Young children's speech: from a simple phonological rule to the murky phonetic reality

Claartje Levelt
LUCL, Leiden University

Word productions of children between 1;0 and 3;0 often deviate phonologically from adult targets. This paper revisits a textbook case of child phonological development — the neutralization of the voicing contrast in onset stops — and shows how it becomes both messier and more interesting once acoustic measurements are allowed into the picture. Strikingly, the direction of this neutralization appears to differ between Dutch- and English-acquiring children: Dutch children favor voiceless realizations, while English children voiced ones. Acoustic measurements explain why Dutch- and English-acquiring children seem to neutralize the contrast in opposite directions, while longitudinal measurements (all measurements using Praat of course) from a Dutch-acquiring child uncover a U-shaped VOT trajectory and, tucked inside it, a covert contrast that impressionistic transcription alone would have missed entirely. A follow-up production experiment shows that toddlers can actively enhance this contrast in a minimal-pair context, hinting at real phonological knowledge beneath the phonetic noise. The moral: categories exist from early on in developing speakers, but are realized by gradient, acoustically subtle means, a conclusion that is right up Paul Boersma's alley.

1 Introduction

During my linguistic education – in the eighties and nineties – phonetics and phonology were considered to be two very separate fields, and in Leiden, where I received my education, phonologists and phoneticians even resided in different buildings. Psycholinguistics was – and might still be? – considered another completely separate field, not to be taken seriously if you were a (generative) phonologist. It was, therefore, initially quite mindboggling and even exhausting for me to interact with Paul Boersma, who did not respect the neat borders between these fields – and who could also argue about this non-stop for nine hours, as I discovered on flights to and back from a conference in California. Another challenge for me in collaborations with Paul was to "let go" of categories, or at least, to welcome "gradient" things. Overcoming both of these challenges has enriched my view of phonological development and my approach to word

Linguistics in Amsterdam, 2026: 8–18
© Claartje Levelt

production data of young speakers, culminating in an NWO-VIDI project where child language phonology was studied using both perception and production experiments, and acoustic measurements of production data - that were of course always carried out in Praat (Boersma & Weenink 2023). A recent NWO-L project together with Paula Fikkert again combines these methods to tease out development at different levels and their interactions even more. I hope to illustrate below how this encompassing view renders a seemingly simple case of phonological development more complex (murky!), but also more interesting and relevant. Inspiration for this illustration came from my contribution "Do young children's realizations reveal rigorous representations?" to the workshop "The reality of phonological representations," March 6, 2014 at the University of Amsterdam, organized by Paul Boersma and Kateřina Chládková in honor of Kateřina's PhD defense.

2 Phonologically deviating productions

It is well-known that word productions from children between 1;0 and 3;0 years of age often deviate phonologically from realizations of those words by more experienced speakers. One of these well-known deviations has to do with voicing, as illustrated in (1) with examples from Dutch (CLPF corpus, Levelt 1994; Fikkert 1994) and English child language (Smith 1973):

(1) Productions of words starting with a plosive consonant

a.	beer	'bear'	[pi]	Robin 1;7.13	Dutch
b.	dier	'animal'	[tiɹ]	Robin 1;10.21	
c.	pain		[be:n]	Amahl ±2;5	English
c.	tell		[dɛl]	Amahl ±2;6	

The question that productions like those in (1) raise is how they come about, and trying to answer this question requires thinking about the nature of (developing) underlying representations, the developing grammar, and restrictions on motor planning and articulation. Are representations initially un(der)specified or are output representations and/or articulations initially heavily constrained, or both? Another question that the data in (1) raise is: why does the direction of change seems to be different for Dutch- and English- acquiring children? Let us first address this last question, which immediately illustrates how acoustic information can inform an apparent phonological dilemma.

3 The voicing contrast in Dutch and English

Smith (1973), in his analysis of the phonological development of his son Amahl (A.), proposes a set of so-called Realization Rules that determine A.'s sound productions, one of which states that all consonants are voiced – "…this rule gets rid of the adult voicing contrast…" (Smith 1973: 22) by neutralizing it to [+voice] realizations, as in (1c, d). For Dutch data like in (1a, b), the rule [-sonorant] → [-voice] could be posited, which also gets rid of the adult voicing contrast, but by neutralizing it to [-voice] realizations. Kager et al. (2007), focusing on toddlers' perception and production of stop consonants, point out that this apparent inverse developmental situation is solved when we take into account the VOT values of stops. While in Dutch the [±voice] contrast is phonetically realized as a voicing lead, with an average VOT of around -80 ms for [+voice] stops, versus a short lag VOT for [-voice] stops (between 0–25 ms), in English the contrast is realized by a short lag VOT for [+voice] stops (between 0–25 ms) versus a long lag VOT (aspiration) for [-voice] stops (25 ms or longer) (Kager et al. 2007). The apparent discrepancy between the acquisition of the voicing contrast in English and in Dutch thus disappears when we focus on the acoustics: a short lag VOT seems to be the default VOT realization for developing speakers cross-linguistically (see among others, Kewley-Port & Preston 1974; Macken & Barton 1980; Scobbie et al. 1996). The problem in phonology, where a short lag VOT is perceived as [+voice] in English and as [-voice] in Dutch, can then be solved if the contrast in English is defined in terms of the feature [delayed release] – where [+delayed release] refers to aspiration - instead of [voice], with a default [-delayed release] realization, while sticking to [voice] in Dutch, with a default [-voice] realization (Kager et al. 2007).

Focusing on stop consonants in Dutch, the contrast that needs to be acquired at the phonological level is thus between [-voice] and [+voice], while at the phonetic level it is between a short lag VOT and a voicing lead. Can we determine whether the *source* of the errors in (1a, b) is the phonological representation or the phonetic/articulatory implementation of the contrast? In the first case, which we call the Phonological Underspecification option (Archangeli 1988; Lahiri & Reetz 2002) there is no [±voice] contrast in the lexical representation resulting in a default [-voice] on the surface, i.e., [Øvoice] → [-voice]. In the second case, which we call the Phonetic Underspecification option (Keating 1988), the contrast might be phonologically specified, but both [-voice] and [+voice] are fitted/articulated with a short lag VOT, neutralizing the contrast at the articulatory-phonetic level. These options can be depicted as in Figure 1:

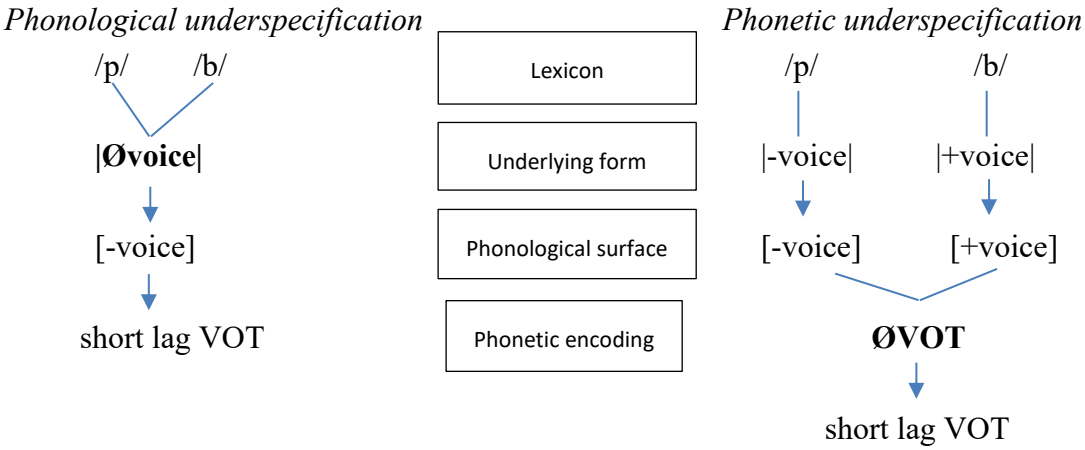


Figure 1: Possible sources of voicing neutralization

However, are we dealing with genuine neutralizations at all here? A third option is that we are in fact dealing with a so-called covert contrast, which Macken and Barton (1980) lucidly describe as "[…] *the contrast in the mouth of the child is somehow inadequate for the ear of the listener.*" That is, there is an underlying phonological contrast, which the young speaker is phonetically expressing, but the acoustic means by which the contrast is expressed are not according to the listener's perceptual norms and it is therefore not noticed.

4 Covert contrast

While the phenomenon of covert contrast in child language productions already received attention in the work by Macken and Barton (1980), and Scobbie et al. (1996) concluded that it is probably a general stage in the acquisition of phonology and phonetics, in-depth acoustic analyses are, to my knowledge, still not customary in studies of sound changes in child language that lead to perceived homophones/neutralizations. Scobbie et al. (1996: 57), in their conclusion, give a very compelling reason to do make this part of these analyses: "*The existence of covert contrast raises the possibility of a secret developmental life inaccessible to impressionistic phonetics. Instrumental analyses of child speech give us access to the child's development of phonetic skills, so we can follow the independent development of the various cues to a contrast (and the gestures underlying the cues), rather than judging the existence of the contrast in an all-or-nothing fashion.*"

However compelling, I don't remember that this registered with me before I started to collaborate with Paul! Taking a little detour before returning to VOT, I want to highlight joint work on a different type of neutralization that occurs in child language, namely between reduced complex onsets and singleton onsets, which illustrates that the exact acoustic property that a developing speaker uses

to express a specific contrast can be very subtle. Toddlers initially reduce complex onsets, like /br/ in the Dutch word *broek* ('trousers') [bruk] to singleton onsets, like [b]. The word *broek* is then produced [buk], and now forms a homophone with the Dutch word *boek* ('book') [buk]. As part of my NWO-VIDI project, Margarita Gulian focused on the development of onset clusters in her PhD research. Together with Paul, in one study (Gulian 2017: Chapter 2[1]) we set out to determine whether apparent homophones like [buk] from /bruk/ and [buk] from /buk/ were acoustically equivalent or would manifest some difference that would establish a covert contrast. Margarita compared 47 pairs of plosive + vowel sequences, each pair produced by a single child, where the plosive in one member of the pair came from a reduced target onset cluster, while the plosive in the other member of the pair came from a singleton target onset. F2 and F3 were measured at two time points in the vowels following these plosives, at the immediate vowel onset and at one quarter of the duration of the vowel. A significant difference was found, namely a raise of F2 in the vowel following the target cluster that was not present in the vowel following the target singleton plosive. This raise resembles the raise of F3 in vowels following adult productions of rhotic + vowel sequences, especially because the F2 in the children's productions is at the same height as F3 productions in adult productions. It might thus constitute a trace of the rhotic that disappeared from the child's production. That this raise indeed established a *covert* contrast was subsequently shown in forced choice experiment where adult listeners were presented with the plosive + vowel sequences and had to determine whether the target word had contained an onset cluster or not. Instead of picking up the "raise/no raise of F2" cue, the answers of the adult participants relied on durational differences between the CV sequences to determine the target. That is, CV-sequences with a slightly longer V were associated with "target word contains a cluster" answers, while duration was in fact not a reliable cue to target words having a cluster or not. Let us now return to VOT.

Using Praat, VOT was measured from the burst of an onset plosive to the first zero crossing in periodicity in the subsequent vowel, in the productions of 125 target onset /b/s and 68 target onset /p/s of the Dutch-acquiring child Robin. These were all the productions of target /b/s and /p/s in the longitudinal dataset of bi-weekly recordings of this child (CLPF database: Fikkert 1994; Levelt 1994) that had a clear burst. Measurements revealed a lot of variability in VOT, specifically for the target voiced labial stops. They also unexpectedly revealed a U-shaped development of the VOT realization of target voiced labial stops, with a predominant voicing lead in the first five recordings, followed by a period of seven recordings with more and more short-lag VOTs, and finally a, now solid, voicing lead in the last three recordings. This might have been influenced by the

[1] Hopefully retirement will finally provide Paul with time to help finish and publish this study!

development of the vocabulary: in the initial recordings only three different /b/-initial target words were produced and two /p/-initial target words. The increase of short-lag realizations of /b/-initial targets thus coincided with an increase in target /b/- and /p/-initial words in the vocabulary, something that merits future research.[2] However, even in the period where target /b/s were predominantly produced with a short-lag VOT, the mean VOT values for target /b/s and target /p/s were statistically different from each other, showing a covert contrast akin to the one Macken and Barton (1980) had found in the English child language data.

5 Phonetic enhancement of contrast: an experiment

The VOT measurements of Robin's production data seem to indicate that there is a representational difference between /b/ and /p/ in Robin's phonology from the start, and that the main developmental challenge is to figure out the exact, language-specific phonetic encoding of this difference. If it is indeed the case that the contrast between voiced and voiceless stops is phonologically represented (in some way) in young speakers, and the development effort lies in the phonetic encoding of the VOT values that express this contrast, then it might be possible to experimentally provoke an enhancement of the VOT contrast in children's productions, just like Baese-Berk and Goldrick (2009) were able to do with English-speaking adults. The question in their study was whether adult speakers would highlight a phonological contrast by means of phonetic enhancement if the sounds displaying this contrast would appear in words that formed a minimal pair, like /k/ and /g/ in *cod–god*. All test items in the Baese-Berk and Goldrick experiment consisted of monosyllabic words with voiceless, i.e., [+delayed release] plosives in the word-onset but they would crucially differ in whether they had a minimal pair lexical neighbor (like in *cod–god*) or not (like in *cat–*gat*). In their Experiment 2, pairs of speakers and listeners were seated on opposite sides of a table, each in front of a computer monitor. Three words would simultaneously appear on both screens, and one word would be highlighted on the speaker's monitor. The speaker then had to ask the listener to click on that word on the listener's monitor, by saying "Click on the *target word*". There were three conditions: *Context*, in which the minimal pair test items appeared with their minimal pair neighbor in the set of three words (like in *cod–god–yell*), *No Context*, in which the minimal pair test items appeared without their minimal pair neighbor (*cod–yell–lamp*), and *No Competitor*, in which the test-item without a minimal pair lexical neighbor formed one of the three words (*cat–lamp–yell*). The VOT

[2] An anonymous reviewer suggests several other angles that could be explored, such as the variability in VOT for voiced stops in adult speakers of Dutch (van Alphen & Smits 2004) and other acoustic cues that highlight the [+voice]/[-voice] contrast (Slis & Cohen 1969).

values of all produced [+delayed release] word-initial test-trials in the three conditions were measured and compared in a statistical analysis. It turned out that the VOT value of the [+delayed release] plosive was most enhanced in the *Context* condition, and that the test items in the *No Context* condition still had significantly longer VOT values than items in the *No Competitor* condition.

The finding that participants increased aspiration in [+delayed release] stops in a minimal pair context indicated to me that speakers could more or less consciously manipulate the VOT value in this context, and because the result was a contrast enhancement, this manipulation must hinge on this contrast being phonologically represented in the lexicon. Therefore (then) bachelor student Laura de Rooij and I set out to create a toddler-friendly, Dutch version of this experiment (De Rooij & Levelt 2017), where we would focus on the VOT values of target voiced stops. Our reasoning was that children's produced VOT values of initial plosives in *Context* versus *No Context* or *No Competitor* conditions could show whether the voicing contrast was represented or not. Since we were testing two-year-old children, we changed the experiment into a picture-naming version with two, instead of three, items per test trial, while initially keeping the *Context*, *No Context* and *No Competitor* conditions.

Stimuli consisted of 29 depictable words that were either familiar to two-year-old children according to Dutch CDI data or that they would be able to learn in a practice session at home. This resulted in 8 minimal pairs of words starting either with a voiced or a voiceless labial plosive, 4 words starting with a voiced labial plosive that did not have a voiceless minimal pair neighbor, and 9 control words that did not start with a plosive and did not rhyme with any of the test items. All items are listed in Table 1 below.

Table 1: The 20 test items and 9 control items in the production experiment.

	Words	**Gloss**
Minimal pairs	baard–paard	beard–horse
	bed–pet	bed–cap
	beer–peer	bear–pear
	boot–poot	boat–paw
	botje–potje	bone(dim)–potty
	boef–poef	thief–hassock
	bakje–pakje	container–present
	bad–pad	bath–toad
Voiced Labial-initial words without a voiceless neighbor	bal	ball
	bus	bus
	boom	tree
	boek	book
Control words	stoel	chair
	hond	dog
	sok	sock
	vis	fish
	aap	monkey
	hand	hand
	kip	chicken
	fiets	bicycle
	koe	cow

Colorful pictures of these words were arranged in pairs on a table-top flipover. In the *Context* condition the two pictures depicting a minimal pair of words occurred next to each other, in the *No-Context* condition the picture of the voiced plosive item of a minimal pair occurred next to the picture of a control item, and in the *No-Competitor* context the picture of a test item without a minimal pair neighbor occurred next to the picture of a control item. Examples of these stimuli are in Figure 2.

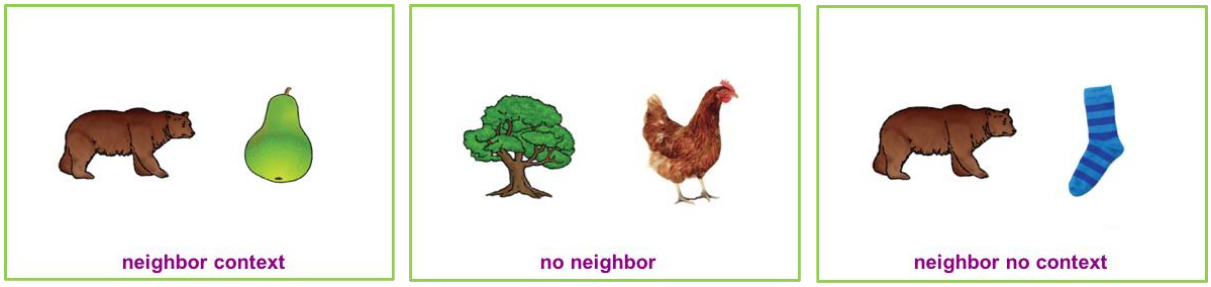


Figure 2: Examples of test trials in the three different conditions.

All pictures of test-items starting with voiced plosives that had a minimal pair neighbor were presented in the *No Context* condition before they were presented in the *Context* condition. In the *Context* condition, participants were asked to

name the picture depicting the word with the voiceless plosive first, reasoning that this production would then trigger a different VOT in the subsequent production of the target voiced counterpart, i.e., it would trigger a VOT manipulation.

We tested 18 children (10 boys, 8 girls) from monolingual Dutch-speaking homes, with a mean age of 23;25 months (range 22;11–24;15). From these participants we collected 300 productions of target /b/s and /p/s and measured their VOT values in Praat. Because the No Neighbor set was quite small, and no significant differences in VOT were found between the /b/s in the No Neighbors context and the /b/s in the No Context condition, we grouped them in a No-Trigger condition and contrasted them to the /b/s in the Context condition, renamed as Trigger condition. A Linear Mixed Effects analysis showed that there was no overall statistical difference between the VOT values of all produced target /b/s – independent of condition - and produced target /p/s (p = .3992). However, in the Trigger condition, target /b/s were produced with significantly reduced VOT values compared to those produced in the no-trigger condition (p = .0285), i.e., the VOT values in this condition moved in the direction of a voicing lead (Levelt & de Rooij 2017). This again shows that the [±voice] contrast is specified. In addition, despite the fact that the phonetic encoding of the contrast is still problematic in general, the manipulation of VOT in the target /b/ productions in the Trigger condition shows that some control over this encoding is present. In the current NWO-L project this experiment has been fine-tuned[3] and the productions of 50 participants are currently being analyzed!

6 Conclusions

The data in (1) seemed to illustrate a clear-cut developmental stage in which young speakers tend to neutralize the voicing contrast in word-onset stops. For Dutch, this stage could be captured by positing the simple phonological rule [-sonorant] → [-voice], or by assuming that stops were underlyingly unspecified for [voice], and receive a default [-voice] interpretation on the surface, i.e., [Øvoice] → [-voice]. It then turned out that this stage might actually be defined by an acoustic rather than a phonological situation, where developing speakers cross-linguistically prefer a short-lag VOT for onset stops. However, measurements in longitudinal data showed that the VOT values of target labial [+voice] stops displayed a U-shaped development; an initial stage with adult-like negative values, i.e., a voicing lead, was followed by a period of short-lag VOTs – coinciding with an increasing number of target words with onset labial stops in the vocabulary – which was followed by a stable voicing lead VOT for target

[3] It now has a 2x2 design for example, where both /p/'s and /b/'s feature in "trigger" and "no-trigger" conditions.

[+voice] stops. In addition, the data revealed the presence of a covert contrast: even in the short-lag VOT production stage a statistically significant difference in VOT values was found for target [+voice] stops and target [-voice] stops, which suggests that the [±voice] contrast is phonologically specified from the start. This was corroborated in a production experiment in which toddlers turned out to be able to manipulate the VOT value of a target [+voice] labial stop, producing a more reduced VOT when they had just produced the target [-voice] labial stop in a minimal pair context (like in: peer 'pear' /per/ - beer 'bear' /ber/).

While on the one hand these studies show that there are phonological categories and contrasts, and that young language learners do seem to be aware of them, they also show that we have to keep our ears and eyes open for all the gradient and acoustically subtle ways in which they are trying to phonetically encode them.

Thank you Paul for your probably unwitting role in my appreciation of this trove of information – and for Praat, which makes it possible to discover these developmental treasures!

7 References

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology 5*(2): 183–207. https://doi.org/10.1017/S0952675700002268

Boersma, P., & Weenink, D. (2023). Praat. *Doing phonetics by computer (Version 6.3.10)*, http://www.praat.org

Fikkert, P. (1994). *On the acquisition of prosodic structure*. [Doctoral dissertation, Universiteit Leiden]. Holland Institute of Generative Linguistics.

Gulian, M. E. (2017). *The development of the speech production mechanism in young children: evidence from the acquisition of onset clusters in Dutch*. [Doctoral dissertation, Universiteit Leiden], LOT Dissertations in Linguistics.

Kager, R., Van der Feest, S., Fikkert, P., Kerkhoff, A., & Zamuner, T. S. (2007). Representations of [voice]: Evidence from acquisition. In J. van de Weijer & E. van der Torre (Eds.), *Voicing in Dutch: (De)voicing – phonology, phonetics, and psycholinguistics* (pp. 41–80). John Benjamins. https://doi.org/10.1075/cilt.286.03kag

Keating, P. A. (1988). Underspecification in phonetics. *Phonology 5*(2): 275–292. https://doi.org/10.1017/S095267570000230X

Kewley-Port, D., & Preston, M. S. (1974). Early apical stop production: A voice onset time analysis. *Journal of Phonetics 2*(3): 195–210. https://doi.org/10.1016/S0095-4470(19)31270-7

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology* 7 (pp. 637–676). Walter de Gruyter.

Levelt, C. (1994). *On the acquisition of Place*. [Doctoral dissertation, Universiteit Leiden]. Holland Institute of Generative Linguistics.

Levelt, C., & de Rooij, L. (2017). To /b/ or not to /b/: do two-year-olds represent the voiced-voiceless distinction in initial stops? Poster presented at the 14th IASCL, Lyon, France.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language 7*(1): 41–74. https://doi.org/10.1017/s0305000900007029

Scobbie, J., Gibbon, F., Hardcastle, W., & Fletcher, P. (1996). Covert contrast as a stage in the acquisition of phonetics and phonology. *QMC Working Papers in Speech and Language Sciences 1*: 43–62.

Slis, I. H., & Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction I. *Language and Speech 12*(2): 80–102. https://doi.org/10.1177/002383096901200202

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

Van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics 32*(4): 455–491. https://doi.org/10.1016/j.wocn.2004.05.001

Claartje Levelt

C.C.Levelt@hum.leidenuniv.nl