheses on the distribution of voicing in obstruent clusters in the child's speech on the basis of how the two processes interact in the developing grammar. §3 presents details of the acoustic study of longitudinal data from J, giving voicing parameter measurements in fricatives and stops in singleton non-assimilatory contexts as well as in clusters. §4 outlines the tenets of the dynamical model, interprets the results in terms of this model and discusses some consequences of adopting this model for understanding the link between categoricity and gradience and the underlying mechanisms of change. Conclusions are summarised in §5.

2 The interacting processes: basic generalisations and hypotheses

2.1 Fricative devoicing

Child-specific devoicing neutralises the voiced–voiceless distinction in obstruents; for example, the mispronunciation of *zęby* /zɛmbɨ/ 'teeth' as [sɛmpɨ] (Łobacz 1996: 180). In standard textbooks on speech development, this is one of numerous processes occurring in normally developing Polish-speaking children in the third year of life (e.g. Kaczmarek 1988). According to Sołtys-Chmielowicz (1998: 136), an impressionistic cross-sectional study based on a sample of more than 1000 children aged 3–7, fewer than 2% of children still devoice obstruents at the age of 3. The process is gradient: the refinement of the earlier acquired surface contrast continues after the age of 3 (Łobacz 1996: 180–181).

There is some indication in the literature that the acquisition of the surface voicing contrast in fricatives may be later than in stops. According to Kaczmarek (1988), the voiceless prepalatal /ɕ/ is the only fricative in the repertoire of 2-year-olds. Its voiced counterpart /ʑ/ emerges at the age of 3. Alveolar and postalveolar (retroflex) fricatives, both voiceless and voiced, are acquired at ages 4 and 5 respectively.[3] In comparison, the acquisition of voicing in stops is complete for labials

---

[3] Ladefoged & Maddieson (1996: 154–155) describe these sounds as 'flat postalveolar (retroflex)'; for a comprehensive discussion see Hamann (2004), and for a study using electromagnetic articulography Lorenc (2018).

and alveolars by the age of 2; it is delayed till the age of 3 only for velars. Bryndal (2015), in her study of acquisition of the inventory of Polish consonants in children aged 3–7, based on 200 subjects tested with a standardised production questionnaire (cf. Krajna 2008), reports non-normative pronunciations of the voiced sibilants in 3- and 4-year-olds. At age 3, none of the three consonants – alveolar /z/, prepalatal /ʑ/ and postalveolar /ʐ/ – reaches the criterion of 90% of normative realisations; at age 4, only /ʑ/ does (2015: 100–101).[4] The acquisition of voicing proceeds faster in stops: for /b/ and /d/ it is complete at age 3; for /g/ at age 4.

Fricative devoicing is also present in J's longitudinal data, which is analysed in the present study. A gradual development towards adult-like surface voicing in fricatives takes place throughout the period of development under investigation; the production of the voicing contrast in stops is predominantly accurate throughout the reported stages. A detailed description of the corpus of the child's speech samples follows in §3.1.

The above asymmetry in the child-specific devoicing pattern is paralleled by cross-linguistic patterns in adult speech. According to Ladefoged & Maddieson (1996: 176–178), most of the world's languages with voiceless fricatives do not have voiced counterparts. The phonetic grounding for this asymmetry is twofold. Voicing is associated with strong low-frequency energy, which may mask the high-frequency frication noise, which has lower amplitude. The vocal tract settings are such that the impedance of airflow at the glottis increases the difficulty of creating turbulence at the point of articulatory constriction. Although many languages (about 32%) have no voicing contrast in either plosives or fricatives, about 33.4% of languages surveyed in Maddieson (2013) have a voicing contrast in plosives but not in fricatives; languages with the opposite pattern, i.e. a voicing contrast in fricatives but not in plosives, are relatively rare – 6.7% of the total.

2.2 Voice Assimilation

The output of the across-the-board process of Voice Assimilation in adult Polish is transparent – uniformly voiceless or voiced obstruent clusters. The process has been widely documented in the standard descriptive sources (e.g. Kowalik *et al.* 1998: 91, Ostaszewska & Tambor 2012: 87–89), and extensively analysed within the generative phonological tradition (e.g. Gussmann 2007, Rubach 2008). The automatic character of the process and its categoricity are widely recognised to be a source of difficulty for Polish learners of English (e.g. Sobkowiak 1996: 50–56). Voice Assimilation effects are also present in loanwords and foreign words (e.g. *football* /fudbɔl/, *Macbeth* /magbɛt/; Gussmann 2007: 291).

---

[4] On most accounts, prepalatals are acquired earlier than other coronal obstruents, and can serve as replacements for the latter (Kaczmarek 1988, Sołtys-Chmielowicz 1998, Bryndal 2015). However, this initial substitution pattern may give way to different patterns in individual children at later stages; cf. the data from a Polish-speaking 4-year-old reported in Łukaszewicz (2000, 2006a, 2007), in which all coronal places reduce to plain alveolars.

Voice Assimilation results in voiced–voiceless alternations both inside words (3a) and across word boundaries (3b). It also has static phonotactic effects (3c). As can be inferred from (a) and (b), Voice Assimilation is regressive: the last obstruent in the cluster determines the surface voicing of the whole cluster. The regressive character of Voice Assimilation is also seen in (3d), where the process interacts with Final Devoicing. Some words, e.g. *już* in (b), do not have realisations which do not involve Voice Assimilation or Final Devoicing.

(3) *Voice Assimilation*

    a. *Alternations inside words*

prośba	/prɔʑ-b-a/	'request (NOM.SG)'	
cf. prosić	/prɔɕ-i-tɕ/	'request (INF)'	UR: ‖prɔɕ-‖
obiadku	/ɔbjat-k-u/	'dinner (DIM.GEN.SG)'	
cf. obiadek	/ɔbjad-ɛk/	'dinner (DIM.NOM.SG)'	UR: ‖ɔbjad‖

    b. *Alternations across word boundaries*

ko/t/ Filemon		'Filemon the cat'	
ko/d/ Bonifacy		'Bonifacy the cat'	
cf. kot	/kɔt/	'cat (NOM.SG)'	UR: ‖kɔt‖
kota	/kɔt-a/	'cat (GEN.SG)	
ko/t/ pocztowy		'postal code'	
ko/d/ dostępu		'access code'	
cf. kod	/kɔt/	'code (NOM.SG)'	UR: ‖kɔd‖
kodu	/kɔd-u/	'code (GEN.SG)'	
ju/ʂ/ podał		'(he) has already given'	
ju/ʐ/ dodał		'(he) has already added'	
cf. już	/juʂ/	'already' (the voicing of the final obstruent always depends on Voice Assimilation or Final Devoicing)	

    c. *Static phonotactic effects*

brzoza	/bʐɔza/	'birch tree'
przełęcz	/pʂɛwɛntʂ/	'mountain pass'

    d. *Combined effect of Voice Assimilation and Final Devoicing*

mózg	/musk/	'brain (NOM.SG)'
cf. mózgu	/muzg-u/	'brain (GEN.SG)'

Final Devoicing affects obstruents at the end of words; however, when the word-final obstruent precedes a voiced obstruent at the beginning of the following word, the effects of Final Devoicing are masked by Voice Assimilation; cf. *ko/d/ dostępu* in (3b). In consequence, Final Devoicing can be inferred only on the basis of voicelessness of obstruents at the end of utterances, or in word-final position when the next word starts with a sonorant.[5] Although the child's data show distributions

[5] This description pertains to the so-called Warsaw Polish standard, the variety relevant from the point of view of the child whose data I analyse in this study. I ignore

corroborating the presence of Final Devoicing in her grammar, I ignore this process in further discussion, because it cannot interact with child-specific fricative devoicing. An OT ranking for this process can be acquired as a part of the ranking established for Voice Assimilation, as sketched below.

From a purely distributional perspective, the contrastive [±voiced] specification is found only in obstruents in presonorant position; elsewhere, it is context-dependent. This simple distribution pattern can be learned at early stages, before children have access to the morphological structure of words and internalise underlying representations of morphemes, which is easily modelled using OT formalism. It can be illustrated using the pure phonotactic learner proposed in Hayes (2004).[6] The algorithm is based on a stringent constraint installation procedure which, among other things, favours markedness and specificity. Consistent voice agreement in the input will result in installing AGREE$_{obstr}$[±vcd] in the top stratum. Because of the surface contrast in voicing in presonorant contexts, the next level in the hierarchy will be occupied by a faithfulness constraint. The 'favour specificity' requirement will ensure that it is a positional constraint, IDENT$_{preson}$[±vcd]. Because of the surface contrast in voicing in presonorant contexts, the next level in the hierarchy will be occupied by a faithfulness constraint (assuming the meta-ranking Markedness ≫ Faithfulness). Note here that standard analyses of regressive Voice Assimilation effects in OT assume either string-based (i.e. presonorant) positional faithfulness or syllable-based (i.e. onset) positional faithfulness (see the discussion in Rubach 2008). Assuming a string-based version of positional faithfulness makes the acquisition of Voice Assimilation independent of prior knowledge of syllable structure. The next two layers will be occupied by the general markedness constraint *OBSTR[+vcd] and the general faithfulness constraint IDENT[±vcd]; the latter constraint ends up lowest in the hierarchy, as it is favoured neither by markedness nor by specificity. The ranking obtained via this procedure is given in (4). The subhierarchy IDENT$_{preson}$[±vcd] ≫ *OBSTR[+vcd] ≫ IDENT[±vcd] automatically generates Final Devoicing.[7]

---

the more intricate voicing patterns of Krakow Polish, as they are irrelevant to the point under discussion.

[6] The predominant hypothesis in experimental studies is that acquisition proceeds bottom-up from global properties discernible in the speech signal to smaller phonological domains. However, most formal algorithms aimed at modelling the emergence of sensitivity to ambient contrasts and phonotactics at early stages of acquisition are word-based. Such algorithms can be put to use to account for the acquisition of utterance-level generalisations such as Voice Assimilation in Polish, if they are allowed to take unparsed phrases as their inputs (cf. Hayes 2004: 197).

[7] Rubach (2008) argues that Final Devoicing requires a two-step analysis: a word-level grammar and an utterance-level grammar. The intricacies of this analysis need not concern us here, as they are of no consequence as far as the acquisition of Voice Assimilation alone is concerned.

(4) $\text{AGREE}_{\text{obstr}}[\pm\text{vcd}] \gg \text{IDENT}_{\text{preson}}[\pm\text{vcd}] \gg *\text{OBSTR}[+\text{vcd}] \gg \text{IDENT}[\pm\text{vcd}]$

In addition to the patterns described above, Polish also has a process of Progressive Devoicing, which an anonymous reviewer raises as a potential issue. However, its highly restricted scope (Gussmann 2007: 308), together with various other characteristics, make it inconspicuous from the perspective of a child. The process involves only two fricatives, /ʂ/ and /f/, whose voicelessness is caused by the preceding voiceless obstruent, as evidenced by the alternations exemplified in (5a). To account for the /r/ ~ /ʂ/ alternation, an idiosyncratic /r/ to /ʐ/ change is also needed, as exemplified in (5b), which operates independently of Progressive Devoicing.

(5) a. *Progressive Devoicing*

    i. /r/ ~ /ʂ/

łotr-a	/wɔtra/	'rascal (GEN.SG)'
cf. łotrz-yk	/wɔt-ʂɨk/	'rascal (DIM.NOM.SG)'
do-cier-a-ć	/dɔʨɛraʨ/	'reach (INF)'
cf. do-trz-e	/dɔt-ʂɛ/	'(he/she/it) will reach'
kr-a	/kra/	'ice-floe (NOM.SG)'
cf. krz-e	/kʂɛ/	'ice-floe (DAT.SG)'

    ii. /v/ ~ /f/

szew-ek	/ʂɛvɛk/	'seam (DIM)'
cf. szw-y	/ʂfɨ/	'seam (NOM.PL)'

  b. *Idiosyncratic* /r/ → /ʐ/

gr-a	/gra/	'game (NOM.SG)'
cf. grz-e	/gʐɛ/	'game (DAT.SG)'
br-a-ć	/braʨ/	'take (INF)'
cf. bierz-e	/bʲɛʐɛ/	'(he/she/it) takes'

At the surface level, these processes do not apply consistently; consider for example /wɔtrɨ/ *łotr-y* 'rascal (NOM.PL)' or /dɔtrɛ/ *dotrę* 'I will reach', in which /r/ is followed respectively by /ɨ/ and /ɛ/, and the examples in (5a.i), in which /ʂ/ occurs in exactly the same contexts. It can further be observed that the /r/ ~ /ʐ/ and /r/ ~ /ʂ/ alternations are often accompanied by /ɛ/ ~ zero alternations inside root morphemes (cf. *docierać – dotrze* in (5a.i) and *brać – bierze* in (5b)); likewise, the /ɛ/ ~ zero alternation is seen in *szewek – szwy* in (5a.ii). In such cases, pre-school children have been reported to have difficulty in perceiving the regularity of the connection between different forms of the same morpheme (Łukaszewicz 2006b: 16–17). Unlike Voice Assimilation, which can be internalised via pure phonotactic learning, Progressive Devoicing requires access to the morphological structure of words. However, words showing these alternations might not be readily available to the child. Not a single instance of the forms *łotrzyk*, *dotrze* and *krze* in (5a.i), which Gussmann (2007: 307) adduces as illustration of surface /ʂ/ alternating with /r/, is found in the Polish child-directed speech corpus (Haman *et al*. 2011). (The same

problem is found with the form *szewek* in (5a.ii), and other examples showing the surface /v/ as an alternant of /f/ in Gussmann 2007.) I conclude that Progressive Devoicing is unlikely to have any relevance at the stages of acquisition reported in this paper.

Another point raised by the anonymous reviewer is the need for acoustic data from adult Polish which would allow us to assess the categoricity of Voice Assimilation and the potential presence of incomplete neutralisation (phonetic gradience of the output reflecting the underlying specification of the target segment; see e.g. Braver 2019 for a discussion of the rich literature on the topic). Below I consider acoustic data from adult Polish, which illustrate the effect of Voice Assimilation on the acoustic parameter most reliably connected with the expression of the voicing contrast in adult Polish, voicing percentage (cf. 'the voicing ratio' in Strycharczuk 2012: 659).[8] The measurements were conducted using Praat's Voice Report function (Boersma & Weenink 2017), with pitch ranges adjusted for adult male *vs*. female speakers (in all other respects, the measurement details were the same as for the child's data in §3.1). The data come from high-quality recordings conducted for the purpose of a previous acoustic study investigating phonetic correlates of word stress in Polish (Łukaszewicz & Rozborski 2008). In that study, children's speech was compared with that of their parents. Speech samples from three adult speakers (one of whom was J's caregiver) were collected in structured interviews. Because of the high level of acquaintance between the experimenter and the participants, the recorded samples (lasting about 90 minutes in total) reflect casual spontaneous speech in informal settings. This sort of speech can be expected to be least prone to incomplete neutralisation, whose extent is known to vary, depending on the communicative context (Port & Crawford 1989; see also the discussion in Gafos 2006: 54–55). The inclusion of the caregiver's data ensures a good approximation of the input to which the child was exposed.

The voicing measurements were conducted for 488 obstruents (244 clusters). The results for each of two obstruents forming a cluster (O_1O_2) are plotted as coordinates in Fig. 2. The axes represent the voicing percentage; the experimental data points can have values between 0 and 100 on either of the axes. O_1 is the target of the process (the horizontal axis), O_2 is in presonorant position – it acts as the Voice Assimilation trigger (the vertical axis). S, Z, T and D denote voiceless fricatives, voiced fricatives, voiceless stops and voiced stops respectively. As expected, the data points corresponding to the predicted voiceless Voice Assimilation outputs (145 cluster tokens), e.g. ‖ZT‖ and ‖ST‖ → /ST/, consistently accumulate in the lower left quarter of the diagram, attracted by the (0,0) point. The

[8] As pointed out by the associate editor, a more complete picture of the realisation of the voicing contrast can be obtained with a combination of electroglottographic, airflow and acoustic data; cf. Smith's (1997) study of the devoicing of /z/ in American English. However, the acoustics of the speech signal seem relevant from the acquisitional perspective, as they constitute the input directly available to the child.
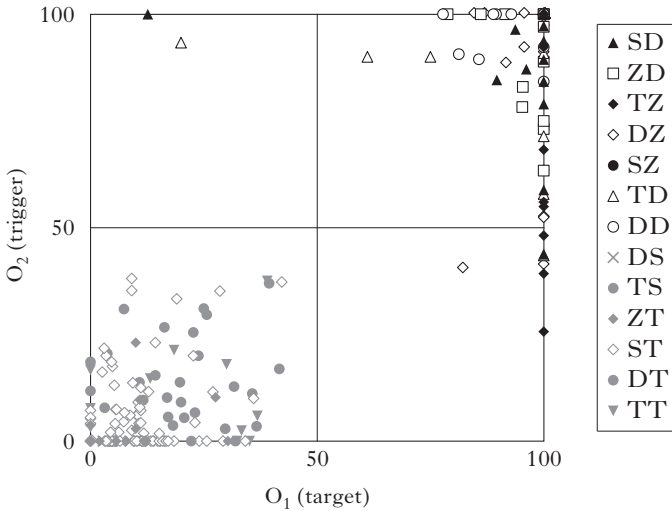
Figure 2

Voice Assimilation effects (voicing percentages) in adult Polish O_1O_2
clusters, depending on the underlying voicing of O_1 and O_2. Voiceless
outputs are grey; voiced outputs black. The two sets do not overlap.

predicted voiced Voice Assimilation outputs (99 cluster tokens), e.g. //SD//
and //ZD// → /ZD/, consistently accumulate in the upper right quarter;
there are 40 data points at exactly (100,100). These cannot be read directly
from the diagram, because its axes are continuous. In 70 out of 99 such
clusters, O_1 (the target of Voice Assimilation) is 100% voiced. The data
are divided depending on the underlying specifications of the obstruents
in the cluster. The two sets, target voiceless and target voiced, do not
overlap.

There are no incomplete neutralisation effects in target voiceless clus-
ters. In target voiced clusters, the incomplete neutralisation effect on O_1
(the Voice Assimilation target) is marginal: there are only four data
points in which the underlyingly voiceless O_1 has voicing below 80%.
These include two clear instances of the suspension of Voice
Assimilation, when the underlyingly voiceless O_1 retains its voicelessness
instead of assimilating to the underlyingly voiced O_2 (cf. the two data
points near the upper left corner). In general, underlying //ZT//, //DS//
clusters do not tend to have more voicing in O_1 than underlying //ST//,
//TS// clusters, i.e. they are not shifted towards the lower right quarter of
the diagram; underlying //SD//, //TZ// clusters do not show less voicing
in O_1 than underlying //ZD//, //DZ// clusters, i.e. they are not shifted
towards the upper left quarter of the diagram. For the vast majority of
tokens, then, the assimilation effect in O_1 is complete. In target voiced
clusters, a small 'gravity' effect can additionally be observed: O_2 (the

trigger of assimilation) is sometimes less voiced than the preceding target. We can ascribe this to the biomechanical difficulty of sustaining voicing throughout the entire cluster.[9]

2.3 Hypotheses

The two alternative hypotheses of the interaction between the child's immature production skills and adult phonology can be stated as follows. Will the child-specific fricative devoicing and Voice Assimilation interact in accordance with the 'feedforward' scenario, producing an output that is illicit from the point of view of the global adult phonotactics predicted by Voice Assimilation, as in (6a)? Or will it reflect the reversed scenario, with the application of Voice Assimilation conditioned by the child-specific pattern, as in (b)? I assume that the voicing distinction is neutralised only in fricatives; stops in contrastive presonorant contexts are realised in the adult-like manner. As shown earlier in Fig. 2 (§2.2), both ‖SD‖ and ‖ZD‖ surface as /ZD/ in adult Polish, and both ‖TZ‖ and ‖DZ‖ surface as /DZ/, so we ignore the differences in underlying representations, which are irrelevant from the point of view of the expected interactions in the child's speech. Also, it is important to note that the hypothesised 'reversed' scenario in (6b) is not possible in adult Polish (incomplete neutralisation can potentially affect the target of the neutralising process in adult speech, but never the trigger).

(6) a. *Feedforward interaction*

		Voice Assimilation	fricative devoicing		
/ZD/ → [SD]		/zgɔda/	→ [ɕgɔda]	zgoda	'all right'
/DZ/ → [DS]		/dɔbʐɛ/	→ [dɔbɕɛ]	dobrze	'all right, well'

b. *Reversed interaction*

		Voice Assimilation	fricative devoicing	Voice Assimilation
/ZD/ → [ST]		/zgɔda/	→ [ɕgɔda]	→ [ɕkɔda]
/DZ/ → [TS]		/dɔbʐɛ/	→ [dɔbɕɛ]	→ [dɔpɕɛ]

The diagrams in Fig. 3a schematise the hypothesised position of the child's data points corresponding to target /ZD/ and /DZ/ clusters in the above scenarios. Target voiceless clusters (/ST/, /TS/) are expected to consistently appear in the lower left part of Fig. 3b. In the 'feedforward' scenario in Fig. 3a.i, target /ZD/ clusters are expected to occupy the upper left quarter of the diagram, and target /DZ/ clusters the lower right quarter. In the 'reversed' scenario in Fig. 3a.ii, realisations of target /ZD/ and /DZ/ are not expected to differ: both are anticipated to appear in the lower left

---

[9] For acoustic illustrations of the Voice Assimilation patterns in adult speech, see Appendix A. The measurement data are available as .xlsx files in the online supplementary materials.
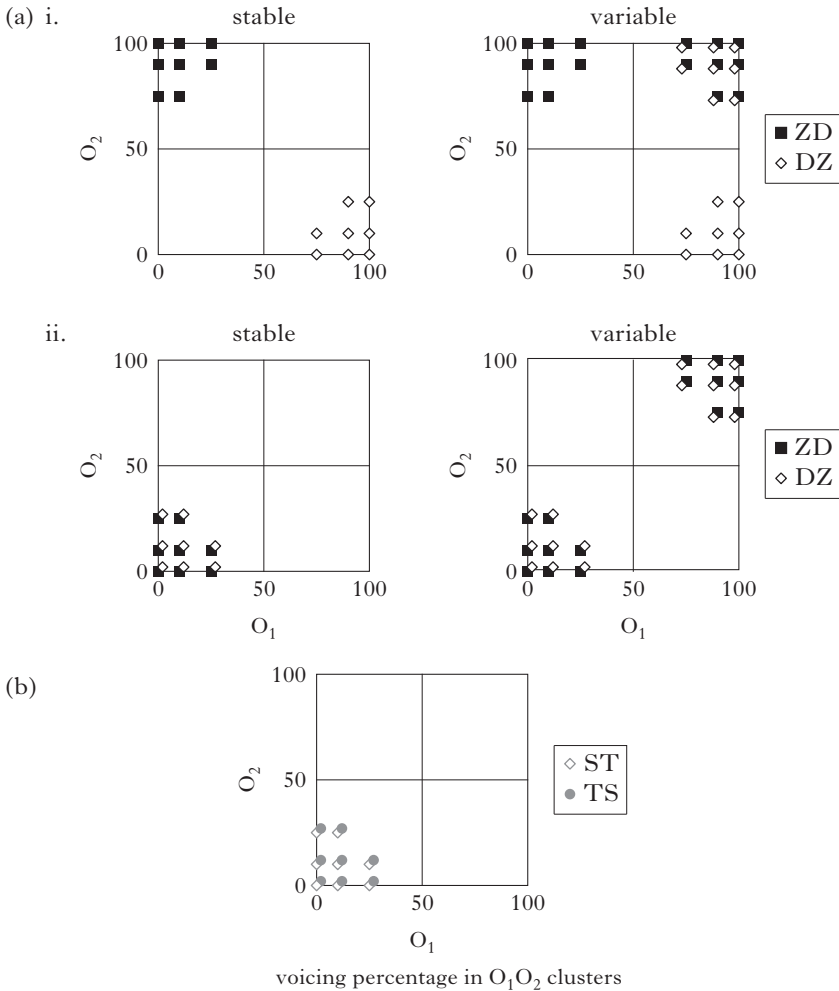
Figure 3

(a) Hypothesised position of the target voiced clusters: (i) feedforward; (ii) reversed.
(b) Hypothesised position of the target voiceless clusters (adult-like).

region of the diagram, like the /ST/ and /ST/ clusters in Fig. 3b. The diagrams on the lefthand side of Fig. 3a are based on the assumption that the effects are absolute; those on the righthand side (more plausibly) admit variability of these effects, i.e. some realisations of /DZ/ and /ZD/ may be adult-like. In the latter case, some data points will end up in the upper right quarter, both in Fig. 3a.i and Fig. 3a.ii; cf. the diagram for the adult data in Fig. 2. (For such data, we additionally expect some biomechanically motivated 'gravity' effect, such as the one reported for parental speech in §2.2 above.)

3 The acoustic study

3.1 The data and acoustic measurements

The corpus of child data used for acoustic analyses consists of 4 hours 40 minutes of audio recordings of naturalistic speech of a normally developing Polish speaking child, interacting with the experimenter during play sessions. The child (J) is female, aged 2;3–2;8 at the time of the recordings and raised in a monolingual environment in which standard Polish (the Warsaw standard) is spoken. The recordings were divided into seven sessions, conveniently covering the span of 21 weeks during which a shift from a complete neutralisation of voicing in fricatives towards a nearly adult-like surface voicing was observed.

The corpus contains rich audio material of quality sufficient to attempt acoustic analyses.[10] However, the spontaneous and naturalistic character of the data poses a challenge to acoustic measurement, and requires a rigorous data preselection procedure. The data were rejected for analysis if they were distorted by background sounds (the experimenter's voice, clattering toys, etc.), or if they had low amplitude caused by the child moving away from the microphone. Tokens were also rejected if the obstruent occurred within a larger span of ceased phonation, which sometimes happened towards the end of an utterance.

Some data had to be rejected during segmentation because of interference from certain reduction and substitution patterns. As typical of development between ages 2 and 3, the child's data did not show variability only on the voiced–voiceless dimension. Two phenomena are relevant here: the occasional realisation of fricatives as affricates (e.g. /z/ → [dʑ] in *język* 'tongue (NOM. SG)' /jɛ̃w̃zɨk/ [jɛɲdʑik]), and erroneous omission or replacement of sounds. As far as the former issue is concerned, analyses both including and discarding such segments were performed, but this did not have a crucial effect on the outcome (see §3.2 for details). Omission of a sound, resulting in reducing a cluster to a singleton, excluded the token from analysis (e.g. the word *gdzie* /ɡdʑɛ/ 'where' was sometimes rendered as [dʑɛ]). Reduction of obstruent clusters was infrequent at all stages. I also discarded tokens in which an obstruent was replaced by a sonorant or *vice versa*, e.g. *zaśpiewać* /zaɕpjɛvatɕ/ [naɕpjɛvatɕ] 'to sing', *piesek* /pjɛsɛk/ [pɛɛsɛk] 'dog (DIM)' (2;2.30). Such examples occurred only sporadically – with the notable exception of the labiodental /v/, which was replaced by [w] in all contexts. Simultaneously, target /f/ was predominantly rendered

[10] The corpus was recorded in 2006 for the purpose of an acoustic study of word stress in spontaneous speech in child and adult Polish (Łukaszewicz & Rozborski 2008). The recordings were made using a DAT (Onkyo Integra R1) recorder. A cardioid condenser microphone (Tonsil MCU-53) was positioned on a stand on the floor, pointing in the direction of the child. The child and the experimenter were sitting on the floor during most of the interaction, which focused on making things out of Play-Doh. All recordings took place in the same room, so the acoustic settings are comparable across the sessions. The research was conducted in agreement with internationally accepted ethical standards.
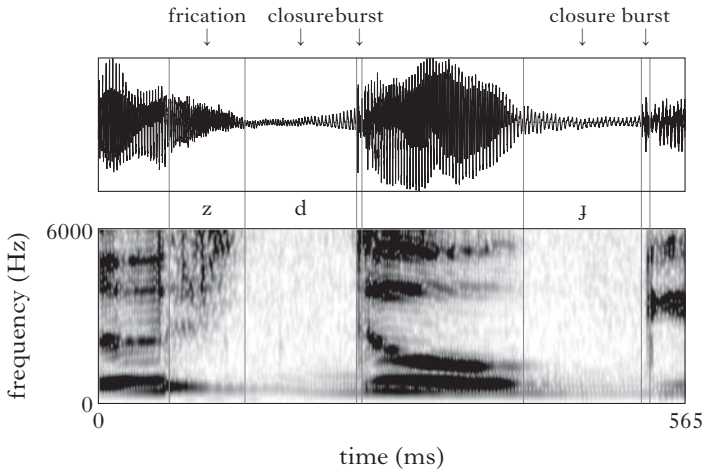
Figure 4

Example of obstruent segmentation: (*to*) *jest drugi* (*mikrofon*)
/jɛz druɟi/ '(it) is another (microphone)' (J 2;5.10).

as [x], which means that labiodental fricatives were generally absent from
J's speech. The focus of the study is thus on clusters containing coronal
fricatives. The remaining 1248 obstruent tokens were segmented in Praat;
three types of phonetic segments were distinguished: closure, frication
and 'burst' (positive VOT), as in Fig. 4.

As is well-known, phonological contrast in voicing can be expressed in
terms of various parameters, depending on the language. The most reliable
cue for voicing in adult Polish is the voicing ratio, i.e. glottal pulsing du-
ration divided by closure duration in the case of stops, and by frication du-
ration in the case of fricatives (Strycharczuk 2012: 659). Measurements of a
related parameter, voicing percentage (i.e. the percentage of closure/frica-
tion during which voicing occurs), were carried out using Praat's Voice
Report function, with a pitch range of 200–450 Hz, cross-correlation anal-
ysis method and default advanced pitch settings.[11] The measurements
were performed in the sound editor window on the part of the sound cor-
responding to the earlier annotated closure or frication portion of an
obstruent. This allowed for visual inspection of pulses. The outcome
of the voicing measurements in fricatives in presonorant position is
exemplified in Fig. 5. All three spectrograms represent the target voiced
fricative /ʐ/, produced with a different degree of voicing depending on the
token: (a) *duża* /duʐa/ 'big (FEM)', (b) *kucharza* /kuxaʐa/ 'cook (GEN.SG)',
(c) *dużo* /duʐɔ/ 'a lot'. For stops, I also measured the portion of the
speech signal corresponding to the release burst plus any subsequent aspi-
ration, i.e. the positive VOT (see Appendix A3).

[11] Voicing value percentages were obtained by subtracting the locally unvoiced frame
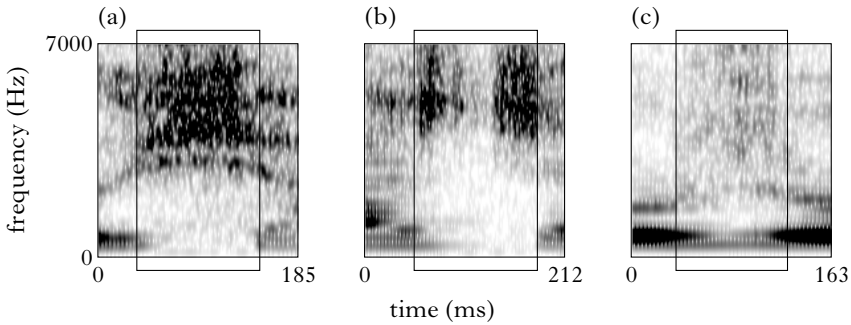percentages reported by Praat from 100.

Figure 5

Example of voicing measurements in target voiced fricatives, using
Praat's Voice Report function: (a) *duša* /duʐa/ 'big (FEM.NOM.SG)'
(J 2;2.30; 23% voicing); (b) *kucharza* /kuxaʐa/ 'cook (GEN.SG)' (J 2;5.10;
60% voicing); (c) *dužo* /duʐɔ/ 'a lot' (J 2;7.19; 100% voicing).

Voicing in fricatives and stops was measured independently in single-
tons in contrastive (presonorant) non-assimilatory contexts and in obstru-
ent clusters (Voice Assimilation contexts); the measurement data are
available in the online supplementary materials.[12] Most target voiced
obstruent clusters come from external sandhi contexts, and consist of a
sequence of a fricative and a stop, as exemplified in (7a). Among these
are tokens containing the word *jest* 'is', in which the final *t* was omitted
(as in casual adult speech). There were also a number of tokens in which
the fricative was preceded by a stop (7b), or where either word-internal
sandhi (c) or static phonotactics (d) were involved. These were also
included, because they were expected to exhibit the same voiced ~ voice-
less variation pattern. In the analyses of clusters, I excluded some tokens
with fricative–fricative sequences (e) and stop–stop sequences (f), as the
lack of clear transitions in such sequences often made their acoustic seg-
mentation infeasible. Both (e) and (f) were rare. Such sequences cannot
provide a compelling argument for the 'feedforward' *vs.* 'reversed' inter-
action between child-specific fricative devoicing and Voice Assimilation
(although we do expect more target-appropriate voicing in target
[+voiced] stop–stop sequences than in target [+voiced] fricative–fricative
sequences, because of the presence of fricative devoicing but no stop de-
voicing in the child's system).

[12] For simplicity, affricates were grouped with stops. Thus, for affricates I report
voicing percentage for closure, but ignore the frication portion in these segments
(which was usually voiceless, regardless of the underlying voicing).

(7) *Examples of obstruent clusters in J's speech at 2;7.19*

    a. musisz dosięgnąć   /muɕiʐ dɔɕɛŋgnɔɲtɕ/   [muɕiʐ dɔɕɛŋgnɔɲtɕ]
       'you have to reach'

      jest głowa   /jɛz(d) gwɔva/   [jɛz gwɔva]
       '(here) is the head'

    b. tak za mało   /tag za mawɔ/   [tak ɕa mawɔ]
       'such (is) too little'

    c. zbadałyśmy   /zbadawɨɕmɨ/   [zbawdawɨɕmɨ]
       '(we) have examined (FEM)'

    d. brzuch   /bʐux/   [bʐux]
       'belly'

      dobrze   /dɔbʐɛ/   [dɔbʐɛ]
       'all right, well'

    e. zniszczysz zaraz   /zɲiʂtʂɨʐ zaraz/   [zɲitɕiʑːaras]
       '(you) are going to break (it) in a moment'

    f. jak będzie   /jag bɛɲdʑɛ/   [jag bɛɲdʑɛ]
       'when it is'

3.2 Results

The results are presented in Fig. 6, divided into two columns. Column (a) consists of boxplots depicting the percentages of voicing present in voiceless *vs.* voiced fricatives and stops at each stage, on the basis of measurements in contrasting non-assimilatory positions. (Affricate-like realisations of fricatives were excluded in these analyses.) These results are juxtaposed against the diagrams in column (b). The latter show the percentages of voicing measured in target voiceless/voiced obstruent clusters (assimilatory contexts) at a given stage. Columns (a) and (b) are based on independent measurements (available in the supplementary materials).

    As seen in column (a), the underlying voiceless–voiced distinction is neutralised in fricatives at stage 1. The boxplots for target voiceless and voiced fricatives are more or less the same at this stage. In contrast, the underlying voiceless–voiced distinction in stops is clearly reflected in J's speech throughout the entire period of development under investigation, including stage 1. The shift in the surface expression of the underlying voiced fricative category – from phonetically voiceless towards phonetically voiced segments – is marked by greater variability, starting at stage 2. At stages 3–6, the means for target voiced fricatives fluctuate around 40% voicing. At stage 7, the mean reaches a considerably higher value of about 64%, and the data begin to exhibit a strongly negative skewness (the distribution expected from the point of view of the adult system). In sum, at stages 2–7, child-specific fricative devoicing gradually disappears, yielding a gradient phonetic output for underlying

voiced fricatives. The ranges for voicing values are wide, occupying nearly the entire 0–100% space at these stages, and indicating developing cognitive control over the phonetic dimension rather than simply the relaxation of some biomechanical constraints on attainable voicing thresholds. The scatterplot in Fig. 7 shows that there is no clearly discernible bimodal distribution of the data points at these stages. In statistical analyses (using Generalised Linear Models fitted in SPSS, with gamma distribution[13] and the Log link function) with voicing percentage as the dependent variable, the difference between voiced and voiceless fricatives does not reach the level of significance at stages 1 and 2 (stage 1: $\chi^2(1) = 0.034$, $p = 0.853$; stage 2: $\chi^2(1) = 1.861$, $p = 0.173$), but is statistically significant throughout stages 3–7 (stage 3: $\chi^2(1) = 5.633$, $p < 0.05$; stage 4: $\chi^2(1) = 5.511$, $p < 0.05$; stage 5: $\chi^2(1) = 6.048$, $p < 0.05$; stage 6: $\chi^2(1) = 17.032$, $p < 0.0001$; stage 7: $\chi^2(1) = 24.643$, $p < 0.0001$). (The p values were obtained via log-likelihood ratio tests comparing the fitted models with intercept-only models.) Voiced stops were significantly different from voiceless stops at all stages ($p < 0.0001$); see Appendix B for these statistics, the SPSS code used to obtain these statistics and the goodness-of-fit details.

In column (b) in Fig. 6, the voicing percentage results for each of two obstruents O_1O_2 forming a cluster are plotted as coordinates on a diagram. At every stage, there are no data points in the upper left quarter of the diagram. This means that there were no unassimilated clusters of the *[ɛgɔda] type hypothesised as part of the 'feedforward' scenario earlier in Fig. 3a.i. Furthermore, most tokens occur in the lower left and the upper right quarters. Thus the data show polarisation similar to the adult system in Fig. 2, suggesting the presence of Voice Assimilation in the child's system. However, the distribution of the target voiced clusters is not exactly adult-like: while all target voiceless obstruents are found in the lower left quarter of the diagram, as expected from the point of view of the adult system, some target voiced clusters also appear in this region; cf. the circled areas in the diagrams in Fig. 6b. The phenomenon is compatible with the hypothesised 'reversed variable' scenario in Fig. 3a.ii above. Target /ZD/ and /DZ/ clusters behave similarly. Further, analogously to the adult input in Fig. 2, target voiced O_1O_2 clusters whose realisations fall outside the circled 'voiceless' regions, and which can be tentatively classified as phonetically 'voiced' outputs, have on average higher percentage voicing values in O_1 than in O_2 (the mean difference between the two segments amounts to 20% voicing; $t(41) = 4.719$, $p < 0.0001$). The problem of lower percentage voicing values in O_2 in such clusters must be connected with the difficulty of sustaining voicing throughout a cluster. Probably, then, it is a biomechanical rather than a cognitive effect.

---

[13] As pointed out by the associate editor, beta regression would be more appropriate for analysing continuous proportional data.
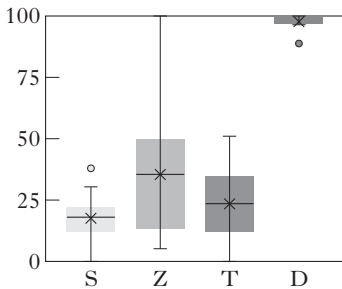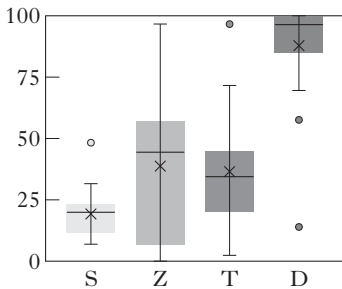
(a) voicing percentage in
contrastive contexts

(b) voicing percentage in
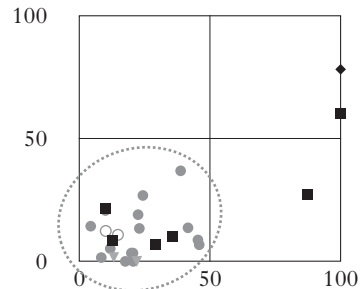O_1O_2 clusters
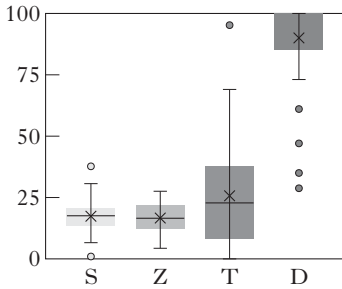
stage 1 (2;2.30)
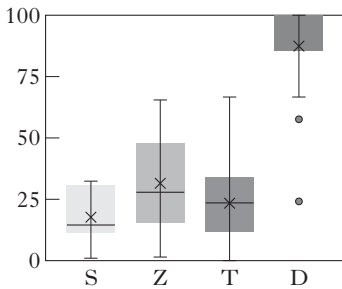
stage 2 (2;3.13)
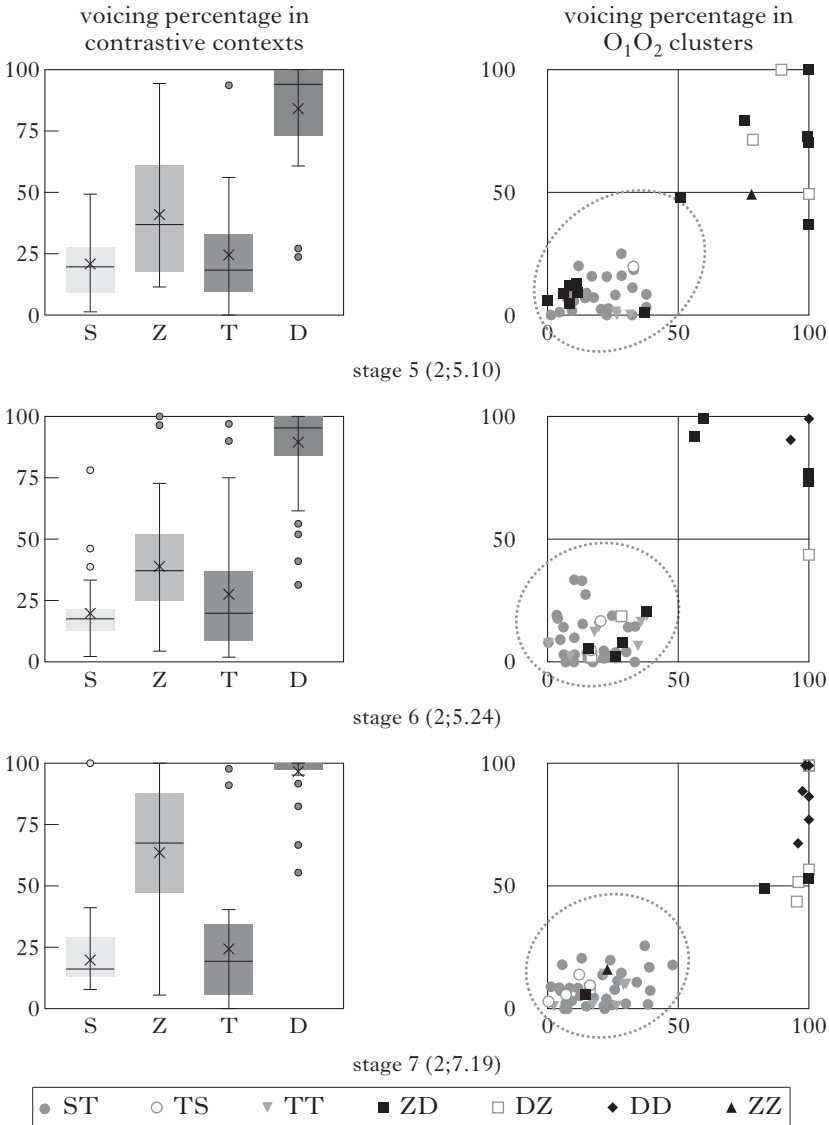
stage 3 (2;4.10)

stage 4 (2;4.24)
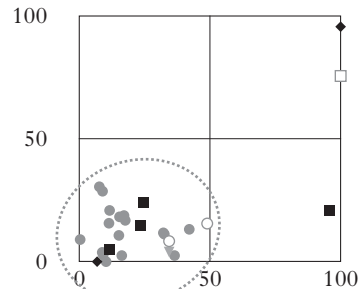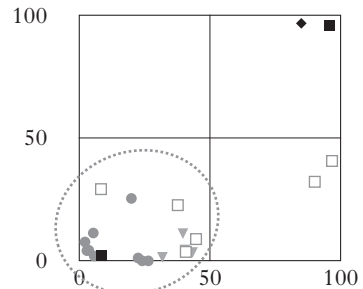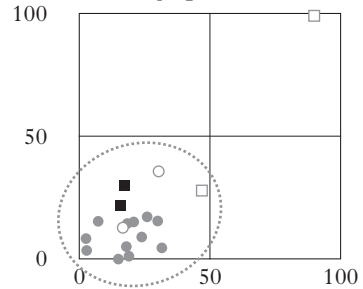
Figure 6

Measurement results in non-assimilatory contrastive and assimilatory contexts.

Finally, as illustrated in Fig. 8, it is of no consequence whether target adult /ZD/ corresponds to the abstract adult underlying representation ‖SD‖ or ‖ZD‖, and /DZ/ to ‖TZ‖ or ‖DZ‖. (Ambiguous cases such as the word-final /ş/ ∼ /z̧/ alternation in the word *już* in (3b) above, which always depends on Final Devoicing or Voice Assimilation, were coded as
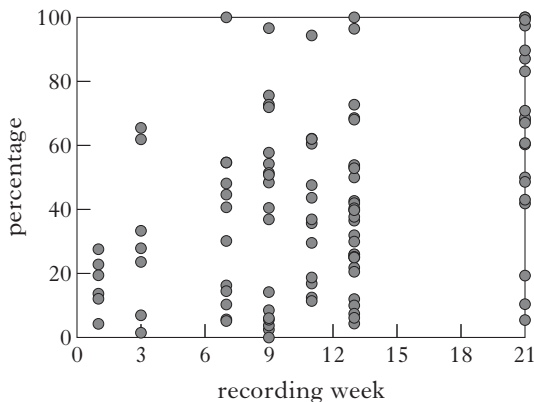
Figure 7

Scatterplots for voicing percentages in target
voiced fricatives for each of the seven stages.

∥SD∥ in this analysis.) Regardless of the underlying specification of O_1 (the Voice Assimilation target) as voiced or voiceless in adult Polish, the variable voiced ~ voiceless realisations of such clusters occur in J's speech.

voicing percentage in O_1O_2 clusters
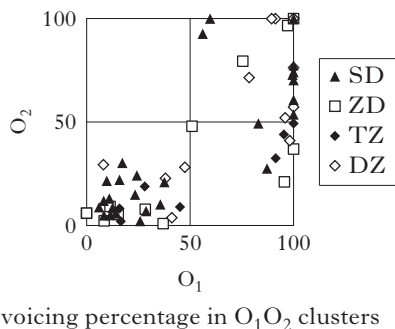
Figure 8

Voicing percentage in target voiced fricative–stop and stop–fricative clusters in J's speech (stages 1–7), depending on the underlying voicing of O_1 in adult Polish.

The monotonic disappearance of child-specific fricative devoicing (or, rather, the growth of the voicing contrast in these segments) is illustrated in Fig. 9. Unlike the realisation of target voiced stops, which is stable and largely adult-like throughout the development, the average extent of voicing in target voiced fricatives is to a substantial degree parallel to the extent of the successful avoidance of 'late repairs' in target voiced obstruent clusters containing fricatives. The inclusion of affricate-like realisations of fricatives does not have a significant impact on the analysis.
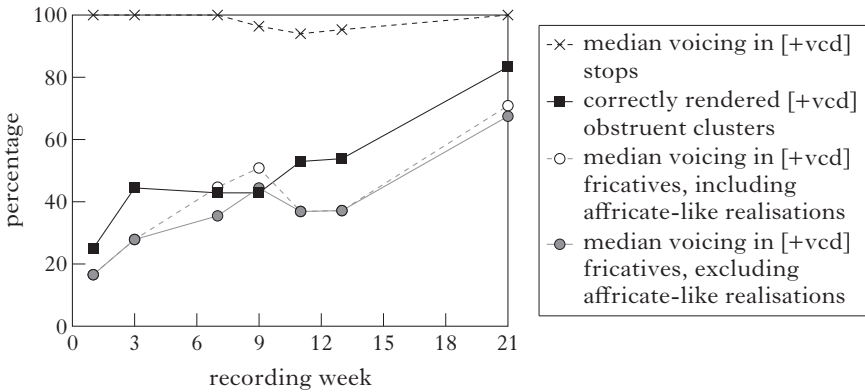
Figure 9

Voicing emerging in target [+voiced] fricatives *vs.* stops
vs. obstruent clusters for each of the seven stages.

As further shown in Fig. 10, the two phenomena are correlated ($r(7) =$ 0.974, $p < 0.001$). This indicates that progress in one domain (voicing in fricatives in non-assimilatory contrastive contexts) is paralleled by progress in the other domain (target-appropriate voicing in obstruent clusters). (Recall that the scores for voicing in fricatives and voicing in clusters are based on datasets which are entirely independent: although they are expressed in percentages, they do not constitute proportions in the sense of compositional data.) The 'success score' for target-appropriate realisations of voiced O_1O_2 clusters was calculated using SPSS on the basis of k-means classification into 'voiceless' and 'voiced', run to fit a two-group model; the clusters classified as 'voiceless' in these analyses correspond to those that are circled in Fig. 7b. The correlation remains significant
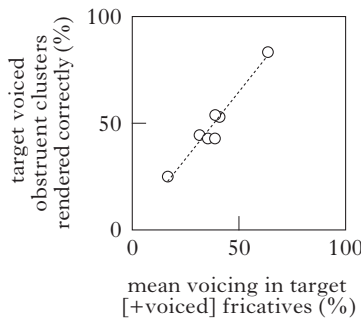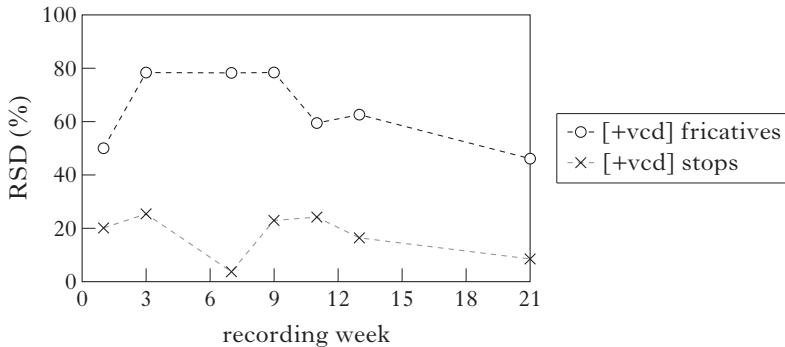

Figure 10

Fricative voicing (gradient) *vs.* Voice Assimilation
(categorical) yielding target-appropriate voiced outputs.

$(r(7) = 0.961, p < 0.001)$, with Voice Assimilation scores recalculated using Laplace's rule of succession, which is useful in estimating underlying probabilities when observations are small in number (or when $s = \{0, n\}$). (The expectation was calculated with the formula $(s + 1)/(n + 2)$, where s is the number of 'successes' and n the number of observations ('opportunities').)

Figure 11
RSD voicing percentage in target [+voiced]
stops *vs.* fricatives for each of the seven stages.

Target voiced stops and fricatives also differ in how their variability profiles change over time, as shown in Fig. 11; variability was measured in terms of relative standard deviation (RSD = $100 \times$ SD/Mean). A shift from voicing contrast neutralisation towards target-appropriate surface contrast in fricatives was marked by enhanced variability; cf. the temporarily increased RSDs, in particular at stages 2–4. Although the dispersion remains considerably high throughout stages 2–7, the RSDs become lower, because of the simultaneously increasing means. The category of voiced stops was relatively target-appropriate and stable throughout the stages (their RSDs did not rise). As highlighted in §4, these differences in variability profiles are meaningful in the dynamical systems approach, in which the enhanced gradience is symptomatic of the grammar undergoing a shift from one organisational mode to another. As the mean voicing percentage increases, the system undergoes a qualitative change from a stable contrast neutralisation to a stable contrast realisation, necessarily passing through an intermediate stage of instability.

4 Symbolic-dynamical representations

The acoustic results in §3.2 suggest an interaction in which the application of an adult-based categorical process (Voice Assimilation) is conditioned

by a child-specific gradient process (fricative devoicing). This runs against the predictions of the standard feedforward modular grammar. In §4.1, I discuss how the behaviour of the developing phonological system can be understood if we assume symbol-like dynamical representations. In brief, the proposal is that child-specific patterns can be formalised in terms of the consequences of scaling the control parameters of the dynamical system. A continuous shift in the control parameter, reflecting the growth of the phonetic capacity in the child, brings about qualitative differences in the dynamical landscape, which is phonological organisation.

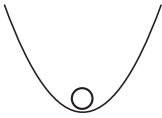## 4.1 Phonological grammar as a dynamical landscape: basic notions

Dynamical systems can be viewed as attractor landscapes, formally expressed through non-linear mathematics (for an overview see Thelen & Smith 1994, especially Chapter 3). Attractors are behavioural modes, which the system develops (via self-organisation) as a function of the interaction of its internal components and their responsiveness to external conditions. Dynamical principles, either stated explicitly in terms of differential equations or only alluded to at the level of metaphor, have been widely applied in natural sciences (for a mathematical introduction, with emphasis on applications, see Strogatz 1994). A wealth of illustrative examples, ranging from modelling the Belousov–Zhabotinsky oscillatory reaction in chemistry to applications in motor and cognitive development research (for example on the ontogeny of treadmill stepping in infants or the puzzle of the Piagetian A-not-B error), can be found in Thelen & Smith (1994). Of interest here is the possibility of conceptualising the relationship between phonological organisation and phonetic parameters using these principles. The symbolic-dynamical representations described below are in terms of first-order differential equations, as formalised in Gafos (2006). The mathematical formalism of dynamical systems has been used in various ways to model phonological–phonetic phenomena in adult speech: articulatory gestures defined as invariant mathematical laws (Kelso *et al*. 1986, Browman & Goldstein 1995), (in)stability of speech perception (Tuller *et al*. 1994), incomplete neutralisation (Gafos 2006, Gafos & Benus 2006), shifts in phonetic indices dependent on qualitative differences in syllable parses (Gafos *et al*. 2014) and perception-induced updates in speech production planning (Roon & Gafos 2016), to mention just a few examples. For a non-mathematical treatment of the shift in phonological organisation from babbling to first words, espousing the dynamical systems approach, see Vihman *et al*. (2015).

As illustrated in (8), the behaviour of a dynamical system can be pictured as a particle landing in potential wells. The position of the particle is the system's state x at some time t. The wells reflect the system's preferred modes, the so-called 'attractors'; the relative depth and steepness of the
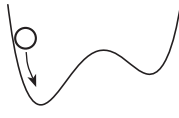
wells reflects the attractors' relative stability. The peaks are unstable points (repellers): the system is unlikely to reside in these positions.
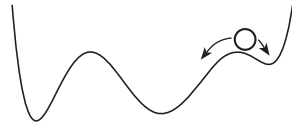
(8) *Examples of dynamical landscapes*
   a. *one attractor*   b. *two attractors*   c. *three attractors*

Attractors are the states towards which the system tends, and thus correspond to mean values of parameters in experimental data. How close x is with respect to those target values at time t depends on its initial position and its trajectory, determined by a differential equation relating dx/dt (the instantaneous velocity of x) to the so-called force function referring to the attractor's slope (for clarification see below). It also depends on the amount of noise by which the system is perturbed. (Internal fluctuations are a consequence of interaction of multiple subcomponents of a complex system.) A relatively stable attractor is less prone to change; (8c) is an intuitive illustration. When perturbed to a relevant degree by internal noise or external forces, the particle is likely to visit the deeper well of the attractor to the left instead of dropping back into the shallow well of the attractor to the right. Globally, i.e. when starting from different initial positions, the system is more likely to reside in the former attractor than in the latter. The measure of relative variability around a mean state is a powerful tool in assessing the degree of the attractor's stability. Thus, a major progress offered by the dynamical approach is that it allows us to treat a certain kind of variation in experimental data as information, not noise; cf. the RSDs for target [+voiced] fricatives *vs*. stops in Fig. 11 above, indicative of unstable *vs*. stable categories.

Characteristically, dynamical systems produce coherent patterns only within a certain range of values of their control parameters. As the control parameter is scaled continuously, it passes through a critical value at which a qualitative transition in the attractor landscape (so-called bifurcation) occurs. Under different conditions, the components of a complex dynamical system are free to reassemble into different stable modes. It is this 'soft assembling' property to which dynamical systems owe their enormous flexibility; it is also why they come with the potential to model developmental data in an insightful way (Thelen & Smith 1994: 60). One of the primary research goals is then to single out factors as potential control parameters that affect the evolution of a dynamical system.

The geometric characterisation of the dynamical landscape in (8) allows us to express the notion of phonological contrast in an intuitively straightforward way. The attractors reflect the macroscopic observable, i.e. the phonological organisation of the system. Given some continuous physical parameter (e.g. the extent of vocal fold vibration), the absence *vs*. presence

of a binary phonological contrast (e.g. a voiced–voiceless distinction) corresponds to a single-attractor *vs.* two-attractor system, as in (8a) and (b) respectively. The system's state x at time t corresponds to measurable phonetic output. In the dynamical view of the phonology–phonetics link, phonology is not antecedent to phonetics, as no translation from discrete symbolic to physical continuous is necessary. The continuous (phonetic) and discrete (phonological) properties of speech are inseparable, yet distinguishable. The qualitative and quantitative aspects of the system are both expressed with non-linear mathematics (cf. Browman & Goldstein 1995, Gafos 2006).

The schematic illustrations of different dynamical systems in (8) are based on polynomials of different degrees.[14] These are potential functions, $V(x)$, defined over some order parameters (e.g. the degree of voicing). Their minima correspond to attractors, i.e. stable equilibrium points at which a particle will eventually land (assuming deterministic dynamical systems), starting from different initial positions. Their maxima are repellers, i.e. unstable equilibrium points. Both attractors and repellers are equilibria in the sense that the rate of change is zero at these points, i.e. $dV(x)/dx = 0$. The number and depth of attractors depend on the degree of the polynomial as well as on the value of its control parameters, which will be discussed below with reference to grammatical $V^G(x)$ and lexical (intentional) potentials $V^I(x)$, as defined in Gafos (2006). Before we turn to exploring the role of these potentials in shaping the attractor landscape in child *vs.* adult phonologies, let us consider a basic characterisation of stability in mathematical terms.

The essence of symbolic-dynamical representations is that they are assembled in time and are understood as mathematically defined laws. They are defined in terms of differential equations, which have the general form $dx/dt = -dV(x)/dx$. (For simplicity, I ignore the noise component which occurs on the righthand side of the equation and introduces indeterminacy in the system's behaviour; see Gafos & Benus 2006: 908–909.) The lefthand side of the equation is the rate of change of x (with respect to t). The righthand side of the equation corresponds to the force function, which is the negative of the first derivative of the potential function $V(x)$. The mathematical relationship between the potential, $V(x)$, and the force function, $F(x) = -dV(x)/dx$, reflects a well-established relationship for oscillating systems in physics. It is the relationship between the potential energy and the restoring force which acts on an oscillating system to bring it back to its equilibrium. The restoring force depends on the degree of displacement – linearly in harmonic, non-linearly in anharmonic systems. Its direction is opposite to the direction of displacement, hence the minus sign in $-dV(x)/dx$. First-order dynamical equations, $dx/dt = -dV(x)/dx$, express the evolution of the system in time. It is important to note that they have equilibria solutions, not

---

[14] Polynomial functions can have multiple local extrema at which the function can assume different values. We need this property to be able to express the fact that attractors may vary in depth.

periodic solutions (see Strogatz 1994: 28). The solution to the equation is a function of x with respect to t, or, to be more precise, an infinite set of functions for an infinitude of different initial conditions (x_0). The solutions are trajectories which converge towards the stable fixed point(s) and diverge from the unstable fixed point(s).
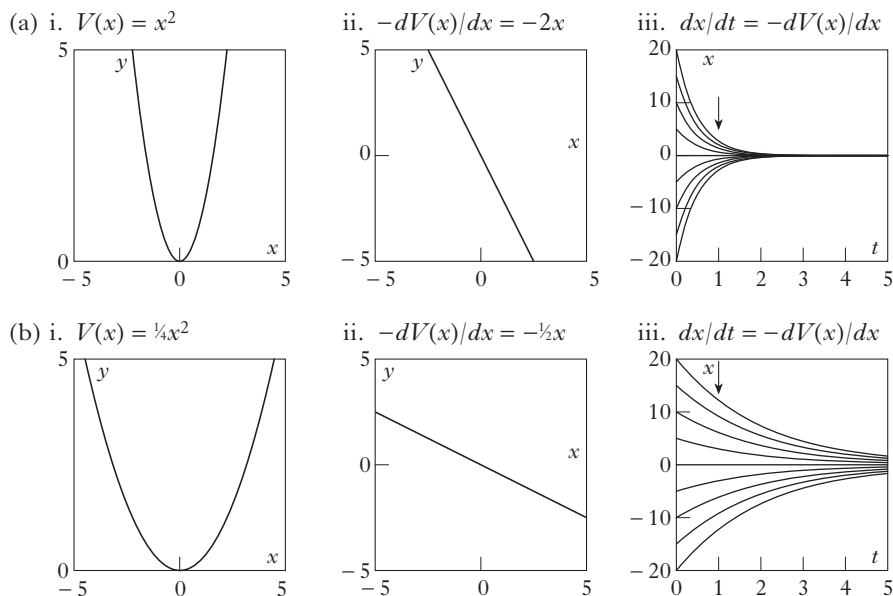
Figure 12
(a) A more stable attractor; (b) a less stable attractor.

Convergence towards the stable fixed points (attractors) is illustrated in Fig. 12 on the basis of relatively simple single-attractor dynamics: $dx/dt = -dV(x)/dx = -2x$ and $dx/dt = -dV(x)/dx = -\frac{1}{2}x$ respectively. As we will see later, single-attractor dynamics generalised as $dx/dt = -dV(x)/dx = -\alpha x$, with $\alpha > 0$, can be considered to be a dynamical equivalent of lexical representations, because they specify the target value of a phonological feature unambiguously, regardless of the shifts in the system's control parameter α. The solutions to these differential equations, which are $x(t) = x_0 e^{-2t}$ and $x(t) = x_0 e^{-\frac{1}{2}t}$ respectively, are plotted in the (iii) panels; in both plots the system's initial states are assumed to take the arbitrarily selected values $x_0 = \{-20, -15, -10, -5, 0, 5, 10, 15, 20\}$.[15] We can readily

---

[15] More often than not, differential equations cannot be solved explicitly and approximation methods need to be used (the simplest of which is Euler's method; see Strogatz 1994: 32–33 for both a discussion of this method and proposed

observe how the differences in the attractors' stability (steepness of the wells) in the (i) panels in Fig. 12 correspond to the differences in slopes (the force functions) in (ii), and ultimately, to the degree of variability in the output data at a given time point in (iii). For example, if we assume that the decision time for assembling representations during the speech planning process is $t = 1$ (indicated by the arrow), this renders output x_t values in Fig. 12a much less variable relative to Fig. 12b.

Differences in stability are one of several expected differences between the adult and child phonological systems, making the latter more liable to change. The systems in Fig. 12 are nevertheless qualitatively the same, i.e. their trajectories converge towards the same stable fixed point.

As proposed in Gafos (2006: 62), a two-way surface phonological contrast (e.g. in voicing) can be represented in terms of the quartic function in Fig. 13, $V^G(x) = kx - \frac{1}{2}x^2 + \frac{1}{4}x^4$, where k stands for a control parameter. (The quartic degree is the lowest polynomial degree allowing us to have two minima and one maximum.) The dynamics of phonological grammar are defined in terms of the differential equation: $G(x) = dx/dt = -dV(x)/dx = -k + x - x^3$. The $G(x)$ formula is referred to as the 'tilted anharmonic oscillator' in Gafos (2006: 66), and proposed as a first approximation of the grammar dynamics.[16] (See Tuller *et al*. 1994: 7 for an earlier application of the same formula in the domain of speech perception.) By scaling the parameter k, we obtain a qualitative shift (bifurcation) between a single-attractor system and a two-attractor system, which corresponds to phonological grammars characterised by the absence *vs*. presence of binary contrast, as in Fig. 13. The bistable organisation emerges only at some critical $(-k_c)$ value of k, and is sustained only within a certain range of $k \in (-k_c, k_c)$. The bistable grammatical potential is symmetrical when $k = 0$: the chances of a particle landing in either of the

---

refinements). The qualitative behaviour of dynamical systems is often visualised using graphic methods. For first-order (one-dimensional) equations these are either a vector field on the real line (so-called 'phase flow'; Strogatz 1994: 28–29) or a potential well in which the particle moves (assuming heavy viscous damping; Strogatz 1994: 30). The latter method will be used in this paper. While differential equations are the legacy of Newton's dynamics, the breakthrough consisting in development of powerful geometric methods to analyse qualitative rather than quantitative questions is due to Poincaré.

[16] Unlike the $-dV(x)/dx$ equations in Fig. 12, this equation contains a non-linear term, x^3. This is because in anharmonic oscillating systems the restoring force is non-linearly proportional to the displacement. This yields the quartic term x^4 in the potential function $V^G(x) = kx - \frac{1}{2}x^2 + \frac{1}{4}x^4$. The presence of the quartic term and absence of a cubic term in $V^G(x)$ ensures a symmetrical effect on x and $-x$ when $k = 0$. The negative *vs*. positive coefficients of the quadratic and quartic terms are kept fixed, so that the change in the landscape depends only on the parameter k, not on these coefficients. Note that reversing the signs in these coefficients would result in obtaining a one-attractor system which would not be able to shift qualitatively to the two-attractor regime depending on k, and thus would be inadequate from the point of view of the expression of phonological contrast. (The behaviour of a particle moving in a symmetrical double-well potential has been extensively studied both in classical and quantum mechanics; see Strogatz 1994.)

two attractors is then equal, as in Fig. 13a. I take this symmetrical potential to represent the adult-like voiced–voiceless distinction.
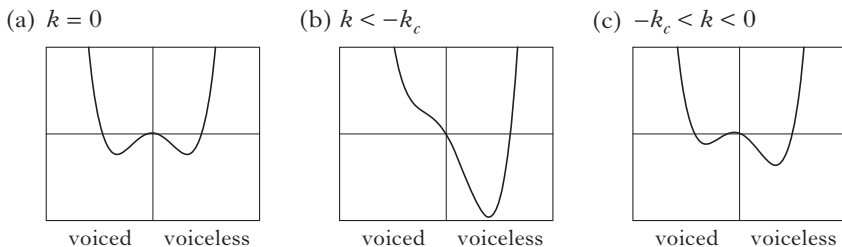
(a) $k = 0$ (b) $k < -k_c$ (c) $-k_c < k < 0$

voiced   voiceless   voiced   voiceless   voiced   voiceless

Figure 13
Grammatical potentials: $V^G(x) = kx - \frac{1}{2}x^2 + \frac{1}{4}x^4$.

Expanding Gafos' hypothesis to the domain of acquisition, there can be a twofold difference between the child's system and the corresponding adult system from the point of view of the grammar dynamics. First, there may be only one attractor in the child's system, resulting in absolute contrast neutralisation, as in Fig. 13b; cf. the behaviour of target voiced fricatives at the initial stage reported in this paper, when they are not distinguished from target voiceless fricatives in J's outputs. (The one-attractor system in (b) can also represent contextual voicing neutralisation in adult speech, e.g. the cross-linguistically common patterns of coda devoicing, as described in Gafos 2006.) Second, there may be two attractors in the child's system, but their strength may be unequal, as illustrated in (c). With the latter representation, variable contrast neutralisation ensues; cf. the unstable behaviour of target voiced fricatives at some stages of J's development reported in this paper. The depth of the well of the 'voiced' attractor gets bigger as the k parameter is scaled up from the critical value $-k_c$ towards 0 (to ultimately match the stability of the 'voiceless' attractor at $k = 0$). Thus, given the three different landscapes in Fig. 13, the predicted order of acquisition is: (b) (initial stage), (c) (intermediate stage), (a) (adult-like stage).

With the grammatical potential supplying two competing attractors for binary contrast (voiced–voiceless in presonorant position), we need a mechanism to ensure that the feature is realised unambiguously in lexical items. In Gafos (2006), lexical features are expressed in terms of intention dynamics, $I(x) = dx/dt = \alpha(x^{REQ} - x)$, supplying attractors at the required voicing values (x^{REQ}); α is a control parameter and expresses the strength of the intention. Accordingly, in Fig. 14 I assume (arbitrarily) that the values $x^{REQ} = -1$ and $x^{REQ} = 1$ correspond to the categories 'voiced' and 'voiceless' respectively. The intention parameter in the illustration in Fig. 14 is assumed to take the value of $\alpha = 1$. The 'voiced' *vs.* 'voiceless' intentional dynamics are mutually exclusive.
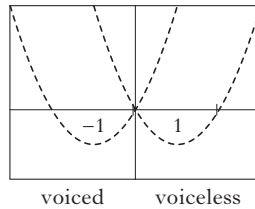
Figure 14
Intentional (lexical) potentials: $V^I(x) = \alpha(\tfrac{1}{2}x^2 - x^{REQ} \times x)$.

In what follows, I assume that every output can be viewed as a combined effect of the grammar dynamics and the intention dynamics, as in (9). For simplicity, the weighting of the grammar dynamics will be kept constant ($w_G = 1$). This idealisation reflects the expected early acquisition of contrasts and phonotactics (recall the discussion in §1 and §2). The relative weighting will be achieved by allowing the control parameter $\alpha = w_I$ of the intention dynamics to assume different values within the range from 0 to 1 (cf. Gafos 2006: 70).

(9) $dx/dt = w_G\,G(x) + w_I\,I(x) = -k + x - x^3 + \alpha(x^{REQ} - x)$;
  $w_G = 1,\ w_I = \alpha \in [0;1],\ x^{REQ} = \{-1, 1\}$

Let us first consider a straightforward case of a fully symmetrical voiced–voiceless contrast in adult speech. In the dynamical interpretation, grammar predicts a voiced–voiceless alternation in presonorant contexts, which needs to be lexically disambiguated. The control parameter in the grammar dynamics is $k = 0$ (as in (Fig. 13a) above). The values $x^{REQ} = \{-1, 1\}$ correspond to voiced and voiceless categories respectively. If the weights of both the grammar and intention dynamics are set to $w_G = w_I = 1$, we find unambiguously voiced or voiceless outputs, because the resultant force function $-dV(x)/dx = G(x) + I(x) = -x^3 + x^{REQ}$ assumes the value of zero only at one point: $x^{REQ} = -1$ for the potential in Fig. 15a *vs.* $x^{REQ} = 1$ for the potential in (b). Note that this function does not have a linear term, which precludes the possibility of having two attractors in the corresponding potential function.

4.2 Contrast emergence in the dynamical landscape

The relative weighting of the grammar and intention dynamics may be different in the child and the adult. Given the order of acquisition, intention may be expected to be weighted lower in children. If the grammar potential is symmetrical (i.e. $k = 0$), and the weight of intention is as low as 0.3, monostable combined potentials still result, as depicted in Fig. 16. Such a system will produce target-appropriate voiced and voiceless fricatives, although the distinction between the two categories will
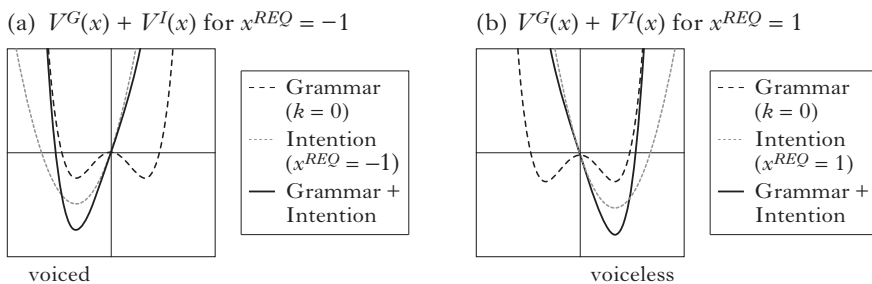
(a) $V^G(x) + V^I(x)$ for $x^{REQ} = -1$

(b) $V^G(x) + V^I(x)$ for $x^{REQ} = 1$

--- Grammar
(k = 0)

···· Intention
($x^{REQ} = -1$)

— Grammar +
Intention

voiced

--- Grammar
(k = 0)

···· Intention
($x^{REQ} = 1$)

— Grammar +
Intention

voiceless

Figure 15

Adult surface contrast as a consequence of combining Grammar and
Intention. The solid black lines represent the combined potentials.

be much less stable than in the adult system in Fig. 15 above. Nevertheless,
the systems in Figs 15 and 16 are qualitatively the same: they are charac-
terised by the same attractors, $x^{REQ} = \{-1, 1\}$.

(a) Target voiced

(b) Target voiceless

--- Grammar
(k = 0)

···· Intention
($x^{REQ} = -1$)

— Grammar +
0.3 ×
Intention

voiced

--- Grammar
(k = 0)

···· Intention
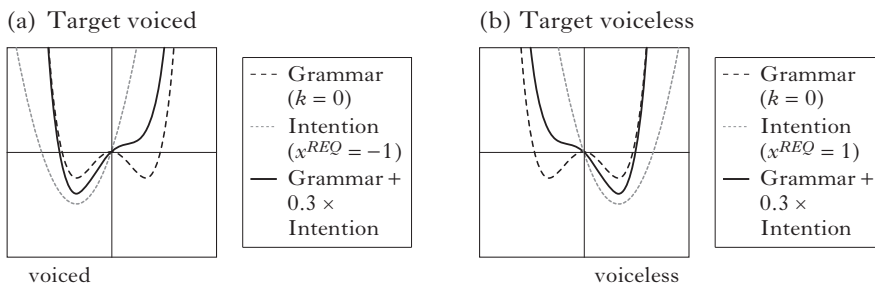($x^{REQ} = 1$)

— Grammar +
0.3 ×
Intention

voiceless

Figure 16

Weaker (less stable) surface contrast due to weaker Intention ($w_I = \alpha = 0.3$).

The mirror images of the voiced and voiceless categories in Fig. 16 are
inadequate from the point of view of the asymmetric behaviour of target
voiced and voiceless fricatives for most of the stages reported in this
paper. The initial bias towards monostable voiceless realisations and the
subsequent emergence of bistability (variation) in the target voiced frica-
tives (and obstruent clusters containing these segments) must follow
from changes in the control parameter k. As the parameter gradually
shifts from negative values upward towards zero, which reflects the
child's growing capacity to produce voicing in fricatives, the landscape
changes from asymmetric to symmetric, passing through a critical point
where a qualitative shift from a monostable (single-attractor) to a bistable
(two-attractor) system occurs. In Fig. 17, I illustrate predicted successive
stages of development of the surface voicing contrast in fricatives, assum-
ing $k = \{-0.5, -0.3, -0.1\}$. The weighting of the intention dynamics is also

assumed to shift, reflecting the expected gradual refinement of lexical representations (hypothetically proportional to the development of production skills): $w_I = \{0.15, 0.2, 0.25\}$; see Appendix C for a summary of these formulas. For simplicity, we model development only at three points: contrast neutralisation (Fig. 17a), variable contrast with weak stability (b) and still variable but more stable contrast (c). The changes in the parameter k crucially influence the modes of stability in target voiced fricatives (the (i) panels), but do not have a major effect on target voiceless fricatives (the (ii) panels). For the latter, the pattern is monostable throughout the reported stages. (The scale is the same in all graphs.)
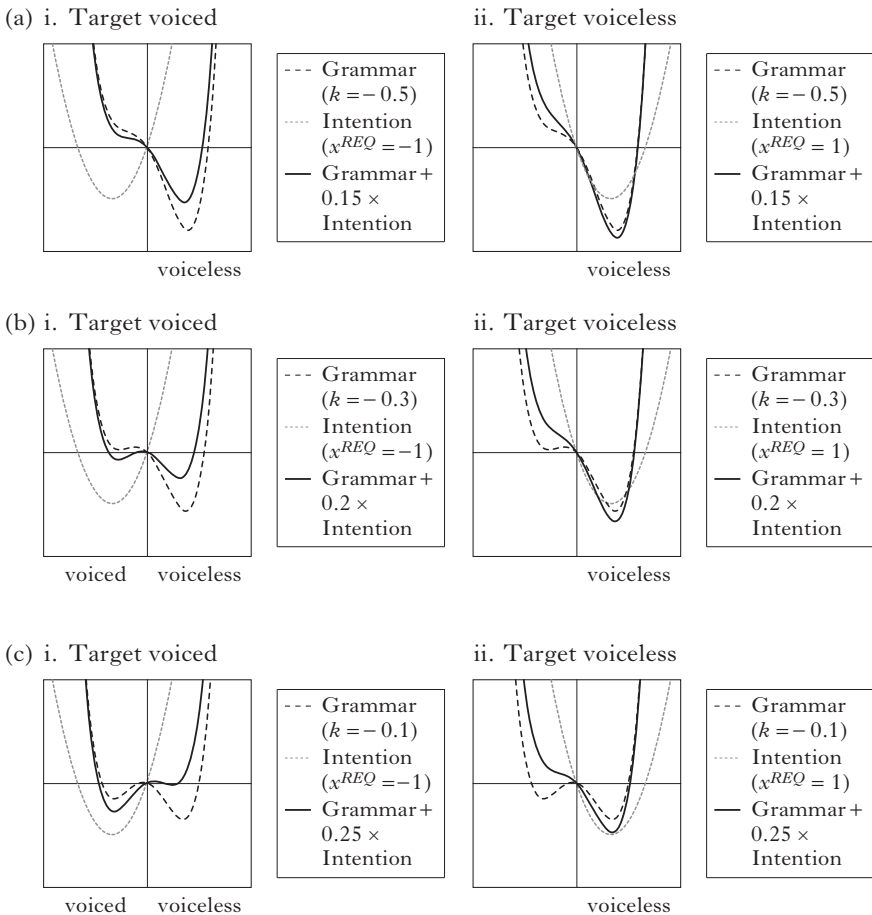

Figure 17
(a) No surface contrast: target voiced fricatives are realised as voiceless.
(b) Loss of stability in the child-specific fricative devoicing pattern: target-appropriate and voiceless renditions of target voiced fricatives concur. (c) Target voiced fricatives are realised in a nearly adult-like manner.

Fig. 17a illustrates the stage at which the voicing contrast in fricatives is neutralised completely in the child's speech. Despite differences in intention (cf. the parabolas drawn as dashed lines), target voiced (i) *vs.* target voiceless (ii) outputs are not rendered contrastively: both are characterised by a single 'voiceless' attractor. The grammatical component ($k = -0.5$), which supplies a single 'voiceless' attractor, is too strong to be outweighed by the contradictory intention ($w_I = 0.15$) supplying a weaker 'voiced' attractor in (a.i). Hence there is no qualitative difference between (a.i) and (a.ii).

Fig. 17b depicts the stage at which target voiced fricatives are produced variably as voiced or voiceless in the child's speech. The parameter k assumes a slightly higher value in comparison with the previous stage ($k = -0.3$). The linear combination of the grammar and the 'voiced' intention allows a second attractor to emerge in the combined potential (b.i). However, the value of k is still too low to yield a monostable 'voiced' pattern. With the same grammar but the opposite, i.e. 'voiceless', intention, we get the monostable 'voiceless' potential in (b.ii).

In Fig. 17c, a further shift in the parameter k towards 0 ($k = -0.1$) produces nearly mirror-image landscape patterns for target voiced (c.i) and voiceless fricatives (c.ii); cf. Fig. 16 above. In (c.i), the voiced attractor is now much more stable; the competing voiceless attractor is almost negligible. Voiceless fricatives, as in previous stages, are represented by a consistently monostable 'voiceless' potential.

The dynamical account of gradient contrast neutralisation in child speech relies on a trade-off between grammar and intention specifying contradictory attractors for target voiced segments. The same mechanism has been applied to account for incomplete neutralisation in adult speech (Gafos 2006). However, the role of grammar *vs.* intention is different in these two accounts. In adult speech, the gradient surface effect is caused by an increase in the intention weight (the lexical factor), while grammar is kept constant. In child speech, grammar is affected by the shift in the control parameter k, which reflects development of a phonetic skill.

I now turn to the question of how combining the grammar and intention dynamics predicts Voice Assimilation 'late repairs' in target voiced obstruent clusters containing fricatives in the child's system. In regressive Voice Assimilation, the voicing specification of the trigger (O_2) is anticipated in the target of assimilation (O_1). The simplest way in which this effect can be formalised within the dynamical systems approach is in terms of a linear combination of the grammatical potential for the entire cluster (with the k parameter reflecting the current capacity of producing voicing in fricatives) and the lexical (monostable) potential of the trigger (O_2). Assembling voicing for the entire cluster, rather than for each segment separately, mimics the effect of the top-ranked AGREE$_{obstr}$[±vcd] constraint in OT. On this account, the predicted development of target voiced obstruent clusters fully parallels the development of target voiced fricatives, as illustrated in Fig. 17. This is because the intention dynamics in Fig. 17 correspond to underlying voicing specifications for an obstruent in

presonorant position, which is the contrastive context, regardless of whether this obstruent is a singleton or part of a cluster, whether it is a stop or a fricative. For example, in *spytać* /spɨtatɕ/ 'to ask (PFV)' *vs.* *zbadać* /zbadatɕ/ 'to examine (PFV)' (but also in *pytać* /pɨtatɕ/ 'to ask' *vs.* *badać* /badatɕ/ 'to examine', as well as in *szyć* /ʂɨtɕ/ 'to sew' *vs.* *żyć* /ʐɨtɕ/ 'to live'), the voicing contrast in /p/ *vs.* /b/ (or /ʂ/ *vs.* /ʐ/) in presonorant position will be the basis for the voiceless *vs.* voiced intention $I(x)$. Combining it with the grammar dynamics $G(x)$ (whose attractor landscape is influenced by the control parameter k) will eventually determine the surface voicing of the entire cluster. In consequence, the same dynamics pertain to target voiced fricatives in non-assimilatory contrastive contexts and in clusters because in both the intention $I(x)$ is voiced, and the grammar $G(x)$ has the same tilt defined by the current value of the control parameter k (the capacity of combining voicing with frication). (This account ignores the lower voicing values for O_2, which were ascribed to biomechanical effects in §2.2 and §3.2 above.)

4.3 Discussion

The dynamical systems approach offers unified accounts of the interaction between categorical and gradient factors in phonological acquisition and of the enhanced gradience and variability accompanying the shift from one stable behavioural mode to another. The continuous shift in the control parameter, reflecting the growth of phonetic capacity, brings about qualitative differences in the dynamical landscape, i.e. phonological organisation.

The dynamical landscape offers a conceptualisation of the phonology–phonetics interface which does not require translating phonological representations into phonetic ones. Crucially, it does so without obliterating the distinction between categorical phonological and gradient phonetic factors. It is worth emphasising here that although symbolic dynamical representations might seem to constitute a major departure from traditional symbolic representations in phonology – as they not only link form and substance, but also structure and process, in this respect hardly resembling phonological primitives such as features – in essence they are no less discrete or symbolic than these standard representations. Like these representations, they comply with 'the tenet of constancy' (in the sense of Jackendoff 1992: 5), and constitute 'a form of mental information … an organized combinatorial space of distinctions available to the brain' (Jackendoff 1992: 3).

The idea that speech sounds are not just physiological acts, but have psychological reality, is fundamental to phonology, and appears in the earliest writings in the field (e.g. Baudouin de Courtenay 1891, Sapir 1925). In Browman & Goldstein's (1995: 177) words, 'the fundamental insight of phonology … is that the pronunciation of the words in a given language may differ from … one another in only a restricted number of ways: the number of degrees of freedom actually employed in this contrastive

behavior is far fewer than the number that is mechanically available'. However, it needs to be underscored that the reduction of the degrees of freedom taking place in the non-linear dynamical landscape goes well beyond grouping physical entities into phonological categories. An essential property of the hypothesised grammatical potential $V^G(x)$ is its capacity to evolve into a number of qualitatively different attractor regimes as a consequence of scaling the system's control parameters. Under this view, a speech sound participates in an organised space of distinctions which has a certain well-defined topology. Organisation then entails not only static relations, but also paradigms for change.

The system's evolution predicted by the grammar dynamics, $G(x)$, has systematic components which seem amenable to rigorous study: the control and order parameters of the system. A continuous shift in the control parameter (in the present study, the child's developing motor skill) results in a qualitative change in the system (here contrast emergence). Enhanced variability and gradience in the order parameter (here voicing percentage) accompany this shift. Although engaging meaningfully in the non-trivial task of identifying these parameters is clearly more challenging for cognitive systems than for mechanical systems (Thelen & Smith 1994), natural candidates for control parameters within the phonology–phonetics domain seem to be speech rate, aerodynamic and motor factors, but also communicative contexts and the lexicon; for example, Gafos (2006) proposes to model the extent of incomplete neutralisation in adult speech by scaling the dynamical system's intentional (lexical) control parameters. It is noteworthy that studying such systematic drifts in the system's behaviour is not possible in other approaches which postulate an intimate phonology–phonetics link (Gafos 2006: 57). Exemplar-based models (e.g. Pierrehumbert 2001, 2002) derive contextual (stylistic, speech rate, etc.) variation in the phonetic output from random variation over the exemplar cloud, but are unable to capture the systematicity of the system's responsiveness to environment as some parameter is varied. A similar problem is encountered within contemporary non-serialist approaches to phonology that model gradient phonetic output (e.g. unified weighted constraints; Flemming 2001), or output variation of a categorical kind (e.g. stochastic OT; Boersma & Hayes 2001). As briefly discussed below, although these approaches offer good approximations of some narrow aspects of the developing system's behaviour, they do not shed light on the central issue: the role of the phonology–phonetics interface in linking the developing phonological grammar with the child's phonetic capacity, with increased variability/gradience being the necessary symptoms of a qualitative change.

In the unified approach of Flemming (2001), weighted constraints can yield intermediate outputs in response to competing demands. In OT, child-specific fricative devoicing can be expressed in terms of a high-ranked *FRIC$_{[+vcd]}$. Assigning progressively less weight to *FRIC$_{[+vcd]}$ (in this approach understood as a MINIMISEEFFORT constraint) relative to IDENT$_{[+obstr, +vcd]}$ will incrementally diminish the influence of the former constraint on the child's output at subsequent stages of development.

(Conversely, assigning more weight to the lexical factor, e.g. via paradigm uniformity, can account for incomplete neutralisation in adult speech; see Braver 2019.) The unified approach can thus predict the gradual increase in the mean voicing percentage in fricatives. However, what falls outside the scope of this approach is the extent of gradience, which suggests that what is at stake is the child's developing psychomotor control of voicing, i.e. honing production skills to combine voicing with frication, rather than relaxing a biomechanical threshold on the voicing percentage that the production system can attain. Furthermore, because the distinction between categoricity and gradience is not expressed in this approach, it also loses sight of the non-linear character of the phonology–phonetics link. As a result, adopting unified weighted constraints is not helpful in predicting the categorical voiced–voiceless switches in target [+voiced] obstruent clusters. Instead of the reported polarised patterns recurring across the stages, we would expect an upward movement of a single collection of data points over time. This is not what happens.

Variable interaction between categorical and gradient factors in child speech reported in this paper also falls outside the scope of classical stochastic OT models designed to account only for categorical variation (Boersma & Hayes 2001). (Likewise, incomplete neutralisation in adult speech, whose extent varies depending on the communicative context, is not covered by these models; see Gafos 2006.) Stochastic OT ranks constraints on a scale and adds a random value drawn from a Gaussian distribution during the evaluation process; the presence of evaluation noise results in variable rankings, hence categorical variation in the output. Assuming Stochastic OT, the grammar shift observed in acquisition can be approximated in the following way. *$\text{FRIC}_{[+\text{vcd}]}$ can be ranked on the top of the hierarchy and then gradually shifted downwards relative to $\text{IDENT}_{\text{preson}}[\pm\text{vcd}]$. (At the same time, the comprehension grammar exhibits the adult-like ranking acquired at the primary stages of acquisition, as outlined in §2.2.) A substantial overlap between the two constraints in terms of their Gaussian distributions will produce variable rankings: *$\text{FRIC}_{[+\text{vcd}]} \gg \text{IDENT}_{\text{preson}}[\pm\text{vcd}]$ and $\text{IDENT}_{\text{preson}}[\pm\text{vcd}] \gg$ *$\text{FRIC}_{[+\text{vcd}]}$, and, in consequence, categorical variation in the output (both in singleton voiced fricatives and in clusters, given that $\text{AGREE}_{\text{obstr}}[\pm\text{vcd}]$ also starts top-ranked). The degree of overlap between the two constraints, and hence of categorical variation in the output, will gradually decrease over time, eventually matching the stable ranking of the two constraints in the comprehension grammar. In this account, the gradual maturation of voicing control in the child is irrelevant in attaining the adult-like grammar. Somewhat paradoxically then, the motivation for the separate production grammar component with the top-ranked *$\text{FRIC}_{[+\text{vcd}]}$, in defiance of the adult stimulus, is the child's initial lack of phonetic skills, yet changes in the production grammar are prompted by the adult stimulus, not by development of the child's psychomotor control of voicing.

In the dynamical landscape, even though child-specific patterns are viewed as having their origin in performance-related factors, they do not fall outside the purview of the phonological grammar. Formalising the development of the phonetic skill as a continuous shift in the control parameter of a dynamical system accounts for a qualitative change from one stable mode to another (in this paper, the absence or presence of contrast in voicing) through an intermediate stage of enhanced variation and gradience. Such discontinuities in development seem to find solid support in the widely reported characteristics of child-specific substitution and reduction patterns. The interaction between child-specific patterns and adult-based phonology requires further empirical research, involving both longitudinal observations and controlled experimentation.

5 Conclusions

This paper has provided an argument for the computational relevance of child-specific patterns, based on the interaction between fricative devoicing, a phonetically gradient child-specific pattern, and Voice Assimilation, a categorical process of adult Polish, reported in the longitudinal data from a Polish-speaking child. The interaction is not predicted by the traditional modular feedforward approach: the phonetic character of the child-specific pattern (lack of phonetic skill) did not prevent it from conditioning the application of the adult-based phonological process. The acquisitional data were analysed applying insights from the dynamical systems approach. The dynamical landscape, expressed formally in the universal language of non-linear mathematics, integrates the discrete and continuous aspects of speech, while keeping them distinct. It also offers us tools for the study of the underlying mechanisms of change. The proposal put forward in this paper was to express child-specific patterns, arising out of performance-related effects (lack of production skills) as consequences of shifts in control parameters of a dynamical system (phonological organisation). The symbolic dynamical representations were found to provide a convenient expression of (i) temporary systematicity, (ii) enhanced gradience during transition towards target-appropriate productions and (iii) the ultimate disappearance of child-specific patterns, all of which are widely known characteristics of these patterns. They also predict a reciprocal relationship between child-specific patterns and adult-based phonological organisation in the child, as reported in this study. Such interactions have largely been overlooked in previous research on phonological acquisition, and await further investigation as an interesting source of information about the phonology–phonetics link in developing systems.

REFERENCES

Baudouin de Courtenay, Jan Niecisław (1891). O zadaniach językoznawstwa. *Prace Filologiczne* **3**. 92–115.

Becker, Michael & Anne-Michelle Tessier (2011). Trajectories of faithfulness in child-specific phonology. *Phonology* **28**. 163–196.

Boersma, Paul & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *LI* **32**. 45–86.

Boersma, Paul & David Weenink (2017). *Praat: doing phonetics by computer*. Version 6.0.21. http://www.praat.org/.

Braver, Aaron (2019). Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology* **36**. 1–36.

Browman, Catherine P. & Louis Goldstein (1989). Articulatory gestures as phonological units. *Phonology* **6**. 201–251.

Browman, Catherine P. & Louis Goldstein (1995). Dynamics and articulatory phonology. In Robert F. Port & Timothy van Gelder (eds.) *Mind as motion: explorations in the dynamics of cognition*. Cambridge, MA: MIT Press. 175–193.

Bryndal, Małgorzata (2015). *Fonologiczna interpretacja procesów doskonalenia wymowy dziecięcej na tle współczesnych teorii fonologicznych*. Gliwice: Komlogo.

Eimas, Peter D., Einar R. Siqueland, Peter Jusczyk & James Vigorito (1971). Speech perception in infants. *Science* **171**. 303–306.

Flemming, Edward (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* **18**. 7–44.

Gafos, Adamantios I. (2006). Dynamics in grammar: comment on Ladd and Ernestus & Baayen. In Louis Goldstein, D. H. Whalen & Catherine T. Best (eds.) *Laboratory phonology 8*. Berlin & New York: Mouton de Gruyter. 51–79.

Gafos, Adamantios & Stefan Benus (2006). Dynamics of phonological cognition. *Cognitive Science* **30**. 905–943.

Gafos, Adamantios I., Simon Charlow, Jason A. Shaw & Philip Hoole (2014). Stochastic time analysis of syllable-referential intervals and simplex onsets. *JPh* **44**. 152–166.

Gussmann, Edmund (2007). *The phonology of Polish*. Oxford: Oxford University Press.

Hale, Mark & Charles Reiss (2008). *The phonological enterprise*. Oxford & New York: Oxford University Press.

Haman, Ewa, Bartłomiej Etenkowski, Magdalena Łuniewska, Joanna Szwabe, Ewa Dąbrowska, Marta Szreder & Marek Łaziński (2011). *Polish CDS corpus*. Available (February 2021) at https://childes.talkbank.org/access/Slavic/Polish/CDS.html.

Hamann, Silke (2004). Retroflex fricatives in Slavic languages. *Journal of the International Phonetic Association* **34**. 53–67.

Hayes, Bruce (1999). Phonetically driven phonology: the role of Optimality Theory and inductive grounding. In Michael Darnell, Edith Moravcsik, Frederick Newmeyer, Michael Noonan & Kathleen Wheatley (eds.) *Functionalism and formalism in linguistics*. Vol. 1: *General papers*. Amsterdam & Philadelphia: Benjamins. 243–285.

Hayes, Bruce (2004). Phonological acquisition in Optimality Theory: the early stages. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in phonological acquisition*. Cambridge: Cambridge University Press. 158–203.

Hayes, Bruce & James White (2013). Phonological naturalness and phonotactic learning. *LI* **44**. 45–75.

Jackendoff, Ray (1992). *Languages of the mind: essays on mental representation*. Cambridge, MA: MIT Press.

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

Kaczmarek, Leon (1988). *Nasze dziecko uczy się mowy*. Lublin: Wydawnictwo Logopedyczne.

Keating, Patricia A. (1988). The phonology–phonetics interface. In Frederick J. Newmeyer (ed.) *Linguistics: the Cambridge survey*. Vol. 1: *Grammatical theory*. Cambridge: Cambridge University Press. 281–302.

Kelso, J. A. S., E. L. Saltzman & B. Tuller (1986a). The dynamical perspective on speech production: data and theory. *JPh* **14**. 29–59.

Kiparsky, Paul & Lise Menn (1977). On the acquisition of phonology. In John Macnamara (ed.) *Language learning and thought*. New York: Academic Press. 47–78.

Kowalik, Krystyna (1998). Morfonologia. In Renata Grzegorczykowa, Roman Laskowski & Henryk Wróbel (eds.) *Gramatyka współczesnego języka polskiego: morfologia*. Warsaw: Wydawnictwo Naukowe PWN. 87–123.

Krajna, Ewa (2008). *100-wyrazowy test artykulacyjny*. Gliwice: Komlogo.

Kuhl, Patricia K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* **50**. 93–107.

Ladefoged, Peter & Ian Maddieson (1996). *The sounds of the world's languages*. Oxford & Malden, MA: Blackwell.

Łobacz, Piotra (1996). *Polska fonologia dziecięca: studia fonetyczno-akustyczne*. Warsaw: Wydawnictwo Energeia.

Lorenc, Anita (2018). Articulatory characteristics of Polish retroflex sibilants: an analysis using electromagnetic articulography. *Logopedia* **47**. 103–121.

Łukaszewicz, Beata (2000). *Phonological analysis of reduction and assimilation processes in the acquisition of English and Polish*. PhD dissertation, University of Warsaw.

Łukaszewicz, Beata (2006a). Acquisition of the ±back contrast in Polish. In Piotr Bański, Beata Łukaszewicz & Monika Opalińska (eds.) *Studies in constraint-based phonology*. Warsaw: Wydawnictwa Uniwersytetu Warszawskiego. 89–109.

Łukaszewicz, Beata (2006b). Extrasyllabicity, transparency and prosodic constituency in the acquisition of Polish. *Lingua* **116**. 1–30.

Łukaszewicz, Beata (2007). Reduction in syllable onsets in the acquisition of Polish: deletion, coalescence, metathesis and gemination. *Journal of Child Language* **34**. 53–82.

Łukaszewicz, Beata & Bogdan Rozborski (2008). Korelaty akustyczne akcentu wyrazowego w języku polskim dorosłych i dzieci. *Prace Filologiczne* **54**. 265–283.

McCarthy, John J. & Alan Prince (1995). Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*. Amherst: GLSA. 249–384.

MacNeilage, Peter F. (1997). Acquisition of speech. In William J. Hardcastle & John Laver (eds.) *The handbook of phonetic sciences*. Oxford & Cambridge, MA: Blackwell. 301–332.

Maddieson, Ian (2013). Voicing in plosives and fricatives. In Matthew S. Dryer & Martin Haspelmath (eds.) *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info.

Menn, Lise (1976). *Pattern, control, and contrast in beginning speech: a case study in the development of word form and word function*. PhD dissertation, University of Illinois, Champaign-Urbana.

Menn, Lise (1978). Phonological units in beginning speech. In Alan Bell & Joan Bybee Hooper (eds.) *Syllables and segments*. Amsterdam: North Holland. 157–171.

Ostaszewska, Danuta & Jolanta Tambor (2012). *Fonetyka i fonologia współczesnego języka polskiego*. Warsaw: Wydawnictwo Naukowe PWN.

Pater, Joe & Jessica A. Barlow (2003). Constraint conflict in cluster reduction. *Journal of Child Language* **30**. 487–526.

Pater, Joe, Christine L. Stager & Janet F. Werker (2004). The perceptual acquisition of phonological contrasts. *Lg* **80**. 384–402.

Pegg, Judith E. & Janet F. Werker (1997). Adult and infant perception of two English phones. *JASA* **102**. 3742–3753.

Peperkamp, Sharon, Rozenn Le Calvez, Jean-Pierre Nadal & Emmanuel Dupoux (2006). The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition* **101**. B31–B41.

Pierrehumbert, Janet B. (2001). Exemplar dynamics: word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*. Amsterdam & Philadelphia: Benjamins. 137–157.

Pierrehumbert, Janet B. (2002). Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology* 7. Berlin & New York: Mouton de Gruyter. 101–139.

Port, Robert F. & Penny Crawford (1989). Incomplete neutralization and pragmatics in German. *JPh* **17**. 257–282.

Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

Roon, Kevin D. & Adamantios I. Gafos (2016). Perceiving while producing: modeling the dynamics of phonological planning. *Journal of Memory and Language* **89**. 222–243.

Rubach, Jerzy (2008). Prevocalic faithfulness. *Phonology* **25**. 433–468.

Sapir, Edward (1925). Sound patterns in language. *Lg* **1**. 37–51.

Smith, Caroline L. (1997). The devoicing of /z/ in American English: effects of local and prosodic context. *JPh* **25**. 471–500.

Smith, Neil V. (1973). *The acquisition of phonology: a case study*. Cambridge: Cambridge University Press.

Smith, Neil V. (2010). *Acquiring phonology: a cross-generational case-study*. Cambridge: Cambridge University Press.

Smolensky, Paul (1996). On the comprehension/production dilemma in child language. *LI* **27**. 720–731.

Sobkowiak, Włodzimierz (1996). E*nglish phonetics for Poles: a resource book for learners and teachers*. Poznań: Bene Nati.

Sołtys-Chmielowicz, Anna (1998). *Wymowa dzieci przedszkolnych*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Stager, Christine L. & Janet F. Werker (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* **388**. 381–382.

Steriade, Donca (1999). Phonetics in phonology: the case of laryngeal neutralization. *UCLA Working Papers in Linguistics: Papers in Phonology* **3**. 25–145.

Strogatz, Steven H. (1994). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering*. Reading, MA: Perseus.

Strycharczuk, Patrycja (2012). Sonorant transparency and the complexity of voicing in Polish. *JPh* **40**. 655–671.

Tamási, Katalin, Cristina McKean, Adamantios Gafos & Barbara Höhle (2019). Children's gradient sensitivity to phonological mismatch: considering the dynamics of looking behavior and pupil dilation. *Journal of Child Language* **46**. 1–23.

Tesar, Bruce & Paul Smolensky (1993). The learnability of Optimality Theory: an algorithm and some basic complexity results. Ms, University of Colorado, Boulder. Available as ROA-2 from the Rutgers Optimality Archive.

Thelen, Esther & Linda B. Smith (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Tuller, Betty, Pamela Case, Mingzhou Ding & J. A. Scott Kelso (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance* **20**. 3–16.

Vihman, Marilyn M., Rory A. DePaolis & Tamar Keren-Portnoy (2015). Babbling and words: a dynamic systems perspective on phonological development. In Edith L. Bavin & Letitia R. Naigles (eds.) *The Cambridge handbook of child language*. Cambridge: Cambridge University Press. 207–229.

Werker, Janet F. & Suzanne Curtin (2005). PRIMIR: a developmental framework of infant speech processing. *Language Learning and Development* **1**. 197–234.

Werker, Janet F. & Richard C. Tees (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* **7**. 49–63.