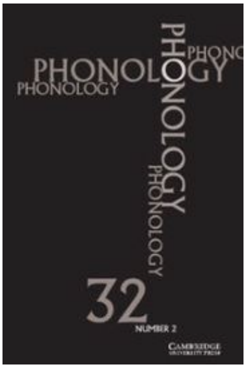

Building phonological lexical representations

Suzanne V. H. van der Feest and Paula Fikkert

Phonology / Volume 32 / Issue 02 / August 2015, pp 207 - 239
DOI: 10.1017/S0952675715000135, Published online: 27 October 2015

Phonology **32** (2015) 207–239.
doi:10.1017/S0952675715000135

*Building phonological lexical representations**

Suzanne V. H. van der Feest
University of Texas at Austin

Paula Fikkert
Radboud University Nijmegen

This paper contributes to the ongoing debate on how much detail young children's word representations contain. We investigate early representations of place of articulation and voicing contrasts, inspired by previously attested asymmetrical patterns in children's early word productions. We tested Dutch-learning 20- and 24-month-olds' perception of these fundamentally different contrasts in a mispronunciation-detection paradigm. Our results show that different kinds and directions of phonological changes yield different effects. Both 20- and 24-month-olds noticed coronal mispronunciations of labials, but not *vice versa*. The 24-month-olds noticed voiced mispronunciations of voiceless stops, but not *vice versa*, while the 20-month-olds failed to notice any voicing mispronunciations. We argue that early lexical representations are specified in very systematic ways, that not all phonological contrasts are encoded at the same time and that the phonological system of a language determines which contrasts are specified first in the representations of early words.

---

1 Introduction

When and how do children store phonetic and phonological detail in their mental representations of words? Views on this issue have changed significantly over the past decades. Arguments have been put forward in both the phonological and the developmental speech-perception literature for fully detailed, as well as underspecified early lexical representations. Traditional phonological theories on the nature of lexical representations have largely been inspired by early production data: when looking at a child's early

* E-mail: SUZANNE@AUSTIN.UTEXAS.EDU, P.FIKKERT@LET.RU.NL.
This research was supported by NWO (Dutch NSF) grant 360-70-100, awarded to Paula Fikkert and René Kager. All experiments were conducted in the Nijmegen Baby Research Center, supported by the NWO Spinoza grant 'Native and non-native listening', awarded to Anne Cutler. We thank Angela Khadar and the rest of the Nijmegen Baby Research Center staff, as well as all participating parents and children for their assistance in the completion of this study. We thank Fieke Noordzij for her help with recordings, Christopher Fennell, and four anonymous reviewers for comments on previous versions of this manuscript. This research was completed as part of the first author's PhD dissertation.

word productions, it is clear that not all contrasts are acquired at the same time (Jakobson 1941, Smith 1973, among many others). Based on children's production patterns, some phonological theories have argued for fully specified early phonological representations of words, with child-specific phonological production rules to account for different outcomes in children's productions (e.g. Smith 1973, Barlow & Gierut 1999, Kager *et al*. 2004). However, other theories have argued for more holistic early representations, with phonological detail being systematically added over time (e.g. Jakobson 1941, Rose & Inkelas 2011). The latter view assumes that if a child does not systematically produce a contrast, it is not yet part of the abstract phonological representation in the mental lexicon. As a result, a contrast not yet acquired in production would also not play a role in word recognition (see Fikkert & Levelt 2008).

More recently, a large body of psycholinguistic speech-perception research with infants has enabled us to investigate children's knowledge of the sound structure of words long before the age at which they may show evidence for such knowledge in their own productions. This line of research has shown that phonological detail may be available for word recognition much earlier than previously assumed (Stager & Werker 1997, Swingley & Aslin 2000, Bailey & Plunkett 2002, Johnson 2005, Nazzi 2005, White & Morgan 2008, among many others). Based on these developmental speech-perception studies, it has often been argued that there is continuity between child and adult representations, and that children use stored phonological detail for word recognition from the beginning of acquisition. For instance, in their seminal study, Juszcyk & Aslin (1995) showed that in a head-turn preference procedure 7·5-month-old infants listened longer to passages that contain words they were first familiarised with (e.g. *cup*) than to the same passages with minimal mispronunciations of the same words (e.g. they heard *tup* after being familiarised on *cup*). The authors argue that infants encode familiar words with phonetic detail, implying that there is continuity between early and later lexical representations. Such a view in turn implies that early production patterns must be accounted for solely by production-specific phonological rules or developing articulatory skills (see Fikkert 2005, 2007).

Another view in the developmental speech-perception literature that assumes continuity between child and adult lexical representations states that phonological detail is stored in the lexicon from the beginning, but not always used. In a series of studies using the 'switch' method, Werker and colleagues argue that at the stage where children acquire their first meaningful words, they may experience cognitive processing overload in certain word-learning situations and experimental tasks (e.g. Stager & Werker 1997, Werker *et al*. 2002, Pater *et al*. 2004). This means that children will show evidence for detailed representations as long as the processing load of the task at hand is sufficiently low, but may fail to use certain details when processing demands are high. As a result, they may confuse phonetically similar words for a short period of time, despite the fact that their auditory discrimination of sound contrasts is unproblematic. According to this view, general cognitive abilities need to develop in order for children to succeed in more demanding tasks. Indeed, recent studies using variations on the switch task have shown that the processing load of a word-learning task can be sufficiently reduced that even 14-month-olds are able to learn minimally different novel words, for example by introducing word–object mappings in a naming phrase to enhance the referential status of a word (Fennell & Waxman 2010), or by pre-exposing infants to the novel objects they are learning names for (Fennell 2012). Taken together, these results support a continuity hypothesis with detailed early lexical representations, where resource limitations account for reported early perceptual difficulties in word-learning tasks. Yoshida *et al*. (2009) also explore this last possibility, testing children in a visual choice task, which arguably has a lower cognitive processing load than a switch task. The authors furthermore discuss a second possible explanation for the initially reported failure of younger infants to learn words in a switch task: they discuss the possibility that gradual learning of general phonological principles of interpretation (such as 'a change from *x* to *y* always signals a lexical difference') may explain why older children are better at using phonological detail for word recognition. Yoshida *et al*. point out that such an account is hard to disentangle from a cognitive processing overload explanation.

In contrast to results supporting the idea of continuity between early and late representations, it has been argued that other studies show that phonological lexical representations may not be fully detailed when children start to focus more on acquiring the meaning of words. For instance, Hallé & de Boysson-Bardies (1996) show that French 11-month-olds tested in the same paradigm as used by Juszcyk & Aslin (1995) preferred familiar words over non-familiar words even when those familiar words were slightly altered (e.g. *ponjour* for *bonjour*). Infants showed the same preference for unaltered *vs*. altered familiar words as compared to unfamiliar words. Some studies show that the prosodic position of the changed consonant matters: Vihman *et al*. (2004) found that a change in the word-initial consonant of a trochee resulted in stronger word-recognition delays in English 11-month-olds, whereas French 11-month-olds' recognition was more distorted by a consonant change in word-medial position in iambs. The view that there may be a difference in representations in different prosodic positions is also supported by data from Dutch-learning 11-month-olds, who preferred lists of correctly pronounced words over lists of mispronounced words, but not when the offset of the words was mispronounced, indicating that less prosodically prominent offsets might be represented with less detail (Swingley 2005).

In sum, results from previous studies suggest that infants at first perceive all detail, and by the time they start building a lexicon seem to ignore some of that detail, at least in some tasks. To explain why this may be, it has been suggested that infants may start with more episodic representations, and learn to generalise over word forms at a later stage (e.g. Jusczyk 1997, Werker & Curtin 2005). This is for instance in line with a model such as PRIMIR (Werker & Curtin 2005), where infants are assumed to start out with these more episodic representations (on the 'Word Form plane'). Next, when lexical learning accelerates, generalisations are made over encountered word tokens that give rise to more abstract phonological representations (on the emerging 'Phonemic plane'). These more abstract representations ensure rapid lexical access and more efficient online word recognition within an expanding lexicon, because the signal-to-word mapping then involves the checking of phonological features in the lexicon rather than all acoustic detail of earlier representations. Such a view assumes that there is a certain amount of discontinuity between early and late lexical representations (see also Ferguson & Farwell 1975, Lahiri & Marslen-Wilson 1991, Hallé & de Boysson-Bardies 1996, Fikkert 2007).

This brings us to the question which we address in the current study: assuming that early lexical representations are detailed to at least some extent, do children show evidence for detailed representations of *all* phonological features and, more specifically, all members of a phonological contrast in words learned under natural conditions? This is an important question, because asymmetries exist in the phonological systems of languages (e.g. Clements 2001, 2003), in adult speech perception (e.g. van Alphen 2004, Eulitz & Lahiri 2004) and in early child production patterns (e.g. Ingram 1974, Levelt 1994, Fikkert & Levelt 2008). In our view, a comprehensive theory of phonological acquisition and the development of lexical representations needs to take into account arguments based on both developmental speech perception and production studies, as well as those based on traditional phonological theories. To do this, we explore here the exact level of phonological detail in early words by examining online word-recognition abilities in two-year-olds. We use a set-up designed to test for possible perceptual asymmetries within a single phonological contrast, inspired by attested asymmetrical patterns in children's productions of early words. If attested asymmetries in production data are due only to production factors such as ease of articulation, we do not expect to find any perceptual asymmetries. In contrast, if asymmetries are also due to differences in phonological lexical representations, we expect to find evidence for this in a word-recognition experiment designed to test for possible asymmetries.

Several studies have investigated representations of familiar words (learned outside the lab under natural circumstances) by means of a 'mispronunciation-detection paradigm' using a language-guided looking procedure (Bergelson & Swingley 2012), also referred to as the 'intermodal preferential looking paradigm' (e.g. Fernald *et al*. 2008) or the 'looking-while-listening procedure' (e.g. Swingley 2010). We use this method for the experiments reported in this paper. Importantly, the processing load of this task is low, and it is considered a sensitive task, in the sense that it is suitable for revealing subtle differences in word recognition as a result of changes in the acoustic signal (see e.g. Swingley 2003 and Mani & Plunkett 2007 for discussion), making it ideal for testing hypotheses on the nature of early word representations. The procedure makes use of children's automatic tendency to look at an object that is named (e.g. Golinkoff *et al*. 1987), and the premise of the task is that longer or faster looking times towards the named target object when it is canonically pronounced compared to when it is slightly mispronounced is evidence for better word recognition. Online word recognition is often described in the psycholinguistic literature in terms of 'activation' of word candidates that are stored in the listener's mental lexicon (e.g. Cutler & Clifton 1999). When a lexical entry is sufficiently activated, the meaning or semantics of the word is accessed. Models of word recognition such as the 'cohort' model (Marslen-Wilson 1987) or the TRACE model (McClelland & Elman 1986) describe activation as a function of the degree to which the speech signal matches the listener's lexical representations. On such a view, a mispronunciation of a word may still activate a word sufficiently for a listener to access its meaning, even if the listener has a detailed representation of a phonological contrast of a segment in a word. However, a mispronunciation will do so less strongly than the correct pronunciation of a word if the acoustic features in the speech signal do not all match the phonological features stored in the listeners' lexical representation of that word. For instance, the perception of a word-initial segment in the speech signal will lead to the activation of all words in the lexicon that contain that word-initial segment. In contrast, if the perceived properties of the initial consonant in the speech signal do not match the represented properties of the initial segment in a specific word representation, this will hinder the activation of that word (although the word may still be activated if the rest of the signal matches the rest of the word). These are not conscious 'decisions' on the part of the listener during the processing of spontaneous speech; rather, the process is automated. We assume that if certain phonological details are not specified in early lexical representations (either features that are never specified in the language or those that have not yet been acquired in a child's early lexicon), these less specified sounds may be more variable in production, leading to the systematic errors typical of early child productions. In perception, a segment that is not specified for a particular feature will allow more variants as candidates for such a segment. For word recognition this means that the more a word is specified for phonological features, the fewer potential word candidates are activated, and *vice versa*: the less specified a word is, the more competing candidates remain (e.g. Lahiri & Marslen-Wilson 1991, Lahiri & Reetz 2002).

Most previous studies which have employed a mispronunciation-detection paradigm have indeed shown that children's looking behaviour generally differs according to whether they are listening to correctly pronounced or mispronounced words. On the basis of this, it has been argued that detailed lexical representations are present as early as 14 months, supporting the notion that early representations are specified, and that there is continuity between early and late lexical representations (e.g. Swingley & Aslin 2000, 2002, Bailey & Plunkett 2002). A number of studies have also looked in greater detail at the effects of different types and positions of mispronunciations. For instance, Dutch 19-month-olds showed sensitivity to various types of mispronunciations in both word-onset and word-medial positions (Swingley 2003). In another study, the cumulative effect of featural mispronunciations in initial consonants on familiar word recognition was tested with American 19-month-olds (White & Morgan 2008). Results of this study provide evidence for graded sensitivity to mispronunciations: mispronunciation effects were larger when more phonological features were mispronounced. However, there are still very few studies that have precisely controlled the nature of all phonological features involved (but see Altvater-Mackensen *et al.* 2014 on toddlers' perception of manner and place of articulation).

In this study, we revisit the investigation of the perception of phonological contrasts, by precisely controlling the phonological dimensions of change: we test for perceptual asymmetries in two experiments by investigating the effects on word recognition of BIDIRECTIONAL mispronunciations of features in word-initial segments. In addition to testing for perceptual asymmetries, we examine the perception of two types of phonological contrasts that are similar in many ways, but can also be characterised as different: (i) place of articulation, a contrast that can be characterised as a 'language-general' contrast used in all human languages, and (ii) voicing, a contrast that is not used contrastively in all languages, and can be characterised as a 'language-specific' or 'emergent' contrast. We will argue and demonstrate below that one crucial difference between the featural representations of these two types of contrast is that, while it is possible for both types that only one member of the contrast is specified in lexical representations, only the language-specific contrasts can be completely absent (or unspecified) in the lexicon.

Importantly, it is widely accepted that children are able to perceive most phonetic detail in the speech signal – an ability that has been demonstrated in many speech-discrimination tasks (e.g. Werker 1995, Jusczyk 1997, Kuhl *et al.* 1997), and we do not argue that children are unable to discriminate between the different members of phonological contrasts that are part of their language (including the ones investigated in this study). For instance, relevant to the voicing contrast under investigation in this paper, English-learning infants are already sensitive to voice onset time (VOT) differences in the first months of life (e.g. Eimas *et al.* 1971, Aslin *et al.* 1981), and Dutch 10-month-olds, as well as 16-month-olds, have no difficulties discriminating between voiced and voiceless stops in word-initial position (Zamuner 2006). However, young children's ability to discriminate a contrast does not automatically mean that it is stored in early lexical representations. Our general prediction is that word recognition will be impaired only when features that are stored and specified in word representations are altered. Based on the specific asymmetries reported in early child productions of voicing and place of articulation, and the assumption that only certain members of a phonological contrast are specified in early lexical representations, we predict that we will also find asymmetries in voicing and place in perception. We will argue that phonological features in children's earliest words are only gradually represented in lexical representations, and hence less specified, which accounts for asymmetries in early production and in comprehension, as demonstrated by mispronunciation-detection patterns.

1.1 Phonological asymmetries in place of articulation and voicing

In Experiments 1 and 2 we test Dutch toddlers' ability to detect bidirectional mispronunciations of place of articulation and voicing. Both place of articulation and voicing are featural contrasts for which one member is typically produced with accuracy earlier than the other member. However, place of articulation and voicing are different in nature, both from an acoustic and a phonological featural perspective, as we will explore in the following sections.

First, place of articulation contrasts can be characterised as 'language-general': place of articulation is one of the contrasts (similar to e.g. manner of articulation) that all human languages need to encode in some way, because by definition any speech sound (with the possible exceptions of [h] and [ʔ]) can be described in terms of place of articulation. Place of articulation is indeed the first phonological contrast that appears in children's early word productions (Jakobson 1941). Both in child language and in adult representations, coronal place of articulation has often been described as the 'default' place of articulation, on the basis of Dutch and many other languages (see e.g. Paradis & Prunet 1989, Lahiri & Marslen-Wilson 1991, Lahiri & Reetz 2002).[1] Children start producing the contrast between labial and coronal place of articulation very early, but their precise realisations of early labials and coronals show asymmetrical patterns whereby it is mainly the labial consonants that are initially produced correctly. However, these early labial consonants are at first restricted to word-initial position. In contrast, there seem to be no restrictions on word position for coronals in early speech (see e.g. Jakobson 1941, Davis & MacNeilage 2000, Fikkert & Levelt 2008). In addition, coronals (unlike labials) are prone to undergo place assimilation in child productions, and to take on the place of articulation of neighbouring segments ('consonant harmony') in many languages, e.g. English (Pater & Werle 2003), French (Rose 2000) and Dutch (Fikkert & Levelt 2008). Consequently, early productions of target coronal consonants show more variation than those of labials. In Experiments 1 and 2 we test and further explore the prediction that when Dutch-learning children have acquired the contrast between coronals and labials, coronals are represented with a 'default' (and underspecified) place of articulation, as opposed to labials, which are 'specified'.

[1] Note that, in contrast, Rice (1996) argues that not only coronal but also velar can serve as default place of articulation. A similar view is expressed in Dresher (2009), who argues that markedness in general is language-specific, and that language-specific contrasts emerge from patterns encountered in the input. For the current study, the notion that the 'default' place could be language-specific does not affect our predictions for the perception of Dutch coronal *vs*. labial.

Our second test case in Experiments 1 and 2 involves testing Dutch toddlers' ability to detect mispronunciations of voicing in word-initial voiced and voiceless stops. Our predictions for the perception of voicing are somewhat different from those for the perception of place of articulation, because of the nature and acoustic properties of voicing features. In contrast to place of articulation, voicing contrasts can be characterised as 'language-specific': they are not found in all languages (Hawaiian, for example, does not use voicing contrastively at all). This means that children acquiring a language that does not use voicing contrasts never need a featural voicing specification for any segment in their sound system: we therefore assume that voicing is a feature which only emerges in a phonological system as the child learns that the language has a voicing contrast. As a result, lexical representations of stop consonants in the initial stages of acquisition will not be specified for voicing (in other words, an emergent feature like voicing will initially be completely *unspecified* in the lexicon, as opposed to an initially *underspecified* place of articulation contrast). Voicing features will remain absent from lexical representations in languages that do not use voicing contrastively. However, because Dutch does have a voicing contrast, at some point in acquisition Dutch-learning children will be able to deduce that voicing is a feature used in their language (which can be described in terms of emerging abstract phonological features or in terms of setting acoustic category boundaries). For instance, it has been argued that children are able to learn which contrasts are used in their language by tracking the distribution of sounds in their linguistic environment (cf. Maye *et al*. 2002).

From an acoustic perspective it is also important to emphasise the difference between 'language-specific' (voicing) and 'language-general' (place of articulation) contrasts. Specifically, it is possible for a speech sound to have 'neutral' voicing properties, which is the case for voiceless stops in Dutch. A segment would be considered to have neutral properties when the VOT is around 0 ms, with no (pre)voicing or aspiration, and no other strong and consistent phonetic cue, signalling a clear difference between voiced and voiceless stops, such as closure, burst or duration of the preceding vowel (cf. van Alphen 2004, van Alphen & Smits 2004). This is in contrast with place of articulation properties, which are never 'neutral': both the 'default' coronal place of articulation and the labial place of articulation are characterised by clearly perceptible cues in the speech signal (with the main cue being the formant transitions at the onset of voicing, as reported by Stevens & Blumstein 1978, for example; see also the papers in Paradis & Prunet 1991). This implies that if a sound is realised with (language-specific) voicing features (like a voiced stop in Dutch) it is possible for that sound to be neutralised, 'removing' its voicing characteristics. This happens in various languages (for example German and Dutch) in word-final position, and to some extent in conversational speech (see Booij 1995, Iverson & Salmons 1995), but it is not possible to neutralise a (language-general) contrast like place of articulation by 'removing' all place of articulation characteristics. This fundamental difference between place-type and voicing-type contrasts leads to different predictions for these two contrasts in the mispronunciation-detection experiments reported here, as we will see below.

In line with the observation that sounds can have neutral acoustic voicing properties, in children's early productions of voicing in various languages all sounds are initially produced with 'neutral' voicing, before they start producing the relevant voicing contrasts of their native language. For example, word-initial labial stops are produced as voiceless unaspirated /p/ in children's early productions in both English and Dutch (Kager *et al*. 2007), despite the fact that these languages phonetically implement voicing contrasts in different ways (the Dutch voicing contrast is between prevoiced 'voiced' stops and 'neutral' unaspirated 'voiceless' stops, similar to languages such as Spanish; see Cho & Ladefoged 1999). The Dutch voicing contrast is acquired relatively late in production (as compared both to other contrasts such as place and to languages with an English-type voicing contrast; see e.g. Kehoe *et al*. 2004). Between the ages of 30 and 36 months children still make errors in their word-initial and word-medial productions of both coronal /d/ and labial /b/ stops, producing them as voiceless /t/ and /p/, but, crucially, these production errors are asymmetrical, as the reverse does not occur (van der Feest 2007, Kager *et al*. 2007). Importantly, these error patterns in early production data cannot be explained solely on the basis of raw input frequencies, since Dutch voiced stops are more frequent than voiceless stops in initial position (Kager *et al*. 2007). Note that one could argue that prevoicing is simply more difficult to produce from an articulatory point of view, and Dutch voiceless (unaspirated) stops are therefore easier to articulate. However, based on articulatory effort, one might expect word-initial plosives to be voiceless, while intervocalic stops would be voiced (due to the natural phonetic process of intervocalic voicing). Instead, we find that Dutch children also prefer voiceless stops in intervocalic position (Kuijpers 1996, van der Feest 2004, Kager *et al*. 2007). On the basis of children's production patterns, Kager *et al*. (2007), as well as Booij (1995), Iverson & Salmons (1995) and van der Feest (2007), argue that for languages with Dutch-type voicing contrasts the specified members of the featural contrast are voiced stops. This corresponds to the acoustic properties of Dutch voiced *vs*. voiceless stops: the acoustic cues for voicing in the signal are much more perceptually salient and informative than the subtle, almost entirely neutral, acoustic cues for voicelessness (see van Alphen 2004 and van Alphen & Smits 2004 for evidence for perceptual asymmetries from lexical priming experiments with adults). This has strong implications for our predictions regarding detection of voicing mispronunciations.

In this study we test the assumption that voicing features are initially absent (or unspecified) in the Dutch lexicon, and when children have acquired the Dutch voicing contrast only voiced stops are specified. Note again that there is a slight but crucial difference as compared to the 'language-general' place of articulation contrast between labials and coronals discussed above: importantly, coronals are never completely *unspecified* for place, but are *underspecified* as [default] from the start of the acquisition process. We consider below the implications of this difference for the predictions of the current study. In sum, assuming that children gradually acquire a system of abstract phonological features and that there is indeed underspecification in (early) lexical representations, we predict perceptual asymmetries for 'language-general' features such as place of articulation that are present in the lexicon from the beginning of acquisition and have a [default] member, as well as for language-specific emergent features such as voicing. The different representations of the contrasts in the lexicon of Dutch learner are given in (1).

(1)

place (*present in all languages*)		voicing (*emergent, language-specific*)	
[default]	[labial]	[ ]	[voiced]
(coronal /d t/)	(/b p/)	(voiceless /p t/)	(/b d/)

In Experiment 1 in this paper, we test Dutch 24-month-olds' perception of place and voicing in familiar words. Around this age we expect children to have acquired the Dutch place and voicing contrasts, because the place contrast is systematically found in production, and the voicing contrast is beginning to appear more systematically (Kuijpers 1996, Kager *et al.* 2007, Fikkert & Levelt 2008). In Experiment 2, we use the same paradigm to test Dutch-learning 20-month-olds. At that age we expect Dutch children to have acquired the place contrasts, but they do not yet reliably produce the (emerging) voicing contrast (van der Feest 2007, Kager *et al.* 2007). Based on the overall results of our experiments and the perceptual asymmetries observed, we will argue that (early) lexical representations are not fully specified. We will further claim that the type of phonological contrasts plays an important role – whether they involve 'general' or 'language-specific' emergent features – as do the exact acoustic characteristics of language-specific phonological contrasts: these properties ultimately determine which phonological features and which members of a phonological contrast are represented first in children's early words.

We will discuss the specific predictions for place and voicing and the difference between them in more detail in the next section. Crucially, if features cannot be unspecified or underspecified in the lexicon, and labial and coronal, as well as voiced and voiceless, features are specified when a child has learned that these contrasts are used functionally in their language, we predict that we will not find perceptual asymmetries, but instead impaired word recognition for all types of mispronunciations in the experiments reported in this paper.

2 Experiment 1

In Experiment 1 we investigated whether we could find evidence for perceptual asymmetries of voicing and place of articulation in 24-month-old

		word-initial sound	feature in lexicon	cue in acoustic signal	word recognition impaired?	rationale
(a)	place	/b p/	[labial]	[coronal]	yes	(i) represented [labial] is not present in the signal, and (ii) represented [labial] conflicts with perceived [coronal] cues
		/d t/	[default]	[labial]	no	(i) represented [default] is not a mismatch with the signal, and (ii) represented [default] is a sufficient match for perceived [labial]
(b)	voice	/b d/	[voice]	[ ] (*neutral*)	no	(i) represented [voice] is not present in the signal, but (ii) does not conflict with perceived neutral voicing cues
		/p t/	[ ]	[voice]	yes	(i) lack of represented [voice] feature is not a mismatch with the signal, and (ii) perceived [voice] feature cannot be matched with the representation

Table I

Experiment 1: Summary of the predicted effects on word recognition of (a) different place mispronunciations and (b) different voice mispronunciations, under a featural underspecification account.

Dutch-learning children. The children were tested in a mispronunciation task specifically designed to test for these potential perceptual asymmetries. Participants were presented with one of four types of test words with different word-initial consonants: /p b t d/ (illustrated in Table I). Mispronunciations never involved the alteration of more than one feature: either the place or the voice feature was changed. This means that each participant was presented either with voiced mispronunciations of voiceless-initial words or with voiceless mispronunciations of voiced-initial words. In addition, each participant heard labial mispronunciations of coronal-initial test items or coronal mispronunciations of labial-initial test items. The experiments included only stop-initial words; fricative-initial words were not included, because the voicing contrast in Dutch fricatives is disappearing in large parts of the Netherlands (Van de Velde & van Hout 2001). Also, since /g/ only occurs as a very low-frequency (loan) phoneme in Dutch, the dorsal stops /k g/ were not included (with the exception of one mispronounced version of a /d/-initial test word, as explained below). If we find asymmetrical patterns in Dutch children's perception of place of articulation and voicing rather than an overall ability to detect word-initial mispronunciations, we argue that this provides support for very specific featural underspecification in early lexical representations.

First, our specific predictions regarding the perception of place of articulation are that if only labial stops are specified when children have acquired the Dutch labial–coronal contrast, a coronal mispronunciation of a specified [labial] should lead to impaired word recognition. The acoustic cues for coronal in the signal do not match the specifications for labials. In terms of phonological features, the speech signal does not contain the required specific [labial] features. Second, we predict that a labial mispronunciation of an underspecified coronal (represented as [default]) may be acceptable and will not hinder word recognition, because a perceived labial in the signal form does not conflict with the represented underspecified [default] feature for coronals. The [default] feature can be activated by more variable acoustic cues, and can be a match for either labial or coronal place of articulation features in the signal. In other words, when the place of articulation representation is [default], more word candidates are left in the online competition, including both coronal-initial and labial-initial ones (Lahiri & Marslen-Wilson 1999, Lahiri & Reetz 2002). Attested results from adult perception studies, where it has been shown that both labial and coronal segments can prime coronals, but only labial segments can prime labials (Lahiri & Reetz 2002), are in line with this prediction. Table Ia summarises these different predicted effects of place of articulation mispronunciations, on the assumption that there is underspecification in the lexicon (in this case, of coronality).

Our specific predictions regarding the perception of the language-specific voicing contrast are different from the predicted pattern for place of articulation, because (i), as explained above, segments can have 'neutral' acoustic voicing properties, and (ii), related to this, it is possible for emergent language-specific contrasts such as voicing to be completely absent from lexical representations. We predict that a mispronunciation of a voiceless stop as [voiced] should lead to impaired word recognition, because in that case the highly salient acoustic cues for prevoicing in the signal do not form a good enough match with the neutral properties of voiceless stops, or, in terms of phonological features, the speech signal contains a [voiced] feature which leads the listener to search for a matching voicing feature in the lexicon, and voiceless stops which are not specified for voicing are not a sufficient match. In contrast, we predict that the mispronunciation of a voiced stop as voiceless may be acceptable and will not hinder word recognition, in line with what we find in adult perception studies, where voiceless stops can 'prime' voiced stops but not *vice versa* (van Alphen 2004, van Alphen & Smits 2004). This is because the neutral voiceless stops are still a good enough match to activate specified voiced stops, i.e. they do not contain conflicting information. In terms of phonological features, the perceived neutral acoustic cues for voiceless stops in the signal do not form a mismatch with the specified feature [voice] (in the way that clearly perceived coronality forms a mismatch with a represented [labial]). Therefore, voiceless mispronunciations of voiced stops do not lead to impaired word recognition, because the listener does not need to keep looking for an exact match for absent (neutral) voicing features in the speech signal (see the summary of the different predicted effects of voicing mispronunciations in Table Ib). Our final prediction is that a child in the initial stages of acquisition does not have representations for voicing in its lexicon, and never needs to look for matching voicing features, in the same way that a listener never needs to look for matching voicing features in a language that does not use voicing contrastively. Therefore, at the earliest stages children will not show impaired word recognition when voicing features are mispronounced. We test this final prediction in Experiment 2, with a group of younger (20-month-old) children.

2.1 Method

2.1.1 *Participants*. Forty-eight 24-month-olds (22 girls, 26 boys), all from monolingual Dutch-speaking families, participated in the experiment. All children were tested at the Baby Research Center in Nijmegen. The children ranged in age from 24;0 to 25;01 months, with a mean age of 24;17 months. An additional 14 children were tested, but excluded from the final analyses, due to general fussiness or failure to stay engaged in the task long enough to complete at least ten of the 14 test trials (11), parental interference (1) or experimenter error (2).

2.1.2 *Stimuli*. The visual stimuli consisted of digital photographs of objects, and were presented side by side in pairs with a white background, on a 192 cm diagonal Sony LCD projection data monitor. The objects were each about 23 cm wide on the screen and 20 cm apart horizontally. The stimuli included four pairs of test items: in the 'p-condition' a cat (*poes* [pus]) and a doll (*pop* [pɔp]), in the 'b-condition' a ball (*bal* [bɑl]) and a tree (*boom* [bom]), in the 'd-condition' a pigeon (*duif* [dœyf]) and a box (*doos* [dos]) and in the 't-condition' a toe (*teen* [ten]) and a tooth (*tand* [tɑnt]). Table II shows the test words and their mispronunciations, with overall word duration and VOT measurements for each word. Each child participated in one of the four conditions; there were twelve children in each condition. Filler items were the same across all four conditions and consisted of a duck (*eend*), a cow (*koe*), a car (*auto*), a baby (*baby*), a bike (*fiets*) and a shoe (*schoen*). They were included to make the test more variable and interesting for the participants. Looking behaviour on filler trials can be analysed post hoc to make sure participants were engaged in the task and able to recognise spoken words in this specific setting. The inclusion of filler trials (which were never mispronounced) also added to the number of correct pronunciations in the experiment, without presenting the children

	CP		MP-voice		MP-place	
b	boom [bom] 487 (−85)	bal [bɑl] 469 (−104)	poom [pom] 563 (2)	pal [pɑl] 525 (6)	doom [dom] 568 (−106)	dal [dɑl] 528 (−119)
p	poes [pus] 581 (14)	pop [pɔp] 435 (9)	boes [bus] 559 (−107)	bop [bɔp] 417 (−114)	toes [tus] 559 (25)	top [tɔp] 424 (17)
d	duif [dœyf] 666 (−19)	doos [dos] 658 (−98)	tuif [tœyf] 639 (20)	toos [tos] 696 (14)	buif [bœyf] 612 (−102)	goos [gos] 677 (−102)
t	teen [ten] 576 (27)	tand [tant] 588 (15)	deen [den] 609 (−123)	dant [dant] 629 (−131)	peen [pen] 582 (16)	pant [pant] 556 (13)

Table II

Test words with transcriptions in the b-condition, the p-condition, the d-condition and the t-condition: the total duration of each test word is given in ms, followed by VOT measurement in ms in parentheses.

with even more repetitions of correct pronunciations of the test items. All words in the experiments are typically known by children of this age; this was confirmed by the parents of all the participants (see also §2.1). All test items were CVC words in Dutch (except for *tand*). All mispronunciations resulted in non-words or in five cases in very low-frequency words that all children were reported not to know (*pal* 'steadfastly', *dal* 'valley', *top* 'tip, peak', *peen* (low-frequency word for 'carrot') and *pand* (formal word for 'building')). In Table II, the correct pronunciations of the words are shown in the CP column, and the mispronunciations of the words in the MP-voice and MP-place columns. For instance, the target words in the p-condition were *poes* [pus] and *pop* [pɔp], which became [bus] and [bɔp] in the MP-voice trials and [tus] and [tɔp] in the MP-place trials. Note that the place mispronunciation of the word *doos* would lead to the word *boos* 'angry' when mispronounced with a labial; this word is known by many 20- and 24-month-old children. We included the item *doos* because this was the most easily depictable and best-known word with a CVC structure from the limited set of words we could choose from for this age group. Hence, the place mispronunciation of *doos* started with the infrequent voiced stop /g/, forming the non-word *goos*. The crucial change here is from coronal to velar, which, like labial, is assumed to be specified in Dutch. Even though voiced velars are rare in Dutch, voiceless velars are not. Hence we did not predict (or find) a different result pattern for this item.

The auditory stimuli were digitally recorded in a soundproof room by a female native speaker of standard Dutch, with a sampling rate of 44,100 Hz. The speaker used a happy but not over-excited infant-directed speaking style, and a slow speaking rate. All test trials started with a 2·5 second silence during which the pictures were shown, followed by the carrier phrase *Kijk naar de* TARGET*!* 'Look at the TARGET!'. Word frequency was as far as possible balanced across conditions, given the limited set of possible test words known to children in our age groups. Table II shows the durations of the target words, including the prevoicing or the voiceless closure before the burst of the initial stop consonant. All voiced segments (in both CP and MP conditions) were produced with prevoicing. The carrier phrase plus the test word were followed by a 750 ms pause and then a second sentence: *Mooi he?* or *Leuk he?* (roughly 'Beautiful/nice, right?'), to space out the time between trials in a child-friendly way. Filler trials included carrier phrases such as *Kun je de koe vinden? Vind je 'm mooi?* 'Can you find the cow? Do you like it?' and *Waar is de fiets? Kun je 'm vinden?* 'Where is the bike? Can you find it?'. Each trial ended four seconds after the onset of the auditory stimuli, giving a total duration of 6·5 seconds.

2.1.3 *Procedure*. After a brief informal play session during which the procedure was explained to the caregiver, children were seated on their caregiver's lap, facing the screen. The experiment took place in a sound-insulated room, in a three-sided enclosure, which was 2 metres tall, 1·3 metres wide and 1·2 metres deep. The parent and child sat at the open end of this enclosure. The speech stimuli were played over TV speakers. Children were videotaped on a DV cassette (using a Sony recorder SR-40P), with a digital video camera (Sony CVX-V18NSP). The camera was placed 30 cm below the screen, hidden by a black curtain with an opening for the lens. The spotlights in the room were dimmed to a preset level. Parents were instructed not to speak or interact with their child during the experiment, and wore Sennheiser Noisegard headphones during the entire experiment. The parents heard music mixed with test trials over the headphones, so that they were unable to hear either the sound of the experiment or when the child was presented with auditory stimuli. Each experiment began and ended with a bouncing yellow duck, and after the first half of the experiment a fish swam across the screen, making a bubbling noise. Each of the four conditions of the experiment contained 14 test trials and 10 filler trials (not including the duck and the fish). Between trials, a flashing light on a black background was shown in the centre of the screen, in order to make the children look again at the centre of the screen between trials, and to aid coding of left and right eye movements. The switching of the background colour from white to black clearly changed the lighting of the whole experimental setting, indicating the beginning and end of a test trial for the coder. The 14 test trials consisted of six correct pronunciation (CP) trials, four mispronunciation of voicing (MP-voice) trials and four mispronunciation of place (MP-place) trials. In this way, children always heard the test words more often in the CP trials than in either of the MP trials. In the p-condition, *pop* was the target for half of the trials and *poes* for the other half, meaning that each target appeared three times in a CP trial, twice in an MP-place trial and twice in an MP-voice trial. Side was counter-balanced by item within the children (for half of the children, *pop* was the target on the left side of the screen three times and on the right four times, for the other half it was the other way around). Trials were presented in four different random orders; order 2 was the reverse of order 1, order 3 was a left-right reversal of order 1 and order 4 a left-right reversal of order 2. In orders 1 and 3, subjects first heard the correct pronunciation of one test word and the voice mispronunciation of the other test word; in orders 2 and 4 this was reversed. Before they visited the Baby Research Center, parents were asked to fill out a Dutch version of the MacArthur Communicative Development Inventory: Toddlers (Fenson *et al.* 1993), i.e. the N-CDI list (*Netherlands-Communicative Development Inventory*; Zink & Lejaegere 2002). Data on receptive as well as productive vocabulary were collected. All children were reported to comprehend all test words. For each participant, we confirmed that the specific test words of the condition the child participated in were part of their receptive as well as their productive vocabulary.

2.1.4 *Coding and data analyses*. The recordings of the children's eye movements were coded offline by trained coders, using SuperCoder (Hollich 2005). For each video frame of 40 ms (videos were recorded with a frame rate of 25 frames per second) the coder indicated where the child was looking: at the left picture, at the right picture, moving between the two, or away from the screen (the latter two were both coded as 'away' from the pictures). The light on the video clearly indicated the beginning and end of each test trial. The coder did not hear any sound, and was unaware of the test order that a particular child had participated in. Coder reliability was determined by comparing codings from two different coders on 10% of the data. The mean agreement between coders was 96%.

The ratio between fixations to the target picture and total fixations to the screen (to either the target picture or the distractor picture) was calculated for each trial of all participants. This measure for fixations to target picture is reported in most previous studies, using a similar set-up (e.g. Swingley & Aslin 2000, Swingley 2003, White & Morgan 2008). Fixations were calculated over a window of analysis which started 360 ms and ended 2000 ms after the onset of the target word, taking into account the average latency for young children to initiate eye movement (Swingley & Aslin 2000, Swingley 2003). The target word onset was defined as the beginning of the stop closure (the end of the preceding schwa) for voiceless stops and the beginning of prevoicing for the voiced stops. Difference scores were computed for each trial, by comparing the post-target onset window with the two-second window *before* the target word onset, during which the pictures were displayed on the screen in silence. Comparing looking times towards the target during these two windows allowed us to evaluate the effect of naming for each trial of every child, and to take baseline visual preferences into account (see e.g. Quam & Swingley 2010 for a similar use of difference scores). Trials on which the child did not pay attention, i.e. did not look at the target or the distracter before and after the target word onset for a period of at least 400 ms, were excluded from analysis. Across all subjects and conditions, approximately 6% of trials were excluded for this reason.

2.2 Results and discussion

We first investigated the effects of mispronunciations of place. The difference scores calculated for each participant were entered into a planned comparisons repeated measures 2 × 1 ANOVA to examine the effects of pronunciation (CP, MP-Place), first for Labial-initial and next for Coronal-initial words. For the Labial-initial words this revealed a significant effect of pronunciation ($F(1,23) = 3{\cdot}7$, $p = 0{\cdot}048$, ${\eta_p}^2 = 0{\cdot}06$). For the Coronal-initial words there was no effect of pronunciation ($F(1,23) = 0{\cdot}04$, $p = 0{\cdot}85$, ${\eta_p}^2 = 0{\cdot}0008$). We then directly compared the effect of pronunciation on both word-types in a 2 × 2 ANOVA with word-type (Labial-initial, Coronal-initial) as between-subjects factor, and pronunciation (CP, MP-place) as within-subject factor. This showed no main effect of word-type ($F(1,23) = 0{\cdot}42$, $p = 0{\cdot}52$, ${\eta_p}^2 = 0{\cdot}004$) or pronunciation ($F(1,23) = 2{\cdot}4$, $p = 0{\cdot}13$, ${\eta_p}^2 = 0{\cdot}002$), nor an interaction between word-type and pronunciation ($F(1,47) = 1{\cdot}7$, $p = 0{\cdot}19$, ${\eta_p}^2 = 0{\cdot}01$). Figure 1 illustrates the difference scores on coronal *vs*. labial targets on the different trial types. To investigate whether the subjects recognised the target words, at least when they were correctly pronounced, planned 2-tailed *t*-tests were conducted comparing difference scores to chance or 0. If a word was recognised, the increase in looking time towards the target picture as expressed by the difference scores should be significantly different from chance. This showed significant increases on CP trials for both Labial-initial targets ($t(23) = 4{\cdot}5$, $p < 0{\cdot}001$, $d = 0{\cdot}68$) and Coronal-initial targets ($t(23) = 4{\cdot}0$, $p < 0{\cdot}001$, $d = 0{\cdot}64$). However, on MP-place trials the increase in looking time towards the target after target word onset was significantly different from chance only for Coronal-initial targets ($t(31) = 2{\cdot}12$, $p = 0{\cdot}04$, $d = 0{\cdot}45$), but not for Labial-initial targets ($t(31) = 1{\cdot}31$, $p = 0{\cdot}20$, $d = 0{\cdot}23$): word recognition was impaired by place mispronunciations only when labial targets were mispronounced as coronal; when coronal words were mispronounced as labial they were still reliably recognised.

We next investigated effects of mispronunciations of voicing: we ran a 2 × 1 ANOVA to examine the effects of pronunciation (CP, MP-voice), first for Voiceless-initial and then for Voiced-initial words. For the Voiceless-initial words there was a significant effect of pronunciation ($F(1,23) = 3{\cdot}8$, $p = 0{\cdot}049$, ${\eta_p}^2 = 0{\cdot}069$). There was no significant effect of pronunciation for the Voiced-initial words ($F(1,23) = 1{\cdot}9$, $p = 0{\cdot}18$,

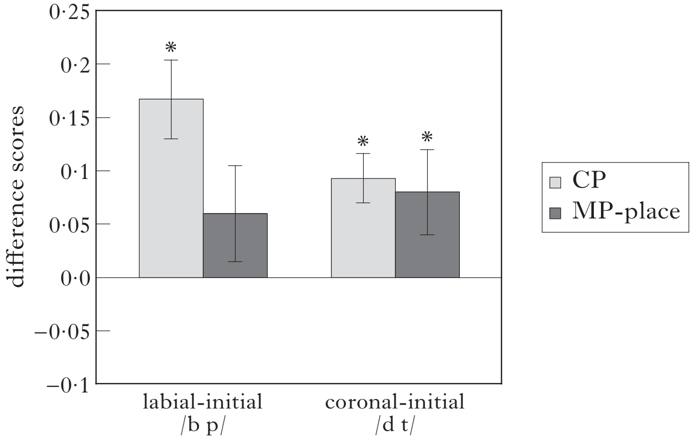

Figure 1

Experiment 1: effects of place of articulation mispronunciations for 24-month-olds: difference scores (target fixations; before *vs*. after target word onset) and standard errors. Data is broken down by pronunciation (correct pronunciations (CP) and place mispronunciations (MP-place)) and by word-type. Asterisks indicate significant differences from chance.

$\eta_p^2 = 0{\cdot}03$). Next, performance on voiced and voiceless words was directly compared in a 2×2 ANOVA with word-type (Voiceless-initial, Voiced-initial) as between-subjects factor and pronunciation (CP, MP-voice) as within-subject factor. There were no main effects of word-type ($F(1,47) = 1{\cdot}51$, $p = 0{\cdot}22$, $\eta_p^2 = 0{\cdot}01$) or pronunciation ($F(1,47) = 0{\cdot}38$, $p = 0{\cdot}53$, $\eta_p^2 = 0{\cdot}004$), but there was a significant interaction between word-type and pronunciation ($F(1,47) = 5{\cdot}35$, $p = 0{\cdot}02$, $\eta_p^2 = 0{\cdot}054$). Figure 2 illustrates the increase in target fixations for Voiceless-initial *vs*. Voiced-initial words on CP and MP-voice trials. Planned 2-tailed *t*-tests were conducted comparing difference scores to chance 0, to investigate whether the subjects recognised the target words reliably. These revealed significant increases on CP trials for Voiceless-initial targets ($t(23) = 5{\cdot}03$, $p < 0{\cdot}001$, $d = 0{\cdot}72$) and for Voiced-initial targets ($t(23) = 3{\cdot}29$, $p = 0{\cdot}003$, $d = 0{\cdot}56$). However, on the MP-voice trials the increase was significant for Voiced-initial targets ($t(31) = 5{\cdot}04$, $p < 0{\cdot}001$, $d = 0{\cdot}72$), but not for Voiceless-initial targets ($t(31) = 0{\cdot}69$, $p = 0{\cdot}50$, $d = 0{\cdot}08$). This indicates that the children recognised all target words when they were correctly pronounced, and that word recognition was impaired when voiceless targets were mispronounced as voiced, but not when voiced targets were mispronounced as voiceless.

The results show that different mispronunciations are not all equal. The 24-month-olds were able to detect mispronunciations in familiar words, in line with previous findings. However, there was no evidence that children detected *all* mispronunciations. Subjects performed differently on the task depending on whether they were presented with voiced or voiceless target

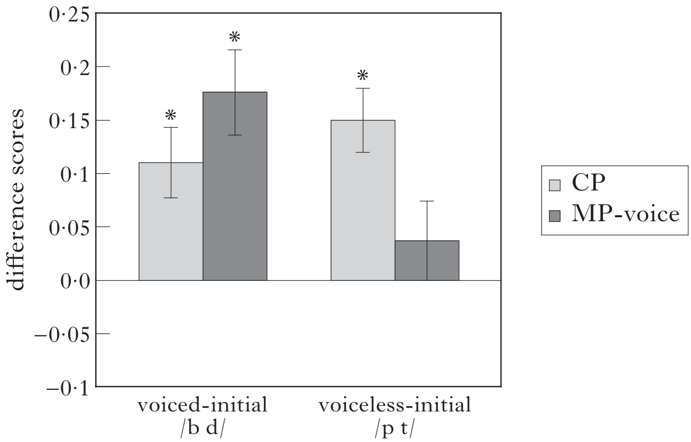

Figure 2

Experiment 1: effects of voice mispronunciations for 24-month-olds: difference scores (target fixations; before *vs*. after target word onset) and standard errors. Data is broken down by pronunciation (correct pronunciations (CP) and voicing mispronunciations (MP-voice)) and by word-type. Asterisks indicate significant differences from chance.

words (the b- and d-conditions compared to the p- and t-conditions), and with labial or coronal words (the b- and p-conditions compared to the d- and t-conditions). Word recognition was not impaired when word-initial voiced stops were produced as voiceless, nor when word-initial coronal stops were produced with a labial place of articulation. Perceptual asymmetries are in line with the asymmetrical patterns in production discussed in Kager *et al*. (2007) and van der Feest (2007), showing that children first produce voiced stops as voiceless stops, and in Fikkert & Levelt (2008), who show that coronals are often produced as labials.

The direction of the perceptual asymmetries corresponds with our predictions. The results indicate that the 24-month-olds have learned that there are voicing and place of articulation contrasts in their language, as voiced targets are treated differently from voiceless targets, and labials differently from coronals. The goal of Experiment 2 is to investigate the abilities to detect these mispronunciations of younger infants, who have been reported to produce the place but not the voicing contrast (e.g. Kuijpers 1996, Kager *et al*. 2007, Fikkert & Levelt 2008). As discussed in §1, voicing is a language-specific contrast that may be completely absent in early representations, before children learn that the contrast is used in their target language. If the absence of voicing contrasts in early productions indicates that the Dutch voicing contrast is indeed still not represented at that stage, we predict that younger 20-month-old children will not detect mispronunciations of voicing; rather, the same perceptual asymmetry for place of articulation will be found as for the 24-month-olds in Experiment 1.

3 Experiment 2

Experiment 2 was identical to Experiment 1, except for the age of the participants: the children in Experiment 2 were 20 months old. The goal of testing this age group was to further investigate the developmental time line of an emerging 'language-specific' contrast such as voicing, which we assume is completely unspecified at the initial stages of acquisition. At 20 months, Dutch children do not yet reliably produce a voicing contrast (van der Feest 2007, Kager *et al*. 2007). If this indeed means that they do not yet have the Dutch voicing contrast represented in their lexicon, we predict that word recognition will not be hindered by any type of voicing mispronunciation. However, as it is often argued that perception precedes production, it is possible that 20-month-olds are already establishing a contrast; we will show evidence of this in a word-recognition task. In that case we predict the same perceptual asymmetry for voicing as found in Experiment 1. Recall that Zamuner (2006) showed that Dutch-learning 10-month-olds and 16-month-olds are able to discriminate between voiced and voiceless stops in word-initial position. Swingley & Aslin (2000) found no age effect for general mispronunciation detection with 18- to 24-month-olds, and Altvater-Mackensen *et al*. (2014) found that Dutch-learning 18-month-olds were able to detect place mispronunciations of labials as coronals. On the basis of those results and the production patterns discussed earlier (and in line with the assumption that language-general contrasts such as place of articulation are acquired earlier than emerging language-specific contrasts such as voicing), we expected Dutch-learning 20-month-olds to be able to detect at least the place mispronunciations of labial-initial target words. Since we predicted that we would find no effects for the voicing mispronunciations, the place mispronunciations also served as a form of control condition, to make sure the Dutch-learning 20-month-olds were able to detect one-feature mispronunciations in this task.

3.1 Method

3.1.1 *Participants*. Forty-eight monolingual Dutch-speaking 20-month-old infants (21 girls, 27 boys) participated in this experiment. The age range of the participants was from 20;03 to 21;02 months, with a mean age of 20;18 months. Sixteen additional children participated, but were excluded from the final analyses because of fussiness or failure to stay engaged in the task long enough to complete at least 10 of the 14 test trials (15) or because of experimenter error (1).

3.1.2 *Stimuli and procedure*. The stimuli and apparatus used in Experiment 2 were identical to those of Experiment 1. All parents were

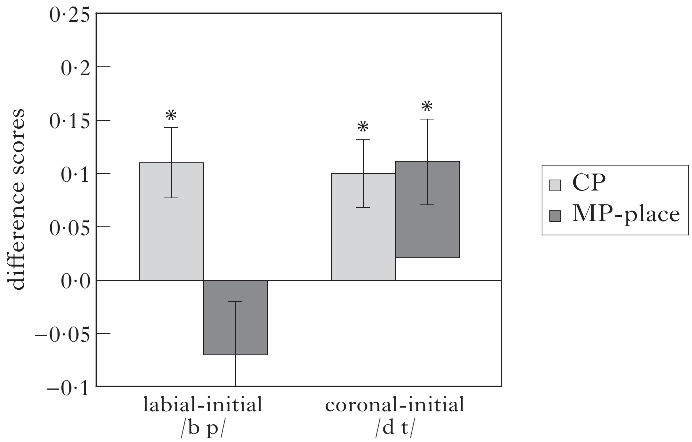

Figure 3

Experiment 2: effects of place of articulation mispronunciations for 20-month-olds: difference scores (target fixations; before *vs*. after target word onset) and standard errors. Data is broken down by pronunciation (correct pronunciations (CP) and place mispronunciations (MP-place) and by word-type. Asterisks indicate significant differences from chance.

again asked to complete the N-CDI list before coming to the lab. We again checked that all children knew the target words of the condition they participated in.

3.2 Results and discussion

For the looking-time analyses, we compared children's target fixations in the 2000 ms pre-target-onset window with the 360–2000 ms window post-target word onset (the same analysis reported for the 24-month-olds in Experiment 1). A 2 × 2 ANOVA with word-type (Labial-initial, Coronal-initial) as between-subjects factor and pronunciation (CP, MP-place) as within-subject factor showed a marginal main effect of pronunciation ($F(1,47) = 3{\cdot}9$, $p = 0{\cdot}05$, $\eta_p^2 = 0{\cdot}04$) and of word-type ($F(1,47) = 3{\cdot}8$, $p = 0{\cdot}05$, $\eta_p^2 = 0{\cdot}04$), and a trend towards an interaction between word-type and pronunciation ($F(1,47) = 3{\cdot}11$, $p = 0{\cdot}08$, $\eta_p^2 = 0{\cdot}03$). This is illustrated in Fig. 3. Given the design of the study, the marginal main effect of pronunciation and the trend towards the word-type and pronunciation interaction, planned follow-up analyses were conducted (the same as for the 24-month-olds in Experiment 1), in which we examined the effects of pronunciation for Labial-initial and Coronal-initial words individually. We found a similar perceptual asymmetry as in Experiment 1: there was a significant effect of pronunciation for the Labial-initial target words ($F(1,23) = 5{\cdot}92$, $p = 0{\cdot}02$, $\eta_p^2 = 0{\cdot}11$), but not for Coronal-initial words ($F(1,23) = 0{\cdot}03$, $p = 0{\cdot}9$, $\eta_p^2 = 0{\cdot}0005$). Planned t-tests were conducted,

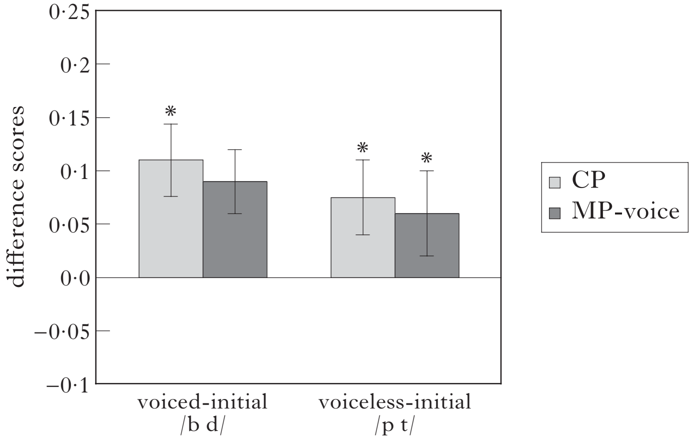

Figure 4

Experiment 2: effects of voice mispronunciations for 20-month-olds: difference scores (target fixations; before *vs.* after target word onset) and standard errors. Data is broken down by pronunciation (correct pronunciations (CP) and voicing mispronunciations (MP-voice)) and by word-type. Asterisks indicate significant differences from chance.

comparing difference scores to chance 0, which revealed significant increases on CP trials for both Labial-initial targets ($t(23)=3{\cdot}33$, $p=0{\cdot}003$, $d=0{\cdot}56$) and Coronal-initial targets ($t(23)=3{\cdot}1$, $p=0{\cdot}005$, $d=0{\cdot}54$). However, on MP-place trials the increase was significant only for Coronal-initial targets ($t(31)=2{\cdot}1$, $p=0{\cdot}49$, $d=0{\cdot}54$), not for Labial-initial targets ($t(31)=-1{\cdot}3$, $p=0{\cdot}2$, $d=0{\cdot}20$). As for the 24-month-olds in Experiment 1, word recognition for the 20-month-olds was hindered by place mispronunciations when Labial-initial targets were mispronounced as coronal, but not when Coronal-initial targets were mispronounced as labial.

A 2 × 2 ANOVA with word-type (Voiceless-initial, Voiced-initial) as between-subjects factor and pronunciation (CP, MP-voice) as within-subject factor showed a marginally significant main effect of word-type ($F(1{,}47)=3{\cdot}6$, $p=0{\cdot}058$, $\eta_p^2=0{\cdot}38$), no main effect of pronunciation ($F(1{,}47)=1{\cdot}92$, $p=0{\cdot}17$, $\eta_p^2=0{\cdot}02$) and no interaction between word-type and pronunciation ($F(1{,}47)=0{\cdot}69$, $p=0{\cdot}40$, $\eta_p^2=0{\cdot}007$; difference scores are illustrated in Fig. 4). We again looked at voiced *vs.* voiceless targets separately, and found no effects of pronunciation for Voiceless-initial targets ($F(1{,}23)=2{\cdot}55$, $p=0{\cdot}12$, $\eta_p^2=0{\cdot}05$) or for Voiced-initial targets ($F(1{,}23)=0{\cdot}15$, $p=0{\cdot}70$, $\eta_p^2=0{\cdot}003$). However, planned *t*-tests comparing difference scores to 0 showed that the subjects did recognise the target words on both Voiced-initial and Voiceless-initial words. There were significant increases in looking time towards the target on CP trials for Voiceless-initial targets ($t(23)=2{\cdot}16$, $p=0{\cdot}04$, $d=0{\cdot}41$) and Voiced-initial targets ($t(23)=3{\cdot}4$, $p=0{\cdot}002$, $d=0{\cdot}57$), as well as on MP-voice trials; the increase was marginally significant for Voiceless-initial targets ($t(23) = 2{\cdot}0$, $p = 0{\cdot}057$, $d = 0{\cdot}38$) and Voiced-initial targets ($t(23) = 2{\cdot}35$, $p = 0{\cdot}03$, $d = 0{\cdot}44$). This indicates that, for the 20-month-olds, word recognition was not significantly impaired by the mispronunciations of voicing.

4 General discussion

Children's early lexical representations are fairly detailed, as has been shown by a growing body of literature on early speech perception and word recognition. However, this does not necessarily mean that early word representations are completely adult-like, or that all phonological featural contrasts are acquired and represented at the same time. In this study we have shown that young children are not always able to detect one-feature mispronunciations of familiar words and, more specifically, that the direction of the same featural change can influence online word recognition. We found perceptual asymmetries for both voicing and place of articulation contrasts in Experiment 1, and an asymmetry in the perception of place in Experiment 2. In Experiment 1, 24-month-old subjects detected prevoiced mispronunciations of voiceless targets, but did not respond differently to voiced correct pronunciations and voiceless mispronunciations of voiced target words. In addition, both the 24-month-olds and the 20-month-olds in Experiment 2 detected place mispronunciations of labial, but not coronal target words. The results from Experiment 2 further showed that, unlike the 24-month-olds, the 20-month-old subjects did not detect mispronunciation of voice on voiced or on voiceless target words. We explore two possible explanations for these perceptual asymmetries and the different patterns found for the 20- *vs*. 24-month-olds: one is based on a perceptual acoustical account, the other on a phonological featural account.

A first explanation for the fact that children do not treat all changes in place or voicing as mispronunciations could be that they do not perceive the distinction sufficiently well. If that is the case, the perceptual boundaries distinguishing labial from coronal and voiced from voiceless sounds may not yet be well enough established. The same labial–coronal asymmetry reported above was also found for Dutch adults, where the asymmetry can be described in terms of confusion matrices (Smits *et al*. 2003), without reference to abstract phonological features. Smits *et al*. showed that coronal targets are more likely to be misperceived as labials than *vice versa*; in other words, coronals in Dutch have less clearly defined category boundaries than labials, even for adults. This means that for Dutch-learning children, the less confusable pronunciations of labials in the input could lead to clearer or stricter category boundaries for labials in lexical representations, earlier than for coronals. In the context of the current study, when our participants in Experiment 1 and 2 heard a labial target mispronounced as coronal, this should have been noticed, impairing word recognition; but when they heard a coronal target (represented with less clear category boundaries, due to greater variability or confusability of coronals in the speech signal) mispronounced as labial, this should have still been considered an acceptable or good enough pronunciation of the target word, without word recognition being significantly impaired. Those predictions are indeed in line with our results. However, it would also be possible to reverse the argument, predicting the opposite pattern: a labial perceived in the signal is clearly a labial, and hence should not be a good match with a coronal target candidate, but should clearly and most strongly activate labial candidates; therefore, at least labial mispronunciations of coronals should be noticed. In contrast, since the acoustic cues for coronals are more variable, the coronal mispronunciations of labial targets could potentially be good enough representations of labial targets. Alternatively, one could argue that both types of mispronunciation should be noticed, because the signal does contain clear place of articulation cues for both labials and coronals in the signal (regardless of the fact that labials are less variable), a prediction which is not confirmed by our results. Thus the exact predictions of an acoustics-based account depend on the interpretation of the results of variability in the signal. As discussed in §1, it has been clearly shown in previous studies that young children are not unable to discriminate labial from coronal stops: therefore, in our view it must be the functional use of the contrast for word recognition that gives rise to asymmetrical patterns. In sum, although an acoustic explanation can account for the place of articulation patterns without making reference to abstract phonological features, it does not necessarily account straightforwardly for the attested asymmetries found in the current word-recognition study, nor for those in perception studies with adults (Eulitz & Lahiri 2004) and in early word production studies (e.g. Jakobson 1941, Davis & MacNeilage 2000, Fikkert & Levelt 2008).

Along the same lines, the results of the perception of voicing could also be explained in terms of developing category boundaries. Under such a view, no voicing contrast exists in the initial stage of acquisition, when the full range of the Dutch VOT contrasts still falls into one perceptual category. The 20-month-olds in Experiment 2 are still in this first stage of the acquisition of voicing. The perceived voicing in the signal therefore does not affect online word recognition in this task, as voicing is not yet used in lexical representations. For the 20-month-olds, both voiced and voiceless pronunciations of stops lead to the same level of activation of word candidates containing those stops, similar to what we would predict for a speaker of a language with no meaningful voicing contrast. An acoustics-based account could also explain the results for voicing mispronunciation in Experiment 1: when 24-month-olds are presented with a clear acoustic cue of prevoicing in the signal, this leads to the activation of fewer potential word candidates, namely only those that contain voiced stops, hindering recognition of words with initial voiceless stops. This asymmetry in the perception of voiced and voiceless stops is also in line with the Dutch confusion matrices proposed by Smits *et al.* (2003): voiceless segments are more easily confused with voiced segments, whereas voiced stops are less likely to be confused with voiceless stops.

However, although an explanation based on incorrect category boundaries may account for the data reported for voicing in Experiments 1 and 2, the experimental results for infants' discrimination of the Dutch voicing contrast may form a problem for a purely perceptual account. As discussed above, Zamuner (2006) showed that Dutch-learning 10- and 16-month-olds are already perfectly able to discriminate between Dutch voiced and voiceless stops in word-initial position in a non-lexical task. This means that Dutch voiced and voiceless stops are not treated acoustically as one group at this age, and there is no reason to assume that Dutch-learning children would have lost this ability at 20 months. Note that shifting of category boundaries may have occurred earlier: there is some evidence that infants prefer the 'English-type' category over the 'Dutch-type' category. The English-type contrast has been shown to be perceived categorically (along the English-type category boundary) by new-borns, both English- and Spanish-learning; Spanish voicing contrasts are very similar to the Dutch (Eimas *et al*. 1971). However, Zamuner's results show that, by the age of 10 months, infants discriminate the 'Dutch-type' categories along the appropriate category boundary. The perception results concerning voicing can therefore not simply be explained by shifting category boundaries, because that would imply that initially (i.e. at 20 months) Dutch voiced and voiceless stops all fall in the same perceptual category. Such an explanation can therefore only work on the additional assumption that children use a different perceptual system or a different level of perception for the discrimination of meaningless sounds and the use of sounds to recognise well-known words. However, in this study we have used a method that has been described as sensitive enough to show phonetic specification of sounds in words; in addition, a mispronunciation-detection paradigm with well-known words learned under natural circumstances has been argued to have a low cognitive processing load, which should enable children to make use of their discriminative abilities. Yet our results show that phonetic detail is used only in some, but not all, contexts.

Our view is therefore that an account referring to abstract phonological features, which may emerge in children's representation over the course of development, is to be preferred to an acoustics-type account based on fuzzy category boundaries. We favour a featural approach because it can account for our findings of perceptual asymmetries in early word recognition for voicing and place of articulation, as well as for perceptual asymmetries found in adults and previously attested patterns in child production, as discussed in §1. On the assumption that children's phonological representations are based on all the (acoustic) detail that they are able to perceive, the attested asymmetries in word recognition are harder to explain, and a purely perceptual account would have to refer only to motor skills to explain children's very systematic production errors in both voicing and place of articulation contrasts. Van Alphen (2004) and van Alphen & Smits (2004), among others, do refer to the articulatory effort of producing voiced stops in Dutch. But, as discussed above, van der Feest (2007) and Kager *et al*. (2007) argue that not all voicing errors in production can be accounted for in an articulatory account. This gives a phonological featural account one major conceptual advantage over a perceptual account: it assumes a level of abstract featural representations in the mental lexicon which can serve as a single intermediate level operating between perception and production. We will summarise our interpretation of the results from this study in turn for each of the four different types of mispronunciation trials that we tested in Experiments 1 and 2, compared to the four correct pronunciation trial types.

Mispronounced coronals. On the basis of our results for the perception of place of articulation, we assume that coronals are the 'default' place of articulation, and are underspecified for place in the lexicon from the beginning of the acquisition process, as discussed in §1 and illustrated in (1). On this assumption, a 'default' underspecified coronal in the lexicon leads by definition neither to a match nor to a mismatch with any place of articulation cues in the speech signal (see for example the 'matching conditions' in a model such as the Featurally Underspecified Lexicon model of Lahiri & Reetz 2002). When coronal-initial words are mispronounced and the children in Experiments 1 and 2 perceive labial features in the speech signal, the labial-initial words in their lexicon remain as potential candidates, as the lexicon is searched for segments specified for the feature [labial]. Crucially though, coronal-initial words are not deactivated when a labial stop is perceived, because their underspecified or 'default' featural representation does not lead to a mismatch with the speech signal, and can be sufficiently activated by a perceived labial; word recognition is therefore not significantly impaired by this type of mispronunciation.

Correctly pronounced coronals. Along similar lines, when coronal-initial words are correctly pronounced and acoustic coronal features are perceived in the speech signal, this also leads neither to a match nor to a mismatch with coronal words in the lexicon, and coronal-initial candidates remain potential word candidates. We therefore find no difference in word recognition in the current task between correct pronunciations and mispronunciations of coronal-initial words.

Mispronounced labials. In contrast, when labial-initial target words are mispronounced, the perception of acoustic coronal features in the speech signal does lead to a mismatch with stored specified [labial] place of articulation features: labial-initial targets are therefore not sufficiently activated as potential candidates, and coronal mispronunciations of labial-initial words do impair word recognition.

Correctly pronounced labials. When labial-initial targets are correctly pronounced, labial cues in the speech signal lead to a perfect match with and immediate activation of labial-initial target words that are specified as [labial] in the lexicon. It is important to note that a mispronunciation-detection paradigm as used here with toddlers is not sensitive enough to measure different gradations of activation of candidates in the way in which a lexical priming experiment (with adults), for example, can. We therefore cannot draw direct conclusions from the current study on whether there is evidence for different degrees of word activation on trials where coronal-initial words are pronounced correctly or mispronounced, as opposed to the degree of activation on trials where labial-initial words are pronounced correctly. (Coronal-initial words – whether pronounced correctly or mispronounced – involve neither a match nor a mismatch between the speech signal and the lexical representation, whereas correctly pronounced labial-initial words lead to a true match between the speech signal and the stored lexical representation.) For a discussion of the potential computation of different degrees of lexical activation in an underspecification model, see e.g. Lahiri & Reetz (2002).

Mispronounced voiced stops. Our results for the perception of voicing lead us to assume that for the 24-month-olds in Experiment 1, voiced target words are specified, in contrast to voiceless stops, which are unspecified, as discussed in §1 and illustrated in Table I and (1). Recall the difference between the language-specific voicing contrast and the language-general place of articulation contrast: the acoustic cues for place of articulation clearly signal coronality, but the acoustic cues for voicelessness in Dutch are neutral, or at best very weak (van Alphen 2004, van Alphen & Smits 2004), and are not used by the 24-month-olds to search the lexicon specifically for voiceless stops. Thus, when voiced stops were mispronounced as voiceless, the 24-month-olds perceived these as neutral in terms of voicing characteristics, and both voiced and voiceless stops remained active potential candidates. Voiceless mispronunciations of voiced stops therefore do not impair word recognition, as opposed to coronal mispronunciations of labial stops, where the signal contains clear acoustic coronal place of articulation cues (*vs*. neutral voicing cues).

Correctly pronounced voiced stops. When voiced stops were correctly pronounced (in Experiment 1), the clear acoustic (pre)voicing cues in the speech signal led to a perfect match with and immediate activation of voiced-initial target words, which are specified as [voiced] in the lexicon. (Note again that the current experiment was not designed to draw conclusions on whether we can see evidence for different degrees of word activation on the trial types where voiced-initial words are pronounced correctly or mispronounced.)

Mispronounced voiceless stops. On the test trials in Experiment 1 where voiceless stops were mispronounced as voiced, we saw that this led to impaired word recognition, even though perceived cues for (pre)voicing in the speech signal do not create a mismatch with the voiceless target words (which are stored with no voicing specification). The presence of prevoicing in the signal is a strong acoustic cue that matches a [voiced] specification in the lexicon: the voiced targets are in this case preferred and activated, as the prevoicing in the signal unambiguously matches the perceived voicing in the signal. This leads to slower and shorter looking times by the 24-month-olds for voiceless targets, because the absence of a representation of voicing features in these candidates means they are not sufficiently activated to be potential word candidates in the lexical competition. The different nature of the place and voicing contrasts therefore leads to a fundamental difference between coronal and voiceless stops, with sufficient activation of [default] coronal-initial candidates when those are mispronounced as labial. In such cases the lexicon is arguably searched for [labial]-initial targets, but the perception of labial place of articulation speech cues does lead to enough activation of the [default] coronal-initial words to remain potential candidates in the competition. In contrast, when voiceless stops are mispronounced as voiced, the lexicon is searched for [voiced]-initial targets, and the perception of (pre)voicing speech cues does not sufficiently activate the voiceless-initial words, because they have a neutral or absent voicing specification.

Correctly pronounced voiceless stops. On the last type of test trial, when voiceless-initial targets are correctly pronounced, the neutral voicing cues in the speech signal display neither a match nor a mismatch with the voiceless stops in the lexicon, which are not specified for voicing, and therefore voiceless-initial targets activate these potential word candidates (in the same way that voiced-initial stops are activated), so that word recognition is unproblematic.

Finally, the patterns for voicing found in Experiment 2 with 20-month-olds, who did not show differences in word recognition on the MP-voice *vs*. the CP trial types, are explained by assuming that these children do not yet have the emerging language-specific voicing contrast represented in their lexicon. Both voiced and voiceless stops are therefore unspecified for voicing at this stage, and because there are no voicing features represented in the lexicon, even (pre)voicing in the signal does not yet lead to the search for voicing features in the lexicon. Thus voicing mispronunciations will not lead to impaired word recognition at that stage of acquisition.

In summary, a featural account combined with the consideration of the acoustic salience of cues in the signal can predict the perceptual asymmetries found in Experiments 1 and 2, on the assumptions that labial stops are specified as [labial] and Dutch voiced stops are specified as [voiced] as soon as children have acquired the voicing contrast. In addition, the difference between the patterns found for place of articulation and voicing are explained not only by the nature of their lexical representations *per se*, but also by the differences between the acoustic cues signalling these contrasts (which in our view follow in part from the different nature of the contrasts): for both labials and coronals, as well as for voiced stops, speech cues are salient and clearly present in the signal, whereas the cues for voicelessness can be considered ‘neutral’ (or absent).

5 Conclusions

We have shown that for both language-general (place of articulation) and language-specific (voicing) types of phonological contrasts, bidirectional mispronunciations yield different effects on children's online word recognition. Regardless of whether one favours a perceptual acoustic account or a phonological featural account to account for the data, this study was designed to show whether there is evidence for perceptual asymmetries in word recognition, and if so, how they work. The asymmetrical results that were found in the present study are important to consider for any study using a mispronunciation-detection method. Importantly, we do not claim that lexical representations of familiar words are in general holistic in nature, as has been argued in previous research, such as the original Hallé & de Boysson-Bardies (1996) study. Instead, we claim that early words are specified in a very systematic way, along the lines of phonological theories on early speech production, and theories assuming that certain features are acquired, and emerge later than others in the lexicon. Our results do not contradict earlier claims that children can detect one-feature mispronunciations of familiar words, as shown in studies such as Swingley & Aslin (2000). Rather, our dataset adds to this body of research, in that our results show that the direction of a sound change can matter. Our study combines psycholinguistic methods for investigating speech perception and production in young children with insights from phonological theory. As suggested by Lahiri & Marslen-Wilson (1991: 154), psycholinguistic models cannot be sufficiently precise without a proper specification of the lexical representations that are the target of lexical access and selection. We have argued that our results support a theory of underspecification of phonological contrasts in the lexicon (in this case, of voiceless stops for 24-month-olds and coronal stops for 20- and 24-month-olds), as well as providing evidence for the absence of emerging language-specific featural contrasts in the lexical representations of younger children, in this case the absence of representations of voicing for Dutch-learning 20-month-olds. Therefore, any account that aims to predict effects of characteristics of the speech signal on online word recognition must always be combined with knowledge of the acoustic characteristics of a specific contrast in order to make contrast-specific predictions. Crucially, our results show that having one segmental contrast or member of a contrast represented does not automatically mean that all contrasts are represented, nor that all perceived phonetic cues are part of a child's phonological lexical representations. As language experience increases and the lexicon grows, children add more phonological detail to their abstract phonological lexical representations, which can then mediate between perception and production.

REFERENCES

Alphen, Petra M. van (2004). *Perceptual relevance of prevoicing in Dutch*. PhD dissertation, Radboud University Nijmegen.

Alphen, Petra M. van & Roel Smits (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *JPh* **32**. 455–491.

Altvater-Mackensen, Nicole, Suzanne V. H. van der Feest & Paula Fikkert (2014). Asymmetries in early word recognition: the case of stops and fricatives. *Language Learning and Development* **10**. 149–178.

Aslin, Richard N., David B. Pisoni, Beth L. Hennessy & Alan J. Perey (1981). Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience. *Child Development* **52**. 1135–1145.
Bailey, Todd M. & Kim Plunkett (2002). Phonological specificity in early words. *Cognitive Development* **17**. 1265–1282.
Barlow, Jessica A. & Judith A. Gierut (1999). Optimality Theory in phonological acquisition. *Journal of Speech, Language and Hearing Research* **42**. 1482–1498.
Bergelson, Elika & Daniel Swingley (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences* **109**. 3253–3258.
Booij, Geert (1995). *The phonology of Dutch*. Oxford: Clarendon Press.
Cho, Taehong & Peter Ladefoged (1999). Variation and universals in VOT: evidence from 18 languages. *JPh* **27**. 207–229.
Clements, G. N. (2001). Representational economy in constraint-based phonology. In T. Alan Hall (ed.) *Distinctive feature theory*. Berlin & New York: Mouton de Gruyter. 71–146.
Clements, G. N. (2003). Feature economy in sound systems. *Phonology* **20**. 287–333.
Cutler, Anne & Charles C. Clifton, Jr. (1999). Comprehending spoken language: a blueprint of the listener. In Colin M. Brown & Peter Hagoort (eds.) *The neurocognition of language*. Oxford: Oxford University Press. 123–166.
Davis, Barbara L. & Peter F. MacNeilage (2000). An embodiment perspective on the acquisition of speech perception. *Phonetica* **57**. 229–241.
Dresher, B. Elan (2009). *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.
Eimas, Peter D., Einar R. Siqueland, Peter Jusczyk & James Vigorito (1971). Speech perception in infants. *Science* **171**. 303–306.
Eulitz, Carsten & Aditi Lahiri (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience* **16**. 577–583.
Feest, Suzanne V. H. van der (2004). Acquiring voicing in Dutch: the role of function words. In Leonie Cornips & Jenny Doetjes (eds.) *Linguistics in the Netherlands 2004*. Amsterdam & Philadelphia: Benjamins. 47–57.
Feest, Suzanne V. H. van der (2007). *Building a phonological lexicon: the acquisition of the Dutch voicing contrast in perception and production*. PhD dissertation, Radboud University Nijmegen.
Fennell, Christopher T. (2012). Object familiarity enhances infants' use of phonetic detail in novel words. *Infancy* **17**. 339–353.
Fennell, Christopher T. & Sandra R. Waxman (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development* **81**. 1376–1383.
Fenson, Larry, Phillip S. Dale, J. Steven Reznick, Donna Thal, Elizabeth Bates, J. P. Hartung, Stephen J. Pethick & Judy S. Reilly (1993). *The MacArthur communicative development inventories: user's guide and technical manual*. San Diego: Singular.
Ferguson, Charles A. & Carol B. Farwell (1975). Words and sounds in early language acquisition. *Lg* **51**. 419–439.
Fernald, Anne, Renate Zangl, Ana Luiz Portillo & Virginia A. Marchman (2008). Looking while listening: using eye movements to monitor spoken language comprehension by infants and young children. In Irina A. Sekerina, Eva M. Fernández & Harald Clahsen (eds.) *Developmental psycholinguistics: on-line methods in children's language processing*. Amsterdam & Philadelphia: Benjamins. 97–135.
Fikkert, Paula (2005). Getting sounds structures in mind: acquisition bridging linguistics and psychology? In Anne Cutler (ed.) *Twenty-first century psycholinguistics: four cornerstones*. Hillsdale: Erlbaum. 43–56.

Fikkert, Paula (2007). Acquiring phonology. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press. 537–554.

Fikkert, Paula & Clara C. Levelt (2008). How does Place fall into place? The lexicon and emergent constraints in children's developing phonological grammar. In Peter Avery, B. Elan Dresher & Keren Rice (eds.) *Contrast in phonology: theory, perception, acquisition*. Berlin & New York: Mouton de Gruyter. 231–268.

Golinkoff, Roberta Michnick, Kathryn Hirsch-Pasek, Kathleen M. Cauley & Laura Gordon (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. *Journal of Child Language* **14**. 23–45.

Hallé, Pierre A. & Bénédicte de Boysson-Bardies (1996). The format of representation of recognized words in infants early receptive lexicon. *Infant Behavior and Development* **19**. 463–481.

Hollich, George (2005). *Supercoder: a program for coding preferential looking* (version 10.5). Software package. Available (May 2015) at http://hincapie.psych.purdue.edu/Splitscreen/.

Ingram, David (1974). Phonological rules in young children. *Journal of Child Language* **1**. 49–64.

Iverson, Gregory K. & Joseph C. Salmons (1995). Aspiration and laryngeal representation in Germanic. *Phonology* **12**. 369–396.

Jakobson, Roman (1941). *Kindersprache, Aphasie, und allgemeine Lautgesetze*. Translated 1968 by Allan R. Keiler as *Child language, aphasia and phonological universals*. The Hague & Paris: Mouton.

Johnson, Elizabeth K. (2005). English-learning infants' representations of word forms with iambic stress. *Infancy* **7**. 99–109.

Jusczyk, Peter W. (1997). *The discovery of spoken language*. Cambridge, Mass.: MIT Press.

Jusczyk, Peter W. & Richard N. Aslin (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology* **29**. 1–23.

Kager, René, Suzanne V. H. van der Feest, Paula Fikkert, Annemarie Kerkhoff & Tania S. Zamuner (2007). Representations of [voice]: evidence from acquisition. In Jeroen van de Weijer & Erik Jan van der Torre (eds.) *Voicing in Dutch: (De)voicing – phonology, phonetics, and psycholinguistics*. Amsterdam & Philadelphia: Benjamins. 41–80.

Kager, René, Joe Pater & Wim Zonneveld (2004). Introduction: constraints in phonological acquisition. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in phonological acquisition*. Cambridge: Cambridge University Press. 1–53.

Kehoe, Margaret M., Conxita Lleó & Martin Rakow (2004). Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition* **7**. 71–88.

Kuhl, Patricia K., Jean E. Andruski, Inna A. Chistovich, Ludmilla A. Chistovich, Elena V. Kozhevnikova, Viktoria L. Ryskina, Elvira I. Stolyarova, Ulla Sundberg & Francisco Lacerda (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* **277**. 684–686.

Kuijpers, Cecile T. L. (1996). Perception of the voicing contrast by Dutch children and adults. *JPh* **24**. 367–382.

Lahiri, Aditi & William Marslen-Wilson (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* **38**. 245–294.

Lahiri, Aditi & Henning Reetz (2002). Underspecified recognition. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory Phonology 7*. Berlin & New York: Mouton de Gruyter. 637–675.

Levelt, Clara C. (1994). *On the acquisition of place*. PhD dissertation, University of Leiden.

McClelland, James L. & Jeffrey L. Elman (1986). The TRACE model of speech perception. *Cognitive Psychology* **18**. 1–86.

Mani, Nivedita & Kim Plunkett (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language* **57**. 252–272.
Marslen-Wilson, William D. (1987). Functional parallelism in spoken word-recognition. *Cognition* **25**. 71–102.
Maye, Jessica, Janet F. Werker & LouAnn Gerken (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* **82**. B101–B111.
Nazzi, Thierry (2005). Use of phonetic specificity during the acquisition of new words: differences between consonants and vowels. *Cognition* **98**. 13–30.
Paradis, Carole & Jean-François Prunet (1989). On coronal transparency. *Phonology* **6**. 317–348.
Paradis, Carole & Jean-François Prunet (eds.) (1991). *The special status of coronals: internal and external evidence*. San Diego: Academic Press.
Pater, Joe, Christine L. Stager & Janet F. Werker (2004). The perceptual acquisition of phonological contrasts. *Lg* **80**. 384–402.
Pater, Joe & Adam Werle (2003). Direction of assimilation in child consonant harmony. *Canadian Journal of Linguistics* **48**. 385–408.
Quam, Carolyn & Daniel Swingley (2010). Phonological knowledge guides 2-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language* **62**. 135–150.
Rice, Keren (1996). Default variability: the coronal–velar relationship. *NLLT* **14**. 493–543.
Rose, Yvan (2000). *Headedness and prosodic licensing in the L1 acquisition of phonology*. PhD dissertation, McGill University.
Rose, Yvan & Sharon Inkelas (2011). The interpretation of phonological patterns in first language acquisition. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.) *The Blackwell companion to phonology*. Malden, Mass.: Wiley-Blackwell. 2414–2438.
Smith, Neilson V. (1973). *The acquisition of phonology: a case study*. Cambridge: Cambridge University Press.
Smits, Roel, Natasha Warner, James M. McQueen & Anne Cutler (2003). Unfolding of phonetic information over time: a database of Dutch diphone perception. *JASA* **113**. 563–574.
Stager, Christine L. & Janet F. Werker (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* **6640**. 381–382.
Stevens, K. N. & S. E. Blumstein (1978). Invariant cues for place of articulation in stop consonants. *JASA* **64**. 1358–1368.
Swingley, Daniel (2003). Phonetic detail in the developing lexicon. *Language and Speech* **46**. 265–294.
Swingley, Daniel (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science* **8**. 432–443.
Swingley, Daniel (2010). Fast mapping and slow mapping in children's word learning. *Language Learning and Development* **6**. 179–183.
Swingley, Daniel & Richard N. Aslin (2000). Spoken word recognition and lexical representation in very young children. *Cognition* **76**. 147–166.
Swingley, Daniel & Richard N. Aslin (2002). Lexical neighborhoods and word-form representations of 14-month-olds. *Psychological Science* **13**. 480–484.
Van de Velde, Hans & Roeland van Hout (2001). The devoicing of fricatives in a reading task. In Ton van der Wouden & Hans Broekhuis (eds.) *Linguistics in the Netherlands 2001*. Amsterdam & Philadelphia: Benjamins. 219–229.
Vihman, Marilyn M., Satsuki Nakai, Rory A. DePaolis & Pierre A. Hallé (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language* **50**. 336–353.

Werker, Janet F. (1995). Exploring developmental changes in cross-language speech perception. In Lila R. Gleitman & Mark Liberman (eds.) *Language*. 2nd edn. Cambridge, Mass.: MIT Press. 87–106.
Werker, Janet F. & Suzanne Curtin (2005). PRIMIR: a developmental framework of infant speech processing. *Language Learning and Development* **1**. 197–234.
Werker, Janet F., Christopher T. Fennell, Kathleen M. Corcoran & Christine L. Stager (2002). Infants' ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy* **3**. 1–30.
White, Katherine S. & James L. Morgan (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language* **59**. 114–132.
Yoshida, Katherine A., Christopher T. Fennell, Daniel Swingley & Janet F. Werker (2009). Fourteen-month-old infants learn similar sounding words. *Developmental Science* **12**. 412–418.
Zamuner, Tania S. (2006). Sensitivity to word-final phonotactics in 9- to 16-month-old infants. *Infancy* **10**. 77–95.
Zink, Inge & Maryline Lejaegere (2002). *N-CDIs: lijsten voor communicatieve ontwikkeling. Aanpassing en hernormering van de MacArthur CDIs van Fenson et al*. Leuven: Acco.