Nature's distributional-learning experiment: Infants' input, infants' perception, and computational modeling

Benders, A.T.

Publication date
2013

4

DUTCH INFANTS' SENSITIVITY TO THE COMBINATION OF VOWEL QUALITY AND DURATION IN A SPEECH SOUND CATEGORIZATION PARADIGM

An adapted version of this chapter is:
Benders, T. & Mandell, D.J. (in preparation).

Abstract

To achieve native-like speech-sound perception, infants need to integrate the multiple acoustic dimensions that signal phoneme contrasts. The present study investigates Dutch 9-month-olds', 15-month-olds' and adults' perception of /ɑ/ and /aː/, which differ in vowel quality and duration. This is done by testing their perception of vowel sounds with typical and atypical combinations of vowel quality and duration. Both categorization behavior in the two-choice categorization task, as measured by reaction times, and attention allocation, as measured by pupil dilations, were investigated. Dutch adults consistently categorized atypical [ɑː] as the vowel /ɑ/, but their categorization of atypical [a] depended on the context that was created during training. Dutch 15-month-old infants' attention allocation changed in reaction to atypical [ɑː] and [a] in comparison to their reaction to typical [ɑ] and [aː]. The influence of context on infants' attention allocation mirrored the effect of context on adults' categorization behavior. Infants' change in attention allocation to the atypical vowel sounds shows that their vowel representations are specified for the combinations of vowel duration and quality. Additionally, infant's receptive vocabulary was related to their attention allocation to the atypical vowel sounds. This study shows that 15-month-old infants can integrate the dimensions of vowel duration and vowel quality in their vowel representations, and that the detailed knowledge of rare and ambiguous cue combinations develops hand in hand with vocabulary size.

4.1 Introduction

Across languages, two of the major phonetic cues that signal vowel contrasts are vowel quality (measured by the first, second, and third formant; F1, F2, and F3) and vowel duration (Maddieson, 2011). How listeners weight these cues in their perception of vowel categories depends on their native language (Gottfried and Beddor, 1988) and native dialect (Escudero and Boersma, 2004). In order to understand infants' developing representations of their native language vowel categories, it is crucial to chart infants' changing sensitivity to these phonetic cues and to their combinations. The current paper investigates whether infants' vowel categories are primarily defined by vowel duration, vowel quality, or the combination of vowel quality and duration.

4.1.1 *Infants' sensitivity to vowel duration and vowel quality*

Newborn infants divide a vowel-quality continuum into categories that roughly correspond to the vowel categories that are found across the languages of the world, even if their language does not use all these categories (Aldridge et al., 2001). Language-specific perception of vowel sounds begins around 6 months when infants begin to lose the ability to discriminate between non-native vowel contrasts (Polka and Werker, 1994) and show stronger prototype effects for native than for non-native vowels (Kuhl et al., 1992; but see Polka and Bohn, 1996). At 12 months of age, infants' neural responses to vowel-quality changes are language specific (Cheour et al., 1998). Infants' ability to discriminate vowel-quality contrasts becomes language specific within the first year after birth.

In contrast, infants remain sensitive to vowel-duration differences for a protracted period of time, independent of their language background. German 6-to-12-month-olds discriminate vowel sounds on the basis of duration differences whenever possible (Bohn and Polka, 2001). Vowel duration differences in German are always accompanied by differences in vowel quality (Heid et al., 1995) and adult native speakers of German do not primarily rely on vowel duration to categorize or discriminate vowel sounds (Sendlmeier, 1981; Bohn and Polka, 2001). English-learning infants distinguish between non-native long and short vowel sounds well past their first birthday (Mugitani et al., 2009). English vowels differ mainly in vowel quality (Hillenbrand et al., 1995) and adult native speakers of English almost exclusively rely on vowel quality to categorize vowel sounds (Flege et al., 1997). Like infants, and unlike adult native speakers, adult second-language learners readily use vowel duration as a cue to distinguish between non-native vowel contrasts (Flege et al., 1997). Vowel duration thus is a psycho-acoustically salient cue (Bohn, 1995), to which infants and adults do not lose sensitivity.

Even though vowel duration is acoustically salient, Japanese infants seem to have difficulty incorporating this cue in their linguistic vowel representations. Vowel duration differences are phonologically contrastive in Japanese in the absence of major vowel-quality differences (e.g., /seki/ 'seat' versus /seːki/ 'century', Vance, 1987; examples from Hirata and Tsukada, 2009). As Japanese 4-month-olds nevertheless do not discriminate between a duration-cued vowel contrast, while they do note a quality-cued difference, Sato et al. (2010) argue that Japanese 4-month-olds do not yet interpret the vowel-duration difference as linguistically relevant. At 18 months of age, Japanese infants' perception of vowel duration seems to differ from the perception of this cue by younger Japanese infants or English infants of the same age (Mugitani et al., 2009). Also in a neuro-imaging paradigm, language-specific duration discrimination was found to be acquired slowly by Japanese infants, as it was not until after their first birthday that their categorical discrimination of short and long vowel sounds on the opposite sides of the category boundary had a neural response indicative of linguistic processing (Minagawa-Kawai et al., 2007).

In contrast, Dutch infants appear to develop language-specific duration perception prior to 18 months of age. In Dutch, duration differences between vowels are always accompanied by vowel-quality differences (Moulton, 1962). Vowel quality is the more important cue for adult native speakers and school-aged children (Van Heuven et al., 1986; Escudero et al., 2009a; Brasileiro, 2009; Giezen et al., 2010). Dutch infants in their first year after birth are sensitive to vowel-duration differences in the same way as English infants, irrespective of the vowel-quality differences between the sounds (Dietrich, 2006). By 18 months, Dutch infants, but not English infants, regard such vowel-duration differences as phonologically contrastive in word learning (Dietrich et al., 2007). Prior to 18 months of age, 15-month-old Dutch infants better discriminate these vowels on the basis of a difference in vowel quality as well as duration than on the basis of a difference in either cue (Chapter 3).

Several aspects of infants' developing sensitivity to vowel duration and quality are still unknown. With few exceptions (Bohn and Polka, 2001; Sato et al., 2010; Chapter 3), studies into infants' vowel perception have investigated infants' perception of either vowel duration or vowel quality and not their perception of both cues. Only Bohn and Polka (2001) and Chapter 3 studied the relative contribution of these cues to infants' perception. Therefore, it is still poorly understood how vowel quality and duration interact during phoneme acquisition, especially after the infants' first birthday.

None of the aforementioned studies have attempted to investigate how infants' acquisition of these phonetic cues relates to their concurrent receptive vocabulary size. Infants' language-specific phoneme perception and word knowledge may develop in mutual dependence as infants' increasing sensitivity to their native-language phoneme categories may enable a better recognition of word forms (Kuhl et al., 2008). Additionally, infants' growing vocabulary may enable them to better select which phonetic information is crucial for word recognition and to make their phoneme representations more precise (Werker and Curtin, 2005). Boersma et al. (2003) attribute an even larger role to infants' word knowledge in phoneme acquisition, as they propose that word knowledge enables infants to integrate acoustic dimensions into phoneme representations. Support for the hypothesis that phoneme perception facilitates word acquisition is provided by the relation between infants' language-specific speech perception at 6 and 7 months of age and their later vocabulary size (Tsao et al., 2004; Kuhl et al., 2005, 2008). At 14 and 17 months of age, infants' vocabulary size is related to their ability to learn similar sounding words (Werker et al., 2002), which is considered evidence that infants' word knowledge has refined their phoneme representations (Werker and Curtin, 2005). If there is a mutual dependence between phoneme perception and vocabulary, infants' concurrent vocabulary size is expected to be related to more fine-grained aspects of speech perception as well, such as the infants' sensitivity to the relevant cues. That prediction is tested in the present study.

The Dutch low vowels /ɑ/ and /aː/ differ in both vowel duration and vowel quality, as /aː/ is longer and has a higher F1 and F2 than /ɑ/ (Adank et al., 2004; Nooteboom and Doodeman, 1980; Rietveld et al., 2003), also in infant-directed speech (Chapter 3). These are the two most frequent full vowels in Dutch child-directed speech (Versteegh and Boves, tion). Therefore, /ɑ/ and /aː/ provide an ideal test case to further investigate the development of language-specific sensitivity to vowel duration and vowel quality. The present study investigates Dutch 9- and 15-month-olds' representation of vowel quality and duration as linguistically relevant cues to the /ɑ/–/aː/ contrast.

4.1.2 *Methods to study infants' phoneme representations*

A discrimination task is not the best choice to investigate which phonetic cues infants find linguistically relevant. If an infant discriminates between two speech sounds that differ in duration, it does not mean that the infant regards the duration difference as linguistically contrastive (Dietrich et al., 2007). On the other hand, if infants do not discriminate between two speech sounds in a simple discrimination task, they may still be able to differentially associate them with a location (Albareda-Castellot et al., 2011). Therefore, the present study used a two-alternative categorization task that required participants to form associations between a sound and a spatial feature (McMurray and Aslin, 2004; Kovács and Mehler, 2009; Albareda-Castellot et al., 2011). In this procedure task, the participant is presented with one of two cueing sounds, after which a visual outcome, a small animation, is presented on either the left or the right of the screen, depending on which cueing sound was played. The participant learns to associate the cueing sounds with the outcome locations. In order for participants to generalize this association to a novel stimulus it is not enough to note that the novel stimulus is different from the previous stimuli. Rather, the participant has to decide which of the learned cueing sounds the novel stimulus is most similar to. Therefore, it asks participants to categorize novel stimuli as one or the other category, this procedure is similar to the two-alternative categorization tasks used to test cue weighting in adults and older children (e.g., Nittrouer, 1992). The exact procedure is a variant on the procedures employed by McMurray and Aslin (2004); Kovács and Mehler (2009); Albareda-Castellot et al. (2011).

In the task employed in the present paper, participants were first presented with outcomes on the left and right of the screen, dependent on the cueing sounds [tɑm] and [taːm]. These words contained vowels with a typical combination of vowel quality and duration of the phonemes /ɑ/ and /aː/[1]. To assess the contribution of vowel duration and quality in participants' representations of the vowels, participants' reaction to the sounds with atypical combinations of vowel duration and quality, [tɑːm] and [tam], were tested.

The first outcome measure in the study was the reaction time (RT) to each outcome location after the cueing sounds were played. If participants relied primarily on the salient vowel-duration cue, as could be expected for the infants (Bohn and Polka, 2001; Mugitani et al., 2009), they would look faster to the [taːm]-location upon hearing the atypical stimulus [tɑːm] and faster to the [tɑm]-location upon hearing the atypical stimulus [tam]. If participants relied primarily on vowel quality, as was expected for the adults (Van Heuven et al., 1986; Escudero et al., 2009a), they would look faster to the [tɑm]-location upon hearing atypical [tɑːm], and faster to the [taːm]-location for atypical [tam]. If participants let neither cue prevail in their representations of the typical vowel sounds [ɑ] and [aː], they would not have a difference in RTs to the atypical vowel sounds. A fourth possibility is that there would be individual differences between infants in their weighting of vowel quality and duration. Anticipatory eye-movement paradigms have the potential of revealing such individual differences (McMurray and Aslin, 2004).

In addition to the RTs, participants' pupil dilations in reaction to the typical and atypical sounds were assessed. It has been proposed

1 In this paper we adhere to the tradition in the phonological literature to present abstract representations of speech sounds or words with / /, and phonetic realizations of these abstract categories with [ ]. In Dutch, the vowel category /ɑ/ is typically realized as [ɑ] , and the vowel category /aː/ is typically realized as [aː].

that pupil dilations in a cognitive task can reflect attention and arousal as well as processing conflict in a decision (Aston-Jones and Cohen, 2005). As our two-alternative categorization task required participants to make a decision as to where they expected the outcome to appear, the pupil dilations during a trial not only tapped participants' general attention to the stimuli, but specifically the processing of the stimuli in order to make that decision. For a participant that is able to categorize the atypical stimuli [tɑːm] and [tam], the pupil dilations may reveal that categorizing atypical stimuli is more difficult than categorizing the typical stimuli. Pupil dilations can be especially informative in infants (Jackson and Sirois, 2009; Gredebäck and Melinder, 2010), as associating a sound with a location is not a trivial task for them (McMurray and Aslin, 2004; Kovács and Mehler, 2009). Even if infants are unable to correctly categorize the typical stimuli, a change in attention allocation to the atypical stimuli would reveal that they regard these cue combinations as atypical.

To conclude, the present study investigates Dutch infants' developing representations of the vowels /ɑ/ and /aː/ in a two-alternative categorization task. The question was whether infants' representations are primarily defined 1) by vowel duration, which is acoustically salient; 2) by vowel quality, which becomes linguistically relevant early in development; or 3) for the combination of vowel duration and quality. We investigate how participants categorize and allocate attention to typical examples of the categories, [ɑ] and [aː], and to tokens with an atypical combination of vowel duration and quality, [ɑː] and [a]. To investigate infants' development just before and after the onset of word acquisition, the performance of 9- and 15-month-old Dutch infants was assessed and compared to that of adults. To investigate the relation between language acquisition and vowel perception at an individual level, the 15-month-olds' performance was releated to their vocabulary size.

4.2 Method

4.2.1 *Subjects*

Participants were 40 (21 females) 9-month-olds (260–297 days), 50 (26 females) 15-month-olds (445–479 days), and 30 (21 females) adults (18–64 years). Participating children were from predominantly Dutch families, born at a gestational age of at least 36 weeks, with no known visual or auditory problems. Participating adults were monolingually raised native speakers of Dutch and reported (corrected to) normal vision and no auditory problems. All adult participants and the parents of all child participants gave informed consent prior to participating.

4.2.2 *Sound stimuli*

The test words used in the experiment can be transcribed as [tɑm], [taːm], [tɑːm], and [tam]. The stimulus design can thus be considered as a two-by-two grid of two vowel quality values (lower [ɑ] and higher [a]) by two duration values (short and long).[2][tɑm] and [taːm] are phonotactically legal word forms of Dutch[3]. The vowel sounds [ɑː] and [a] feature combinations of vowel quality and duration that do not typically occur in Dutch.

For the sound stimuli, a voice-trained female native speaker of Dutch was recorded producing the words /tɑm/ and /taːm/. The recordings were made in a sound-proof booth, using a Sennheiser HF condenser microphone MKH-105 on a Tascam CD-recorder, sampled at 44100 Hz.

Three recordings were selected of /tɑm/ and of /taːm/, on the basis of a close match in perceived pitch level and pitch contours. The three natural recordings of /tɑm/ served as the basis for the tokens of [tɑm] and [tɑːm], while the three natural recordings of /taːm/ served as the basis for the tokens of [taːm] and [tam]. The duration of the vowels in all six tokens was changed to 120 ms to create the six tokens with a short vowel duration, [tɑm] and [tam], and to 240 ms to create the six tokens with a long vowel duration, [taːm] and [tɑːm]. The resulting twelve tokens formed three of the earlier mentioned two-by-two stimulus grids of the two vowel quality and two duration values.

The consonantal frames remained unaltered in the duration manipulation. Irregular waveform periods, which occurred as a consequence of the duration manipulation, were manually removed from the signal, so that the duration of the short and long vowel sounds was usually somewhat shorter than 120 ms and 240 ms, respectively. Table 10 gives the durations and vowel qualities of the tokens after manipulation.

Each stimulus presented one of the test words ([tɑm], [taːm], [tɑːm], or [tam]) in the form of a sequence of the three different tokens of the test word.[4] Per test word, that is, for [tɑm], [taːm], [tɑːm], and [tam], three such stimuli were made, with different orders of the tokens.

All manipulations were done with Praat (Boersma and Weenink, 2010). The resynthesis was performed using the overlap-add procedure (Moulines and Charpentier, 1990, as implemented in Praat).

2 Thanks to Bob McMurray for the wording suggestion.

3 [taːm] is a pseudo-word in Dutch. [tɑm] is a real word that means 'tame' and is unlikely to be known by our child participants. Stimuli based on the word [tɑm] have been used previously by Dietrich et al. (2007).

4 The first token started after a silence of 265 ms, and ended in between 634 ms (for the shortest token) and 816 ms (for the longest token); the second token started in between 1682 ms (for the longest token) and 1863 ms (for the shortest token), and ended at 2233 ms; the third token started at 3124 ms and ended in between 3494 ms and 3675 ms.

Test word	To-ken	Vowel quality (same for short and long) F1	F2	F0 measures maximum	range	Vowel duration (differs between short and long) short	long
[tɑm]	1	877	1252	131	25	110	217
&	2	841	1261	139	23	120	227
[tɑːm]	3	841	1310	130	23	112	232
[tam]	1	910	1521	200	99	120	240
&	2	898	1527	200	100	120	230
[taːm]	3	944	1549	200	99	120	222

Table 10: **Acoustic measurements of the 12 tokens used in the present experiment.**

In the exit interview after the experiment, the majority of the adult participants transcribed the sounds they had heard during the experiment as "tam" and "taam", which are the Dutch spellings of [tɑm] and [taːm]. The remaining participants wrote "tan" and "taan". These transcriptions indicate that the stimuli contained clear examples of the intended vowels.

The experiment started with the Dutch pseudo-words /tibi/ and /drukəl/ (Swingley, 2007) as cueing sounds. The recordings for these words were made by a different female native speaker of Dutch. One token was selected per word and three copies of that token were combined into a stimulus.

4.2.3 *Visual stimuli*

All visual stimuli were 150 by 150 pixel yellow or pink rectangular boxes with white stripes. The outcome presented at the end of the trial was an animation in a 150 by 150 pixel box of either a dancing panda, a pink elephant with balloons, or Teletubbies' Tinky Winky throwing a ball.

4.2.4 *Set-up and procedure*

Prior to the experiment, parents were instructed not to interact with their child during the procedure. Adult participants were instructed that they would participate in an experiment for small children and received no further instructions. Either before or after the experiment, parents of 15-month-old infants filled out the short version of the Dutch adaptation (N-CDI, Zink and Lejaegere, 2002) of the

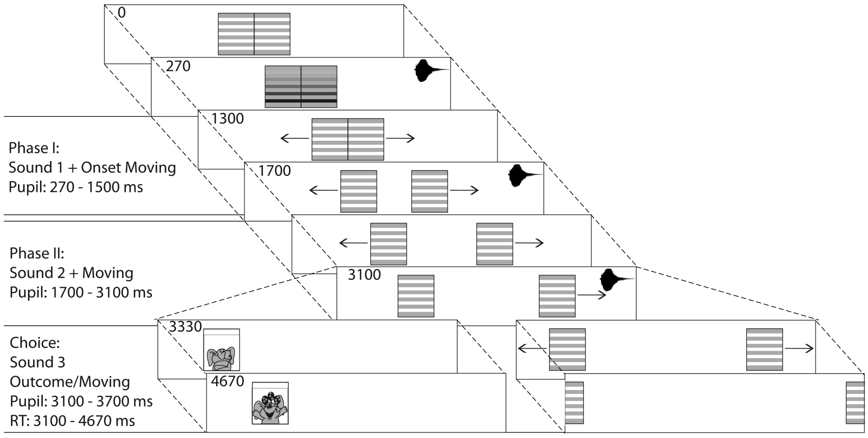


Figure 9: **The sequence of visual events in the trials in the two-alternative categorization task.** The visual events are identical across outcome trials and away trials until 3330 ms into the trial. The bottom two boxes on the left give the last visual events on outcome trials. The bottom two boxes on the right give the last visual events away trials. The numbers in the corner of the boxes give the timing of the visual events in ms. The waveforms indicate the approximate onsets of the three auditory tokens.

MacArthur Communicative Development Inventory (Fenson et al., 1993).

The experiment was conducted in a sound-proofed booth at the University of Amsterdam. Black curtains hid the equipment from view. Children were seated in an elevated car seat with their parent sitting on a chair behind them. Adult participants were seated on a chair. The experimenter was in a different room but could observe the participant through a webcam. The auditory stimuli were presented at a level of 65 dB(A). The visual stimuli were presented on the screen of a Tobii T120 Eye Tracker system, which was mounted on a movable arm. The monitor was placed 60 cm from the adult participants' eyes and 65 cm from the child participants' eyes. The eye-tracker was calibrated using an age-appropriate 9-point calibration from the Tobii Studio software and for the stimulus locations for which the Tobii Sudio software recorded no look on the first run a recalibration was attempted. The experiment was programmed in E-Prime and run on a personal computer.

The sequence of events in a trial is outlined in Figure 9. At the beginning of each trial, two striped boxes appeared side by side in the center of the screen. After 270 ms, the boxes began to flash with rainbow colors and the first auditory token was played. At 1300 ms into the trial, the boxes stopped flashing and began moving horizontally across the screen in opposite directions. Then the second auditory token was played, which had its offset at 2233 ms. The third auditory token was played at 3100 ms. There were two types of trials. On outcome trials, the boxes stopped moving at 3330 ms and an outcome video was played in one box until the end of the trial at 4670 ms. On away trials, both boxes continued to move across the screen towards the edge of the screen until the end of the trial.

Within each trial, the presentation of multiple auditory tokens was intended to ensure that infants had sufficient opportunity to process the sound before making a decision. This better processing of the stimulus within each trial was hoped to transfer to better learning of the sound–side associations. The presentation of moving blocks during the complete trial were intended to engage the infants' attention.

The experiment consisted of 40 trials, divided over four blocks. A summary of the trials per block is given in Table 11.

Block	Outcome trials		Away trials			
1	tibi (4)	drukəl (4)	tibi (1)		drukəl(1)	
2	tɑm (4)	taːm (4)	tɑm (1)		taːm(1)	
3	tɑm (3)	taːm (3)	tɑm(1)	taːm (1)	tɑːm (1)	tam (1)
4	tɑm (3)	taːm (3)	tɑm(1)	taːm (1)	tɑːm (1)	tam (1)

Table 11: **A summary of each of the four blocks in the experiment:** The number of outcome trials per stimulus word and the number of away trials per stimulus word.

The first block was designed so that subjects could become accustomed to the procedure. Participants first saw six outcome trials with the words /tibi/ (left) and /drukəl/ (right) as cueing sounds. These associations were then tested on two away trials, one with /tibi/ and one with /drukəl/, which were then followed by two more outcome trials.

In block 2, participants were shown the sound–side associations with the typical test words [tɑm] and [taːm] as cueing sounds. Eight outcome trials were presented and then two away trials. In blocks 3 and 4, the sound–side associations with [tɑm] and [taːm] were reinforced on six outcome trials per block. The remaining four trials in block 3 and in block 4 were away trials, one with each of the typical test words [tɑm] and [taːm], and one with each of the atypical generalization test sounds [tɑːm] and [tam].

Whether [tɑm] or [taːm] was the cueing sound for an outcome on the left or right of the screen was counterbalanced between participants within each age group. The first two trials of block 2 always presented [tɑm] and [taːm], the order of which was randomized across participants. The order of all other trials was randomized for each participant, with the restrictions that each cueing sound was presented on no more than three trials in a row and there were no more than three away trials in a row. Before the first trial and in between trials, a green dot appeared in the center of the screen. The experimenter could prompt looming of the dot with a bell sound to redirect the participant's attention to the screen.

4.2.5 *Analysis plan*

The data were divided into three phases. Phase I started 270 ms into the trial, which marks the onset of the first sound and of the flashing boxes, and ended at 1500 ms, after the movement began. Phase II began at 1700 ms, just before the onset of the second sound, and ended at 3100 ms. The choice phase began at 3100 ms, just before the onset of the third sound. For the RT analysis, the choice phase continued until 4670 ms, the end of the trial. For the pupil analysis, the choice phase ended at 3700 ms, that is, 400 ms after the outcome would have appeared.

The data were cleaned by identifying missing segments shorter than 500 ms, which were classified as tracking errors. Missing segments longer than 500 ms were classified as a look away from the screen.

For the RT analysis, the XY-coordinates of where the participant was looking were classified as being in the [aː]-outcome Area of Interest (AOI), the [ɑ]-outcome AOI, or the elsewhere AOI. The [aː]-outcome AOI was defined as the area on the screen where the outcome would appear on outcome trials with the cueing sound [taːm]. The [ɑ]-outcome AOI was the area where the outcome would appear on outcome trials with [tɑm]. The segments that were missing due to tracking errors were assigned to the last valid AOI before the missing data occurred.

The maximum possible RT for an actual look during the trial toward the [ɑ]-outcome AOI or the [aː]-outcome AOI after the onset of the third sound is 1540 ms. If the participant looked to only one outcome AOI on a trial, the RT for the other AOI was given an RT of 2000 ms. No RTs were computed if the participant did not look at either outcome AOI on a given trial. By assigining a ceiling value of 2000 ms to the trials on which the participant was involved in the task but not looking at the outcome, we respect the fundamental difference between trials on which the participant made a choice and random missingness. The analyses of the RTs to the [ɑ]-outcome AOI and the [aː]-outcome AOI on away trials are both reported, but only the RTs to the [ɑ]-outcome AOI are interpreted.

Pupil data were cleaned for each participant separately, with all pupil sizes more than three standard deviations away from the participant's mean excluded. This resulted in less than 3% of each participant's data being excluded. For each gaze point, the pupil sizes of both eyes were averaged into 50-ms time bins. Missing 50-ms time bins that were due to tracking errors were replaced with linear interpolation. The data were not interpolated if the missing data were at a visual transition point (from flashing to stable colors at 1300 ms or from moving boxes to outcome at 3330 ms) of if the data were missing due to a look away.

The average dynamic pupil response across the entire trial was computed on the initial learning trials for [tɑm] and [taːm] (trials 11 through 18). The pupil response on all away trials in blocks 3 and 4 was baselined by subtracting the average response to [taːm] on the initial learning trials in each 50-ms time bin from the pupil response at that point in the away trial[5]. The dependent variable of all pupil analyses was the attention allocation on away trials to words with typical and atypical vowel sounds in reference to each infant's attention to the initial [taːm] trials.

The data were analyzed using multi-level modeling (MLM). Each age group was analyzed separately, but in order to facilitate comparison across the age groups, a specific effort was made to fit the same equation to each age group's data. For the RT analyses, the vowel sound that the participant heard (typical [ɑ] or [aː], or atypical [ɑː] or [a]) was included in the equation. The participant was the subject level variable and sequence number, which refers to the 1st, 2nd, 3rd, etc... time that the participant was tested on an away trial, was included as the repeated measure. These models were fit using an identity covariance structure.

The pupil data were analyzed separately for each of the three phases. Pupil dilations can reflect processing of the stimulus or conflict in decision (Aston-Jones and Cohen, 2005). In the present task, pupil dilations during the first two phases were thought to reflect stimulus processing. Pupil dilations during the choice phase were interpreted to reflect choice conflict. In the pupil analyses, trial and timing within the trial were modeled as a two-level repeated measure. Those continuous independent variables were preferably modeled as a random factor because these factors will not necessarily have identical effects among subjects. Therefore, it is more conservative to model them as a random effect whenever possible. If the model did not converge, timing within the trial was modeled as a fixed factor and this will be mentioned explicitly. For the 15-month-old children additional analyses were conducted to assess the relation between the outcome measures and the raw receptive-vocabulary score from the N-CDI (henceforth: CDI-score).

4.3 Results

The effect of the initial training trials with /tibi/ and /drukəl/ was assessed. Across all age groups, it was clear that the training trials af-

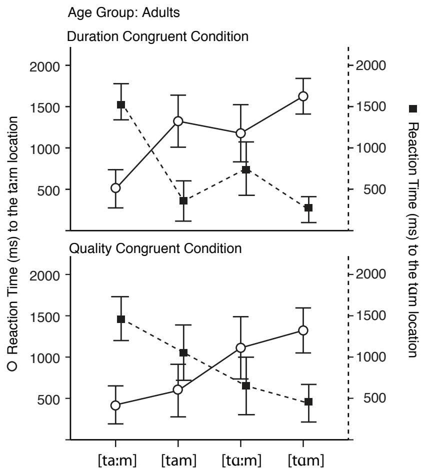
5 [taːm] was chosen as the baseline because [aː] is the more peripheral vowel in the Dutch vowel space and may therefore serve as the referent in perception (Polka and Bohn, 2003, 2011)

fected the RT measures as well as the pupil measures. All participants had to associate /tibi/ with the left side of the screen. In one situation, which we refer to as the duration-congruent condition, the shorter words /tibi/ and [tɑm] were associated with one side of the screen, and the longer words /drukəl/ and [taːm] were associated with the other side. In the other situation, which we refer to as the quality-congruent condition, the words with a front vowel, /tibi/ and [taːm], were associated with one side of the screen, and the words with a back vowel, /drukəl/ and [tɑm], were associated with the opposite side of the screen. To account for the effect of the training condition, a main effect of training condition and an interaction between condition and vowel sound were included in all the analyses. As the research question concerned the vowel sounds, only the main effects of vowel sound and the interactions between condition and vowel sound were interpreted.

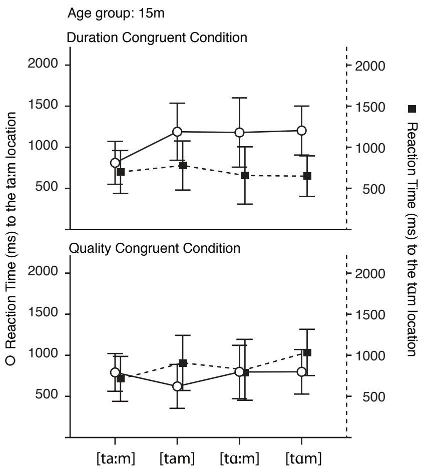
A subset of infant participants appeared to have a bias for one side, the same substantive result patterns were found in the data with and without those side-biased infants. Only the analyses without the side-biased infants are reported.

Group	Effect	RTs to [ɑ]-outcome AOI			RTs to [aː]-outcome AOI		
		F	df	p	F	df	p
A	Int	300.50	1,278	<.001	397.20	1,278	<.001
	C	5.97	1,278	.015	6.16	1,278	.014
	V	36.46	3,278	<.001	22.77	3,278	<.001
	V∗C	3.70	3,278	.012	2.14	3,278	.096
15	Int.	222.00	1,227	<.001	307.50	1,227	<.001
	C	2.74	1,227	.099	10.61	1,227	.001
	V	0.50	3,227	.681	0.91	3,227	.435
	V∗C	0.70	3,227	.551	1.39	3,227	.246
9	Int.	390.60	1,299	<.001	290.00	1,299	<.001
	C	1.49	1,299	.224	0.26	1,299	.608
	V	0.14	3,299	.939	2.15	3,299	.094
	V∗C	1.07	3,299	.362	0.30	3,299	.827

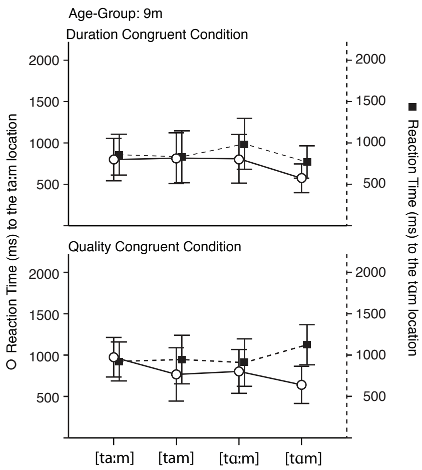
Table 12: **Analysis of reaction times.** Fixed effects (Int.=Intercept, C=Condition, V=Vowel sound) for reaction times to the /ɑ/-outcome AOI (left columns) and the /aː/-outcome AOI (right columns) from the final MLMs fit to each age-group (A=Adults, 15=15-month-olds, 9=9-month-olds).



(a)



(b)



(c)

Figure 10: **Mean reaction times** to the [aː]-outcome AOI (left y-axis, solid line, white circles) and the [ɑ]-outcome AOI (right y-axis, striped line, black squares) in the duration-congruent condition (top graphs) and the quality-congruent condition (bottom graphs) in the Adults (subfigure a), 15-month-olds (subfigure b) and 9-month-olds (subfigure c). Reaction times are given for (from left to right): typical [taːm], atypical [tam], atypical [tɑːm], and typical [tɑm]. The means are the means over participants and the error bars give 95% CIs for the mean RT to that sound.

4.3.1 *RT analysis*

4.3.1.1 *Adults – RT analysis*

There was a significant vowel by condition interaction for adults' RTs to the [ɑ]-outcome AOI ($F[3, 278] = 3.70, p = .012$), as can be seen in Table 12. Figure 10a shows adults' reaction times to the [ɑ]-outcome AOI in reaction to each of the four test words. Adults in both conditions were significantly faster to look at the [ɑ]-outcome AOI on trials with the typical vowel sound [ɑ] than on trials with [aː].

The atypical vowel sound [ɑː] patterns with typical [ɑ] for adults in both conditions. However, adults in the duration-congruent condition had a slower RT to the [ɑ]-outcome location when hearing [ɑː] than those in the quality-congruent condition. The atypical vowel sound [a] patterns with [ɑ] for adults in the duration-congruent condition, but with [aː] for adults in the quality-congruent condition.

These results suggest that adults readily categorized the typical vowel sounds [ɑ] and [aː]. Their categorization of the atypical vowel sound [ɑː] was consistent across the training conditions with both groups categorizing it as /ɑ/. This shows that adults relied on vowel quality to categorize [ɑː]. However, their categorization of the atypical vowel sound [a] depended on their training history. Adults thus did not rely automatically on vowel quality for their categorization of [a].

4.3.1.2 *Infants – RT analysis*

There were no significant differences in the 9- and 15-month-olds' RTs to the vowel sounds or significant interactions between the conditions and vowel sounds. The results of this analysis, which are shown in Table 12, do not show that infants were able to form associations between the typical words [tɑm] and [taːm] and the outcome locations. The mean RTs in Figures 10b and 10c show that the infants' RTs to the outcome locations were not significantly different for the four test words.

4.3.2 *Pupil analysis*

4.3.2.1 *Adults – pupil analysis*

For the analysis of adults' pupils when the first and second sound were played, time was included as a fixed effect. For the analysis of the choice phase, when the third sound was played, time was included as a random effect. The results from these analyses are given in Table 13 and adults' attention allocation per phase and condition is displayed in Figure 11(a).

During the first phase there was a significant effect of vowel on adults' pupil responses ($F[3, 196.50] = 3.43, p = .018$). While this

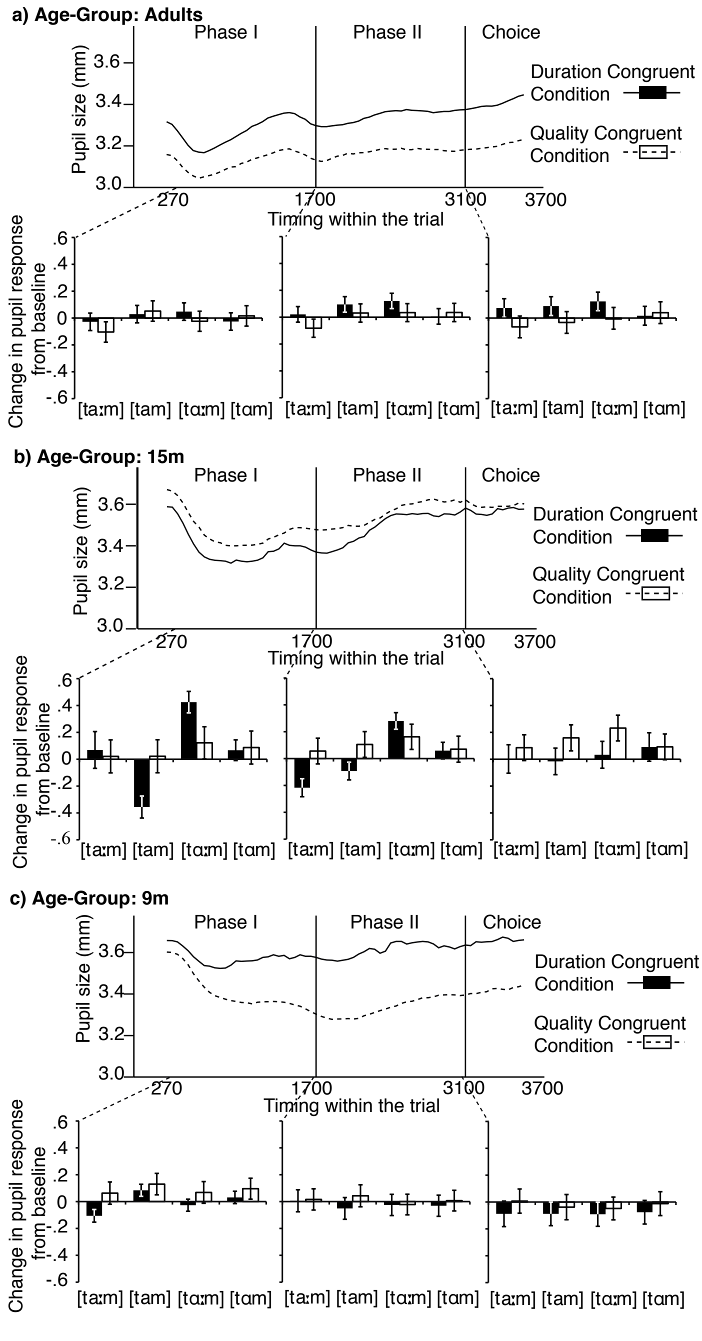


Figure 11: **Mean pupil dilations** averaged over all trials to demonstate the average dynamic response to the task (top figures). The mean pupil response, baselined to the initial [taːm]-trials, averaged for phase I, phase II, and the choice phase (bottom figures). The dashed lines represent the beginning of each phase in the trial. Results are separated for the duration-congruent condition (black bars, solid line) and the quality-congruent condition (white bars, striped line) and given for (from left to right): typical [taːm], atypical [tam], atypical [tɑːm], and typical [tɑm]. Results are reported separately for Adults (subfigure a), 15-month-olds (subfigure b) and 9-month-olds (subfigure c). The reported means are the estimated means from the analyses with time as a random factor. The error bars give 95% CIs. For these error bars, the points where the error bar does not cross the horizontal line are significantly different from the pupil dilations in the baseline trials with [taːm].

.

Group	Effect	Phase I F	Phase I df	Phase I p	Phase II F	Phase II df	Phase II p	Choice Phase F	Choice Phase df	Choice Phase p
A	Int.	8.60	1, 349.68	.004	0.54	1,1369.91	.463	2.17	1, 39.31	.149
	C	0.81	1, 196.50	.371	6.25	1, 211.99	.013	4.91	1, 39.31	.033
	V	3.43	3, 196.50	.018	4.67	3, 211.99	.004	1.36	3, 176.47	.256
	V∗C	1.65	3, 196.50	.180	1.72	3, 211.99	.164	3.75	3, 176.47	.012
	(T)	28.59	1,3861.79	<.001	0.442	1,3167.37	.506			
15	Int.	6.40	1, 168.90	.012	9.02	1, 165.15	.003	10.60	1, 28.44	.003
	C	0.07	1, 168.90	.794	6.66	1, 165.15	.011	4.85	1, 28.44	.036
	V	30.11	3, 186.58	<.001	22.09	3, 227.14	<.001	1.85	3, 132.04	.142
	V∗C	18.10	3, 186.58	<.001	10.88	3, 227.14	<.001	2.67	3, 132.04	.050
	(T)									
9	Int.	8.32	1, 242.71	.004	0.85	1,1122.54	.357	37.04	1,2178.44	<.001
	C	10.17	1, 242.71	.002	1.23	1, 237.52	.268	4.93	1, 207.70	.027
	V	6.80	3, 275.87	<.001	0.49	3, 237.50	.689	0.35	3, 207.69	.790
	V∗C	1.60	3, 275.87	.190	0.29	3, 237.52	.832	0.09	3, 207.70	.967
	(T)				3.61	1,5633.10	.058	41.00	1,2157.71	<.001

Table 13: **Analysis of pupil dilations.** Fixed effects (I=Intercept, C=Condition, V=Vowel sound) for pupil dilation from the final MLMs fit to each age group (A=Adults, 15=15-month-olds, 9=9-month-olds) for each of the three phases of the trial. Only if time (T) was included as a fixed instead of a random factor, the last row is filled.

shows that the response differed between the vowels, the pupil response for none of the vowel sounds significantly differed from baseline. During the second phase, there was a significant main effect of condition ($F[1, 211.99] = 6.25, p = .013$) and of vowel ($F[3, 211.99] = 4.67, p = .004$). Adults had a large pupil response to the two atypical vowel sounds [ɑː] and [a] relative to baseline. For the choice phase, there was a sound by condition interaction ($F[3, 176.47] = 3.75, p = .012$). Adults in the duration-congruent condition had larger than baseline pupils to the typical vowel [aː] and both atypical sounds. Adults in the quality-congruent condition did not have a pupil response that differed from baseline. Together, the results from all three phases suggest that adults needed more attention to process the atypical vowel sounds than the typical vowel sounds.

4.3.2.2 *15-month-olds – pupil analysis*

The results from the analyses on the 15-month-olds' pupil responses can be found in Table 13 and Figure 11(b). These results show that when the first word was played there was a significant vowel by condition interaction ($F[3, 186.58] = 18.10, p < .001$). Infants in both conditions had an increase in attention over baseline to the atypical vowel sound [ɑː]. Infants in the duration-congruent condition had a significantly smaller pupil than baseline to the atypical vowel sound [a], whereas the pupil response of infants in the quality-congruent condition to this sound was not significantly different than baseline. During the second phase, there was also a significant vowel by condition interaction ($F[3, 227.14] = 10.88, p < 0.001$). Infants in both conditions showed a larger than baseline pupil response to words with the atypical [ɑː]. Infants in the duration-congruent condition had a significantly smaller than baseline pupil response to the typical vowel sound [aː] and the atypical vowel sound [a]. Infants in the quality-congruent condition had a larger than baseline pupil response to the atypical vowel sound [a]. During the choice phase, the 15-month-old infants had a significant vowel by condition interaction ($F[3, 132.04] = 2.67, p = .050$). Only infants in the quality-congruent condition had a larger than baseline pupil response to the words with the atypical vowel sounds.

In both conditions, 15-month-olds showed increased attention allocation to the atypical vowel sound [ɑː] over baseline as compared to the typical vowel sounds, indicating that irrespective of the early training trials infants viewed this sound as unusual. This showed that infants reacted to the unusual combination of vowel quality and duration in this vowel sound and did not fully rely on either its familiar vowel-quality characteristics or its familiar duration characteristics. Whether the atypical vowel sound [a] was viewed as unusual depended on the infants' training history.

4.3.2.3 *9-month-olds – pupil analysis*

For the analysis of the 9-month-olds' pupils in the first phase time was included as a random effect. For the analysis of the second and the choice phase time was included as a fixed effect. The results from these analyses can be found in Table 13 and Figure 11(c).

The first phase showed main effects of condition ($F[1, 242.71] = 10.17, p = .002$) and vowel sound ($F[3, 275.87] = 6.80, p < .001$) on the 9-month-olds' pupil responses. For both conditions the 9-month-olds showed an increase over baseline in attention to the words with the atypical vowel sound [a]. There were no significant differences between the conditions or the vowel sounds in the second phase. In the choice phase, there was a significant difference between the conditions ($F[1, 207.70] = 4.93, p = .027$), but this factor did not interact

Effect	RTs to [ɑ]-outcome AOI			RTs to [aː]-outcome AOI		
	F	df	p	F	df	p
Int	14.06	1, 39.12	.001	39.16	1, 41.82	.001
C	1.97	1, 38.39	.168	6.23	1, 41.68	.017
V	0.40	3,207.39	.754	0.18	3,206.15	.910
V∗C	0.69	3,206.51	.557	1.56	3,205.41	.201
CDI	2.00	1, 42.31	.165	0.53	1, 44.48	.472
V∗CDI	0.31	3,207.92	.821	0.67	3,206.62	.571

Table 14: **Analysis on 15-month-olds' reaction times and CDI-score.** Fixed effects (Int.=Intercept, C=Condition, V=Vowel sound, CDI=CDI-score) for reaction times from the final MLMs relating 15-month-olds' CDI-score to their performance in the task.

Effect	Phase I			Phase II			Choice phase		
	F	df	p	F	df	p	F	df	p
Int	1.01	1,158.81	.317	1.55	1,160.15	.215	.01	1, 96.58	.911
C	0.61	1,159.12	.437	0.03	1,160.47	.856	2.87	1, 93.39	.094
V	1.21	3,159.48	.307	1.18	3,160.80	.321	2.74	3,148.68	.045
V∗C	9.43	3,168.49	<.001	9.14	3,169.52	<.001	2.44	3,147.80	.067
CDI	10.80	1,158.96	0.001	4.68	1,160.10	.032	1.36	1, 96.34	.247
V∗CDI	6.30	3,162.13	<.001	6.07	3,163.37	.001	10.56	3,172.17	<.001

Table 15: **Analysis on 15-month-olds' pupil dilations and CDI score.** Fixed effects (Int.=Intercept, C=Condition, V=Vowel sound, CDI=CDI-score) for pupil sizes from the final MLMs relating 15-month-olds' CDI-score to their attention allocation in the task, for each of the three phases of the trial.

with vowel, nor was there a main effect for vowel. These results show that despite an initial increase in attention to the atypical vowel sound [a], there was no evidence that the 9-month-olds sustained their attention throughout the trial.

4.3.3 *15-month relation between CDI-scores and RTs and pupil sizes*

In order to investigate the relation between vocabulary size and vowel perception, the MLMs were run again on the data of the 15-month-olds, but with the CDI-score and the interaction between vowel sound and CDI-score entered as fixed effects. The results from these analyses are given in Tables 14 and 15.

For the analysis of RT, there was no significant interaction between vowel sound and CDI-score, nor was there a main effect of CDI-score on 15-month-olds' general RT. For the analysis of pupil dilation, vowel sound significantly interacted with CDI-score during the first phase ($F[3, 162.13] = 6.30, p < .001$). Infants with a higher CDI-score had a larger pupil response to the atypical vowel sound [ɑː] than those with a lower CDI-score ($\beta = 0.008, p = .020$). There was a negative relation between CDI-score and the pupil response to the atypical [a] ($\beta = -0.007, p = .044$). During the second phase, there was also a significant vowel sound by CDI-score interaction ($F[3, 163.37] = 6.07, p = .001$). As in the first phase, there was a positive relation between CDI-score and pupil response to the atypical vowel sound [ɑː] ($\beta = 0.008, p = 0.026$) and a negative relation between CDI-score and pupil response to the atypical vowel sound [a] ($\beta = -0.008, p = .042$). In the choice phase, the significant vowel by CDI-score interaction was maintained ($F[3, 172.17] = 10.56, p < .001$), with a positive relation between CDI-score and the pupil response to [ɑː] ($\beta = 0.006, p = .006$). In the choice phase, the negative relation between CDI-score and the pupil response to [a] was not significant, but there was a significant negative relation between CDI-score and the pupil response to the typical vowel sound [aː] ($\beta = -0.006, p = .006$).

4.4 Discussion

The central aim of this paper was to investigate whether infants' vowel categories are primarily defined by vowel duration, vowel quality, or the combination of vowel quality and duration. The results of the present study show that by 15 months of age, infants are combining these cues in their vowel representations, because they react differently to atypical than to typical combinations of vowel quality and duration. An unexpected finding was that both adults' categorization of and infants' attention allocation to the atypical combinations of vowel quality and duration were influenced by the experimental context. As will be explained later, the context only influenced the adults' and infants' interpretation of the atypical combination that was a possible but ambiguous vowel sound, but not their interpretation of the combination that was a very infrequent vowel sound.

Only the adults reliably predicted the outcome locations for the words with the typical vowel sounds [ɑ] and [aː] on away trials. The atypical vowel sound [ɑː], which had the vowel quality of /ɑ/ and duration of /aː/, was consistently categorized by the adults as /ɑ/. This finding shows that adults rely more on vowel quality, the cue that is generally found to dominate their perception of /ɑ/ and /aː/ (Van Heuven et al., 1986; Escudero et al., 2009a; Giezen et al., 2010), as well as their categorization of other vowels (Van Heuven et al., 1986). Vowel sounds like [ɑː] are infrequent in Dutch infant-directed speech (Chapter 3) and Dutch listeners give vowel sounds like [ɑː] low acceptability ratings (Van Heuven et al., 1986)[6]. The present results show that Dutch adults nevertheless consistently categorize [ɑː] as /ɑ/, which indicates that they use their default categorization strategy, reliance on vowel quality, to categorize this infrequent vowel sound.

Adults' categorization of the atypical vowel sound [a], with the duration of /ɑ/ and the vowel quality of /aː/, was dependent on the initial training trials with /tibi/ and /drukəl/. Adults in the duration-congruent condition categorized [a] as /ɑ/, whereas adults in the quality-congruent condition categorized [a] as /aː/. The effect of the training condition shows that adults rely on the cue that is favored by the context to determine how [a] should be categorized. The atypical vowel sound [a] is ambiguous between /ɑ/ and /aː/. In Dutch IDS, the phonemes /ɑ/ and /aː/ are both sometimes produced as the vowel sound [a]. Vowel sounds like [a] can be found as a realization of /ɑ/ in Northern Dutch (Adank et al., 2007) and in Amsterdam Dutch before some coronal codas (Faddegon, 1951) and as a realization of /aː/ before a stressed syllable (Rietveld et al., 2003). The adults' inconsistent categorization of [a] as both /ɑ/ and /aː/ is most likely due to the ambiguity of this vowel sound. Adult listeners thus take the context of the situation into account when categorizing this ambiguous vowel sound.

Although the 15-month-olds did not reliably categorize the typical cueing sounds [ɑ] and [aː], the infants did show evidence of combining vowel duration and quality in their perception by increasing their attention to the atypical cueing sound [ɑː]. Adults consistently categorized [ɑː] as /ɑ/, but there are good reasons to assume that adults recognize that [ɑː] is an infrequent vowel sound. Infants' increased attention allocation to [ɑː] shows that they similarly recognize that [ɑː] is uncommon in their language environment. If either vowel duration or quality had dominated infants' perception and vowel representations, infants would have recognized [ɑː] as familiar, either because it has the familiar vowel quality of /ɑ/ or because it has the familiar vowel duration of /aː/. Only if infants attended to both cues could they notice that these familiar vowel quality and duration characteristics were incorrectly combined in [ɑː], which is what was found. Therefore, these results confirm those in Chapter 3 by showing that by 15 months of age, Dutch infants have representations for /ɑ/ and /aː/ that involve vowel duration as well as vowel quality.

6 Informally, we observed in the exit interviews that participants more readily noted [ɑː] as a deviant vowel than they reported on [a] as sounding unfamiliar. In Dutch, the vowel sound [ɑː] marginally appears in loanwords from English (e.g. [mɑːstər], 'master') and in that respect forms a third infrequent phoneme category. Lengthening of /ɑ/, does not typically occur in Dutch and [ɑː] is therefore an unlikely realization of /ɑ/. In Amsterdam Dutch, vowel sounds that resemble [ɑː] can be a realization of /aː/ (Brouwer, 1989).

Just as the training conditions differentially influenced adults' *categorization* of [a], they differentially affected the infants' *attention allocation* to [a]. At the group level, the 15-month-olds in the duration-congruent reduced their attention to [a], whereas infants in the quality-congruent condition increased their attention to [a]. As the effect of the training condition on adults' categorization of [a] was a result of the linguistic ambiguity of this vowel sound, we hypothesize that the effect of the training condition on infants' attention allocation to [a] is also evidence of the infants' linguistic processing of the sound. These results show that if infants acquire a contrast that is signaled by the early acquired vowel-quality cue and the later acquired duration cue, they are able to combine vowel duration and vowel quality in their representations before turning one and a half years of age, and are sensitive to the relative frequency and ambiguity of atypical cue combinations.

The 9-month-old infants did not reliably predict the outcome locations. Although they allocated more attention to the atypical vowel sound [a] in the beginning of the trials, this was not sustained throughout the trials. From these results we cannot draw any conclusions about 9-month-old infants' representations of /ɑ/ and /aː/.

Importantly, the 15-month-olds' attention allocation to the atypical vowel sounds was related to the infants' vocabulary size: Infants with a larger vocabulary allocated more attention to [ɑː] and less attention to [a] than infants with a smaller vocabulary. Linguistically more advanced infants thus better recognize that [ɑː] is an infrequent vowel sound. The adult results showed that they recognize [a] as a potential realization of both /ɑ/ and /aː/. The finding that infants with a larger vocabulary react with less surprise to [a] shows that they have begun to recognize that [a] is a possible vowel sound in their language. These infants thus went beyond noticing the acoustic differences between the typical and atypical vowel sounds and reacted selectively to the atypical combinations of vowel duration and quality in [ɑː] and [a], which have a different linguistic status and frequency in their native language.

Several studies to date have reported a relation between infants' phoneme perception and language development.[7] Language-specific speech discrimination skills in the second half of infants' first year have been found to be related to later vocabulary size (Tsao et al., 2004; Kuhl et al., 2005; Rivera-GAxiola et al., 2005; Kuhl et al., 2008). Conboy et al. (2008) report a relation between speech perception and concurrent vocabulary size. To the best of our knowledge, the present study is the second result that indicates a concurrent relation between speech perception and vocabulary. Infants' speech perception skills are often measured in looking-time procedures, which tend to give

7 See for an overview the Individual Variability in Infancy project on sites.google.com//site/invarinf/

binary rather than continuous outcomes (Aslin and Fiser, 2005). Furthermore, such studies mostly test infants' perception of very typical exemplars, which is well established by the time infants start learning their first words (Kuhl et al., 1997; Polka and Werker, 1994). By using the continuous pupil dilation measure and testing infants' perception of atypical examples (cf. Tsao et al., 2004), the present study could reveal subtle relations between infants' perception of phonemes and their language development. Infants' speech perception skills and vocabulary size might be independently influenced by the amount of input they receive (cf. Huttenlocher et al., 1991). However, this finding also lends support to accounts of infant language acquisition that propose a tight connection between the development of these two skills (Boersma et al., 2003; Werker and Curtin, 2005; Kuhl et al., 2008).

Infants' inability in the present study to associate the stimuli [tɑm] and [taːm] with the two outcome locations cannot be due to their inability to discriminate between the vowels /ɑ/ and /aː/. Chapter 3 has found that 15-month-old Dutch infants can discriminate between /ɑ/ and /aː/ in a simple discrimination task, and the change in infants' attention allocation to atypical [ɑː] and [a] reveals fine-grained sensitivity to the possible realizations of these vowels. The present results therefore confirm once more that it is difficult for infants to use their speech perception abilities to learn arbitrary audio-visual associations (cf. McMurray and Aslin, 2004; Kovács and Mehler, 2009; Albareda-Castellot et al., 2011). Possibly the most important of such arbitrary audio-visual associations that infants must acquire are word-object associations. Infants of 14 months old can discriminate between [bɪ] and [dɪ] in a speech discrimination task, but have difficulties using this ability in a word learning task with the minimal pair [bɪ] and [dɪ] (Stager and Werker, 1997; but seeYoshida et al., 2009). It has been proposed that infants' limited processing capacaties prevent them from listening carefully to the shape of the speech sounds when they have to form word-object associations (Werker et al., 2002; Fennell and Waxman, 2010; Fennell, 2012). In the present two-alternative categorization task the pupil dilations revealed that infants were processing the speech sounds in detail and in accordance with the distribution of such speech sounds in their language environment. Therefore, infants' difficulties with forming audio-visual associations must not be automatically ascribed to their inability to listen to the exact shape of the speech sounds. Rather, the present results suggest that infants always listen to the details of speech sounds and relate these to their emerging phoneme representations.

4.5 Summary

In this study we have shown that Dutch infants of 15 months old associate their vowel categories of /ɑ/ and /aː/ each with a combination of vowel quality and duration. Infants furthermore react differently to the infrequency of one atypical token, namely [ɑː], and the ambiguity of a second atypical token, namely [a]. This detailed insight in infants' category structure could only be obtained in a task that included typical as well as atypical category examples, and tested infants' recognition both in overt behavior and unconscious attention allocation.