Buckler, Helen and Fikkert, Paula (2015) Dutch and German 3-year-olds' representations of voicing alternations. Language and Speech, 59 (2). pp. 236-265. ISSN 1756-6053

Article

Dutch and German 3-year-olds' representations of voicing alternations

Helen Buckler[1*] & Paula Fikkert[2]

[1] University of Toronto Mississauga, Canada

[2] Radboud University Nijmegen, The Netherlands

* Corresponding author:

University of Toronto,

3359 Mississauga Road

Mississauga

ON, L5L 1C6, Canada

helen.buckler@utoronto.ca

Abstract

The voicing contrast is neutralised syllable and word finally in Dutch and German, leading to alternations within the morphological paradigm (e.g. Dutch 'bed(s)', *be[t]-be[d]en*, German 'dog(s)', *Hun[t]-Hun[d]e*). Despite structural similarity, language-specific morphological, phonological and lexical properties impact on the distribution of this alternation in the two languages. Previous acquisition research has focused on one language only, predominantly focusing on children's production accuracy, concluding that alternations are not acquired until late in the acquisition process in either language. This paper adapts a perceptual method to investigate how voicing alternations are represented in the mental lexicon of Dutch and German 3-year-olds. Sensitivity to mispronunciations of voicing word-medially in plural forms was measured using a visual fixation procedure. Dutch children exhibited evidence of overgeneralising the voicing alternation, whereas German children consistently preferred the correct pronunciation to mispronunciations. Results indicate that the acquisition of voicing alternations is influenced by language-specific factors beyond the alternation itself.

Keywords

Language acquisition, lexical representation, cross-language, alternations, morphophonology

Introduction

In this paper we use a perception-based task to investigate how morphophonological alternations are represented in the toddler’s mental lexicon. Morphophonological alternations are changes in the surface phonetic form of the stem or affix that arise due to the application of inflectional morphology, ensuring the surface form adheres to language-specific phonotactic patterns. The name derives from the position that these alternations occupy at the boundary of phonology/phonetics and morphology.

While the acquisition of language-specific phonotactics (e.g. Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Mattys & Jusczyk, 2001), and the acquisition of inflectional morphology (e.g. Cazden, 1968; Clahsen, Rothweiler, Woest, & Marcus, 1992; Mervis & Johnson, 1991) have often been studied, there has been little research into the interface of these two domains. Morphophonological alternations are acknowledged as being one of the most cognitively complex processes to acquire, with acquisition not being achieved until adolescence (Kiparsky & Menn, 1977; Pierrehumbert, 2003). Despite the long history of this observation there has been little experimental investigation into the acquisition of these processes. Existing papers on the acquisition of morphophonological alternations have primarily been interested in children’s productions, and their ability to generalise alternation patterns to novel forms (Kerkhoff, 2007; Van de Vijver & Baer-Henney, 2011; Van Wijk, 2007; Zamuner, Kerkhoff, & Fikkert, 2011), using methods based on Berko (1958), or using artificial language learning paradigms (e.g. Finley & Badecker, 2009; Pater & Tessier, 2003). In contrast, we use a perception-based task, focusing not on the process of acquisition itself, but on how morphophonological alternations are represented in the developing mental lexicon.

Although the precise combination of a phonotactic constraint and its interaction with morphophonology is language-specific, similar patterns are attested in multiple languages. This paper focuses on the voicing alternation in Dutch and German that occurs due to final devoicing (e.g. Dutch 'bed(s)' [bɛt]~[bɛdən]). Considering how morphophonological alternations occur at the intersection of domains, their acquisition must be studied within the context of both the morphological and phonological literature, specifically, morphological models of lexical access and the phonology of the voicing contrast. Comparing the same phenomenon and conducting parallel experiments in two languages enables us to examine the influence of language-general and language-specific factors on toddlers' lexical representations.

Both Dutch and German have a two-way voicing contrast between voiced and voiceless obstruents, as shown in Table 1. In both languages both voiced and voiceless obstruents occur in onset and medial positions, but only voiceless obstruents are permitted syllable- or word-finally. The voicing contrast is neutralised in this position. Neutralisation of voicing word-finally is a phonotactic constraint that occurs across the lexicon without regard for factors such as word-class or affix type. In this paper we focus on voicing alternations in singular-plural noun pairs.

[INSERT TABLE 1 ABOUT HERE]

Voicing alternations occur in morphological paradigms where a complex form contains a stem-final, but word-medial, voiced obstruent. When this stem occurs in isolation the voiced obstruent will be word-final and therefore voiceless. Consider, for example, the Dutch stem (and singular form) *bed* ('bed') with a word-final [t]. The plural *bedden* comprises the stem *bed* and the plural suffix *–en*. The stem-final segment is not word-final in the plural, and surfaces as [d]. Not all morphological paradigms contain voicing alternations, instead they have a voiceless segment throughout the paradigm. A comparable Dutch example is the word *pet* ('cap') which contains a [t] in both the singular *pet* and plural *petten*; [pɛt] and [pɛtən].

A comparison of voicing and voicing alternations in Dutch and German

In this study we aim to establish the impact of language-specific factors on the acquisition of a morphophonological alternation. Dutch and German are typologically related languages, both of which have a voicing alternation. Despite these apparent similarities, there are a number of differences between the languages affecting the acoustic realisation and phonological representation of the voicing contrast, the distribution of the voicing contrast, and the lexical frequency of alternations. We hypothesized that these differences would impact on the acquisition of voicing alternations by learners of each language.

The acoustics of the voicing contrast differ between Dutch and German, where Dutch is classified as a prevoicing language and German an aspirating language. This classification is based on Voice Onset Time (VOT) in obstruents. Lisker and Abramson (1964) identified VOT, the "timing relation between voice onset and the release of occlusion" (p. 387), as the primary marker of the voicing contrast cross-linguistically. Typically languages make use of three points on the VOT continuum; voicing lead (prevoicing), short-lag VOT (voiceless unaspirated) and long-lag VOT (voiceless aspirated). Languages with a two-way voicing contrast utilise two of the three points. In Dutch voiced stops are prevoiced, and voiceless stops are voiceless and unaspirated. Prevoicing is the strongest perceptual cue used by Dutch listeners, however there are individual differences in duration of prevoicing in production (Van Alphen & Smits, 2004). German distinguishes between voiceless unaspirated stops and voiceless aspirated stops. This cross-linguistic contrast is most pronounced in word-onset position. In intervocalic position German stops are often slightly voiced, however, this can be argued to be passive voicing due to the surrounding vocalic context, and is not as strong as the degree of voicing in stops in this position in prevoicing languages (Jessen & Ringen, 2002).

Acoustic differences contribute to debate about how the voicing contrast is represented phonologically at the feature level. The *Single Feature Hypothesis* (Kager, Van der Feest, Fikkert, Kerkhoff, & Zamuner, 2007) argues that the voicing contrast is phonologically the same in all languages, regardless of phonetic implementation. Voiced segments are marked as [voice], contrasting with either [-voice] or [ ], depending on whether features are assumed to be mono- or bivalent (Lombardi, 1995; Mester & Itô, 1989; Wetzels & Mascaró, 2001). The alternative view, the *Multiple Feature Hypothesis* (Kager et al., 2007), maintains that phonological representations reflect the phonetics more closely. Iverson and Salmons (1995) for example, argue that there are two features, [voice] and [spread glottis], only one of which is active in languages with a two-way voicing contrast. Following this argument, the relevant feature in Dutch is still [voice]; voiced segments are marked as [voice] and voiceless segments are unspecified. In aspirating languages, such as German, voiceless segments are specified as [spread glottis] and voiced segments are unspecified (Iverson & Salmons, 1995; Jessen & Ringen, 2002;

Jessen, 1998; Petrova, Plapp, Ringen, & Szentgyörgyi, 2006). Kager et al. (2007) present evidence from children's production errors supporting the Multiple Feature Hypothesis. They find that Dutch children make more devoicing errors whereas German children make more voicing errors. In both cases, Kager et al. (2007) argue, the error pattern can be attributed to neutralisation of the contrast to its unmarked value.

Differences in the voicing contrast between Dutch and German extend beyond the phonetic and phonological realisations of the contrast. There are also a number of differences in usage patterns and lexical distribution that may play a role in children's acquisition of voicing alternations.

Firstly, the reliability and variability of the voicing contrast in each language differs. According to Clements' (2003) theory of feature economy, German uses voicing to maximal effect, maintaining a voicing contrast for labial, alveolar and velar plosives, and labiodental and alveolar fricatives (/p/-/b/, /t/-/d/, /k/-/g/, /f/-/v/, /s/-/z/) in word-initial and word-medial position. An exception to this is that /s/ and /z/ are not contrastive word initially. The voicing contrast is more restricted in Dutch. For example, according to CELEX counts (Baayen et al., 1993, accessed via Reelex, the Reetz-CELEX interface, Reetz, 2010) there are only 110 minimal word-pairs that differ only in the [voice]-specification of the fricative (e.g. *fee* 'fairy' vs. *vee* 'cattle'). Moreover, the fricative voicing contrast has been neutralised in many regions of the Netherlands, making *fee* 'fairy' and *vee* 'cattle' indistinguishable; [feː] (Ernestus, 2000; van de Velde, Gerritsen, & van Hout, 1996). In contrast, there are 172 minimal pairs in German (also CELEX) that differ in fricative voicing, and this contrast is maintained by speakers. With regard to plosives, /g/ is not a native phoneme of Dutch, only occurring in a few loan words, e.g., *buggy, goal*. Finally, there are very few items with a final /b/, minimising morphological paradigms with a labial plosive voicing alternation. CELEX includes only 19 nouns with a [p]~[b] alternation between the singular and plural forms (e.g. *krab-krabben* 'crab(s)'). Consequently, many Dutch learners must glean their knowledge of voicing alternations from the alveolar plosives /t/ and /d/. German children on the other hand receive evidence from the whole class of obstruents.

There are also differences in the complexity of voicing assimilation patterns. Although both Dutch and German display voicing assimilation across word and morpheme boundaries, Dutch voicing assimilation is arguably more complex because it can be either progressive or regressive, depending on manner of articulation. In contrast, only progressive voicing assimilation is commonly attested in German. Booij (1995) describes two assimilation rules at play in Dutch: (1) before a voiced stop voiceless obstruents will be voiced, e.g., *voetbal* 'football' /tb/ will be realised as [db] due to regressive voicing assimilation. (2) Following a voiceless obstruent a voiced fricative will be devoiced, e.g., *opvallend* 'remarkable' /pv/ becomes [pf] due to progressive (de)voicing assimilation. Analysis of spontaneous speech corpora has indicated that this assimilation pattern is frequently attested, though not as strictly adhered to as previously believed (Ernestus, Lahey, Verhees, & Baayen, 2006), thereby adding further variation to the Dutch child's input. German predominantly displays progressive devoicing assimilation; following a voiceless obstruent, voiced plosives will be devoiced, e.g., *wegbringen* 'to take away' /kb/ becomes [kp] (Kohler, 1977). Thus, when a word-final voiceless obstruent is followed by a word-initial voiced obstruent, in German progressive devoicing occurs but in Dutch the speaker must also track manner of articulation as this determines the direction of the assimilation. Arguably the simpler system is easier for the German-learner to acquire.

The third cross-linguistic difference concerns other phonological alternations. Dutch has an optional alternation between /d/ and [j] or [ʋ]. For example, *rode* 'red' may be pronounced [rodə] or [rojə], and *oude* 'old' may be [ɔudə] or [ɔuʊə] (Booij, 1995). This alternation occurs in adjectives and verbs, but not nouns. Whether or not this alternation will impact on Dutch learners' acquisition of voicing alternations depends on whether children are sensitive to the word-class restriction. German also has an additional alternation on obstruents, namely between /g/ and [ç] following [ɪ], e.g., *König* 'king', [køːnɪç]. The limited context of this alternation leads us to believe that it is likely to be less disruptive to acquisition of the voicing alternation than the Dutch /d/~glide alternation.

Together these differences indicate that German-learning children may have an advantage over their Dutch peers when learning about voicing alternations. Voicing is not a reliable or robust cue in Dutch, and it is possible that Dutch learners pay little attention to it (cf. Warner, Smits, McQueen, & Cutler, 2005). In addition, German children have more experience with alternations because they encounter them across the whole class of obstruents, assimilation processes are clearer because assimilation goes in one direction only, and evidence for a voicing alternation is not masked by a conflicting alternation.

The final contrast between Dutch and German that we consider relates to differences in the lexical frequency of voicing alternations in the two languages. Corpus studies of voicing alternations in the two languages have previously been published (Kerkhoff, 2007; Van de Vijver & Baer-Henney, 2011), however, these analyses take different approaches and the data are not directly comparable across the two languages. We conducted our own corpus analysis of the frequency and distribution of voicing alternations in Dutch and German using a corpus of child directed speech available through the CHILDES database (MacWhinney, 2000). It is likely that there are differences between child-directed and adult-directed speech, therefore we do not claim that these data necessarily reflect the patterns in the languages as a whole[1], however, they provide an indication of the input that the child receives.

From the CLPF (Fikkert, 1994; C. C. Levelt, 1994) and Van Kampen (Van Kampen, 1994) corpora for Dutch, and the Leo and Rigol (Behrens, 2006) corpora for German we took all transcripts where the child was under 3;6[2]. This age limit corresponds to the age of the children participating in the experimental task. Following van de Vijver and Baer-Henney (2011) we extracted all singular nouns ending in an obstruent that take a vowel-initial plural suffix from the adult speaker tiers of our sub-corpus. In line with the prior discussion about the limited voicing contrast in Dutch, only the coronal and labial plosives were included for Dutch, though we refer to obstruents in both languages. Each word-final obstruent was labeled as alternating or not. Total type and token counts, and the proportion of alternations, are presented in Table 2.

[INSERT TABLE 2 ABOUT HERE]

There is no relationship between language and proportion of word types with an alternating final obstruent (χ^2 [1, N=1087]=0.57 p=.45). Of the singular nouns with a final obstruent that a German or Dutch child hears, approximately one third exhibit voicing alternation in the plural. Considering token frequency, a difference in distribution is attested

[1] However see van de Vijver & Baer-Henney, 2011 for similarities between child-directed and adult-directed speech.
[2] Transcripts
CLPF corpus: all transcripts. Van Kampen corpus: laura01-laura41, sarah01-sarah34. Leo corpus: le011112-le030529. Rigol corpus: cs000013-cs030513, pa000012-pa030519, sb000017-sb030519.

between Dutch and German (χ^2 [1, N=25197]=46.61 p<.05). If a Dutch-learning child hears a singular noun with a final obstruent, there is an equal chance of it alternating in the plural or not. For a German-learning child there is a slightly greater likelihood that it will not alternate. It should be noted that our result for Dutch differs slightly from Kerkhoff (2007), where 60% of Dutch tokens were found to be alternating. In contrast to our data she excluded stems where the final segment was preceded by an obstruent or liquid and these are included in our data set.

Stem forms alone do not provide the child with enough evidence to establish whether the final obstruent is alternating or not. It is only in the context of the morphological paradigm that alternations become apparent. From the same corpora we also extracted plural forms of each noun stem providing us with all singular-plural pairs in the corpus where the noun stem has a final obstruent (cf. Kerkhoff, 2007). This data is presented in Table 3.

[INSERT TABLE 3 ABOUT HERE]

Proportionally there is little difference between the type frequency of noun plurals that contain an alternation in Dutch and German: 38.6% of Dutch plural types are alternating and 36.7% of the German plural types. Again, a chi-square test was performed and no relationship was attested between language and type frequency, χ^2 (1, N=243) = 0.01, p=.92. A significant relationship was found between language and token frequency (χ^2 [1, N=2988] = 160.58, p<.05). In Dutch, only 32% of plural tokens are from a morphological paradigm with a voicing alternation, whereas in German the proportion is 63%. This result indicates that in German plurals from paradigms with alternations have some of the highest token frequencies. Of the top 10 most frequent plural forms in this corpus study, in German 8 of the 10 are from alternating paradigms. For comparison, in Dutch only 4 of the 10 are.

It is debatable whether type or token frequency is of greater importance for the child's ability to establish intraparadigmatic links. There is evidence that points to the role of type frequency (e.g. Ernestus & Baayen, 2003, 2007), but hearing multiple tokens presents the child with phonetically variable evidence, for example from different speakers of in different auditory situations, and such variability is also known to aid the formation of abstract representations (Pierrehumbert, 2003; Richtsmeier et al., 2011). If token frequency is important for acquisition of alternations, Dutch children do not receive as much evidence for alternations in nominal paradigms as German children do. Although the proportion of plural types containing alternations heard by German and Dutch children is similar, in German these items are encountered more frequently.

In sum, German children receive more cues in their input that may be useful in the acquisition of voicing alternations. We predict that in age-matched samples before full adult-like competency is acquired, German children will have more robust knowledge of lexical items with voicing alternations than their Dutch-learning peers. This prediction is grounded in the patterns of the phonological system and lexical frequency of alternations in their input.

This study

The aim of this study is to investigate how morphophonological alternations are represented in the developing mental lexicon, taking the voicing alternation in singular-plural pairs as a test case. We compare learners of two typologically related languages in order to assess the influence of language-specific factors on the representation and acquisition of morphophonological alternations. Using a mispronunciation detection paradigm we investigate the specificity of 3-year-olds’ representations of voicing in word-medial position in plural and monomorphemic words. Although primarily interested in representations of voicing alternations, it is necessary to establish how the voicing contrast is represented phonologically, in relation to the Single and Multiple Feature Hypotheses, by toddlers of each language group. For this reason, monomorphemic words were included where a voiced or voiceless obstruent occurs in the same phonological, but different morphological, context as in the plural words.

Previous research in both Dutch (Kerkhoff, 2007; Zamuner et al., 2011) and German (Van de Vijver & Baer-Henney, 2011) has used elicitation tasks to measure how accurate children are in producing alternations in familiar words. The common finding is that the production of voicing alternations, even in known words, remains difficult until at least school age, though there is a striking difference between the performance of Dutch and German children. German children’s productions are more accurate than their age-matched Dutch peers’. These studies also tested children’s ability to generalise voicing alternations to unknown or novel words in wug-test style experiments (Berko, 1958). In neither language did children produce many voicing alternations when inflecting nonsense words, although German children had a tendency to do so more often than Dutch children.

On the basis of production data alone it is difficult to ascertain how alternations are represented in the child’s mental lexicon (Zamuner et al., 2011). Even if a child correctly produces alternations in *bed* and *bedden* (Dutch 'bed(s)') it is impossible to determine whether they have an accurate representation of the voicing contrast and alternations, or have simply learned this word form. Similarly, it is not possible to conclude whether inaccurate productions stem from an inaccurate representation in the mental lexicon, or articulatory or task demands. Requiring an overt response from the child requires a willingness to cooperate and lack of shyness on their behalf (cf. Mills & Neville, 1997). It could be that children's lexical representations are immature, and their production abilities are an accurate reflection of their lexical representation (cf. Ferguson & Farwell, 1975; Fikkert, 2010; Vihman & Croft, 2007). On the other hand, children's lexical representations may be fully specified, but inaccuracies arise due to immature articulatory control, or difficulties mapping representations to articulatory gestures (cf. Inkelas & Rose, 2007; MacNeilage & Davis, 2000; Pierrehumbert, 2003).

For these reasons we used a perception task, the Intermodal Preferential Looking Task (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 2009), to measure 3-year-olds' sensitivity to mispronunciations of voicing word-medially in familiar words. Since Swingley and Aslin (2000) this procedure has often been used to test the phonetic specificity of young children's lexicons (e.g. Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Fennell & Werker, 2003; Mani & Plunkett, 2007; Mills et al., 2004; Swingley, 2003, 2009; Van der Feest, 2007). It is assumed that if the target word is familiar, and the mental lexicon contains a detailed phonetic representation of the word, then mispronunciations will be disruptive to word recognition.

Previous studies have primarily investigated mispronunciations of voicing in word-onset position, where mispronunciations are likely to be highly salient due to the importance of this position for lexical access. However, this paradigm has also successfully been used to demonstrate the specificity of representations word-finally (Swingley, 2009) and word-medially (Bowey & Hirakis, 2006; Swingley, 2003). Swingley (2003) found that Dutch 19-month-olds can detect mispronunciations of place of articulation from *baby* to **bady* or **bagy.* Frauenfelder, Scholten and Content (2001) also found that adults' lexical access is disrupted by word-medial mispronunciations. Indeed, Cole and Perfetti (1980) argued that listeners might be more sensitive to mispronunciations word-medially as they have already accessed the intended word from the correct first syllable, thereby making the second syllable more predictable and the mispronunciation more prominent. This explanation is likely to be more applicable in natural speech, where the context is not as restricted and the intended target less predictable than to an experimental paradigm.

We compared children's sensitivity to mispronunciations of stem-final obstruents occurring word-medially in plural forms (e.g. Dutch **pedden* for *petten* 'caps'or **betten* for *bedden* 'beds') and bisyllabic monomorphemic forms (e.g. Dutch **kedding* for *ketting* 'necklace' or **latter* for *ladder* 'ladder'). All word forms contained a word-medial obstruent in the same phonological context, but the morphological context differs. In plural forms the critical obstruent occurs in a potentially alternating context, but the absence of a morpheme boundary in monomorphemic words makes this a position where alternations cannot occur.

We predicted that all children would know the target voicing value of word-medial obstruents in monomorphemic words. Sensitivity to mispronunciations in these forms should therefore reflect the specificity of their phonological representations of voicing in word-medial position. If feature representations are fully specified, then both Dutch and German children are predicted to be sensitive to mispronunciations of voicing in both directions in these words. If feature representations are underspecified (Lahiri & Reetz, 2002), asymmetrical sensitivity to mispronunciations are predicted and the direction of the asymmetry is dependent upon how the voicing contrast is specified at the level of the feature. If the specified feature of the voicing contrast [voice], as is predicted for Dutch and German by the Single Feature Hypothesis (Lombardi, 1995; Mester & Itô, 1989; Wetzels & Mascaró, 2001), mispronunciations of voicing should be noticed (i.e. /t/ as *[d]), but not vice versa. This pattern was attested by (Van der Feest, 2007) when using a similar task to investigate Dutch toddlers' representations of voicing in word-onset position. If, as is predicted by the Multiple Feature Hypothesis (Kager et al., 2007), German represents voicing with the feature [spread glottis], the asymmetry will be reversed and participants are predicted to be sensitive to devoicing (i.e. /d/ as *[t]), but not voicing, mispronunciations.

Results for monomorphemic words, which provide insight into the participants' representations of the voicing contrast, affect predictions for how mispronunciations in plural word forms will be detected. We identified three possible hypothetical strategies that children may adopt. The first, the *Robust Representation Hypothesis*, assumes that children's representations of voicing alternations are accurate. Participants have knowledge of voicing alternations and which lexical items require an alternation in the plural. This predicts that children may be sensitive to mispronunciations in both directions (i.e. /t/ as *[d] or vice versa), however, the interaction of phonology and morphology plays a role. As such, mispronunciations of voicing in plural words are predicted to have the same effect on gaze behaviour as monomorphemic words, the direction of which is dependent upon the phonological specification of the voicing contrast (cf. Single vs. Multiple Feature Hypotheses). The second scenario, labeled here the *Open Hypothesis*, predicts that participants are aware that voicing alternations occur in morphological paradigms, but either have not specified which lexical items require an alternation, or have interpreted alternations as optional. This predicts that plural word-forms with both [t] and [d] would be considered acceptable, regardless of whether they are correct or not, and participants do not consider any of the mispronunciations as "wrong". This prediction does not bear on the phonological representation of voicing, and therefore predicts a difference in sensitivity to mispronunciations of plural and monomorphemic words. The final hypothesis is *Paradigm Uniformity,* which assumes that participants' representations contain a voiceless segment in all plural forms. This is the pattern attested in children's productions, where they fail to produce alternations (Kerkhoff, 2007; Van de Vijver & Baer-Henney, 2011; Zamuner et al., 2011). In this scenario, all words with a medial [t] conform to the child's lexical representation, whereas [d] mismatches. Using Dutch example words, *petten* ('caps') and **betten* ('beds') match, and **pedden* and *bedden* mismatch. This hypothesis interacts with the phonological representation of voicing. The Single Feature Hypothesis predicts that mispronunciations of /t/ as *[d] will be noticed, and therefore, predicts differences in looking behaviour to correct pronunciations and mispronunciations in all plural words. The Multiple Feature Hypothesis, however, predicts that the relevant feature for German is [spread glottis], and that mispronunciations of /t/ as *[d] will not be noticed. Therefore, the Paradigm Uniformity Hypothesis, in conjunction with the Multiple Feature Hypothesis, predict that German children will not be sensitive to mispronunciations in any plural words because in all cases they consider [t] to be the correct form, and [d] the incorrect form, and they are not predicted to be sensitive to mispronunciations in this direction. Note also that the Open Hypothesis and Paradigm Uniformity make the same predictions for sensitivity to mispronunciations of voicing in plural words by German children if they are representing the voicing contrast with the feature [spread glottis], namely, that they will not detect mispronunciations.

We expected German children to have a more robust lexical representation voicing alternations due to the properties of their linguistic environment discussed. As such, we hypothesized that their gaze behaviour would most likely adhere to the *Robust Representations Hypothesis.* Dutch children were predicted to have less knowledge of voicing alternations and adhere to either the *Paradigm Uniformity* or *Open* hypotheses. Plural words are the test case for distinguishing these hypotheses, together with monomorphemic words, which provide evidence for the feature specification of the voicing contrast in each language, i.e. the Single vs. Multiple Feature Hypotheses.

Experiment 1

Method

Participants. 37 Dutch-speaking children, with an average age of 37 months and 29 days (range: 37 months and 7 days to 38 months and 25 days, 19 girls), were included in the analysis. A further three children were tested but excluded from this analysis for fussiness or not participating in at least 8 of the 16 test trials. Children were recruited through the Baby Research Center of the Max Planck Institute for Psycholinguistics and Radboud University Nijmegen.

Materials. The test stimuli consisted of 16 bisyllabic nouns, half with word-medial /t/ and half with a word-medial /d/ (Table 4). Half of the words were plural forms and half were monomorphemic (singular) forms. Mispronunciations were created by changing the voicing value of the word-medial /t/ or /d/, i.e. /t/ became [d] and /d/ became [t].

Each test target item was yoked with a distractor item that should be familiar to 3-year-olds. The label of the distractor item had the same onset consonant as the target in order to delay participants' ability to make a decision until later in the word.

Items were selected on the basis of the following five criteria: (1) the medial obstruent should be intervocalic; (2) items should be easily depictable; (3) items should be familiar to children of this age; (4) mispronunciations should result in non-words; (5) targets should have a higher token frequency in the singular than the plural.

Criterion 5 limits the possibility that children are not interpreting highly frequent plurals, for example *tanden* 'teeth', as morphologically complex but instead treating them as non-decomposable units (Tesar & Prince, 2003). Frequency counts were obtained from CELEX, accessed via the Reetz-CELEX interface (Baayen et al., 1993; Reetz, 2010). One item, *noten* 'nuts', violates this condition in both CELEX and CHILDES counts, however, as it conformed to all other criteria and, in the absence of a more appropriate item, we nevertheless included it as a target word. Another item, *botten* 'bones', violates this condition in the CELEX count only, however, there is a difference in use of the item between child and adult language. In adult language *botten* is more frequently used to refer to bones as found in a skeleton and therefore in the plural. It is this information that is captured by CELEX. In the lexicon of a child the word *bot* is more often used in the context of a dog's bone, and occurs in the singular more often than the plural. This information is captured by the CHILDES frequency count, and conforms to our inclusion criteria.

The notion of familiarity was addressed by selecting items that were likely to be known by 3-year-olds. These were chosen from the Dutch version of the MacArthur

Communicative Development Inventory (Zink & Lejaegere, 2002). As this list does not provide information about inflected forms of individual lexical items, which is where our specific interest lay, we also considered whether an item appeared in a corpus of Child Directed Speech. We used the same corpora as before, namely the CLPF (Fikkert, 1994; C. C. Levelt, 1994) and the van Kampen (Van Kampen, 1994) corpora up to the age of 3;6 accessed through the CHILDES database (MacWhinney, 2000). We did not distinguish whether the word was uttered by the child or an adult, assuming that if an item occurs in the corpus it has been uttered in the presence of the child and may form part of their receptive lexicon.

Familiarity to each target and distractor item was measured per child, with items removed that individual children were reported not to know. One week before participating in the experiment parents were sent a picture book containing 64 colour photographs and their written names. Three adult native Dutch speakers had verified that the images were typical exemplars of the labeled category as would be understood by a young child. Parents were asked to read the book with their child and indicate on the accompanying questionnaire whether their child produced a given word, or understood it but did not say it. They were further asked whether their child recognised the image as its intended referent. We considered an item to be known by the child if the answer to any of these three questions was yes. Gaze data was removed from analysis for test trials where the child was unfamiliar with either the target or distractor item.

The 64 items in the picture book would all appear in the experiment as the target of a test trial (n=16), its yoked distractor (n=16), or as a target or distractor in a filler trial (n=32). The 32 filler items were paired into 16 filler trials. These items were selected based on their familiarity to 3-year-olds, and were always presented with their correct pronunciation. Although filler trials were not analysed, they were included in the picture book so as not to draw undue attention to the test items.

[INSERT TABLE 4 ABOUT HERE]

Audio stimuli were produced by a female Dutch speaker in a child-directed manner. Recordings were made in a sound-treated recording booth and digitised at a sampling rate of 44.1kHz and resolution of 16 bits in Adobe Audition. Stimuli were edited using Praat (Boersma & Weenink, 2011).There were no systematic differences in the duration ($t(15)=-.19$; $p=.86$) or pitch ($t(15)=.43$; $p=.67$) of correctly and incorrectly pronounced words. Intensity was equalised to 65dB.

Procedure. Children sat on their caregiver’s lap 60 cm away from the eye tracker monitor in a dimly lit room. Throughout the experiment the caregiver listened to masking music interspersed with speech through closed headphones. Stimuli were presented using Tobii-Studio software, and auditory stimuli were presented through centrally located loudspeakers below the screen. The test began with a 9-point calibration procedure. If not all points were calibrated in the first attempt, individual points were recalibrated a second time. The test began immediately after calibration.

In each trial participants saw two pictures on the screen. Visual stimuli were colour photographs of objects on a grey background, presented side by side on the 17-inch TFT monitor of a Tobii T60 eye tracker. A thin black vertical line divided the screen in two, and each photograph was positioned in the centre of the screen half.

Each child was presented with 32 trials divided into four blocks of eight trials. Half of the trials were test trials and half were filler trials. Test trials were equally divided over correct and mispronounced trials, counterbalanced for voicing and morphology. Six stimuli lists were created, varying which items were presented as correct pronunciations and which as mispronunciations. There was no repetition of the same item appearing in both a correct and mispronunciation trial. Filler trials served to increase the ratio of correctly pronounced to mispronounced trials over the whole experiment to 3:1.

A fixation cross was displayed in the centre of the screen for 500ms prior to each trial. After a silent preview of the images lasting 1600 ms the child heard *kijk!* ("look!"). 900ms later, 2500ms from the trial onset, the target word was presented. The trial ended after a further 2500ms. Images were presented on a grey background. A thin black vertical line divided the screen in two, and each image was positioned in the middle of either screen-half.

Data Analysis. A number of pre-defined exclusion criteria were applied to the data. Despite the loss of statistical power that results from removing data, these exclusion criteria were considered necessary due to the subtlety of the mispronunciation and the adverse effect that noise in the data may have on the results. In addition, we analyse our data using Growth Curve Analyses (cf. Mirman, Dixon, & Magnuson, 2008) which are more robust to missing data points (cf. Curran, Obeidat, & Losardo, 2010).

Firstly, unreliable measurement points were removed. The eye tracker codes each measurement point for validity or reliability from 0 (certain) to 4 (data missing or definitely incorrect). Following the recommendation of the manufacturer ("Tobii Studio 1.X User

Manual," 2008), measurement points with a validity code of 2 or higher were removed. This includes points where either the child was not looking at the screen, or points where they were looking to the screen but the tracking quality was poor.

Secondly, we removed data from whole trials if (1) the child did not look to the screen at all during the trial; (2) they did not look to both displayed images during the 2500 ms prenaming window; or (3) they did not look to either the target or distractor for at least 100 ms in the 2500 ms after the target onset. This ensured that we only included trials where the child was participating in the task.

Thirdly, trials were removed on the basis of parental report. Using the data from parent's questionnaires we removed trials from the analyses in which the child was unfamiliar with either the target or yoked distractor. 136 trials were removed for this reason.

The final criterion applied was to remove the participant from further analysis if, following all exclusion criteria, there were fewer than 50% of test trials remaining (fewer than 8 out of 16 trials). Data from three children were removed. On average each child contributed 12.73 trials, out of a possible 16, to the analysis (*SD* = 2.4, range = 9-16).

Two areas of interest (AOIs) were defined in the display. Each AOI corresponded to half of the display, excluding a 10 pixel-wide vertical line down the centre. Large AOIs were used to compensate for variability in children's looking behaviour or miscalibration of the eye tracker. The screen was blank apart from the two images. Fixations within either of the screen halves were considered to be object fixations. Fixations falling outside either AOI were considered as off-screen and not included in the analysis. Looks to the AOIs were coded for whether they were looks to the target or distractor.

We used Growth Curve Analysis (GCA) with orthogonal polynomials to quantify differences in the time-course of gaze behaviour towards the target picture in the different test conditions. GCA is a multi-level modeling framework designed to analyse change over time at group and individual levels (Singer & Willett, 2003). The time over which change is measured could be months or milliseconds, making this method suitable for analysing time course of fixations in an eye tracking study (see Mirman, Dixon, & Magnuson, 2008 for details of this method as applied to eye tracking data). The sampling rate of the eye tracker was 60 Hz, resulting in one data point per 16.7 ms.

The time window of analysis was 1000 ms in duration, starting 300 ms after the onset of the target word. Studies using the Visual World Paradigm with adults assume a latency of 200 ms (e.g. Allopenna, Magnuson, & Tanenhaus, 1998) whereas the assumed mean latency for infants is 367 ms (e.g. Swingley & Aslin, 2000). As noted by Swingley and Aslin (and references therein), minimum latencies to mobilise an eye movement in children can be as short as 233 ms, and they assume 367 ms as an "educated guess". Since the publication of Swingley and Aslin (2000), 367 ms has become the standard assumption in the field. This latency may well be a fair assumption for younger infants typically tested using this experimental paradigm, often between the ages of 18 and 24 months (e.g. Mani, Coleman, & Plunkett, 2008; Swingley & Aslin, 2000), and even as young as 12 months old (Mani & Plunkett, 2010). The children in our study were 38 months old, and, as it is well established that eye movement latency decreases with age (Miller, 1969), it is logical to assume that 3-year-olds will be faster in programming an eye movement than 18-month-olds.

GCA captures the pattern of the gaze behaviour data using two hierarchically related submodels. The first submodel, Level 1, captures the effects of time on fixation proportions using third-order orthogonal polynomials. A third-order polynomial was necessary to capture the S-shape of the data; the initial 50% fixations on the target, the following increase in fixations to the target and the final plateau. Other polynomials capture different elements of the shape of the data. By introducing orthogonal polynomials the intercept reflects the average height of the curve, making it analogous to more traditional analyses that average fixations over a specific time-window. The linear term reflects the overall angle of the curve (a straight line), and the quadratic term reflects a symmetric rise and fall rate around a central inflection point.

The Level 2 submodel captures the effects of experimental manipulation on the Level 1 intercept and linear time terms. Fixed effects of Pronunciation (correct or mispronounced), Morphology (monomorphemic or plural) and Target Voicing (canonical /d/ or /t/), and the interaction of these effects were included. We did not include effects of experimental manipulation on all Level 1 time terms as the cognitive interpretation of such effects is unclear (Mirman et al., 2008).

The Level 2 submodel also includes random effects for individual participants and items. We include random effects of individual participants and items on all three time terms, allowing for variation in the intercept, slope and curvature of the line. These were included to account for certain variation that is unknown in our data relating to individual participants and items. We do not know exactly when participants will initiate a shift of gaze to the target, or the speed with which their gaze will shift. Items are time-locked to the word onset, and the timing of critical obstruent relative to this varies per word depending on the duration of the first syllable. Differences in the intercept, slope and curvature resulting from these factors are accounted for in the random effects of our model.

The fixed and random effects of our model were justified by our experimental design and hypotheses. Following the recommendations Barr et al. (2013) random effects were "maximized", on the basis of the theoretically-relevant variation between participants and items.

The analysis was run in R (R Core Team, 2012) using the lmer function from package lme4 (Bates, Maechler, & Bolker, 2013). The data is binomial, with the dependent variable either 1 or 0 as the participant's gaze can either be on target or not. The reference levels were correct pronunciation, plural and underlyingly voiced. The model performs comparisons between each level of a factor and the baseline reference level, but not among the levels. Post hoc pairwise comparisons were conducted with the function glht from the package multcomp (Hothorn, Bretz, & Westfall, 2008) to quantify the effect of mispronunciations on each word type. This package simultaneously performs multiple comparisons on the model, and provides z-values of each comparison, and p-values corrected for multiple comparisons.

Results and Discussion

Effects on the intercept term capture differences in the overall average curve height (i.e. a higher intercept term reflects a higher proportion of fixations to the target). Effects on the linear time term reflect overall differences in the gradient of the slope (i.e. higher linear time term indicates a steeper slope, that is, faster shift of gaze to the target). We expected lower intercept and linear time terms for mispronunciation trials relative to correct pronunciation trials, indicating participants look less overall and are slower to identify the target when its label is mispronounced compared to when it is correctly pronounced.

Interactions in the model involving the factor Pronunciation are most relevant to our hypotheses. Complete results of the GCA are presented in Appendix 1. The three-way interaction between Pronunciation, Morphology and Target Voicing was significant on both the intercept and linear time terms (Intercept: β=-0.91; *SE*=0.13; p<.001. Linear Time: β=-16.34; *SE*=4.03; p<.001) indicating that the difference in the magnitude of the mispronunciation effect is greater between /d/ plural and monomorphemic words than between /t/ plural and monomorphemic words, in relation to both the height and slope of the curve. That is, variation in the effect of mispronunciations was greater for /d/ words than /t/ words, and this affects both recognition speed and overall time spent on looking to the target. In addition, the two-way interaction of Pronunciation and Morphology was also significant on both the intercept and linear time terms (Intercept: β=0.29; *SE*=0.09; p<.01. Linear Time: β=10.94; *SE*=2.81; p<.001), as was the two-way interaction of Pronunciation and Target Voicing (Intercept: β=0.26; *SE*=0.09; p<.01. Linear Time: β=6.97, *SE*=2.8, p<.05). Together these interactions indicate that both morphological structure and target voicing modified the effect mispronunciations had on word recognition. Pair-wise comparisons allow us to further investigate the effect of mispronunciations on each word type. Results of the pair-wise comparisons are shown in Table 5. Figure 1, which plots target fixations over time, also summarises the results of the pair-wise comparisons.

[INSERT FIG. 1 ABOUT HERE]

[INSERT TABLE 5 ABOUT HERE]

Monomorphemic Words. We predicted that children would be sensitive to mispronunciations in monomorphemic words, as this is a non-alternating context, with the possibility of an asymmetry due to underspecification of the feature [voice]. Note that both the Single and Multiple Feature Hypotheses assume the relevant feature in Dutch to be [voice]. Participants were sensitive to mispronunciations of /t/ in monomorphemic words. As expected, both the intercept and linear time terms are lower when the target was mispronounced compared to when it was correctly pronounced. This indicates that they were slower, and looked less overall, when the target word was mispronounced, suggesting that 3-year-olds have a robust representation of the voicing specification of the medial obstruent in these lexical items, and mispronunciations are disruptive to word recognition. For monomorphemic words with /d/, participants' gaze behaviour does show some differentiation between correct and mispronounced words. However, this is in the opposite direction to our predictions. The intercept term is higher for mispronunciations than correct pronunciations, indicating that they looked on average more to the target when it was mispronounced than when it was correctly pronounced. There was no difference in the speed of looks to the target, as shown by a non-significant effect on the linear time term. Interpretation of this result is less straightforward. One could argue that there is a statistical difference between the two conditions, therefore participants did notice that they were "different," reflecting some sensitivity to pronunciation. This is a somewhat unsatisfactory interpretation, as the direction of the result is counter to expectation. However, responses to correct pronunciation trials should be taken into account when judging the effect of a mispronunciation. In this case, participants were poor in their recognition of the target in its correctly pronounced form. Fixations to correctly pronounced monomorphemic /d/ words do not exhibit an S-shaped curve that would typically be associated with word recognition. In light of this, the meaning of a statistically significant effect is difficult to interpret reliably.

Monomorphemic words were the test case for determining whether asymmetries in sensitivity to voicing mispronunciations were found that would contribute to the discussion of how voicing is represented at the feature level in Dutch. Dutch data cannot support or refute the Single or Multiple Feature Hypotheses, as both assume the feature to be [voice], however, this data can provide insight into whether features are fully specified or underspecified (e.g. Lahiri & Reetz, 2002). If the feature [voice] is underspecified, then we expect to find an asymmetry in children's sensitivity to mispronunciations. In line with the predictions of underspecification, Van der Feest (2007) found that Dutch 2-year-olds were sensitive to mispronunciations of voiceless obstruents in word-onset position, but not of voiced obstruents. Our data for monomorphemic /t/ words are in line with Van der Feest's (2007) data; toddlers were sensitive to mispronunciations of /t/ word-medially in monomorphemic words. However, in order to attest asymmetries a clear result is needed for mispronunciations in both directions. Because of issues with word recognition in monomorphemic /d/ trials, and subsequent difficulty in interpreting whether children were sensitive to mispronunciations of /d/ or not, it is not possible to conclude whether our data support the underspecification hypothesis or not.

Plural Words. We hypothesized that Dutch toddlers would not have a fully developed lexical representation of whether a voicing alternation is found within a given morphological paradigm. We described two possible scenarios that toddlers may be adhering to, depending on how developed their knowledge of alternations is. On the one hand, the Paradigm Uniformity Hypothesis predicted that they may not realise that alternations occur, and therefore treat all plural forms with [d] as unacceptable, and accept all plural forms with [t], regardless of whether the presented form is a correct pronunciation or a mispronunciation. Alternatively, the Open Hypothesis predicted that they may have some knowledge that voicing alternations occur within morphological paradigms, but not specific knowledge of which morphological paradigms this applies to. In this case we predicted that plurals with both [t] and [d] would be considered acceptable, i.e. gaze behaviour to correct and mispronounced plural forms would not differ. Neither hypothesis is upheld by the data.

For plural /t/ words there was no difference on the linear time term, the gradient of the curve, between correct and mispronounced trials. There was a significant difference on the intercept term, where, counter to expectation, it was higher for mispronunciations compared to correct pronunciations. Participants looked to the target more when it was mispronounced than when it was correctly pronounced. Contrary to the uninterpretable result of monomorphemic /d/ words, which also elicited a similar statistical pattern, in the case of plural /t/ words looks to correct pronunciations display a typical S-shaped recognition curve. We can therefore interpret the increased looks to mispronunciations of the target word with more confidence. This result suggests that toddlers have a preference for mispronunciations over correct pronunciations.

For plural /d/ words, there was no difference on the intercept term between correct and mispronunciation trials; on average participants looked to the target image equally in the 1000ms time window. However, the time course of their gaze behaviour differed, as captured by the significant effect on the linear time term. Participants were faster to fixate on the target image when its label was correctly pronounced than when it was mispronounced. For correctly pronounced trials the time window captures the initial shift of gaze to the target, an increase in the height of the curve, a plateau where they are fixated on the target image, followed by a decrease in looks to the target as they start to look away. For mispronunciation trials the initial increase in looks occurs later, and the end of the time-window falls at a point just before they start to look away.

Considering the results for both plural /t/ and /d/ words together, in both cases toddlers show greater recognition for plural words with [d], whether this is implemented through the speed or duration of their looks to the target image. This could be interpreted as an overgeneralization of voicing alternations; Dutch toddlers expect plural words to contain a medial voiced obstruent. Although not a pattern we specifically predicted, this result is in line with our prediction that Dutch toddlers will have some knowledge of the occurrence of voicing alternations, but not yet have a mature representation of which morphological paradigms require a voicing alternation and which do not. We return to this point in the General Discussion.

One of our key hypotheses is that language-specific factors will have an impact on the course of acquisition of morphophonological alternations, specifically predicting that German toddlers will have a more robust representation of alternations than their Dutch-learning peers. Experiment 2 replicates Experiment 1 with German-learning children. We predicted that German children would display sensitivity to mispronunciations of both monomorphemic and plural words. In addition, German data is needed to test the predictions of the Single and Multiple Feature Hypotheses. The Multiple Feature Hypothesis predicts that German represents the voicing contrast with the feature [spread glottis]. If this feature is underspecified (c.f. Lahiri & Reetz, 2002) then we expect to find asymmetric sensitivity to mispronunciations by German children, whereby they are sensitive to mispronunciations of /d/ to [t], but not the reverse.

Experiment 2

Method

Participants. 23 German-speaking children with an average age of 37 months and 22 days (range: 36 months and 1 day to 39 months and 1 day, 12 girls) participated in this experiment. One further child was excluded from the analysis for fussiness. Children were recruited through the BabyLab of the University of Potsdam.

Materials. Materials were selected according to the same criteria as Experiment 1, and are presented in Table 6. It was not possible to find 16 target nouns with the target obstruent appearing in intervocalic position. We included 5 words where the target obstruent appeared after a sonorant (either [r], [l] or [n]). Note, however, that /r/ is vocalised following a long vowel in German. We did not expect these different contexts to have an influence on results. In addition, three target items did not fulfill the criterion that the mispronunciation should result in a non-word; the mispronunciations of *Leiter, Feder* and *Weide* are, for some speakers, the same as the real words *leider* 'unfortunately', *Väter* 'fathers' and *Weite* 'width'. These items were nevertheless selected as they were the best possible matches of other inclusion criteria and predicted to have little effect on the result as *Väter* and *Weite* are infrequent in the child's vocabulary, and *leider* is not a noun.

[INSERT TABLE 6 ABOUT HERE]

Speech stimuli were recorded by a female speaker of German in a child friendly manner in a sound-treated recording booth. Stimuli were recorded and prepared using the same equipment and in the same manner as Experiment 1. There were no systematic differences in the duration ($t(15)=.32$; $p=.75$) or pitch ($t(15)=.55$; $p=.59$) of the correctly and incorrectly pronounced words. Intensity was equalised to 65dB.

Visual stimuli also conformed to the same criteria as applied in Experiment 1. Three adult native German speakers verified that all images were typical exemplars of the labeled category as would be understood by a young child.

Procedure. The task was identical to Experiment 1, with a few minor alterations due to the different testing laboratories. Children sat independently on a chair or on their caregivers' laps, with their face 60-70 cm away from the Tobii monitor in a dimly lit room. If the child was on their caregiver's lap, the caregiver wore blacked-out glasses so they could not see the images displayed on the screen and influence their child's behaviour. If the child sat alone, the parent sat on a chair approximately 1m behind the child. Stimuli were presented using ClearView software on a Tobii 1750 eye-tracker. This eye tracker has a sampling rate of 50Hz. Auditory stimuli were presented through speakers located centrally beneath the screen.

The procedure began with a five-point calibration procedure, with second calibration of individual points that were not calibrated the first time. The test began immediately after the calibration procedure.

Each child was presented with 32 trials divided into four blocks of eight trials. Half of the trials were test trials, and half filler trials. The time-course of a trial was identical to Experiment 1. A trial lasted 5000 ms, and target and distractor images were displayed for the duration of the trial. The target word was presented after 2500 ms. Before the target was labeled, the child heard *Schau mal!* ('look').

Data Analysis. Data were prepared and analysed in the same manner as Experiment 1. Following the application of all exclusion criteria data remained for analysis from 23 participants, who contributed an average of 13.5 trials, out of a possible 16 (*SD* = 1.85, range = 11-16).

Results and Discussion

As in Experiment 1, interactions in the model that involve the factor Pronunciation are of most interest to our hypotheses. For clarification, effects on the intercept time term reflect differences in the average height of the curve, analogous to mean looking time. Effects on the linear time term reflect differences in the gradient of the curve, reflecting differences in the speed of looks to the target. Results of the pair-wise comparisons are presented in Table 7, and Figure 2 graphically presents the time-course information. Complete results of the GCA are presented in Appendix 2.

The three-way interaction of Pronunciation, Morphology and Target Voicing was significant on the intercept, but not the linear time term (Intercept: β=-0.56; *SE*=0.16; p<.001. Linear Time: β=-6.72; *SE*=5.01; p=.18). The difference in the magnitude of the mispronunciation effect (the difference in looking behaviour in correct and mispronunciation trials) was greater between monomorphemic and plural words with /t/ than /d/. The difference in size of the mispronunciation effect between monomorphemic and plural words with /d/ is reflected in the two-way interaction of Pronunciation and Morphology on the intercept term, which is only marginally significant (β=-0.2; *SE*=0.11; *p*=.08), indicating that differences between looks to the target in correct and mispronounced trials is similar, whether the trial is monomorphemic or plural /d/ words. On the linear time term the Pronunciation by Morphology interaction does reach significance (β=15.8; *SE*=3.59; *p*<.001), indicating that the difference in the size of the mispronunciation effect on the gradient of the curve is greater for plural /d/ words than monomorphemic /d/ words. That is to say, mispronunciations of plural words with /d/ are more disruptive to speed of recognition than are mispronunciations of monomorphemic /d/ words. The interaction of Pronunciation and Target Voicing does not reach significance on either the intercept or linear time terms (Intercept: β=0.08; *SE*=0.11; *p*=.46. Linear Time: β=-3.73; *SE*=3.57; *p*=.3). This result indicates that the size of the mispronunciation effect does not differ between plural words with /d/ or /t/.

[INSERT FIGURE 2 ABOUT HERE]

[INSERT TABLE 7 ABOUT HERE]

Monomorphemic Words. Toddlers were sensitive to mispronunciations of monomorphemic words with /t/. Both the intercept and linear time term are significantly higher for correct than mispronunciations. That is, as predicted, toddlers were faster to locate the target and spent on average more time looking to it, when its label was correctly pronounced than when it was mispronounced. For monomorphemic words with /d/ there was no difference in the gradient, i.e. speed, of shifts to the target image when its label was correctly pronounced or mispronounced. However, looking at this data, in both cases the curve is essentially flat. Participants failed to recognise the target word, regardless of its pronunciation. Statistically, participants looked on average more to the target when its label was mispronounced, but in the absence of recognition this effect is difficult to interpret. Note that a similar pattern was attested in Experiment 1, and we return to this in the General Discussion.

Even though results for monomorphemic /d/ words are inconclusive, the presence of an effect for monomorphemic /t/ words speaks against the Multiple Feature Hypothesis with underspecification (e.g. Iverson & Salmons, 1995; Jessen, 1998; Kager et al., 2007). This hypothesis assumes that the German voicing contrast is marked by the underspecified feature [spread glottis]; /t/ is marked as [spread glottis] and /d/ as [ ]. It predicts that speakers will not notice mispronunciations of /t/ to [d], as this mispronunciation involves the removal of a feature. Addition of a feature, such as occurs when /d/ is mispronounced as [t], should be noticed. Participants in Experiment 2 were sensitive to mispronunciations of /t/, therefore suggesting one of two possibilities. It may be that the Multiple Feature Hypothesis is correct, and [spread glottis] is the relevant feature, but it is not underspecified and we should not expect to find asymmetries. Alternatively, in line with the Single Feature Hypothesis, the relevant feature may be [voice] and not [spread glottis].

Plural Words. In plural trials with both /t/ and /d/ the linear time term is greater for correct pronunciations than mispronunciations. The gradient is steeper if the target word is correctly pronounced, indicating faster recognition. However, in both cases the average height of the curve, the intercept term, is higher for trials where the target word is mispronounced compared to when it is correctly pronounced. This goes against our predictions about the response a mispronunciation should elicit when compared to correct pronunciations. Looking at the plot, gaze behaviour when the target is correctly pronounced adhere to a typical S-shaped pattern; upon hearing the onset of the target word there is a slight delay before an increase in looks to target which subsequently plateaus. Therefore, unlike monomorphemic words with /d/, we can be certain that participants are recognising the target images as their intended referent when correctly pronounced. The differences we find between looks to correct and mispronounced trials can therefore be attributed to the effect of a mispronunciation on word recognition, rather than issues with item or picture recognition.

In both plural /t/ and /d/ words, the curve of looks to mispronunciations is shallower, yet on average higher, than to a correct pronunciation. Participants started to look to the target when it was mispronounced earlier than when it was correctly pronounced, but not with the same speed of conviction as when it was correctly pronounced, resulting in this crossed pattern where one line goes sharply from low to high, and the other starts higher, and gradually increases though never reaching the peak of the other. Considering the mispronunciation occurred word-medially, looks to correct and mispronunciations should be similar during the first half of the word as the two forms are supposedly acoustically identical until this point. This raises the question of why participants’ shifted their gaze to the target in mispronounced trials earlier than in correctly pronounced trials. The method employed during recording of stimuli may play a role. Mispronunciations were recorded in the form that they were presented in, that is, the speaker was required to utter **Bedden* (‘beds’ mispronunciation) or **Hunte* (‘dogs’ mispronunciation). We chose this method rather than cross-splicing tokens as it results in more natural sounding stimuli. However, for the native speaker recording these forms they are obvious non-words. Although we matched tokens for pitch and vowel duration between correct and mispronounced forms as much as possible it is nevertheless feasible that there were subphonemic cues present in the first syllable that participants were sensitive to. Crucially for our research question, the same pattern is attested in mispronunciations of plural words with /t/ or /d/, and looks to mispronunciations differ from looks to correct pronunciations in a similar way. This indicates that German toddlers are sensitive to mispronunciations of voicing in plural word forms.

We hypothesized that German toddlers would have more robust representations of voicing alternations within morphological paradigms, according to the Robust Representation Hypothesis, and this is supported by our data. They know which lexical items require an alternation and which do not, and mispronunciations of voicing affect word recognition. They are equally sensitive to mispronunciations of voicing in both directions in plural words. Therefore, we find no asymmetry that would enable us to distinguish between the Single and Multiple Feature Hypotheses. From this data we cannot conclude whether the relevant feature in German is [voice] or [spread glottis].

General discussion

This paper set out to investigate Dutch and German children's lexical representations of voicing alternations using an online mispronunciation detection paradigm. Previous literature has shown that voicing alternations are difficult to acquire, and 3-year-olds make many errors in their productions (Kerkhoff, 2007; Van de Vijver & Baer-Henney, 2011; Zamuner et al., 2011). We predicted that a more sensitive method than production data would indicate that their knowledge of voicing alternations is advanced of their production ability. We further predicted that German children would outperform their age-matched Dutch peers because the phonological system and lexical distribution of the voicing contrast and voicing alternations provide them with more robust cues.

Both predictions were upheld in Experiments 1 and 2. Despite making frequent errors in their productions, Dutch- and German-learning 3-year-olds' lexical representations do contain information about whether a stem-final obstruent should be voiced in the plural or not. Furthermore, German children's lexical representations are more robust than those of the Dutch children. This result provides evidence that this non-allophonic alternation is acquired earlier than previously believed, and well before adolescence, as has been claimed (Pierrehumbert, 2003). These data demonstrate that it is not (only) the cognitive complexity of morphophonological alternations that makes them difficult to acquire, but the properties of the native language exert a strong influence.

Three hypotheses were presented, demonstrating the possible ways in which toddlers may respond to mispronunciations of voicing in plural words. These responses would be a reflection of the robustness of their lexical representations. The *Robust Representation Hypothesis* predicted that children know which lexical items require an alternation, and their word recognition is disrupted by mispronunciations in both directions (either from /t/ to [d] or vice versa). The *Open Hypothesis* predicted that children may have some knowledge of voicing alternations, without having specified which paradigms require an alternation. In this situation toddlers were predicted not to be sensitive to mispronunciations in either direction, but to accept all forms as potentially valid pronunciations. Finally the *Paradigm Uniformity Hypothesis* predicted that toddlers would not expect voicing alternations to occur between the singular and plural forms of a noun, and the presence of a voiced segment would be unexpected and disruptive to word recognition. It was predicted that properties of the German language, including the reliability of voicing as a phonological cue and the frequent occurrence of noun plurals requiring an alternation, would aid children in their acquisition of voicing alternations. As such, German toddlers were hypothesized to adhere to the *Robust Representation Hypothesis.* Dutch does not provide learners with such reliable cues to learn voicing alternations, and therefore Dutch children were expected to follow either the *Open Hypothesis* or the *Paradigm Uniformity Hypothesis* in this task.

Data from Experiment 2 support the prediction that German children have robust representations of voicing in plural forms. In both plurals with /t/ and /d/ they noticed mispronunciations. When the plural was correctly pronounced they recognised the word and their looks to the target image increased. If the plural was mispronounced they displayed some evidence of word recognition, but were significantly slower in shifting their gaze to the target image.

Dutch participants in Experiment 1 displayed a different pattern of results in plural word trials, however the attested pattern did not conform to any of the three hypotheses, or any other prediction. In all cases it was predicted that sensitivity to mispronunciations would manifest itself in fewer and/or slower looks to the target, and this was the case for plural words with /d/ where they were slower to identify the target when its label was mispronounced. In plural words with /t/, however, participants looked more to the target on average when it was correctly pronounced than mispronounced. That is, they displayed a preference for plural forms produced with [d] rather than [t]. This could be interpreted in two ways. On the one hand, it might be evidence for overgeneralisation of voicing alternations; the child is aware that voicing may alternate between the singular and plural form and assumes that this should happen in all cases. Production data speaks against this interpretation. Children of this age, when asked to produce the plural of, for example, *bed* ('bed'), will consistently reply **betten*, and not *bedden* (Kerkhoff, 2007; Zamuner et al., 2011). Why do children adhere to Paradigm Uniformity in their speech, but not in an online perception task? It could be argued that children's productions are limited by their articulatory abilities, however Zamuner et al. (2011) rule this possibility out by assessing Dutch 3-year-olds' ability to imitate plural words with medial [t] or [d]. Toddlers were highly accurate in their imitations, suggesting that failure to produce alternations in plural forms is not due to speech or hearing constraints. It could also be argued that voiced segments are more natural, or less marked, than voiceless segments intervocalically (cf. Westbury & Keating, 1986), and children have a preference for the least marked form. However the predictions of markedness theory extend to production too, predicting that children should produce voiced segments in plural forms.

An alternative interpretation of the results for plural /t/ words in Experiment 1 is not that children are overgeneralising voicing alternations, but that it is a surprise effect; children expected to hear one form, and when they encounter something else the confusion causes them to spend longer looking at what they thought the target was going to be. It is only in this one word-type that participants display a preference for the mispronounced form, and one would expect the same effect to be found across the different word-types. However, if we consider how words enter and are represented in the mental lexicon, mispronunciations of plural /t/ words are somewhat unusual. Thinking in terms of a dual-route model of the mental lexicon (Baayen et al., 1997, 2003; Caramazza et al., 1988; Clahsen, 1999; W. J. Levelt et al., 1999; Marcus, 1995; Pinker, 1991) complex words are assumed to have two forms. On the one hand, the mental lexicon contains an accurate, fully listed form identical to the word's pronunciation, e.g. *petten* ('caps') /pɛtən/. On the other hand the mental lexicon contains the morphemes (stems and affixes) needed to compute complex words on the fly, e.g. *petten* ('caps') is concatenated from the stem *pet* and the plural suffix *–en*. The child encounters words in their environment, and can store these without necessarily needing to conduct any morphological analysis, i.e. their parents say *bedden* ('beds') with a [d] and *petten* ('caps') with a [t], and the child stores these forms as such. However, children are also able to generate their own complex forms according to generalisations that they make over the input they receive. For example, they may have a generalisation that the plural is formed by suffixing *–en* to the stem, with no reference to voicing alternations. For plural /d/ words they are able to access or generate forms with both [t] and [d]; accurate *bedden* is listed, and the inaccurate **betten* can be generated from the stem and suffix. For plural /t/ words on the other hand, both forms converge on the correct form with [t]. A mispronunciation of a plural /t/ word results in a form that has no place in the child's lexicon, as they have neither heard nor generated it. In this sense, mispronunciations of plural /t/ words could be considered an even greater violation of expectation than mispronunciations of plural /d/ words. This model, however, would make the same predictions for German, yet the German data of Experiment 2 indicate that mispronunciations of plural words with /t/ and /d/ are equally disruptive. In addition, no previous mispronunciation detection study finds evidence for increased looks to the mispronunciation compared to the correct pronunciation. Even though there is reason to argue that mispronunciations of plural /t/ words are different, there is little reason to expect that this difference alone would trigger the opposite pattern of gaze behaviour.

When comparing representations of voicing between Dutch and German it is also necessary to consider how voicing is represented separately from the issue of alternations. Monomorphemic words were included with the aim of establishing how the voicing contrast is represented at the level of the feature in each language. The Single Feature Hypothesis predicts that the relevant feature is [voice] in both Dutch, a true voicing language, and German, an aspirating language (Lombardi, 1995; Mester & Itô, 1989; Wetzels & Mascaró, 2001). The Multiple Feature Hypothesis maintains that German represents the voicing contrast with the feature [spread glottis] and not [voice] (e.g. Iverson & Salmons, 1995; Jessen, 1998; Kager et al., 2007). Underspecification theory (Lahiri & Reetz, 2002) further predicts asymmetries in the direction of sensitivity to mispronunciations, whereby mispronunciations from the underspecified to the specified value are noticed, but not vice versa. Accordingly, the Multiple Feature Hypothesis predicts that Dutch children should notice mispronunciations of /t/ to *[d] but not /d/ to *[t]. Conversely, German speakers should be sensitive to mispronunciations of /d/ to *[t], but not the reverse. German participants in Experiment 2 were sensitive to mispronunciations of /t/ in monomorphemic words, indicating that the relevant feature is not underspecified [spread glottis]. Dutch participants in Experiment 1 were also sensitive to mispronunciations of /t/ in monomorphemic words, which is in line with the predictions of both the Single and Multiple Feature Hypotheses. Results for monomorphemic /d/ words are needed to identify asymmetries in sensitivity to mispronunciations that would be indicative of underspecification. Unfortunately the results for these words were inconclusive.

Sensitivity to mispronunciations of monomorphemic words with /d/ were difficult to determine. In both Experiment 1 and 2 participants had issues recognising the target when it was correctly pronounced. This was unexpected as familiarity of the items to each child was controlled for. Of the 99 items that were removed for their unfamiliarity in Experiment 1, 38 were monomorphemic /d/ words. Similarly, in Experiment 2 38 items were removed on the basis of parental reports, of which 21 were monomorphemic /d/ words. In Experiment 1, a similar number of items were unfamiliar from the monomorphemic /d/ category and the plural /t/ category (38 and 37 respectively), yet plural words with /t/ were recognised better in the task. The by-item estimates of our model output reveal that there is no one item that stands out as not being recognised, however the class of monomorphemic /d/ words behave differently from all other categories. It seems, therefore, that in Experiment 1 removing items on the basis of parental report was justified in plural /t/ trials and the trials remaining in the analysis did come from children who were familiar with these items and recognised them in the task. It does not explain why children were poor to recognise monomorphemic /d/ trials.

A final possibility is that the monomorphemic /d/ words are less easily depictable than other conditions and therefore less recognisable for participants. According to parental reports, a number of children knew the words *schaduw* ('shadow') and *pudding* ('pudding') but did not associate the presented image with these word forms. Similarly, in German parents indicated that their child would be likely to label the image *Erde* ('earth') as *Welt* ('world'), and were unsure of how to label the image *Weide* ('meadow'). Because the word form was indicated as being known by the children we included these trials in the analysis, but it seems that there were issues with the clarity of the images that we failed to control for sufficiently. It is difficult to remove these trials reliably as there are inconsistencies in how parents completed the questionnaire.

Even if we discount monomorphemic /d/ words due to their poor recognition, the fact that children were sensitive to mispronunciations of /t/ is evidence that any asymmetry does not go in the opposite direction to Dutch. We find no evidenced in our data to support Multiple Feature Hypothesis with underspecification of [spread glottis], despite production data to the contrary (Kager et al., 2007). However, our data do not rule out the Multiple Feature Hypothesis with fully specified feature values, as this interpretation would not predict the presence of asymmetries in perception. Without reliable data from monomorphemic /d/ words, from this data we are unable to conclude whether the Multiple Feature Hypothesis (with no underspecification) is accurate and the relevant feature in German is [spread glottis], or, in line with the Single Feature Hypothesis, the relevant feature is [voice].

Taken together, the data indicate that German children have increased knowledge of voicing alternations when compared to age-matched Dutch children. We have argued that this advantage comes from differences in the phonological system of the two languages and lexical frequency of voicing alternations. A cross-linguistic study on two typologically related languages highlights the impact of subtle linguistic differences on the acquisition of similar phenomenon. Our results emphasise the role of the native language and how the frequency or saliency of supposedly "difficult" structures can influence acquisition. Previous literature has hinted at the role of frequency and native language as factors in the acquisition of morphophonological alternations. For example, Fikkert and Freitas (2006) argued that variation in the input allows children to acquire alternations in the European Portuguese vowel system at an early age, and Bals (2004) reported evidence for the acquisition of phonological and morphological relationships in North Saami by the age of 2;5. However, these studies compared the age-of-acquisition of different morphophonological alternations across different languages. By comparing the acquisition of the same morphophonological alternation by children learning two typologically related languages in similar cultural environments the role of language-specific factors are highlighted, allowing us to more confidently conclude that native language properties impact on children's acquisition of morphophonological alternations.

Previous literature on the acquisition of morphophonological alternations has not accounted for the role of input variation despite the emphasis that is placed on input in current understanding of language acquisition more generally. The prevalent view is that infants are born as "universal listeners" and during the first year of life their universal abilities diminish and language specific abilities are emphasised (see Cutler, 2012, Chapter 8 for overview). Infants' sensitivity to their native language develops through a variety of statistical mechanisms that allow the infant to learn from the speech stream alone, in the absence of top-down knowledge such as a lexicon. For example, infants are able to track, and make use of, the frequency of occurrence of segments (Saffran, Aslin, & Newport, 1996), how often they co-occur (Mattys & Jusczyk, 2001), or what the predominant stress patterns is (Jusczyk, Houston, & Newsome, 1999). Previous theories of the acquisition of non-allophonic morphophonological alternations have claimed that the system cannot be acquired without top-down knowledge from morphology and semantics (Peperkamp & Dupoux, 2002; Tesar & Prince, 2003). These theories claim that phonotactic knowledge can help infants initially in identifying that voicing is not contrastive in final position, but knowledge of which morphological paradigms contain an alternation can only be derived through the addition of morphological and semantic knowledge. Compared to these theories, results here show that bottom-up knowledge can help learners in learning morphophonological alternations to a greater extent than previously believed. German children performed better than their Dutch peers and we have little reason to believe that there are substantial differences in the general linguistic or cognitive capabilities of the two groups although they cannot entirely be ruled out. The difference specifically relates to how robust their knowledge of the voicing contrast and voicing alternations is. This knowledge is underpinned by the higher functional load of the voicing contrast in German; voicing and alternations are occur across the whole class of obstruents, and there are more lexical items with alternations. These cues can be condensed down to properties of variability and frequency, two cues that are known to be beneficial for learning.

Funding

This work was supported by an International Max Planck Research School for Language Sciences doctoral fellowship and a Deutsche Akademischer Austausch Dienst research grant for doctoral candidates, both awarded to the first author.

Acknowledgements

We thank Linda Drijvers, Linda Hormes and Sho Tsuji for recording the stimuli, Anne Beyer and Imme Lammertink for research assistance, Tom Fritzsche for technical assistance, Titia Benders, Christina Bergman, Anne Cutler, Evelien Krikhaar and Stefanie Ramachers for helpful comments and Barbara Höhle for access to lab facilities.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. doi:10.1006/jmla.1997.2558

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, *37*(1), 94–117. doi:10.1006/jmla.1997.2509

Baayen, R. H., McQueen, J. M., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: revisiting Dutch plurals. In R. H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 355–390). Berlin: Mouton de Gruyter.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX Lexical Database (CD-ROM)*. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early 2ords. *Cognitive Development*, *17*, 1265–1282.

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, *32*(1), 159–173.

Bals, B. A. (2004). The acquisition of grade alternation in North Saami. *Nordlyd*, *32*(1), 1–27.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., & Bolker, B. (2013). Package "lme4." Retrieved from http://lme4.r-forge.r-project.org/

Behrens, H. (2006). The Input-Output relationship in first language acquisition. *Language and Cognitive Processes*, *21*, 2–24.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.

Boersma, P., & Weenink, D. (2011). Praat: Doing Phonetics by Computer (version 5.2.33).

Booij, G. (1995). *The Phonology of Dutch*. Oxford: Clarendon Press.

Bowey, J. A., & Hirakis, E. (2006). Testing the protracted lexical restructuring hypothesis: the effects of position and acoustic-phonetic clarity on sensitivity to mispronunciations in children and adults. *Journal of Experimental Child Psychology*, *95*(1), 1–17. doi:10.1016/j.jecp.2006.02.001

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*, 297–332. doi:10.1016/0010-0277(88)90017-0

Cazden, C. (1968). The acquisition of noun and verb inflections. *Child Development*, *39*(2), 433–448. doi:10.2307/1126956

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, *22*(6), 991–1060.

Clahsen, H., Rothweiler, M., Woest, A., & Marcus, G. F. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition*, *45*(3), 225–255.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology*, *20*(3), 287–333.

Cole, R. A., & Perfetti, C. (1980). Listening for mispronunciations in a children's story: The use of context by children and adults. *Journal of Verbal Learning and Verbal Behavior*, *19*, 297–315.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about Growth Curve Modeling. *Journal of Cognition and Development*, *11*(2), 121–136. doi:10.1080/15248371003699969.Twelve

Cutler, A. (2012). *Native Listening: Language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.

Ernestus, M. (2000). *Voice Assimilation and Segment Reduction in Casual Dutch: a Corpus-Based Study of the Phonology-Phonetics Interface*. Free University, Amsterdam.

Ernestus, M., & Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, *79*(1), 5–38. doi:10.1353/lan.2003.0076

Ernestus, M., & Baayen, R. H. (2007). Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch. *Language and Cognitive Processes*, *22*(1), 1–24. doi:10.1080/01690960500268303

Ernestus, M., Lahey, M., Verhees, F., & Baayen, R. H. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America*, *120*(2), 1040–1051. doi:10.1121/1.2211548

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, *46*(2-3), 245–264.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, *51*(2), 419–439. doi:10.2307/412864

Fikkert, P. (1994). *On the acquisition of prosodic structure*. Holland Academic Graphics, The Hague.

Fikkert, P. (2010). Developing representations and the emergence of phonology: Evidence from perception and production. *Laboratory Phonology*, *10*(4), 227–255.

Fikkert, P., & Freitas, M. J. (2006). Allophony and allomorphy cue phonological acquisition: Evidence from the European Portuguese vowel system. *Catalan Journal of Linguistics*, *5*, 83–108.

Finley, S., & Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language, 61*(3), 423-437. doi: 10.1016/j.jml.2009.05.002

Frauenfelder, U. H., Scholten, M., & Content, A. (2001). Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*, *16*(5-6), 583–607. doi:10.1080/01690960143000146

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (2009). The eyes have it: lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(01), 23–45. doi:10.1017/S030500090001271X

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363. doi:10.1002/bimj.200810425

Inkelas, S., & Rose, Y. (2007). Positional neutralization: A case study from child language. *Language*, *83*(4), 707–736. doi:10.1353/lan.2008.0000

Iverson, G. K., & Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, *12*, 369–396. doi:10.1017/S0952675700002566

Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam / Philadelphia: John Benjamins.

Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology*, *19*(2), 189–218. doi:10.1017/S0952675702004311

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*, 402–420. doi:10.1006/jmla.1993.1022

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3-4), 159–207. doi:10.1006/cogp.1999.0716

Kager, R., Van der Feest, S., Fikkert, P., Kerkhoff, A., & Zamuner, T. S. (2007). Representations of [voice]: Evidence from acquisition. In J. van de Weijer &

E. J. van der Torre (Eds.), *Voicing in Dutch* (pp. 41–80). Amsterdam: John Benjamins.

Kerkhoff, A. (2007). *The Acquisition of Morpho-Phonology: The Dutch voicing alternation*. Utrecht University.

Kiparsky, P., & Menn, L. (1977). On the Acquisition of Phonology. In *Perspectives in Neurolinguistics and Psycholinguistics* (pp. 47–78). New York: Academic Press.

Kohler, K. J. (1977). *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt Verlag.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. *Laboratory Phonology*, *7*, 637–675.

Levelt, C. C. (1994). *On the acquisition of place*. The Hague: Holland Academic Graphics.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38; discussion 38–75.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384-422.

Lombardi, L. (1995). Laryngeal neutralization and syllable wellformedness. *Natural Language & Linguistic Theory*, *13*, 39–74. doi:10.1007/BF00992778

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, *288*, 527–531. doi:10.1126/science.288.5465.527

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk.* (Third Edit.). Mahwah, NJ: Lawrence Erlbaum Associates.

Mani, N., Coleman, J., & Plunkett, K. (2008). Phonological specificity of vowel contrasts at 18-months. *Language and Speech*, *51*(1-2), 3–21. doi:10.1177/00238309080510010201

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, *57*(2), 252–272.

Mani, N., & Plunkett, K. (2010). Twelve-month-olds know their cups From their keps and tups. *Infancy*, *15*(5), 445–470. doi:10.1111/j.1532-7078.2009.00027.x

Marcus, G. F. (1995). Children's overregularization of English plurals: a quantitative analysis. *Journal of Child Language*, *22*, 447–459.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91–121. doi:10.1016/S0010-0277(00)00109-8

Mervis, C. B., & Johnson, K. E. (1991). Acquisition of the plural morpheme: A case study. *Developmental Psychology*, *27*(2), 222–235. doi:10.1037/0012-1649.27.2.222

Mester, R. A., & Itô, J. (1989). Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language*, *65*(2), 258–293. doi:10.2307/415333

Miller, L. K. (1969). Eye-movement latency as a function of age, stimulus uncertainty, and position in the visual field. *Perceptual and Motor Skills*, *28*, 631–636. doi:10.2466/pms.1969.28.2.631

Mills, D. L., & Neville, H. J. (1997). Electrophysiological studies of language and language impairment. *Seminars in Pediatric Neurology*, *4*(2), 125–134. doi:10.1016/S1071-9091(97)80029-0

Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H. J., & Werker, J. F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience*, *16*(8), 1452–1464.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. doi:10.1016/j.jml.2007.11.006

Pater, J., & Tessier, A. (2003). Phonotactic knowledge and the acquisition of alternations. In M. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Conference of Phonetic Sciences* (pp. 1177-1180).

Peperkamp, S., & Dupoux, E. (2002). Coping with phonological variation in early lexical acquisition. In I. Lasser (Ed.), *The Process of Language Acquisition* (pp. 359–385). Berlin: Peter Lang Verlag.

Petrova, O., Plapp, R., Ringen, C., & Szentgyörgyi, S. (2006). Voice and aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish. *The Linguistic Review*, *23*(1), 1–35. doi:10.1515/TLR.2006.001

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2-3), 115–154. doi:10.1177/00238309030460020501

Pinker, S. (1991). Rules of language. *Science*, *253*(21), 530–535. doi:10.1126/science.1857983

R Core Team. (2012). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

Reetz, H. (2010). Reelex – Reetz-CELEX interface (Version 0.4.2) Retrieved from http://web.phonetik.uni-frankfurt.de/simplex.htm

Richtsmeier, P. T., Gerken, L., & Ohala, D. K. (2011). Contributions of phonetic token variability and word-type frequency to phonological representations. *Journal of Child Language*, *38*(5), 951–78. doi:10.1017/S0305000910000371

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. doi:10.1126/science.274.5294.1926

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis*. Oxford: Oxford University Press.

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech*, *46*, 265–294. doi:10.1177/00238309030460021001

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, *60*(2), 252–269. doi:10.1016/j.jml.2008.11.003

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166. doi:10.1016/S0010-0277(00)00081-0

Tesar, B., & Prince, A. (2003). Using phonotactics to learn phonological alternations. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* (Vol. 39, pp. 241–269). Chicago Linguistic Society.

Tobii Studio 1.X User Manual. (2008). Tobii Technology. Retrieved from www.tobii.com

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

Van de Velde, H., Gerritsen, M., & van Hout, R. (1996). The devoicing of fricatives in Standard Dutch: A real-time study based on radio recordings. *Language Variation and Change*, *8*, 149–175. doi:10.1017/S0954394500001125

Van de Vijver, R., & Baer-Henney, D. (2011). Acquisition of voicing and vowel alternations in German. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th Annual Boston University Conference on Language Development* (pp. 603–615). Somerville, MA: Cascadilla Press.

Van der Feest, S. (2007). *Building a Phonological Lexicon*. LOT.

Van Kampen, J. (1994). The learnability of the left branch condition. *Linguistics in the Netherlands*, 83–94.

Van Wijk, J. (2007). *The acquisition of the Dutch Plural*. Utrecht University.

Vihman, M., & Croft, W. (2007). Phonological development: toward a "radical" templatic phonology. *Linguistics*, *45*(4), 683–725. doi:10.1515/LING.2007.021

Warner, N., Smits, R., McQueen, J. M., & Cutler, A. (2005). Phonological and statistical effects on timing of speech perception: Insights from a database of Dutch diphone perception. *Speech Communication*, *46*(1), 53–72. doi:10.1016/j.specom.2005.01.003

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, *22*, 145–166. doi:10.1017/S0022226700010598

Wetzels, W. L., & Mascaró, J. (2001). The typology of voicing and devoicing. *Language*, *77*(2), 207–244. doi:10.1353/lan.2001.0123

Zamuner, T. S., Kerkhoff, A., & Fikkert, P. (2011). Phonotactics and morphophonology in early child language: Evidence from Dutch. *Applied Psycholinguistics*, *33*(03), 481–499. doi:10.1017/S0142716411000440

Zink, I., & Lejaegere, M. (2002). *N-CDIs: Lijsten voor communicatieve ontwikkeling. Aanpassing en hernormering van deMacArthur CDIs van Fenson et al. [A CDI user’s manual with normative and validity data]*. Leusden, The Netherlands: Acco.

Appendix I

Table 8. Experiment 1 - Growth Curve Analysis of target fixation proportions.

Effect	*Estimate*	*SE*	*z-value*	*p-value*
Intercept	1.01	0.22	4.54	<.001 ***
Pronunciation (CP vs. MP)	-0.05	0.06	-0.88	.38
Morphology (plural vs. monomorphemic)	-0.23	0.27	-0.86	.39
Voicing (/d/ vs. /t/)	-0.23	0.27	-0.85	.39
Linear Time	27.59	4.49	5.6	<.001 ***
Quadratic Time	-6.1	2.28	-2.68	<.01 **
Cubic Time	-6.33	1.27	-5.0	<.001 ***
Pronunciation * Morphology	0.29	0.09	3.21	<.01 **
Pronunciation * Voicing	0.26	0.09	2.88	<.01 **
Morphology * Voicing	0.5	0.39	1.3	.2
Pronunciation * Linear Time	-7.19	1.88	-3.83	<.001 ***
Morphology * Linear Time	-13.23	4.76	-2.78	<.01 **
Voicing * Linear Time	-1.09	4.73	-0.23	.82
Pronunciation * Morphology * Voicing	-0.91	0.13	-7.1	<.001 ***
Pronunciation * Morphology * Linear Time	10.94	2.81	3.89	<.001 ***
Pronunciation * Voicing * Linear Time	6.97	2.8	2.49	<.05 *
Morphology * Voicing * Linear Time	15.34	6.74	2.28	<.05 *
Pronunciation * Morphology * Voicing * Linear Time	-16.34	4.03	-4.06	<.001 ***

Note. ***p <.001. **p<.01. *p<.05.

Appendix 2

Table 9. Experiment 2 - Growth Curve Analysis of target fixation proportions.

Effect	*Estimate*	*SE*	*z-value*	*p-value*
Intercept	0.56	0.32	1.78	0.08
Pronunciation (CP vs. MP)	0.46	0.08	5.58	<.001 ***
Morphology (plural vs. monomorphemic)	0.07	0.28	0.23	.82
Voicing (/d/ vs. /t/)	-0.36	0.28	-1.28	.2
Linear Time	24.66	7.07	3.49	<.001 ***
Quadratic Time	3.79	2.26	1.67	.09
Cubic Time	-1.33	2.1	-0.631	.53
Pronunciation * Morphology	-0.2	0.11	-1.77	<.08
Pronunciation * Voicing	0.08	0.11	0.74	.45
Morphology * Voicing	0.1	0.4	0.26	.79
Pronunciation * Linear Time	-16.25	2.6	-6.24	<.001 ***
Morphology * Linear Time	-13.38	7.01	-1.91	.06
Voicing * Linear Time	-9.51	6.93	-1.37	.17
Pronunciation * Morphology * Voicing	-0.56	0.16	-3.55	<.001 ***
Pronunciation * Morphology * Linear Time	15.8	3.59	4.41	<.001 ***
Pronunciation * Voicing * Linear Time	-3.73	3.57	-1.04	.3
Morphology * Voicing * Linear Time	12.64	9.82	1.29	.2
Pronunciation * Morphology * Voicing * Linear Time	-6.72	5.01	-1.34	.18

Note. ****p* <.001. ***p*<.01. **p*<.05.

Tables & Figures

Table 1. The voicing contrast in Dutch and German.

		Initial			*Medial*			*Final*		
Dutch	/t/	*tak*	[tak]	‘branch’	*ketting*	[kɛtɪŋ]	‘necklace’	*pet*	[pɛt]	‘cap’
	/d/	*dak*	[dak]	‘roof’	*ladder*	[lɑdər]	‘ladder’	*bed*	[bɛt]	‘bed’
German	/t/	*Teich*	[taɪç]	‘pond’	*Beutel*	[bɔʏtl̩]	‘bag’	*Brot*	[broːt]	‘bread’
	/d/	*Dach*	[dax]	‘roof’	*Feder*	[feːdɐ]	‘feather’	*Hund*	[hʊnt]	‘dog’

Table 2. Child Directed Speech – Dutch and German singular nouns with stem final obstruents.

	Nouns with final obstruent (total)		*Nouns with final alternating obstruent (total)*		*Proportion nouns with alternation (%)*	
	Types	*Tokens*	*Types*	*Tokens*	*Types*	*Tokens*
Dutch	257	4410	85	2112	33.1	47.9
German	830	20787	298	9252	35.9	44.5

Table 3. Child Directed Speech – Voicing alternations in Dutch and German singular-plural pairs.

	Plural forms		*Plurals with alternation (count)*		*Plurals with alternation (%)*	
	Types	*Tokens*	*Types*	*Tokens*	*Types*	*Tokens*
Dutch	57	493	22	158	38.6	32
German	196	2495	72	1572	36.7	63

Table 4. Experiment 1 – Dutch test stimuli.

Word Type	*Item*		*Gloss*	*Yoked*	*Gloss*
Plural /t/	*botten*	[bɔtən]	bones	*bomen*	trees
	fluiten	[flœytən]	flutes	*fietsen*	bikes
	noten	[notən]	nuts	*neuzen*	noses
	petten	[pɛtən]	caps	*peren*	pears
Plural /d/	*bedden*	[bɛdən]	beds	*boeken*	books
	broden	[brodən]	breads	*brillen*	glasses
	hoeden	[hudən]	hats	*handen*	hands
	kleden	[kleɪdən]	rugs	*klokken*	clocks
Monomorphemic /t/	*boter*	[botər]	butter	*beker*	cup
	gieter	[xitər]	watering can	*glijbaan*	slide
	ketting	[kɛtɪŋ]	necklace	*kussen*	cushion
	sleutel	[sløʏtəl]	key	*spiegel*	mirror
Monomorphemic /d/	*ladder*	[lɑdər]	ladder	*lepel*	spoon
	pudding	[pʏdɪŋ]	pudding	*puzzel*	puzzle
	ridder	[rɪdər]	knight	*robot*	robot
	schaduw	[sxadyw]	shadow	*schouder*	shoulder

Table 5. Experiment 1 - Post-hoc comparison of the effect of Pronunciation on different word types.

Word Type		*CP Estimate*	*MP Estimate*	*Difference between CP and MP*	*SE*	*z-value*	*p-value*
Mono. /t/	Int.	1.05	0.63	-0.42	0.06	-6.69	<.001 ***
	L.T.	28.61	22.98	-5.63	1.95	-2.88	<.05 *
Mono. /d/	Int.	0.78	1.01	0.23	0.07	3.53	<.01 **
	L.T.	14.36	18.1	3.75	2.1	1.79	.45
Plural /t/	Int.	0.78	0.99	0.2	.07	3.1	<.05 *
	L.T.	26.5	26.28	-0.22	2.08	-0.11	1.0
Plural /d/	Int.	1.01	0.96	-0.05	0.06	-0.88	.98
	L.T.	27.59	20.4	-7.19	1.88	-3.83	<.01 **

CP = Correct Pronunciation.

MP = Mispronunciation.

Int. = Intercept, reflecting the height of the curve.

L.T. = Linear Time, reflecting the gradient of the curve.

Table 6. Experiment 2 - German test stimuli.

Word Type	*Item*		*Gloss*	*Yoked*	*Gloss*
Plural /t/	*Betten*	[bɛtən]	beds	*Bretter*	boards
	Brote	[broːtə]	breads	*Boote*	boats
	Schwerter	[ʃveːɐtɐ]	swords	*Schwänze*	tails
	Zelte	[tsɛltə]	tents	*Zehe*	toes
Plural /d/	*Hunde*	[hʊndə]	dogs	*Hände*	hands
	Kleider	[klaidɐ]	dresses	*Klaviere*	pianos
	Monde	[moːndə]	moons	*Münder*	mouths
	Pferde	[pfeːɐdə]	horses	*Pflaster*	plasters
Monomorphemic /t/	*Beutel*	[bɔytəl]	bag	*Becher*	beaker
	Garten	[gartən]	garden	*Gabel*	fork
	Leiter	[laɪtɐ]	ladder	*Lampe*	lamp
	Schulter	[ʃʊltɐ]	shoulder	*Schlüssel*	key
Monomorphemic /d/	*Erde*	[eːɐdə]	earth	*Erdbeere*	strawberry
	Feder	[feːdɐ]	feather	*Fenster*	window
	Nadel	[naːdəl]	needle	*Nase*	nose
	Weide	[vaidə]	meadow	*Wolke*	cloud

Table 7. Experiment 2 - Post-hoc comparison of the effect of Pronunciation on different word types.

Word Type		*CP Estimate*	*MP Estimate*	*Difference between CP and MP*	*SE*	*z-value*	*p-value*
Mono. /t/	Int.	0.38	0.16	-0.22	0.08	-2.79	.04 *
	L.T.	14.42	3.51	-10.9	2.5	-4.36	<.001 ***
Mono. /d/	Int.	0.63	0.88	0.26	0.08	3.25	<.01 **
	L.T.	11.29	10.83	-0.45	2.48	-1.83	1
Plural /t/	Int.	0.2	0.75	0.54	0.08	7.08	<.001 ***
	L.T.	15.15	-4.83	-19.98	2.44	-8.17	<.001 ***
Plural /d/	Int.	0.56	1.02	0.46	0.08	5.58	<.001 ***
	L.T.	24.66	8.41	-16.25	2.6	-6.24	<.001 ***

CP = Correct Pronunciation.

MP = Mispronunciation.

Int. = Intercept, reflecting the height of the curve.

L.T. = Linear Time, reflecting the gradient of the curve.

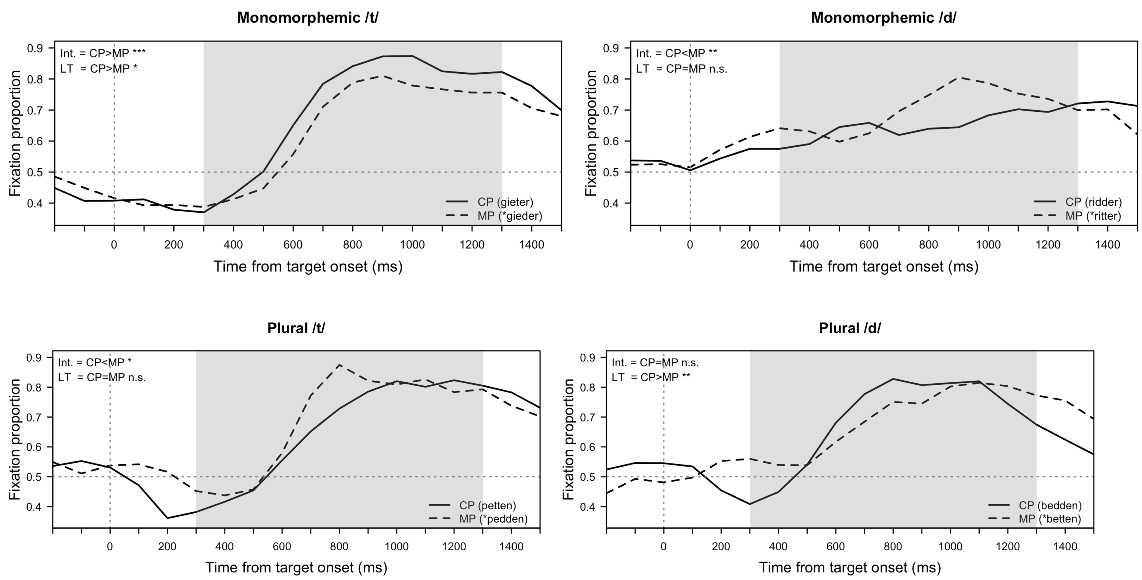

Figure 1. Experiment 1 Dutch - Target fixations to different trial types. Solid lines indicate gaze behaviour during correctly pronounced trials, and the dashed lines mispronunciation trials. The dashed vertical line corresponds to the onset of the target word, and the shaded area represents the 1000ms time window of analysis, starting 300ms after target word onset. Fixation proportions above 50% indicate looks to the target rather than the distractor. The abbreviations "Int." and "LT" stand for "Intercept" and "Linear Time" respectively, and indicate statistical differences between the two lines during the analysis window, corresponding to the pair-wise comparisons displayed in Table 5. For example, "Int. = CP>MP *" indicates that the intercept of the correct pronunciation line is significantly higher than the intercept of the mispronunciation line. Similarly, "LT = CP>MP *" indicates that the slope of the correct pronunciation line is significantly steeper than the mispronunciation line.

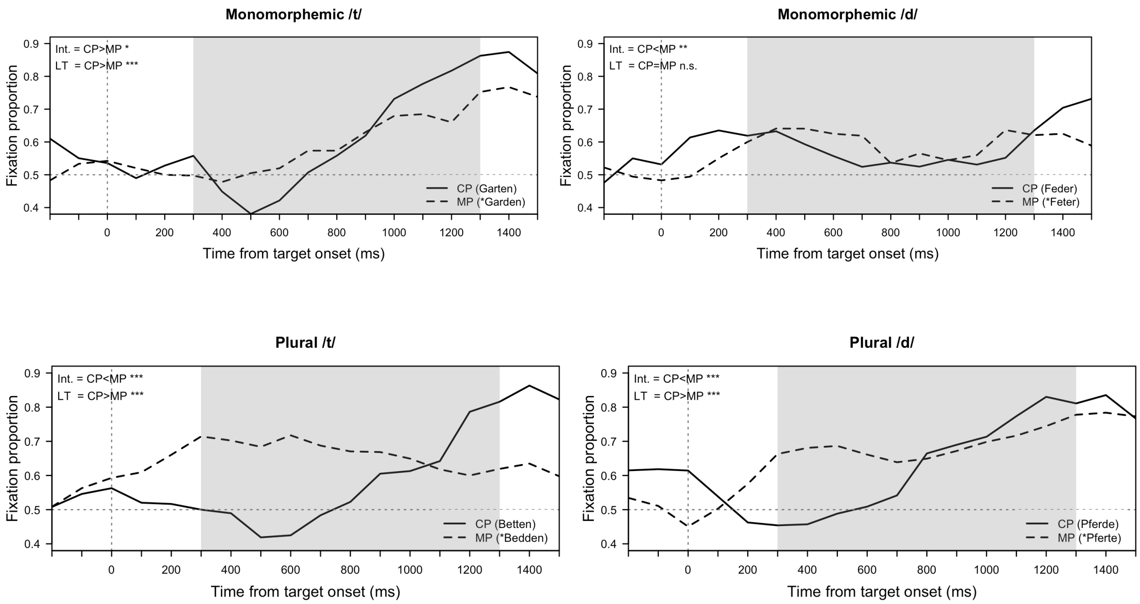

Figure 2. Experiment 2 German - Target fixations to different trial types. Solid lines indicate gaze behaviour during correctly pronounced trials, and the dashed lines mispronunciation trials. The dashed vertical line corresponds to the onset of the target word, and the shaded area represents the 1000ms time window of analysis, starting 300ms after target word onset. Fixation proportions above 50% indicate looks to the target rather than the distractor. The abbreviations "Int." and "LT" stand for "Intercept" and "Linear Time" respectively, and indicate statistical differences between the two lines during the analysis window, corresponding to the pair-wise comparisons displayed in Table 7. For example, "Int. = CP>MP *" indicates that the intercept of the correct pronunciation line is significantly higher than the intercept of the mispronunciation line. Similarly, "LT = CP>MP *" indicates that the slope of the correct pronunciation line is significantly steeper than the mispronunciation line.