The development of the speech production mechanism in young children : evidence from the acquisition of onset clusters in Dutch

Gulian, M.E.

Author: Gulian, Margarita
Title: The development of the speech production mechanism in young children : evidence from the acquisition of onset clusters in Dutch
Date: 2017-10-31

Chapter 2. From toddlers' mouths to adults' ears: production and perception of reduced onset clusters occurs on different channels[1]

2.1. Introduction

Phonetic or phonological cluster reduction is a common phenomenon in young children's speech productions. In this chapter cases of cluster reduction in word onsets are discussed in which the child apparently omits the second consonant, as in [dʌk] for *truck* and [siːp] for *sleep*. The research discussed here addresses two questions. First, we want to find out whether toddlers intend to express a complex onset despite the apparent omission of the second consonant. Does the lexical representation of a reduced cluster contain information about the omitted consonant or not? For this purpose we compare children's productions of onset clusters that have been phonetically transcribed as reduced forms, to their productions of similar words that do not contain a cluster in the target adult form, by means of an acoustic analysis. The purpose of performing a detailed analysis of the reduced form is to help to determine the source of the deviation from the adult target form. Our acoustic analyses indeed reveal traces of the omitted consonant. This leads to our second question, namely whether adults can distinguish children's words with reduced onsets from words starting with an identical simple onset when these are presented next to each other. In other words, when adults are asked to pick from a child's minimal pair the production that has an onset cluster in the adult language, can they use the acoustic trace of the "omitted" consonant as a reliable cue? Here we find that adult listeners use different cues for their decisions than the cues that the child provides.

---

[1] This Chapter is identical to the manuscript: Gulian, M, Levelt, C. & Boersma, P. *From toddlers' mouths to adults' ears: production and perception of reduced onset clusters occurs on different channels*. It therefore uses the first person plural instead of singular. The manuscrpt is ready for submission to a linguistic journal.

2.1.1. Theoretical background

One of the goals of the present study is to get better insight in the way consonant clusters are stored and handled in the toddler's mental lexicon and speech production mechanism. Do toddlers store adult cluster words as CV- (consonant–vowel) sequences or as CCV- sequences underlyingly, and if this is the case, where in the production process does the reduction take place?

To explore the possibilities, we suggest the heuristic model of speech production in Figure 1, which combines phonological and psycholinguistic views of the levels of representation involved (Levelt et al., 1999; Boersma, 2011). In Figure 1, speech production involves the step-wise retrieval of information and application of knowledge in different modules. The production of a single word requires the activation of a lemma in the mental lexicon. Each lemma activates its corresponding phonological underlying form, which contains the stored information about the word's sounds. From this information a phonological surface form is created in the phonological production process. Subsequently, phonetic implementation may convert this surface form to an auditory-phonetic target (for adults: MacNeilage, 1981, Gay et al., 1981; for children: Oller & MacNeilage, 1983), which is then translated by sensorimotor knowledge to an articulatory-motor program that controls the speech muscles. The precise steps in the whole process are subject to debate, but Figure 1 will help us formulate hypotheses about the localization and causes of reduction.

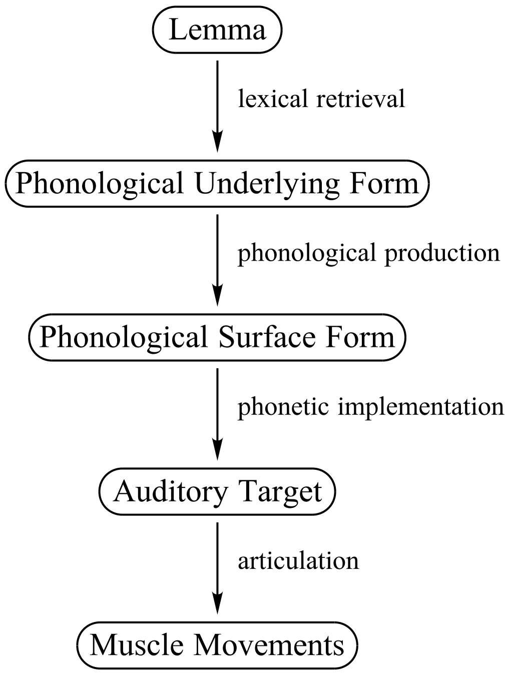

Figure 1: Heuristic speech production mechanism

Figure 1 suggests at least seven potential locations or causes for cluster reduction.

The acoustic signal will have different characteristics depending on the locus of reduction. Consider the Dutch adult word pair [bʀoːt] 'bread' versus [boːt] 'boat', and assume that the child stores 'boat' as /boːt/ in her underlying form. The question now is: where does the child reduce the adult's /bʀ/ in [bʀoːt] ('bread') If the child's underlying form for 'bread' is /boːt/, identical to the one for 'boat', then the child appears to have reduced the cluster either (1) already somewhere in her comprehension of the adult word, or (2) when storing the word in her lexicon for the first time, perhaps as a result of a morpheme-structure constraint; in these cases, we predict that the child will pronounce 'bread' in an identical way to 'boat' at the acoustic level. If the child's underlying form for 'bread' is /bʀoːt/, but her surface form is /boːt/, then either (3) her phonological grammar dictates that underlying /bʀ/ should correspond to a surface /b/, or (4) the surface form is restricted by a structural constraint such as */CC/; in these cases, the reduction is again discrete (i.e. all or none), so that complete acoustic homophony with the production of 'boat' is predicted. If the child's underlying and surface forms are /bʀoːt/, it is possible that (5) she has trouble mapping the surface /bʀ/ to the appropriate auditory cues, thus targeting something close to, but not necessarily identical to, [boːt]; in this case, the reduction is not discrete at the acoustic level, but a transcriber may classify the sound as the phonological surface form /boːt/ with her adult Dutch perception system. In this case we predict that the child may object to an adult pronouncing 'bread' as [boːt] (i.e. the *fis* phenomenon: Berko & Brown, 1960). If the auditory target is a full-fledged [bʀoːt], the articulatory result may still be close to [boːt] as a result of (6) a sensorimotor mapping that does not yet link the auditory cues with the appropriate muscle gestures (Ferguson & Macken, 1983) or (7) developmental restrictions on the planning or timing of muscle gestures (Studdert-Kennedy, 1987); in these cases we may find an acoustic trace of /ʀ/, although a Dutch transcriber might not notice this. Therefore, if we analyze the child's acoustic productions of 'bread' and do find a trace, then we can conclude that reduction has taken place by one of the mechanisms (4) through (7); if there is no trace at all, the cause may lie in mechanisms (1) through (3).

Gradient versions of these mechanisms are also possible. It could be the case, for instance, that (due to a comprehension restriction, a lexical restriction, or a surface restriction) the child's surface structure is the reduced segment sequence /CV/ but does exhibit in the vowel an extra feature, for instance rhoticity, that somehow expresses the reduced C2. Thus, 'bread' could be represented as /bo$^{+\text{rho}}$ːt/. The extra feature would typically come with fewer auditory cues for the adult listener than a segment would, so that an intended /bo$^{+\text{rho}}$ːt/ will be perceived by an adult listener as a complete homonym of /boːt/

'boat'. If this is the case, an acoustic trace of /ʀ/ may be found in the child's realization of 'bread'.

2.1.2. Covert contrasts in the literature

Studying the acoustic waveforms of toddlers' productions is an interesting way to find out more about the lexical representations of early words. Up until now, young children's lexical representations have mostly been studied using perception experiments (e.g. Fennell & Werker, 2003; Swingley, 2003; Swingley & Aslin, 2000, 2007; White & Morgan, 2008; for an overview see Newman, 2008). However, a detailed analysis of children's productions gives a different perspective on the issue, and directly confronts the difference that exists between detailed representations and reduced productions (Pater & Barlow, 2003; Smolensky, 1996).

Acoustic analyses have led to the discovery of a number of "covert contrasts" in toddler's productions (for an early overview see Scobbie, 1998). McLeod et al. (1998) showed that Australian English two-and-a-half-year-olds pronounce a [k] reflecting a target /sk/ cluster with a shorter VOT than a [k] reflecting a target singleton /k/ onset. Carter and Gerken (2004) analyzed truncations in two-year old children who had to repeat sentences like *He kissed Lucinda* – *Lucinda* being a ready target for reduction in toddler speech – and *He kissed Cindy* and found a larger time gap between *kissed* and reduced *cinda* than between *kissed* and correct *Cindy*. Song and Demuth (2008) recorded longitudinally three children (1;6 – 2;6) and found in their utterances differences between reduced target coda clusters and similar correctly produced target singleton forms: compensatory vowel lengthening was found in case the coda cluster was reduced. Lowenstein and Nittrouer (2008) showed that American-English two-year-olds produce voiceless target plosives with longer VOTs than voiced target plosives, although the two transcribers could not perceive this difference. Gulian and Levelt (2011) found that Dutch two-year-olds pronounced reduced article-noun phrases with a reduced cluster differently from singleton counterparts.

The authors compared phrases like *een peen*, where *peen* [peːn] was the reduced form of *speen* (/speːn/ 'pacifier') with *een peek*, where *peek* [peːk] was the intended singleton nonword *peek* /peːk/. They found that there was a larger time interval between the nasal in *een* and the plosive in *peen* as compared to the same interval in *een peek*.

All of these studies thus reveal knowledge that language learners have, but do not make audible in a way that adult listeners can perceive.

In the two studies below, we focus on two clusters that are very often reduced in Dutch child language productions, namely /Cr/ (plosive + rhotic[2]) and /kn/. In study 1, word productions with reduced renditions of these target clusters are analyzed acoustically and compared to productions of corresponding words with singleton onsets. Thus, an adult onset cluster /Cr/, apparently produced by the toddler as [C-] is compared to the toddler's production of a phonetically similar word with an adult singleton onset /C-/. For instance, the utterance [boːt] for *brood* 'bread' is compared to *boot* [boːt] 'boat'. An example of the other cluster type is *knippen* (adult target [knɪpə]) 'to cut', produced by the child as [kɪpə], which is compared to *kippen* [kɪpə] 'chickens'. In study 2 we test the way adults perceive these minimal pairs in toddler speech.

2.2. Study 1: Child production of /Cr/~/C/ and /kn/~/k/ word pairs

In order to answer the question where in the production model cluster reduction originates, we concentrate on /kn/ and /Cr/ cluster types in Dutch. Specifically, we look for the productions of minimal pairs of singleton and cluster targets, e.g. for cases in which the same child produced both 'bread' (adult target [bʀoːt]) and 'boat' (adult target [boːt]), or for cases in which the same child produced 'chickens' (adult target [kɪpə]) as well as 'to cut' (adult target [knɪpə]).

[2] In this position, Dutch has only one rhotic phoneme, which can be realized as [ʀ], [r] or [ɾ] (Sebregts, 2015).

Any small systematic acoustic difference between the members of a produced pair could indicate that the child intends to make a difference between the word forms, even if both members were transcribed identically by adult researchers (namely with a single consonant).

2.2.1. Participants

For this study we looked for young Dutch monolingual children who reduced /Cr/ and/or /kn/ cluster words in their speech. We found such children in two separate datasets, namely in the existing CLPF database (Levelt, 1994; Fikkert, 1994), available in Childes/Phonbank (Rose & MacWhinney, 2014), and in our own new recordings at day-care centers collected specifically for the present purpose. The CLPF database consists of longitudinal recordings of 12 children acquiring Dutch as their first language, aged roughly between one year and two and a half years at the start of the data-collecting period. Currently, audio files are available for 6 of the 12 children, and in the data of 4 of these children we could find the necessary word-pairs for comparison. The day-care-center dataset was collected by the author of this thesis by recording 30 toddlers with a mean age of 2;1 years at four Dutch day-care centers. Nine of these children already produced /kn/ and /Cr/ target clusters in an adult-like manner and were excluded from the analyses. The data for analysis thus included the productions of four children from the CLPF database and 21 children from the day-care-center recordings, forming a total of 25 children. Eight of these 25 children reduced both /Cr/ and /kn/ clusters, while the remaining 17 children reduced either one cluster or the other (see Appendix 2).

2.2.2. Method: /Cr/~/C/ word pairs

We start out by discussing the acoustic analysis of target /Cr/~/C/ word pairs.

2.2.2.1. Participant selection

Here we consider data from those four children from the CLPF database whose speech exhibited the phenomenon of /Cr/ cluster reduction and who also produced a singleton counterpart in the same session or in a closely related session, and data from those 11 day-care-center children who reduced the /Cr/ cluster at least once and who produced at least one singleton counterpart. The mean age of the 4 database children at the times of the recordings that are used here was 2;1.6 (age range 1;8.10 - 2;4.26). The mean age of the 11 day-care-center children was 2;0.29 (age range 1;6.0 - 3;0.1).

2.2.2.2. Data selection

In order to detect acoustic traces of reduced /Cr/ clusters, word pairs were compared for each child separately. For every child, we paired a word production with a reduced target cluster with a word production with a singleton onset consonant that matched the onset consonant of the reduced form as closely as possible.

At the day-care centers, toddlers were asked to repeat a list of Dutch words with initial clusters and matching words (or sometimes non-words) with a simple onset (i.e. *trein* ‘train’ matched to *Thijs* (a common Dutch boy’s name). If possible, pictures of the words-to-be-repeated were used to encourage production. A list of the words that the children had to repeat is given in Appendix 1. The children’s utterances were recorded with 16-bit 44100-Hz sampling with a Microtrack II digital recorder and an external Microtrack II microphone. At the time of each recording, the responses by the 30 children were also transcribed online. Later, the online transcriptions containing cluster reductions were selected for more detailed off-line phonetic transcriptions, which I checked first, and were subsecuently checked by an experienced phonologist.

Word productions were determined to be reduced cluster words if in the data the target word contained an onset cluster and according to the phonetic transcription of the word produced by the child, the second consonant of the consonant cluster was omitted[3], such as in the transcription [boːt] for intended *brood* ‘bread’ [bʀoːt]. In the CLPF database a search was carried out to find all utterances that contained a cluster in its adult target form but missed the second consonant of that cluster in the phonetic transcription of the child’s actual production. The matching singleton-onset word would ideally form a minimal pair with the target cluster-onset word, differing only in the absence of a second consonant. For *brood* [bʀoːt], for example, the ideal match was *boot* ‘boat’ [boːt].

From now on the target singleton-onset words such as *tijd* and *boot* will be referred to as /C/ words. If no ideal match could be found for a /Cr/ word, a /C/ word was selected that shared as many features as possible with the onset plosive and the subsequent vowel, e.g. the /Cr/ word *trein* ‘train’ /tʀɛin/, produced as [tɛin], was paired with *Thijs* (Dutch boy’s name) /tɛis/ in the analysis. If the utterances selected for analysis were polysyllabic, such as *draaimolen* ‘merry-go-round’ /ˈdʀaːiˌmoːlə/, they were always stressed on the first syllable. The /Cr/ word and the matching /C/ word were always produced by the same child and originated from the same recording session, or from recording sessions that were no more than 1 month apart.

After the strict selection criteria for matching word pairs, in the end the analysis of the /Cr/ clusters is based on 47 word pairs, i.e. 47 target /Cr/ words matched to 47 /C/ words. Of these 47 word pairs, 21 came from the CLPF database, and 26 from the day-care center set. A list of the reduced cluster words and their matching singleton consonant forms is given in Appendix 3.

[3] The authors of the CLPF database, C. Levelt and P. Fikkert, were the primary transcribers of the CLPF dataset, and I made the transciption of the day-care-center dataset.

2.2.2.3. Measurement method

In this study we looked at words with target onset /Cr/ clusters where the second consonant was apparently omitted, and compared them acoustically to similar /C/ words. All the acoustic measurements presented in this chapter were made using Praat 5.0.10 (Boersma & Weenink, 2008).

To minimize the chance of subjective measures, the assistant who carried out the acoustic measurements was blind to the actual transcriptions of the words produced by the children. All 94 utterances (which consisted of the 47 reduced cluster words and 47 matching simple onset words) were anonymized[4]. If necessary, the utterances were trimmed back to only the initial consonant-vowel sequence; for example in the utterances [boːt] (from *brood* 'bread') and [boːtə] (from *boten* 'boats') the final part was removed, leading to [boː] and [boː] respectively, to prevent them from revealing the word meaning to the assistant, who was told to determine the vowel and utterance boundaries as if the word was a /C/ word.

Two formants, F2 and F3, were measured at two points in time in the spectrum by means of a band-filter analysis method that had been used before for the description of infant vowel productions (Wempe, 2001; Van der Stelt et al., 2005). This method, which takes into account the child's pitch to estimate a spectral envelope representation of an utterance, has the advantage above LPC (linear predictive coding, the most widely used formant analysis method) of being less sensitive to the incorrectly chosen parameters that are likely to occur with LPC if there are high pitches in the data, which is often the case in child speech.

The approximant rhotics [ɹ] and [ɻ], such as occur in English and in Dutch codas, tend to come with a low F3 and to some extent a low F2 (Lindau, 1985; Plug & Ogden, 2003; Scobbie & Sebregts, 2011). The adult realization in Dutch /Cr/

[4] After all analyses were carried out, the anonymous utterances were linked to the original transcriptions again.

clusters is more likely the uvular trill [ʀ] or the alveolar trill or tap [ʀ, ɾ]. Gulian (in prep.) found modest F3 and F2 values for these variants: in going from the rhotic to the vowel there was a rise in F2 and F3 for front vowels and a lowering in F2 for back vowels, both for the alveolar and for the uvular variants, with the alveolar trill exhibiting a slightly lower F3 than the uvular trill. If we detect similar formant movements in the vowel onset in reduced cluster words in toddler speech, this could therefore be considered a likely trace of the omitted rhotic.

In order to be able to measure F2 and F3 movement in the vowel, the two formants were first measured at the immediate vowel onset (t_1), followed by a measurement at one quarter of the entire duration of the vowel (t_2). Figure 2 gives an example of a child production of the word *kraan* 'faucet' /kʀaːn/, where the target rhotic is fully produced and the raising of F2 and F3 can be clearly discerned.

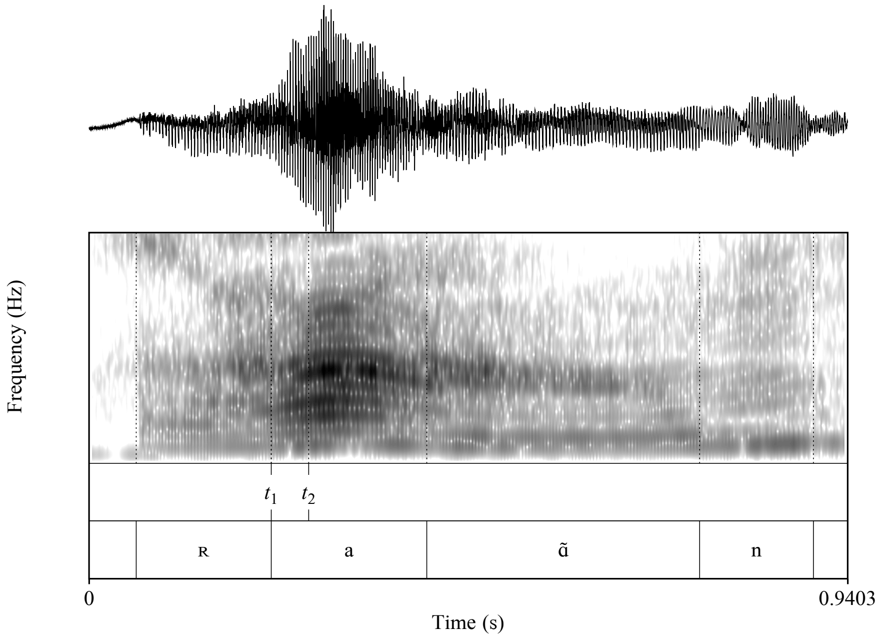

Figure 2: Waveform and spectrogram of the word *kraan*, produced [ʀaãn] by a child aged 3;3 (t_1 and t_2 are the time points where formants are measured).

In order to capture a possible upward movement in the vowel onset, the values for F2 and F3 at times t_1 and t_2 are used for a simple calculation: the Hz value at t_1 is subtracted from the Hz value at t_2. When the obtained value is positive, the formants have moved upward, and this is interpreted as indicating the intended presence of a preceding rhotic. For example, in Figure 2, the calculations for F2 and F3 both result in positive values (a difference of respectively 268 Hz and 24 Hz). This confirms that both F2 and F3 are rising in the vowel onset, although /a/ is not a prototypical front vowel. If the value is negative, the formant in question appears to have lowered in the vowel onset, which is interpreted as indicating a traceless omission of the target rhotic in the production.

In addition to F2 and F3 movement in vowel onsets, we measured vowel and utterance duration. Since we were interested in acoustic traces of the second consonant in reduced onset clusters, we had to be on the look-out for forms of compensatory lengthening. For the utterance duration measure, the duration of the vowel plus preceding and following coda consonants (in the cases where the coda was not trimmed) was measured.

To summarize, for the /Cr/~/C/ minimal pairs, such as the words *trein* and *Thijs*, where *trein* is produced as [tɛin], four different measures were taken. All 47 pairs (94 utterances) were measured for their vowel duration, their utterance duration, and the F2 and F3 movement in the vowel onset. In the result section (2.4) we turn to the outcomes of these four acoustic measures and determine whether any of these measures was distinctive for the minimal pairs. Below we first discuss the different measures we took with the other cluster type in this study.

2.2.3. Method: /kn/~/k/ word pairs

This section discusses the analysis of the reduction of target /kn/~/k/ word pairs.

2.2.3.1. Participant selection

Utterances from both the CLPF database and the day-care-center recordings are analyzed. We use utterances from two children from the CLPF database, with a mean age of 2;3 (age range 2;0 – 2;6)[5]. These children overlap with the children that exhibited /Cr/ cluster reductions in this study (Study 1, see Appendix 1). From the 30 children who were recorded at day-care centers, 16 children (mean age 1;11, range 1;8 - 3;0) reduced the /kn/ clusters. In total we analyze the /kn/~/k/ utterances of 18 children.

---

[5] As in the /Cr/~/C/ acoustic analysis, recordings from different sessions from the same children were used.

2.2.3.2. Data selection

In order to detect acoustic traces of reduced /kn/ clusters for each child, again pairs of words were compared. As in the previous study, a reduced onset-cluster utterance was selected and compared to a singleton-onset utterance that was closely matched to the reduced cluster production. The selection criteria for the matching word are identical to those mentioned in 2.1. Here /kn/ words, such as *knippen* 'to cut' and *knoop* 'button', which according to their transcriptions were produced as [kɪpə] and [koˑp], were compared to words like *kip* 'chicken' [kɪp] and *kopen* 'to buy' [koːpə].

For the present analysis, 37 target cluster words were matched to 37 target words starting with a singleton /k/. From these 37 pairs, six were produced by the two children from the CLPF database, while the remaining 31 pairs were produced by the children from the day-care center. A list of the /kn/~/k/ word pairs is given in Appendix 4.

2.2.3.3. Measurement method

As in the /Cr/ study, the assistant who carried out the acoustic analyses was blind to the transcriptions and to the intended form of the utterances. Three duration measures were taken: vowel duration, utterance duration, and an additional duration measure called "vowel complexity". The vowel complexity measure was first suggested in a pilot study (Gulian &| Levelt, 2008), which observed that the vowel onset in reduced /kn/ words often exhibited an atypical diphthongization. An example of an utterance containing a diphthongized vowel onset is given in Figure 3. In order to determine the vowel complexity measure we use the following criteria: if the first half of the vowel exhibits a changing vocalic pattern, i.e. a diphthongization, then the duration of the first part (labeled as "$part_1$" in Figure 3) constitutes the vowel complexity value for this item; otherwise, if the first half of the vowel does not exhibit a change, the vowel complexity value is taken to be zero. In case of doubt, the vowel complexity measure is taken to be zero as well. In order to become acquainted with the determination of this duration measure the assistant was familiarized with productions from the pilot study, containing similar diphthongized vowel onsets, which were not part of the list of 74 stimuli from the present data set.

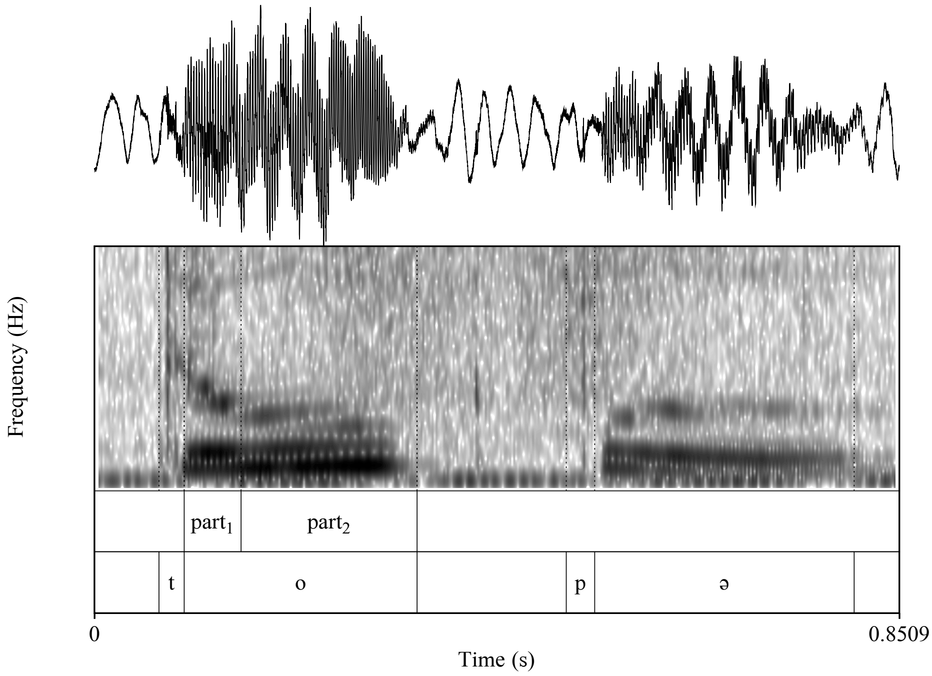

Figure 3: Waveform and spectrogram of the word *knopen* (actual production [topə]) by a child aged 2;5. $Part_1$ and $part_2$ stand for the two parts within the vowel that are revealed in the spectrogram.

In addition to the three duration measures, we took three other measures that could point to traces of the omitted nasal in the vowel. In order to do this we selected three measures from a list of acoustic characteristics of nasalized vowels described in Pruthi et al. (2007). According to Pruthi et al. (2007), nasalized vowels exhibit, among others, reduction in the first formant (F1) amplitude[6] and in the overall intensity of the vowel, and a movement of the low-frequency center of gravity towards a neutral vowel configuration (besides these three, Pruthi et al. studied another four acoustic correlates of nasality, but we chose not to measure them because they were all related to the F1 measure).

The three measures we took (mean F1, overall vowel intensity and mean center of gravity of the vowel following the plosive) were analyzed by means of a script in *Praat*. To measure mean F1, the settings for formant measurement in *Praat* were adapted to toddlers' voice quality, namely to search for up to 5 formants in the range from 0 to 6,000 Hz. All measures were carried out for both reduced /kn/ words and singleton /k/ words.

To sum up, for the reduced nasal clusters in Dutch we measured six acoustic characteristics, which consisted of comparing the word pairs in terms of utterance duration, vowel duration, "vowel complexity duration", mean F1, overall vowel intensity and center of gravity.

2.2.4. Results of Study 1

2.2.4.1. Results: /Cr/~/C/ word pairs

In the case of /Cr/~/C/ word pairs, we carried out four acoustic measures: vowel duration, word duration, F2 and F3 movement in the vowel onset. Our question was whether the target complex onset words would differ from target simple onset words in any of these four acoustic measures, even though they would all be produced with a simple onset. For this purpose we conduct a repeated-measures multivariate analysis of variance with word type (reduced cluster vs. simple onset) as repeated factor and four acoustic measure types (vowel duration, utterance duration, F2 movement and F3 movement) as dependent variables.

[6] The reduction of F1 amplitude is especially true for low vowels (Pruthi, 2007; p. 3871), while for high vowels, nasality brings F1 higher.

The analysis of variance reveals that word type has a significant effect on one of the acoustic measures, namely F2 movement (F [1,46] = 4.97, p = .031). This significant effect shows that F2 movement tends to be positive (thus upwards) in the vowel onset of reduced /Cr/ cluster words as compared to simple onset words (M = 81 Hz, SD = 298 Hz, and M = -43 Hz, SD = 282 Hz, respectively); see Figure 4. Word type did not show a significant effect on the other dependent variables (for all three: $p \geq .359$).

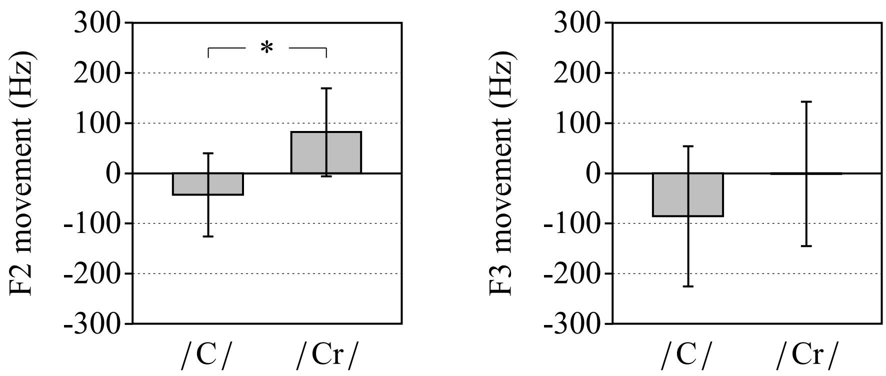

Figure 4: F2 and F3 movements in the vowels of target /C/~/Cr/ word pairs, showing means and 95% confidence intervals.

2.2.4.2. Results: /kn/~/k/ word pairs

The /kn/~/k/ word pairs are compared using six different acoustic measures: vowel duration, word duration, vowel complexity duration[7], center of gravity, overall intensity and mean F1. Because the type of the following vowel determines the height of F1 of the nasal (when present), we take into account

[7] Section 2.3.3 gives a clarification of this unconventional measure.

whether the vowel in the word pairs is high or not. From now on we refer to these two word types as /kni/ and /ki/ words versus /kna/ and /ka/ words.

A repeated-measures MANOVA with word type (reduced cluster vs. simple onset) as the repeated factor, vowel type (onset + /i/ vs. onset + /a/) as the between-items factor, and six dependent variables (vowel duration, word duration, vowel complexity duration, center of gravity, intensity and F1) reveals a substantial main effect of vowel type (F [6,30] = 6.633, p < .001) and a main effect of word type (F [6,30] = 2.336, p = .057). We find no interaction between vowel type and word type (all $p \geq .164$). The univariate ANOVAs reveal an effect of vowel type on one of the six acoustic measures: mean F1 (F [1,35] = 32.8, p < .001). If we consult the descriptive statistics for this factor, we conclude that /ki/~/kni/ words tend to have lower F1 than /kna/~/ka/ words (see Table 1 and Figures 5 and 6).

Table 1: Descriptive statistics of the acoustic measures that show a significant interaction either with vowel type (the first pair) or with word type (the last two pairs).

Measure and word type	**Mean (SD)**
F1: /ki/, /kni/ words vs. /ka/, /kna/ words	554 (105) Hz vs. 735 (130) Hz
vowel complexity: /kn/ words vs. /k/ words	0.018 (0.025) s vs. 0.006 (0.019) s
center of gravity: /kn/ words vs. /k/ words	656 (267) Hz vs. 760 (323) Hz

Another statistically significant effect that is found is the effect of word type on vowel complexity: F [1,35] = 5.884, p = .021, and a nearly significant effect on center of gravity: F [1,35] = 3.414, p = .073, see Table 1 and Figures 5 and 6. Word type did not show an effect on the other dependent variables (all $p \geq .405$).

As the mean values of these measures show, the vowel complexity duration tends to be longer in reduced /kn/ words than in simple onset /k/ words, while center of gravity tends to be lower for reduced clusters than for simple onsets.

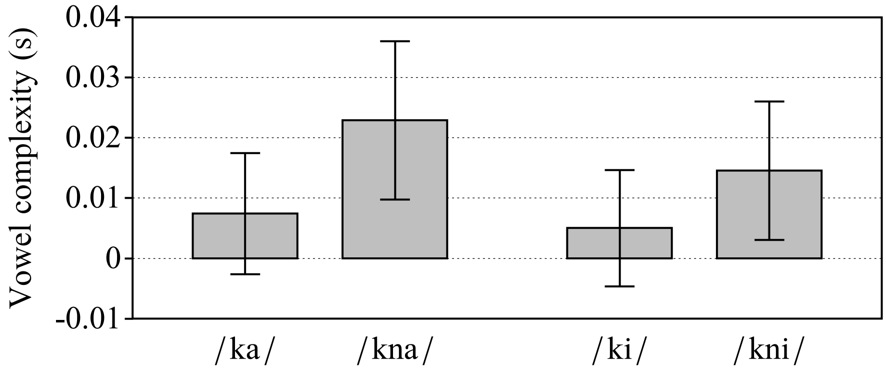

Figure 5: Vowel complexity measure in low (left) and high (right) vowels of target /k/~/kn/ word pairs. The values on the *y*-axis are presented in seconds.

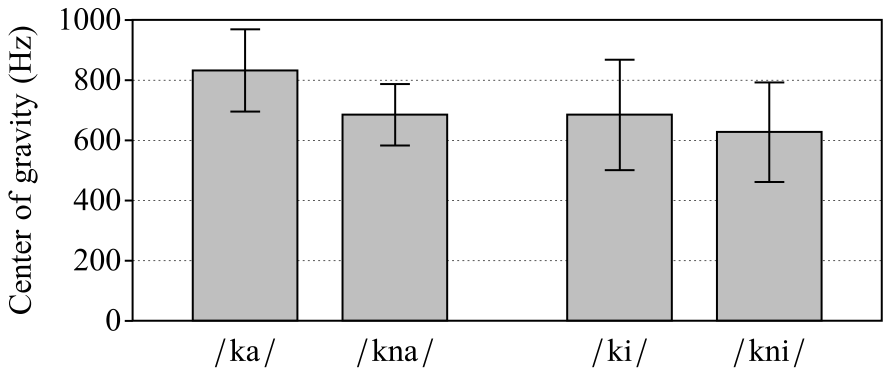

Figure 6: Center of gravity measure in low (left) and high (right) vowels of target /k/~/kn/ word pairs. The values on the *y*-axis are presented in Hz.

2.2.4.3. Summary of the results

Regarding the acoustic measures, the following can be concluded. For reduced /Cr/ words, the acoustic characteristic that seems to distinguish them from simple onset /C/ words is F2 movement in the vowel onset, showing a rise in words with reduced clusters. Word, vowel duration and F3 movement were not distinctive between the two sets of words (see Figure 4). As for reduced /kn/ words, vowel complexity duration was longer for the reduced cluster words than for the simple onset words and center of gravity showed a trend of being lower for the reduced /kn/ words (see Figures 5, 6). None of the other acoustic measures of nasality were distinctive. For both types of cluster reduction, then, Dutch toddlers produce some of the acoustic characteristics of the target second consonant of the cluster.

Given the fact that acoustic traces of the second consonant of both target /Cr/ and /kn/ words can be found in the productions of Dutch toddlers, the next question is whether or not adult listeners are able to pick up on these subtle cues. Since the productions of target words with onset clusters that were used for the acoustic measurements had been transcribed – by trained linguists – with singleton onsets, this does not seem likely in a natural context. However, would it be possible to find evidence for listeners picking up on the cues in an experimental setting? This question was addressed in Study 2.

2.3. Study 2: Adult perception of reduced target clusters /Cr/ and /kn/

In this study Dutch adult listeners participated in a forced-choice identification task, where the word-pair stimuli used in the first study were presented without the situational and word context that could help to disambiguate the two utterances. We asked participants to try to identify which of the two words that formed a minimal pair was originally an onset cluster word. The question was whether adult listeners would be able to rely on the acoustic cues provided by the child to correctly identify the reduced-cluster [CV] sequence from its singleton counterpart [CV] sequence.

2.3.1. Method

2.3.1.1. Stimuli: Word pairs with onset clusters /Cr/ and /kn/

The stimuli were the same CV sequences as those in Study 1. See the description of the stimuli in 2.2.2.2 and 2.2.3.2.

2.3.1.2. Procedure

The forced-choice identification task was carried out using the computer program *Praat*. Before starting, the test participants received information about how the minimal pairs had been obtained and why they heard only the first CV(C) sequence[8] of the child utterances. They were told in advance that perceptually the two utterances hardly differed from one another and that a consonant cluster was not perceivable, but that nevertheless in each test trial one of the two utterances was a target onset-cluster word. Finally, the adults were instructed that their task was to try their best to identify or, if necessary, guess which of the two stimuli corresponded best to a target word with an onset cluster.

Each participant was seated in front of a computer where he or she saw a gray screen with the instruction in the upper part of the screen saying "Choose the word that seems to start with a consonant cluster" and the words "first" and "second" that appeared on two yellow buttons in the screen center (all written in Dutch). The participants were not told what the possible target words were. In other words, when they heard [bo:], [bo:], they were not told to choose between the words *brood* and *boot*. Each participant simply heard two stimuli in a row and had to click on the left button ("first") when the participant thought the first of the two utterances was more likely to be the cluster word and on the right button ("second") when the participant thought the second utterance was more

[8] From now on we refer to all trimmed stimuli as words.

likely to be the cluster word. The participants heard the stimuli through a Sennheiser headphone set and the stimuli were presented to them only once. After hearing both utterances they were “forced” to make their decision.

The first three trials familiarized the participants with the test procedure. The experiment itself consisted of two parts, one for /Cr/~/C/ minimal pairs and one for /kn/~/k/ minimal pairs. The order of the two parts of the experiment was randomly distributed among the participants, so that 17 of them started with the /Cr/~/C/ part of the test and 18 of them started with the /kn/~/k/ part of the test. The /Cr/~/C/ part consisted of 47 trials and the /Cr/~/C/ part of the experiment consisted of 36 trials. In total, the experiment lasted about 10 minutes.

2.3.1.3. Participants

In total 35 native speakers of Dutch (19 women, 16 men) performed the forced-choice identification task. Mean age was 35.8 years (age range 23–60). Of these, 12 participants were exposed to the speech of toddlers regularly.

2.3.1.4. Analysis

Each response in the forced-choice perception task was labeled as correct or incorrect depending on whether the participant choice matched the toddler’s intention. For instance, if in the ‘bread’ ~ ‘boat’ pair the participant chose the intended ‘bread’ word as the cluster word, this choice was deemed correct. Pooling the results over all 35 listeners, we obtained a ‘correct’ score between 0 and 35 for each item pair.

As we were interested in the factors that influence the choice of adult Dutch listeners when they are forced to decide between a cluster and a singleton utterance produced by a toddler (although these two utterances sound almost the same), we submitted the correctness scores from the forced-choice identification test to a logistic-regression analysis. The predictors (factors) in this logistic regression are the differences between the acoustic measures of the cluster word and the acoustic measures of the singleton word. For instance, for the /Cr/~/C/-word set, Study 1 measured four acoustic cues (F2 movement in the vowel onset, F3 movement in the vowel onset, vowel duration and word duration) for both members of each of the 47 utterance pairs, so that the logistic regression for the /Cr/~/C/ word pairs involves the following four factors: F2 movement difference (the F2 movement of the cluster word minus the F2 movement of the corresponding singleton word), F3 movement difference, vowel duration difference, and word duration difference. This results in four difference measures for each of the 47 pairs. Likewise, the /kn/~/k/ word set leads to six difference measures for each of the 37 utterance pairs.

The interpretation of the coefficients that result from the regression analysis is as follows. If, for instance, the estimated coefficient for the vowel duration difference is positive, this would indicate that listeners have a greater chance of scoring correct if the cluster word is longer than the singleton word than if the cluster word is shorter than the singleton word; we could then conclude that listeners associate cluster words with longer duration and singleton words with shorter duration.

Following the same line of reasoning, if the estimated coefficient for the vowel duration difference is negative, this would indicate that listeners have a greater chance of scoring correct if the cluster word is shorter than the singleton word than if the cluster word is longer than the singleton word. If this would be the case, then we could conclude that listeners associate cluster words with a shorter duration and singleton words with longer duration.

2.3.2. Results

2.3.2.1. /Cr/~/C/ word pairs

Here we discuss the overall scores in the identification test, which are calculated in percentages. As explained in the previous section, /Cr/~/C/ word pairs were presented to the listeners and their task was to identify the target /Cr/-word from the two. The answers were identified as being either correct or incorrect. Higher 'correct' scores indicate a higher sensitivity of the participant to the acoustic cues that distinguish the child's /Cr/ production from her /C/ production. The results are sorted by item and are used in the logistic regression analysis where the acoustic predictors for correct perception are investigated.

In the identification task, the mean result for correct identification per item was 51%, the range being 2.8% to 91.4% (N = 47). The large variation results from the fact that some reduced cluster words were easy to identify, while others were incorrectly taken for simple onset words. Our next question then was: are the correct identifications, although at chance level, correlated with any of the acoustic cues that were found to be distinctive in Study 1? In order to find an answer to this question we obtained a difference value for the four cues from Study 1 (F2, F3, utterance duration and vowel duration) for each of the utterance pairs from the forced-choice identification task, by subtracting the value obtained for the simple onset word from the value obtained for the reduced cluster word.

In order to find out which of the four cues are good predictors for the correct judgment of a certain word pair, we conduct a logistic regression analysis (with the function *glmer* in *R*). The response variable was whether the participant gave a correct or incorrect answer, where an answer was considered correct if the participant chose the intended cluster word by the toddler as the cluster word. The potential fixed factors in the logistic regression were the F2 movement difference (dMovF2), the F3 movement difference (dMovF3), the vowel duration difference (dDurV) and the word duration difference (dDurW). Finally, the participants are included in the regression model as a random factor.

With only participant as a factor, the intercept of the model becomes +0.04038. This means that the average log odds in favor of a correct choice is +0.04038, i.e. a correct choice is exp(+0.04038) = 1.0412 more likely than an incorrect choice. This corresponds to the average percentage correct mentioned above (51%/49% = 1.04). The standard error of the log odds is 0.06940, so the 95% confidence interval of the log odds runs from -0.09567 to +0.17643 (i.e. the odds correct could be anywhere between 0.909 and 1.193), which includes zero, so we have not shown that listeners can differentiate the average pair better than chance.

Of the four fixed factors, entering dDurW yields the best improvement of the model ($p = 1.18 \cdot 10^{-8}$), although entering dDurV would have yielded a substantial improvement ($p = 5.4 \cdot 10^{-4}$) as well. After adding dDurW, adding ddF2 into the model yields another improvement, and after this no additions, not even of dDurV, improve the model any further.[9]

In the final model, dDurW has a coefficient of 1.4368345/s (standard error 0.2463989/s; c.i. = +0.95380 .. +1.91987/s; p from zero = $5.5 \cdot 10^{-9}$). This means that every increase of 10 ms in the duration difference between cluster word and singleton word increases the odds for a correct score by a factor of exp(0.014368345) = 1.0145, i.e. by 1.45% (c.i. = +0.96% .. +1.94%). This means that the longer the cluster word lasts, the more often it is correctly chosen to be the cluster word, and the longer the singleton word lasts, the more often this is incorrectly chosen to be the cluster word. The unsurprising interpretation is that

[9] The cause that dDurV cannot be included in the model, although it had such a good p value when entering it on its own, is that dDurV and dDurW are strongly correlated, so that entering dDurW takes away dDurV's chances.

listeners associate cluster words with a longer duration and singleton words with a shorter duration.

The other factor in the final model, dMovF2, has a coefficient of -0.0002874/Hz (standard error 0.0001350/Hz; c.i. = -0.00055205. -0.00002275/Hz; *p* from zero = 0.0332). This means that every *decrease* of 10 Hz in the F2 movement difference between cluster word and singleton word increases the odds for a correct score by 0.287% (c.i. = +0.023% .. +0.554%). This means that the more F2 falls (or the less it rises) in the cluster word, the more often it is correctly chosen to be the cluster word, and the more F2 rises (or the less it falls) in the singleton word, the more often this is incorrectly chosen to be the cluster word. The interpretation is that listeners associate cluster words with a falling F2 and singleton words with a rising (or less falling) F2.

It thus appears that word duration[10] is the main successful predictor for the participants' judgments. Given that the correlation found is positive, it means that correct results are correlated with higher difference values for word duration. Remember that the difference value was obtained by a subtraction of the simple onset word duration from the reduced cluster word duration. A positive value then means that the reduced cluster word was longer than the simple onset word. Adult listeners appear to judge the longer of the two utterances to be the target /Cr/ word. Given that word duration is not a distinctive acoustic characteristic of the /Cr/~/C/ minimal pairs (recall the results from the MANOVA, 2.2.4.1), the participants leaned on the wrong acoustic cue when making up their mind about which of the two utterances in the forced-choice identification task had an onset cluster in its target form.

[10] Word duration refers to the duration of the utterance that was used in the perception test. The utterance could have either a CV or a CVC structure, dependent on whether it was trimmed or not.

The same goes for the second predictor suggested by the model. According to the results, the F2 difference score was also a good predictor of the variation in the participant's judgments. Given that here a negative correlation was found, it means that out of the two utterances the participants perceived, they chose for the one with lower F2 in the vowel onset. If we recall the results from the MANOVA in 2.2.4.2, we can see that this is opposite to the intention of the toddlers, who tend to use the rise of F2 in the vowel onset to "mark" the omitted liquid in their utterance.

2.3.2.2. /kn/~/k/ word pairs

For the /kn/ dataset, the fraction correct perception, averaged over the 37 items (utterance pairs), is 713/1295 = 55.06%, with the worst item scoring 8.6% and the best item scoring 88.6%. For the correlation analysis we are interested in seeing whether a higher correct perception of an item is correlated with one of the acoustic characteristics of nasality or with a measure of a covert contrast like duration, which were measured in advance.

As with the /Cr/~/C/ pairs, we conduct a logistic regression analysis. The response variable is again whether the participant gave a correct or incorrect answer, but the potential fixed factors in the logistic regression are the differences (between cluster and singleton) in mean F1 (dF1), in "vowel complexity" (dComplex), in overall vowel intensity (dIntens), in center of gravity (dCoG), in vowel duration (dDurV), and in word duration (dDurW), and the vowel type, i.e. whether the onset is followed by /i,ɪ/ or by /a,ɑ,o/ (VowTyp)[11].

With only participant as a (random) factor, the intercept of the model becomes +0.20301. This means that the average log odds in favor of a correct choice are

[11] In the /i,ɪ/ word pair set, /ɪ/ was the predominant vowel (16 out of 18), while in the /a,ɑ,o/ set the predominant vowels were /ɑ/ and /o/, 8 out of 19 and 10 out of 19, respectively. In this way we could determine our data set as consisting of front vs. back vowels.

+0.20301, i.e. a correct choice is exp(+0.20301) = 1.2251 times more likely than an incorrect choice. This corresponds to the average percentage correct mentioned above (55.06%/44.94%). The 95% confidence interval of the log odds runs from +0.09370 to +0.31273 (i.e. the odds correct could be anywhere between 1.098 and 1.367), which does not include zero, so we have shown that listeners differentiate the average pair better than chance.

Of the six fixed factors, entering dF1 yields the best improvement of the model (χ^2 = 36.985, df = 1, p = $1.2 \cdot 10^{-9}$). After adding dF1, adding dComplex gives the best improvement (χ^2 = 22.264, p = $2.4 \cdot 10^{-6}$). After adding dComplex, adding dDurV into the model yields another improvement (χ^2 = 11.844, p = 0.00058), although adding dDurW would have been almost equally good (χ^2 = 9.955, p = 0.0016). After adding dDurV, only VowTyp can be added (χ^2 = 11.970, p = 0.00054).[12]

In the final model, dF1 has a coefficient of -0.00289/Hz (c.i. = -0.00378 .. -0.00202/Hz; p from zero = $1.1 \cdot 10^{-10}$). This means that every decrease of 10 Hz in the mean-F1 difference between cluster word and singleton word increases the odds for a correct score by a factor of exp(0.0289) = 1.0293, i.e. by 2.93% (c.i. = +2.04% .. +3.85%). This means that the lower the F1 of the cluster word is, the more often it is correctly chosen to be the cluster word, and the lower the F1 of the singleton word is, the more often this is incorrectly chosen to be the cluster word. The interpretation is that listeners associate cluster words with a closed vowel realization and singleton words with an open vowel realization.

The second factor in the model, dComplex, has a coefficient of -5.95/s (c.i. = -9.99 .. -1.96/Hz; p from zero = 0.0036). This means that every decrease of 10 ms in the vowel complexity difference between cluster word and singleton word

[12] The cause that dDurV cannot be included in the model, although it had such a good p value when entering it on its own, is that dDurV and dDurW are strongly correlated, so that entering dDurW takes away dDurV's chances.

increases the odds for a correct score by a factor of exp(0.00595) = 1.0613, i.e. by 6.13% (c.i. = +1.98% .. +10.50%). This means that the more complex the vowel is, the more often it is correctly chosen to be the cluster word, and the less complex the vowel is, the more often it is incorrectly chosen to be the cluster word. The interpretation is that listeners associate cluster words with complex vowels and singleton words with simple vowels.

The third factor in the final model, dDurV, has a coefficient of +1.304/s (c.i. = +0.615 .. +2.019/s; *p* from zero = 0.000266). This means that every increase of 10 ms in the vowel duration difference between cluster word and singleton word increases the odds for a correct score by 1.01% (c.i. = +0.62% .. +2.04%). This means that the longer the vowel in the cluster word is, the more often it is correctly chosen to be the cluster word, and the longer the vowel in the singleton word is, the more often this is incorrectly chosen to be the cluster word. The interpretation is that listeners associate cluster words with longer vowels and singleton words with shorter vowels.

The fourth factor in the final model, VowTyp, has a coefficient of +0.414 for /i,ɪ/ over /a,ɑ,o/ (c.i. = +0.179 .. +0.651; *p* = 0.00057). This means that /k(n)i,k(n)ɪ/ words have a higher odds for a correct score than /k(n)a,k(n)ɑ,k(n)o/ words by a factor of exp(0.414) = 1.513, i.e. by 51.3% (c.i. = +19.6% .. +91.7%). This means that if a cluster word contains /kni,knɪ/, it is more likely to be correctly chosen to be the cluster word than if a word contains /kna,knɑ,kno/. The interpretation is that listeners differentiate the /kn/~/k/ word pairs containing a front vowel better than they differentiate the /kn/~/k/ word pairs containing a back vowel.

2.4. Discussion

2.4.1. /Cr/~/C/ word pairs

In Study 1 we carried out four acoustic measurements with the /Cr/~/C/ word pairs, namely word duration, vowel duration, and F2 and F3 movement in the vowel onset. It turned out that only the F2 measure showed a significant difference between the word pairs, the F2 slightly rising in the vowel onset of target /Cr/ words. The other three measures did not reliably differentiate between the word pairs.

For adult production it is known that rhotics exhibit a low F3 together with a small F2–F3 separation (Kent & Read, 2002; Lindau, 1985). In adult speech, Gulian (in preparation) found a rise of F2 in the vowel onset only for front vowels following /Cr/ clusters. However, in the child language data the effect of F2 in the vowel onset found here is not restricted to front vowels. We should therefore consider the possibility that the rising F2 in the child language data is not a trace of the target rhotic, but of another sound that substitutes this rhotic. Since F2 is rising, this segment should have a low F2. In English we would find a ready candidate for this substitute, namely the glide /w/. A salient effect of the glide /w/ (in English) on the following vowel is a rising F2 (Cooper, Liberman & Borst, 1951). Further, Chaney (1988) recorded adult-like articulations of target segments /r, l, w, j/ by four-year-olds and found that among these segments /w/ exhibited the lowest F2 value (829–1179 Hz). Chaney reported a large F2 transition from the glide to the vowel for all vowel types but this transition was faster for front vowels. More evidence from the acquisition of English points to the fact that in cluster acquisition, target /tw/ and /kw/ are the first clusters to be realized correctly and that /Cw/ is the designated substitute for word-initial consonant clusters (Smit, 1993, Smith 1981). However, in Dutch the glide /w/ only occurs between vowels, in unstressed syllables, and it is hardly ever used as a substitute for liquids in word-initial clusters – only twice in the CLPF database for example. It is thus not likely that the rising F2 is a trace of a (later) substitute /w/ for the target rhotic segment. Interestingly, the F2 in the children’s productions has a frequency comparable to that of the F3 in adult speech. A more likely explanation for the rising F2 in the children’s productions, then, is that it acoustically mimics the rising F3 in post-rhotic vowels in adult speech, a strategy that can arise under circumstances of underdeveloped speaker normalization skills. This could also explain why we do not find a rising F3 as a trace of the target rhotic in the children's productions. The rising F2 might thus be a genuine trace of the rhotic in the target /Cr/ cluster, and shows a first discrepancy between the production and perception of adults and children.

The results of the adult perception experiment show another discrepancy, in this case between the cues that the adult participants relied on to make a decision, and the cues that the children provided. When the two words in the forced-choice identification task perceptually differed in duration, the adult listeners tended to choose the longer one as the target /Cr/ word. This choice might be based on an expected compensatory lengthening of the vowel. While this does occur in reduced coda clusters (Gilbers & Den Ouden, 1994), for target onset clusters the children in our study did not systematically use word duration to differentiate between the minimal pairs. This discrepancy between the children's intentions and the adults' interpretation explains the poor performance in the forced-choice identification task.

2.4.2. /kn/~/k/ word pairs

We measured six acoustic cues for the /kn/~/k/ word pairs, namely, mean F1, vowel complexity duration, overall vowel intensity, center of gravity, vowel duration and word duration. We found that target /kn/ words tend to contain a different vowel complexity duration (i.e. a vowel consisting of more than one vowel element) from target /k/ words. Besides, target /kn/ words probably show lower center of gravity than their /k/ counterparts; this is especially true if a low vowel follows. The vowel complexity trace is a child-specific cue to the intended realization of the /n/ that is not found in adult productions (Gulian & Levelt, 2009). A tentative interpretation of the "vowel complexity measure" effect could be that toddlers' first step into learning to master /kn/ clusters occurs through diphthongizing the vowel. However, the exact nature of this trace needs to be explored further.

In the perception test, participants performed only slightly better than chance. The factor that best predicted the participants' correct choice for the /kn/ target word, was the score for difference in mean F1. That is, listeners associate cluster words with a closed vowel realization and singleton words with an open vowel realization. Furthermore, the perception scores appeared to be associated with the vowel complexity score, where listeners associate cluster words with complex vowels and singleton words with simple vowels. The perception scores were correlated with vowel duration too, where listeners associate cluster words with the words containing longer vowels. Finally, listeners appear to differentiate the /kn/~/k/ word pairs containing a front vowel better than they differentiate the /kn/~/k/ word pairs containing a back vowel.

Two of the difference scores (difference mean F1 and difference vowel duration) which the participants thought to point to target /kn/ clusters, did not significantly differentiate between /kn/~/k/ word pairs in the children's productions. As in the /Cr/~/C/ case, then, there might be a discrepancy between the distinctive cues that can be found in the productions of Dutch children and the cues that adults seem to rely on to distinguish between these productions. Again, this discrepancy is a likely explanation for the poor performance of adult listeners in the perception test.

However, in the /kn/~/k/ case there was also one successful predictor of correct perception, namely the difference in vowel complexity, which indeed distinguishes the children's /kn/~/k/ target word productions. In a perception test, listeners correctly associate target /kn/ words with complex vowels and target /k/ words with simple vowels. However, despite the correctly detected complex vowel in cluster words by the adult participants, the overall perception scores were nearly at chance level due to the apparently higher reliance on misleading cues.

The /kn/~/k/ word pairs were divided into a front vowel word set and back vowel word set. The front vowel word set consisted predominantly of word pairs containing /ɪ/, while the back vowel word set consisted of /ɑ/ words on the one hand and of /o/ words on the other hand. The mean F1 of the front vowel word set was 554 Hz, and 737 Hz of the back vowel word set. We found that words like *knikker*, which were reduced to [kɪkəɹ], were easier to associate with cluster words than words like *knakworst* reduced to [tɑkvɔst] or words like *knoopjes* reduced to [koːpəs]. Moreover we found that the listeners were better at associating reduced cluster words with the words containing vowels with low F1 than with words containing vowels with higher F1. This might explain the high scores in the identification test for the /kni/~/ki/ word pairs, despite the fact that our data set does not consist of prototypical front nor of prototypical back vowels.

2.5. Conclusion

In this chapter we analyzed the acoustic characteristics of pairs of toddler utterances that differed in the presence or absence of an onset cluster in the target words they were producing. We looked at minimal pairs of /Cr/~/C/ and /kn/~/k/ target words, which were realized in such a similar way that they received identical transcriptions in the CLPF database. We found that acoustic traces of the omitted consonants were present in the children's productions (Study 1), but that adult listeners were not very good at picking up on these acoustic cues (Study 2). Even though Dutch children's target rhotics in the complex cluster /Cr/ were not perceived by adult transcribers, the children did tend to realize a rising F2 in the vowel onset, which is reminiscent of the rising F3 that we see in adult speech. As for target /kn/ clusters in which the nasal sound was not perceived by adult transcribers, we found that the subsequent vowel in the child's production did show a moving formant pattern, and a lower center of gravity. These new covert contrasts add to those discovered before, for VOT by McLeod et al. (1998) and for omitted target coda consonants by Song and Demuth (2008).

The existence of an acoustic trace sheds light on the locus of the omission in the production model (see Figure 1). Since a phonetic correlate is present, the consonant must have been available in the lexical representation as well as in the phonological surface representation. The consonant must therefore exist at every stage in Figure 1, although its presence in the acoustic representation may be minimal. One could say that the ultimate "deletion" of the consonant takes place in the subsequent perception by the adult listener.

This communication problem is apparently caused by the acoustic "near merger" that the child creates. As mentioned in the introduction, the source for this near merger might lie in an incomplete mapping from phonological surface form to the appropriate set of auditory target cues; for instance, the child might only implement a low F1 cue for the segment /n/ in the cluster, which for an adult observer may constitute nothing more than a slight non-phonemic diphthongization on the vowel. Alternatively, the source might lie in an incomplete sensorimotor mapping from auditory cues to articulatory gestures; for instance, the child may be able to implement a lowering of the velum for the /n/ in a cluster, but not the required tongue-tip gesture. Or it might lie in developmental restrictions on motor behavior; executing all the required articulatory movements for two subsequent consonants within a certain time requires practice, and therefore for some time only a partial realization succeeds. Alternatively, the information about the adult /r/ or /n/ in clusters that the child stored in her underlying form may have been incomplete, e.g. in the form of a phonological feature on the vowel, which at the phonetic level comes with auditory cues that the adult listener cannot interpret. This requires further study.

In the adult perception experiment, the main predictive factor for perceiving a cluster was duration. Relying on duration is a rational strategy because the correct production of a word starting with a consonant cluster is likely to last longer than its minimal pair starting with a singleton consonant. Unfortunately for the participants in the study, the children in this study did not systematically realize the durational aspect of the target clusters. Adult listeners apparently use different cues in order to disambiguate toddlers' productions than the cues that toddlers themselves provide to contrast between words that on the surface sound like minimal pairs. This "mistuning" between toddlers' intention and adults' perception will of course disappear over time because the child's own perception and articulation abilities improve. Being misunderstood by adult listeners might be a trigger to improve on their word productions.

Appendix 1: List of the words used for acoustic analysis in the present study in their target adult form. The words are followed by a phonetic transcription and an English translation.

1. Beertje [beːɹtjə] 'bear'
2. Bomen [boːmə] 'trees'
3. Boot [boːt] 'boat'
4. Borstel [bɔɹstəl] 'brush'
5. Bril [bʀɪl] 'glasses'
6. Brokje [bʀɔkjə] 'piece'
7. Bromfiets [bʀɔmfits] 'motor bike'
8. Brood [bʀoːt] 'bread'
9. Daar [daːɹ] 'there'
10. Die [di] 'that'
11. Dit [dɪt] 'this'
12. Draait [dʀaːit] 'it turns'
13. Draaimolen [dʀaːimoːlə] 'merry-go-round'
14. Draak [dʀaːk] 'dragon'
15. Drinken [dʀɪŋkə] 'to drink'
16. Gat [χɑt] 'hole'
17. Kijk [kɛik] 'look'
18. Kat [kɑt] 'cat'
19. Kaak [kaːk] '(nonword)'
20. Kinderboerderij [kɪndəɹbuːɹdəʀɛi] 'children's farm'
21. Kindje [kɪntjə] 'little kid'
22. Kikker [kɪkəɹ] 'frog'
23. Kip [kɪp] 'chicken'
24. Koom [koːm] '(nonword)'
25. Kopen [koːpə] 'to buy'
26. Knaak [knaːk] 'old currency (functions as nonword)'
27. Knap [knɑp] 'clever'

28. Knakworst [knɑkʋɔɹst] ‘frankfurter sausage’
29. Knie [kni] ‘knee’
30. Knippen [knɪpə] ‘to cut’
31. Knikker [knɪkəɹ] ‘marble ball’
32. Knopen [knoːpə] ‘buttons’
33. Knuppel [knʏpəl] ‘base-ball bat’
34. Noot [noːt] ‘nut’
35. Paat [paːt] ‘(nonword)’
36. Praten [pʀaːtə] ‘to talk’
37. Taak [taːk] ‘task (functions as nonword)’
38. Tak [tɑk] ‘branch’
39. Tas [tɑs] ‘bag’
40. Tekenen [teːkənə] ‘to draw’
41. Tik tak [tɪk tɑk] ‘tick tock of a clock’
42. Thijs [tɛis] ‘(proper boy's name)’
43. Tractor [tʀɑktɔɹ] ‘truck’
44. Trein [tʀɛin] ‘train’
45. Trekken [tʀɛkə] ‘to pull’

Appendix 2: List of children producing cluster reduction; (m) stands for boy and (f) stands for girl.

Database children		Children at day-care centers	
/Cr/ cluster reduction	/kn/ cluster reduction	/Cr/ cluster reduction	/kn/ cluster reduction
Cato (f)	**Cato** (f)	Daniel (m)	Daniel (m)
Robin (m)	**Robin** (m)	Herman (m)	Herman (m)
Jarmo (m)		Mick (m)	Mick (m)
Tirza (f)		Taeke (m)	Taeke (m)
		Matteo (m)	Matteo (m)
		Hannah (f)	Hannah (f)
		Lars (m)	Jan (m)
		Gyula (f)	Daniel C (m)
		Pieter (m)	Denim (m)
		Micky (m)	Evio (m)
		Boaz (m)	Amber (f)
			Mauro (m)
			Han (m)
			Julian (m)
			Sterre (f)
			Sem (m)

Appendix 3: List of /Cr/~/C/ word pairs used for acoustic analysis. The pairs consist of a cluster reduction and singleton utterances from children from the database and the day-care center sets. The word pairs from the database set appear in bold.

Utterance number	Cluster word			Singleton word		
Name child	age	Intended production	Actual production	age	Adult target	Actual production
1. **Cato**	1;10	draak [dʀa:k]	[da:k]	1;11	daar [da:ɹ]	[da:ɹ]
2. **Cato**	2;0	drinkt [dʀɪŋkt]	[dɪːŋkt]	1;11	dit [dɪt]	[tɪt]
3. **Cato**	2;1	drinken [dʀɪŋkə]	[dɪːŋkə]	2;1	dit [dɪt]	[dɪt]
4. **Cato**	2;1	trekken [tʀɛkə]	[tɛkə]	2;0	tekenen [te:kənə]	[te:kənə]
5. **Cato**	2;4	draait [dʀait]	[d:ait]	2;3	daar [da:ɹ]	[da:ɹ]
6. **Tirza**	1;8	broek [bʀuk]	[buk]	1;8	boek [buk]	[buk]
7. **Tirza**	1;8	broek [bʀuk]	[buk]	1;8	boek [buk]	[buk]
8. **Tirza**	1;8	broek [bʀuk]	[buk]	1;8	boek [buk]	[buk]
9. **Tirza**	1;11	draaimolen [dʀaimo:lə]	[da:i]	1;11	daar [da:ɹ]	[da:ɹ]
10. **Tirza**	1;11	draaimolen [dʀa:imo:lə]	[da:i]	1;11	daar[da:ɹ]	[da:ɹ]
11. **Tirza**	2;0	draaimolen [dʀa:imo:lə]	[da:i]	2;1	daar [da:ɹ]	[da:]
12. **Tirza**	2;0	draai [dʀa:it]	[da:it]	2;0	daar [da:ɹ]	[ta:]
13. **Tirza**	2;0	trein [tʀɛin]	[tɛin]	2;0	kijk [kɛik]	[kɛi]

14. **Tirza**	2;2	kraan [kʀaːn]	[kaːn]	2;2	taart [buk]	[taːt]
15. **Tirza**	2;2	trein [tʀɛin]	[tɛin]	2;2	kijk [kɛik]	[tɛi]
16. **Tirza**	2;4	broek [bʀuk]	[buk]	2;4	boek [buk]	[buk]
17. **Tirza**	2;4	broodje [bʀoːtjə]	[boːtjə]	2;4	boot [buk]	[boːt]
18. **Robin**	2;0	drinken [dʀɪŋkə]	[dɪːŋkə]	1;11	die [ti]	[tɪ]
19. **Robin**	2;0	drinken [dʀɪŋkə]	[dɪːŋkə]	1;11	die [ti]	[tɪ]
20. **Jarmo**	2;3	bril [bʀɪl]	[bɪl]	2;2	beertje [beɹtjə]	[biːcə]
21. **Jarmo**	2;2	brokje [bʀɔkjə]	[bɔkjə]	2;2	bomen [boːmə]	[boːmə]
22. Boaz	2;0	traktor [tʀɑktɔɹ]	[tɑktɔɹ]	2;0	tak [tɑk]	[tɑk]
23. Daniel	1;6	praten [pʀaːtə]	[paːtə]	1;6	paat [paːt]	[paːt]
24. Daniel	1;6	trein [tʀɛin]	[tɛin]	1;6	Thijs [tɛis]	[tɛis]
25. Daniel	1;6	traktor [tʀɑktɔɹ]	[tɑtɔ]	1;6	tak [tɑk]	[tɑk]
26. Gyula	2;1	brood [bʀoːt]	[boːt]	2;1	boot [boːt]	[boːt]
27. Herman	3;0	brood [bʀoːt]	[boːt]	3;0	noot [noːt]	[poːt]
28. Herman	3;0	praat [pʀaːt]	[paːtə]	3;0	kaak [kaːk]	[kaːk]
29. Mick	2;2	traktor [tʀɑktɔɹ]	[tɑktɔɹ]	2;2	kat [kɑt]	[kɑt]
30. Micky	2;1	traktor [tʀɑktɔɹ]	[tɑktɔɹ]	2;1	kat [kɑt]	[kɑt]
31. Taeke	1;8	traktor [tʀɑktɔɹ]	[tɑktɔɹ]	1;8	kat [kɑt]	[kɑt]

32. Pieter	2;0	trein [tʀɛin]	[tɛin]	2;0	Thijs [tɛis]	[tɛis]
33. Pieter	2;0	brood [bʀoːt]	[boːt]	2;0	boot [boːt]	[boːt]
34. Pieter	2;0	traktor [tʀɑktɔɹ]	[tɑktɔɹ]	2;0	kat [kɑt]	[kɑt]
35. Hannah	2;1	broek [bʀuk]	[buk]	2;1	boek [buk]	[buk]
36. Hannah	2;1	draakje [dʀaːkjə]	[daːkjə]	2;1	taak [taːk]	[taːk]
37. Hannah	2;1	kraan [kʀaːn]	[kaːn]	2;1	taak [taːk]	[taːk]
38. Hannah	2;1	trein [tʀɛin]	[tɛin]	2;2	Thijs [tɛis]	[tɛis]
39. Matteo	2;2	trein [tʀɛin]	[tɛin]	2;2	Thijs [tɛis]	[tɛis]
40. Matteo	2;2	draakje [dʀaːkjə]	[taːkjə]	2;2	taak [taːk]	[taːk]
41. Matteo	2;3	trein [tʀɛin]	[tɛin]	2;3	Thijs [tɛis]	[tɛis]
42. Matteo	2;3	draak [dʀaːk]	[daːk]	2;3	taak [taːk]	[taːk]
43. Matteo	2;3	kraan [kʀaːn]	[kaːn]	2;3	gat [xɑt]	[kaːt]
44. Matteo	2;3	draak [dʀaːk]	[daːk]	2;3	taak [taːk]	[taːk]
45. Lars	2;6	trein [tʀɛin]	[tɛin]	2;6	Thijs [tɛis]	[tɛis]
46. Lars	2;6	kraan [kʀaːn]	[kaːn]	2;6	taak [taːk]	[taːt]
47. Lars	2;6	draak [dʀaːk]	[daːt]	2;6	kaak [taːk]	[taːt]

Appendix 4: List of /kn/~/k/ word pairs used for acoustic analysis. The pairs consist of a cluster reduction and singleton utterances from children from the database and the day-care center sets. The information about the children from the database appears in bold.

Utterance number	Cluster word			Singleton word		
Name child	age	Intended production	Actual production	age	Intended production	Actual production
1. **Cato**	2;4	knie [kni]	[kiː]	2;3	kikker [kɪkəɹ]	[kɪkəɹ]
2. **Cato**	2;0	knie [kni]	[ki]	1;11	kikker [kɪkəɹ]	[kɪkəɹ]
3. **Cato**	2;6	knikker [knɪkəɹ]	[kɪkəɹ]	2;5	kikker [kɪkəɹ]	[kɪkəɹ]
4. **Robin**	2;4	knip [knɪp]	[kɪp]	2;3	kip [kɪp]	[kɪp]
5. **Robin**	2;4	knippen [knɪpə]	[kɪpə]	2;3	kind [kɪnt]	[kɪnt]
6. **Robin**	2;4	knuppel [knʏpəl]	[kɪpɔ]	2;0	kijk [kɛik]	[kɪk]
7. Daniel	1;6	knakworst [knɑkʋɔɹst]	[kɑkvɔst]	1;6	tak [tɑk]	[tɑk]
8. Daniel C	1;6	knakworst [knɑkʋɔɹst]	[tɑkvɔst]	1;6	tak [tɑk]	[tɑk]
9. Daniel C	1;8	knakworst [knɑkʋɔɹst]	[tɑkvɔst]	1;8	tas [tɑs]	[tɑs]
10. Daniel C	1;8	knoopjes [knoːpjəs]	[toːpjə]	1;8	boot [boːt]	[boːt]
11. Evio	2;0	knopen [knoːpə]	[koːpə]	2;0	boot [boːt]	[boːt]
12. Evio	2;0	knip [knɪp]	[kɪp]	2;0	kindje [kɪnt͜jə]	[kɪnt͜jə]
13.Herman	3;0	knaak [knaːk]	[kaːk]	3;0	kaak [kaːk]	[kaːk]
14.Herman	3;0	knap [knɑp]	[kɑːp]	3;0	kat [kɑt]	[kɑt]
15.Herman	3;0	knippen [knɪpə]	[kɪpə]	3;0	kippen [kɪpə]	[kɪpə]
16. Jan	1;10	knakworst [knɑkʋɔɹst]	[kɑkvɔst]	1;10	tas [tɑs]	[tɑs]

17. Jan	1;10	knakworst [knɑkʋɔɹst]	[kɑkvɔst]	1;10	kat [kɑt]	[kɑt]
18. Jan	1;10	knippen [knɪpə]	[kɪpə]	1;10	kippen [kɪpə]	[kɪpə]
19. Denim	2;0	knippen [knɪpə]	[kɪpə]	2;0	kippen [kɪpə]	[kɪpə]
20. Denim	2;0	knopen [knoːpə]	[koːpə]	2;0	kopen [koːpə]	[koːpə]
21. Mick	2;2	knippen [knɪpə]	[kɪpə]	2;2	kippen [kɪpə]	[kɪpə]
22. Mick	2;2	knopen [knoːpə]	[koːpə]	2;2	kopen [koːpə]	[koːpə]
23. Mick	2;2	knakworst [knɑkʋɔɹst]	[kɑkvɔst]	2;2	kat [kɑt]	[kɑt]
24. Taeke	1;8	knippen [knɪpə]	[kɪpə]	1;8	kippen [kɪpə]	[kɪpə]
25. Taeke	1;8	knopen [knoːpə]	[koːpə]	1;8	kopen [koːpə]	[koːpə]
26. Matteo	2;2	knoopjes [knoːpjəs]	[koːpjəs]	2;2	koon [koːn]	[koːn]
27. Matteo	2;2	knoop [knoːp]	[koːp]	2;2	kopen [koːpə]	[koːpə]
28. Matteo	2;2	knoop [knoːp]	[koːp]	2;2	koom [koːm]	[koːm]
29.Hannah	2;1	knoopjes [knoːpjəs]	[koːpjəs]	2;1	boten [boːtə]	[poːtə]
30.Hannah	2;1	knoopjes [knoːpjəs]	[koːpjəs]	2;1	boot [boːt]	[boːt]
31.Hannah	2;1	knoopjes [knoːpjəs]	[koːpjəs]	2;1	toon [toːn]	[toːn]
32. Amber	2;2	knikker [knɪkəɹ]	[kɪkəɹ]	2;1	kikker [kɪkəɹ]	[kɪkəɹ]
33. Mauro	2;0	knikker [knɪkəɹ]	[kɪkəɹ]	2;0	kikker [kɪkəɹ]	[kɪkəɹ]

34. Han	2;0	knikker [knɪkəɹ]	[kɪkəɹ]	2;0	kikker [kɪkəɹ]	[kɪkəɹ]
35. Julian	1;11	knikker [knɪkəɹ]	[kɪkəɹ]	1;11	kikker [kɪkəɹ]	[kɪkəɹ]
36. Sterre	2;0	knikker [knɪkəɹ]	[kɪkəɹ]	2;0	kikker [kɪkəɹ]	[kɪkəɹ]
37. Sem	2;2	knikker [knɪkəɹ]	[kɪkəɹ]	2;2	kikker [kɪkəɹ]	[kɪkəɹ]