Speech Perception in Children with a Cleft Palate: Preliminary Data

Paula Fikkert and Imme Lammertink

1. Introduction

Research into how speech production and perception are related has a long history. Proponents of the motor theory of speech perception argued in the sixties that speech sounds are perceived in terms of the articulatory features or gestures required to produce those sounds (Liberman et al*.,* 1967; Liberman, 1996). The prediction was that infants, who do not yet produce speech sounds, would have difficulty with speech perception. Testing this claim set of a vibrant research field on infants' speech perception. Contrary to the motor theory of speech perception, the seminal study by Eimas et al. (1971) showed among others that (categorical) speech perception was part of the biological endowment for language and that production was not necessary to perceive speech. When experimental work showed that infants were able to discriminate virtually all speech contrasts they were tested on, the theory lost in popularity. Nowadays, the common assumption is that speech perception precedes and even predicts speech production (e.g. Kuhl, 2004, Tsao, Liu & Kuhl, 2004).

Recently, however, there is renewed interest for the relationship between speech perception and production. The discovery of mirror neurons and the identification of active motor areas involved in both speaking and listening to speech (Wilson, Saygin, Sereno & Iacoboni, 2004) suggests that speech perception and production are closely linked. Imada and colleagues (2006) demonstrated that infants' motor and premotor areas in the cortex are activated when listening to speech, just as in adults, (Hickok & Poeppel, 2007). More recently, Bruderer et al. (2015) argued that sensorimotor information from the articulators influences speech perception. In their study six-month-old infants were tested on the same dental-retroflex contrast as used in their earlier 1984

* Paula Fikkert (corresponding author), Radboud University Nijmegen, Centre for Language Studies and Department of Dutch Language and Culture, PO Box 9103, 6500 HD Nijmegen, The Netherlands, p.fikkert@let.ru.nl; Imme Lammertink, The University of Amsterdam, Amsterdam Center for Language Studies, I.L.Lammertink@uva.nl. We would like to thank Ronald Admiraal for collaborating, and Lisa Cheng, Claartje Levelt and Capi Wever for their inspiration; Angela Khadar for help with running the subjects, and all parents and children who participated in this study. We would also like to thanks the audience at BU for their inspiring questions and sharp remarks.

 Proceedings of the 40th annual Boston University Conference on Language Development, ed. Jennifer Scott and Deb Waughtal, 130-142. Somerville, MA: Cascadilla Press.

study (Werker & Tees, 1984). In their earlier work Werker and colleagues showed that both Hindi and English six-month-olds are able to discriminate the contrast, and that English infants only lose this ability at the end of their first year. Bruderer and colleagues added two conditions to the original replication to test for the link between speech production and perception. In one condition they used a pacifier that inhibited the movement required to make a retroflex; in the other condition the pacifier did not inhibit that movement. Precisely in the condition in which infants were experimentally 'impaired' with the pacifier to make the contrast, infants failed to discriminate the contrast. The authors concluded: "before infants are able to speak, their articulatory configurations affect the way they perceive speech, suggesting that the speech production system shapes speech perception from early in life" (Bruderer et al, 2015).

A number of other studies have established correlations between early speech perception (discrimination, segmentation or word recognition) and later speech production (often production scores on parental checklists such as the Communicative Development Inventories (CDI) (Fenson et al. 2007), standardized language tests (e.g. Tsao, Liu, Kuhl, 2004) or children's own preferential production patterns in babbling and first words (e.g. Majorano, Vihman & DePaolis, 2014). These correlations suggest for a link between speech perception and production. It remains however difficult to assess the precise relationship. These studies all concerned infants, and it remains to be seen to what extent these results extend to later ages. It also remains unclear how perception and production are related when it concerns recognition of words stored in the mental lexicon. To recognize words the perceived features, whether acoustic, articulatory, or in terms of motor plans need to be matched to the (phonological) features that make up the phonological representation of a word. Some have argued that representations contain detailed acoustic features; others have argued that they are stored with articulatory gestures to aid production; yet others have argued that the features stored in mental representations are abstract phonological features.

The current study aims to gain a better understanding of the relationship between language perception and production by comparing word recognition in typically developing children with children with a cleft palate. Children with a cleft palate have an articulation apparatus that is impaired, making the articulation of certain sounds hard or impossible. Children with a cleft palate produce fewer stops in general and fewer (labial and) coronal stops in particular (Howard & Lohmander, 2011, Hutters & Brønsted, 1987; Harding & Grunwell, 1996; Chapman, Hardin-Jones & Halter, 2003). For this reason we studied the perception of voiceless labial and coronal stops. We used a looking-while-listening procedure using a mispronunciation paradigm, replicating Van der Feest and Fikkert (2015). We tested children's recognition of well-known words that were either correctly pronounced, or mispronounced, following the paradigm in Swingley and Aslin (2000). The mispronunciations either changed a target coronal stop /t/ to a labial stop /p/, or vice versa.

In the original study, Dutch children of 20-, and 24-months-old looked significantly longer to correct pronunciations than to mispronunciations. This effect was due, however, to the labial condition: a mispronounced labial target hindered word recognition in typically developing children, but a mispronunciation of coronal targets did not. We argued that coronal is the default place of articulation, which shows both more variation in the target language (as it is often involved in assimilation processes), as well as in children's own productions, as coronal targets are often subject to 'consonant harmony' (Fikkert & Levelt, 2008). In the current study, we were interested to see whether children with a unilateral cleft palate and without hearing problems, would show more difficulty in the recognition of well-known words that start with target forms that they find difficult to produce (coronal and labial stops). All children have undergone surgery and have (had) speech therapy, although currently we have no insight into their further development at this time. The prediction is that if production affects the perception of those sounds children with a cleft palate will perform significantly worse on speech recognition tasks than typically developing children (control group).

2. Methods

2.1. Participants

Data collection is still on going, but so far 103 native Dutch-speaking children participated in the experiment. Of these twenty-four children were excluded due to interference by parents (three children; all 18-month-old controls) or inattention to the screen during the experiment (20 children, see section 2.5 Data pre-processing). As a result, eighteen 18-month-olds with a cleft palate (Female = 5, range = 16.2 months – 19.18 months), twenty-three age-matched controls (Female = 10, range = 17.97 – 19 months), fifteen 30-month-olds with a cleft palate (Female = 5, range = 25.20 - 32.15 months) and twenty-four age-matched controls (Female = 11, range = 30.29 – 30.9 months) were included in the final analysis. Cleft palate children were recruited via *Center for Schisis and Craniofaciale afwijkingen* team of Radboud University Medical Centre. All Cleft palate children had a non-syndromic cleft and were reported to have no significant hearing of vision problems. For this study we can therefore only use a small subset of the population, as many children with a cleft palate suffer also from hearing problems, or additional health problems. Age-matched controls were recruited via the database of the Baby Research Center (BRC) in Nijmegen. Again, no hearing or vision problems were reported.

2.2. Apparatus

All children were tested at the BRC in Nijmegen. The data were obtained with a 17-inch Tobii T120 eye-tracker (Tobii Technology AB; binocular infrared light reflection, 120 Hz sampling frequency, accuracy: 0.5^{o}, recovery is typical 300ms). Eye-tracker calibration and stimulus presentation were controlled by Tobii studio software. Audio speakers were placed at either side of the screen, hidden from participant view. Participants sat approximately 60cm from the monitor, with toddlers sitting on their parent's lap.

2.3. Procedure

Each session began with a 9-point infant-friendly calibration procedure. Data collection started when good calibration for both eyes was obtained for at least five locations on the screen (every corner and the center). Children then participated in a mispronunciation detection task using a looking-while-listening procedure (Swingley & Aslin, 2000). The task comprised of thirty-two trials (12 filler trials and 20 target trials; see section 2.4 Stimuli). After eleven, nine and twelve trials respectively, children saw a 4-9s animated video (a running chick, a train, and a skating dog). The experiment took 4 minutes in total. Two versions of the experiment were created, with trial order reversed. Children were randomly assigned to one of the two versions.

2.4. Stimuli

The experiment consisted of twenty target, and twelve filler trials. Target trials fell in one of the following four conditions: (1) Labial correct pronunciation; (2) Labial mispronunciation; (3) Coronal correct pronunciation; or (4) Coronal mispronunciation. The labial targets were *poes* [pus] ('cat') and *pop* [pɔp] ('doll'), either correct pronounced (CP) or mispronounced (MP; Table 1). Mispronunciations involved a change in place of articulation of the initial labial consonant [p] into a coronal consonant [t]; resulting in *toes* [tus] and *top* [tɔp] respectively. Coronal targets were *tand* [tɑnt] ('teeth') and *teen* [ten] ('toe'), again either correct pronounced or mispronounced. Coronal mispronunciations involved a change in place of articulation of the initial coronal consonant [t] into a labial consonant [p], resulting in *pand* [pɑnt] and *peen* [pen] respectively. The resulting mispronunciations were either non-words or low frequent real words unknown to children of this age. Correct and mispronounced targets from each condition were evenly distributed across test items so that each correct target was presented three times (resulting in 6 CP labials and 6 CP coronals) and each mispronounced targets were presented two times (resulting in 4 MP labials and 4 MP coronals). Filler items were the same across the four conditions and included a baby, a shoe, a cow, a duck, a bike and a car, each presented twice. The fillers were added to keep the children interested in the video and to prevent sequence effects that might be induced by repeated presentation of the targets.

The visual stimuli consisted of test pairs displayed next to each other on a white background. The trials always showed the same pair of test items and each test item served as a target in half of the test trials and as distracter in the other half of the test trials. The side at which the target and distracted appeared was balanced over trials.

A female native speaker of standard Dutch recorded the audio stimuli in a sound-attenuated booth. The speaker was asked to read aloud in an infant-directed register. All trials started with 2.5 seconds during which the pictures were showed in silence. The the target word was presented embedded in the carrier phrase "*kijk naar de [target]!*" ("Look at the [target]!"). The carrier phrase was followed by a 750ms silence and a second sentence which was either "*leuk he?*" or "*mooi he?*" ("Nice, isn't it?"). Fillers were embedded in carrier sentences like "*Kun je de [filler] vinden? Vind je hem leuk?*" (Can you find the [filler]? Do you like it?") or *"Waar is de [filler]? Kun je hem vinden?"* ("Where is the [filler]? Can you find it?").

Table 1: Correct pronunciations and mispronunciations of the test words.

Place of articulation	CP	MPplace
Labial	1. [p]oes, [p]op	2. [t]oes, [t]op
Coronal	3. [t]and, [t]een	4. [p]and, [p]een

2.5. Data-pre processing

Before analyzing children's gaze behavior, the raw dataset was pre-processed to remove unreliable tracker output and to prepare the gaze measurements for the right analysis. We only counted participants' gaze measurements when the Tobii output marked the look as valid (validity score >= 2) in at least one eye. Individual target trials were excluded if the participant looked anywhere at the screen for less than 25% in the last second before onset of the target word (-1000-0ms) or following word onset (500-1500ms). If more that 10 target trials (50%) of a participant were removed, the participants' data was completely excluded from analysis. In total twenty participants were excluded by this criterion (8 eighteen-months-cleft palate, 5 eighteen-months-control, 4 thirty-months-cleft palate, 3 thirty-months-control). From the remaining participants 270 (17%) target trials were excluded.

3. Results

The complete pre-processed datasets for the 18-month-olds (Cleft palate and control together) and the 30-month-olds (Cleft palate and control together) were analyzed using generalized linear mixed effect models (lme4; Bates et al., 2012) in the statistical programming language R (R Development Core Team, 2013).

3.1. Pre-analysis: naming effect

A first step in the analysis was to check for a naming effect in the correct pronounced targets for the 18-, and 30-month-olds separately. As common in studies with a looking-while-listening paradigm, gaze data was averaged over a given pre-, and post-naming window. Average looking time to the target in the pre-naming window is compared to average looking time to the target in the post-naming window. A significant increase in looking time towards the target between pre-, and post-naming window is assumed to be evidence for a naming effect. In the current study, the pre-window is defined as the 2000-0ms preceding word onset. The post-naming window is defined as the 500-1500ms following word onset. The latter is comparable to the window used in a similar looking-while-listening study with 18-month-old Dutch toddlers by Tsuji and colleagues (2016). To assess the presence of a naming effect for both age groups, two separate models were fit. The models were fit on correct pronunciations only. The dependent variable was the average looking time to the target. Predictor variables included *window* (pre vs. post) and *group* (cleft palate vs. control). The predictor variables were contrast coded (-1; 1) and the intercept was allowed to vary by subject and item.

In both age groups the average looking time to the target was larger in the post- compared to pre-window (18m: 146816 observations, N = 41, $\beta = .077$, $z = 8.37$, $p < 0.001$; 30m: 133226 observations, N = 39, $\beta = .25$, $z = 24.48$, $p < .001$). These results indicate the presence of a naming effect for both ages.

3.2. Effect of mispronunciation: 18-month-olds

As a naming effect was established with our first analysis, we looked for effects of mispronunciation in the post-window (500 – 1500ms from target offset) only. It was hypothesized that mispronouncing the target words hinders word recognition and therefore that the average looking time to the target would be smaller in the MP versus CP condition. A linear mixed effect model, including both correct, and mispronounced targets, but restricted to the 18-month-olds and post-window only, was fit to assess the effect of a mispronunciation on word recognition in both typically as well as cleft palate children (82476 observations, $N = 41$, Figure 1). The predictor variables were *condition* (MP vs. CP) and *group* (cleft palate vs. control). Again, the predictor variables were contrast coded (-1; 1) and the intercept was allowed to vary by subject and item. Model coefficients showed a significant interaction between *group* and *condition* ($\beta = 0.020$, $z = 2.60$, $p < .01$). Visual inspection of the data (Figure 1) suggests that the interaction between *group* and *condition* comes from a difference between the groups of children, in which control children are hindered by the mispronunciation (larger difference between CP and MP condition) whereas cleft palate children are not. To test this explanation statistically, we fitted two models for the control and cleft palate children separately. The model restricted to data from the control children (50066 observations, $N = 23$) showed a significant effect of *condition,* with more looks to the correct pronounced than the mispronounced target ($\beta = .18$, $z = 2.11$, $p < .05$). These results show that 18-month-old controls look less to the target when it is mispronounced compared to when it is pronounced correctly. This suggests that a mispronunciation hinders word recognition in these children. No other coefficients reached significance. Contrary, the model restricted to data from the 18-month-old cleft palate children (32410 observations, $N = 18$) showed no significant coefficients at all. Thus, children with a cleft palate did not look less to mispronounced targets than correct pronounced targets. This finding suggests that mispronunciation does not hinder word recognition in children with a cleft palate.

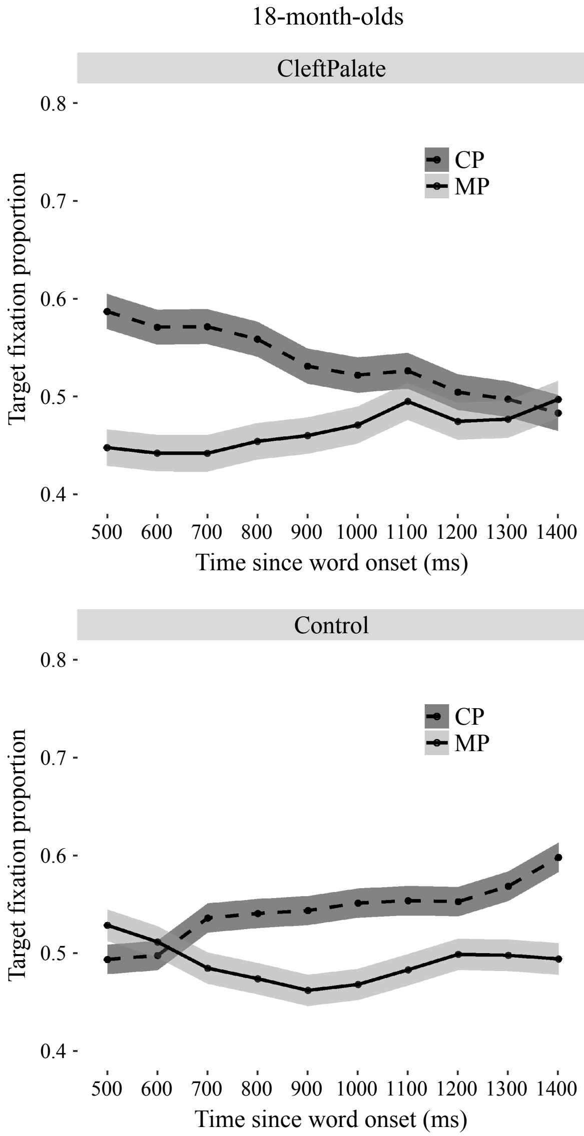

Figure 1. Time-course of proportion target fixation after word onset for correct pronunciation (CP, dark grey and dashed lines), and mispronunciation (MP, light grey and solid lines), separately for 18-month-olds CleftPalate (top) and Control (bottom). Shaded areas indicate ±1 SE.

3.3. Effect of mispronunciation: 30-month-olds

To assess whether word recognition is hindered in 30-month-old children with a cleft palate and their typically developing peers by a mispronunciation of the targets, similar analyses as with the 18-month-olds were performed. First a model including data from both groups together was fit (72046 observations, N=39, Figure 2). No main effect of *condition* or *group* was found. The model showed, however, a significant interaction between *group* and *condition* ($\beta = .26$, $z = 7.50$, $p < .001$). Similar as for the 18-month-olds, visual inspection of the data (Figure 2) suggests that the interaction between *group* and *condition* comes from a difference between the groups of children, in which control children are hindered by the mispronunciations whereas Cleft palate children are not. Again, we fitted two models for the control (45710 observations, $N = 24$) and cleft palate (26336 observations, $N = 15$) children separately to test for this explanation statistically. No significant effect of *condition* was found in either the control or cleft palate model. Therefore our explanation for the interaction was not verified with the statistical model.

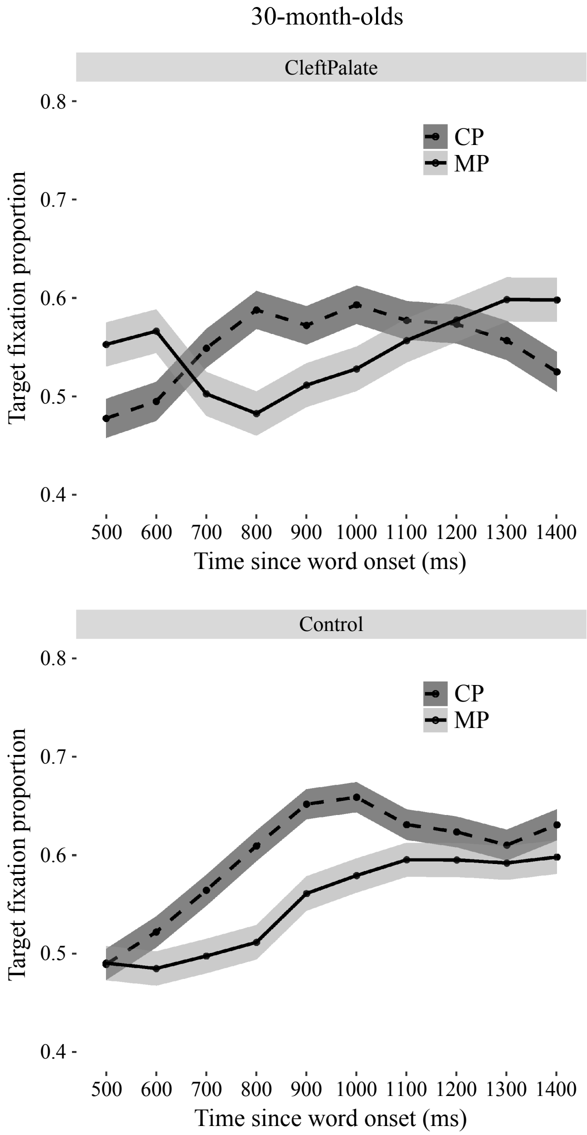

Figure 2. Time-course of proportion target fixation after word onset for correct pronunciation (CP, dark grey and dashed lines), and mispronunciation (MP, light grey and solid lines), separately for 30-month-olds CleftPalate (top) and Control (bottom). Shaded areas indicate ±1 SE.

4. Discussion and Conclusions

In this study, we tested whether children with a cleft palate, who have difficulties producing coronal and labial stops, perform differently in a word recognition task in which we compared looking times to correct versus mispronunciations of well-known words. Our prediction was that the inability to produce labial and coronal stops would affect word recognition and that children with a cleft palate would perform worse than the control children.

For the control children we found that mispronunciations hinder word recognition, as they show significantly longer looking times to the correct picture when the word is pronounced correctly versus when it is mispronounced. However, the children with a cleft palate showed no difference in looking time in the two conditions.

There could be several explanations for this finding. First, it could be that children with cleft palate have problems hearing the difference between /p/ and /t/ because of impairments to their articulatory apparatus. In the case, this contrast would also not be well represented in their mental representations of words either. Although all children have been undergone surgery and received speech therapy, the differences still persist in the 30-month-old children. In our study we only tested well-known words, which presumably have been acquired while the articulatory deficit still existed. Thus, the early-acquired words may have representations that are not as robust and detailed, as those of the control group. This raises a number of questions for further research. Do these children catch up with the control group when their articulatory apparatus is restored and they are able to produce the relevant sounds? Another question is whether early-acquired words that have been learned while the articulation was impaired are treated differently from late-acquired words, i.e. words acquired after the surgery and speech therapy? In that case, the difference may be due to the nature of their phonological representations.

Another question is whether the cause of the perceptual difficulties lies in the nature of their perception of the input they hear, or whether it is the lack of articulatory feedback, both auditory, as their own output also functions as input, and in terms of proprioception. After all, their early productions of target /t/ and /p/ will sound very different from the input sounds, and they are produced differently as well. Unfortunately, our study is unable to tease those two explanations apart.

Our study also raises another point. In the recent study by Bruderer and colleagues (2015) the tongue movement required to make a retroflex was artificially made impossible. Moreover, even without the pacifier, which made the articulation of the retroflex impossible, the 6-mont-old infants would probably not be able to articulate the retroflex. In other words, in their study there was no direct relation between experience with production and perception. In our study, the impairment was not related to tongue movement, but to problems with making the constriction required to make a coronal or labial stop. Our subjects were in all likelihood intending to produce the targets as well as they possibly could, but without reaching the adult target.

In our study, there were no interactions with place of articulation, suggesting that labials and coronals were treated the same way. This is surprising at least for the control subjects, as we have seen this in several earlier studies (e.g., Van der Feest & Fikkert, 2015; Dijkstra & Fikkert, 2011; Tsuji, et al. 2015, Altvater-Mackensen, Van der Feest & Fikkert 2014). There was however a trend for control children in the right direction suggesting that mispronunciations of coronal targets did not hinder word recognition. For the children with a cleft palate, the trend was in the opposite direction, suggesting that they treat coronal targets very differently than the controls. It is too early to make conclusions. We will continue testing, although given the strict selection criteria (we only include children with a cleft palate without hearing problems) it will take some time. Our study does seem to show that production shapes the perception of well-known words in children with a cleft palate.

References

Altvater-Mackensen, Nicole, Suzanne V.H. van der Feest & Paula Fikkert (2014). Asymmetries in early word recognition: the case of stops and fricatives. *Language Learning and Development* **10**(2): 149–178.

Bates, Douglas, Martin Maechler & Ben Bolker (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. http://CRAN.R-project.org/package=lme4.

Bruderer, Alison G., D. Kyle Danielson, Padmapriya Kandhadai & Janet F. Werker (2015). Sensorimotor influences on speech perception in infancy. *PNAS* **112** (44): 13531–13536.

Chapman Kathy L., Mary Hardin-Jones & Kelli-Ann Halter (2003). The relationship between early speech and later speech and language performance for children with cleft lip and palate. *Clinical Linguistics & Phonetics* **17**: 173–197.

Dijkstra, Nienke & Paula Fikkert (2011). Universal Constraints on the Discrimination of Place of Articulation? Asymmetries in the Discrimination of ‘paan’ and ‘taan’ by 6-month-old Dutch Infants. In: Nick Danis, Kate Mesh, and Hyunsuk Sung (Eds.), *Proceedings of the Boston University Conference on Child Development* 35. 170–182.

Eimas, Peter D., Einar R. Siqueland, Peter J. Jusczyk, & James Vigorito (1971). Speech perception in infants. *Science* **171** (3968): 303–306.

Fenson, Larry, Virgina A. Marchman, Donna J. Thal, Philip S. Dale, J.Steven Reznick, & Elizabeth Bates (2007). *MacArthur-Bates Communicative Development Inventories: User’s Guide and Technical Manual* - Second Edition. Baltimore: Brookes Publishing.

Fikkert, Paula & Clara C. Levelt (2008). How does place fall into place? The lexicon and emergent constraints in the developing phonological grammar. In: Avery, Peter, B. Elan Dresher & Keren Rice (Eds.), *Contrast in phonology: Perception and Acquisition*. Berlin: Mouton. 219–256.

Harding, Anne & Pamela Grunwell, (1996). Characteristics of cleft palate speech characteristics. *European Journal of Disorders of Communication*, **31**(4): 331–358.

Hickok, Gregory & David Poeppel (2007). The cortical organization of speech processing. *Nat Rev Neurosci*. **8**(5): 393–402.

Howard, Sara & Annette Lohmander (2011). *Cleft Palate Speech: Assessment and Intervention*. Wiley.

Hutters, Birgit & Kirsten Brøndsted (1987). Strategies in cleft palate speech—with special reference to Danish. *Cleft Palate Journal* **24**:126–136.

Imada Toshaki, Yang Zhang, Marie Cheour M, Samu Taulu, Antti Ahonen & Patricia K. Kuhl. (2006). Infant speech perception activates Broca's area: A developmental magnetoencephalography study. *Neuroreport* **17**(10): 957–962.

Kuhl Patricia K. (2004). Early language acquisition: Cracking the speech code. *Nat Rev Neurosci*. **5**(11): 831–843.

Liberman, Alvin M., Franklin S. Cooper, Donald P. Shankweiler & Michael Studdert-Kennedy (1967). Perception of the speech code. *Psychological review* **74** (6): 431–461.

Liberman, Alvin M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.

Majorano Marinella, Marilyn M. Vihman & Rory A. DePaolis (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development* **10**(2): 179–204.

R Development Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Swingley, Daniel & Richard N. Aslin (2000). Spoken word recognition and lexical representation in very young children. *Cognition* **76**: 147–166.

Tsao, Feng-Ming, Huei-Mei Liu & Patricia K. Kuhl, (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development* **75** (4): 1067–84.

Tsuji, Sho, Reiko Mazuka, Cristia, Alejandrina & Paula Fikkert (2015). Even at 4 months, a labial is a good enough coronal, but not vice versa. *Cognition* **134**: 252–256.

Tsuji, Sho, Paula Fikkert, Naoto Yamane & Reiko Mazuka (2016). Language-general biases and language-specific experience contribute to phonological detail in toddlers' word representations. *Developmental Psychology* **52**(3).

Van der Feest, Suzanne V. H. & Paula Fikkert (2015). Building phonological lexical representations. *Phonology* **32**: 207–239.

Werker Janet F. & Richard C. Tees (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav Dev*. **7**(1): 49–63.

Wilson, Stephen M., Ayşe Pinar Saygin, Martin I. Sereno & Marco Iacoboni (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience* **7** (7): 701–702.

Proceedings of the 40th annual Boston University Conference on Language Development

edited by Jennifer Scott
and Deb Waughtal

Cascadilla Press Somerville, MA 2016

Copyright information

ISSN 1080-692X
ISBN 978-1-57473-066-1 (2 volume set, paperback)
ISBN 978-1-57473-166-8 (2 volume set, library binding)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com