INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

On the role of temporal variability in the acquisition of the German vowel length contrast

Felicitas Kleber[1]

[1]Institute of Phonetics and Speech Processing, LMU Munich, Germany
kleber@phonetik.uni-muenchen.de

Abstract

This study is part of a larger project investigating the acquisition of stable vowel-plus-consonant timing patterns needed to convey the phonemic vowel length and the voicing contrast in German. The research is motivated by findings showing greater temporal variability in children until the age of 12. The specific aims of the current study were to test (1) whether temporal variability in the production of the vowel length contrast decreases with increasing age (in general and more so when the variability is speech rate induced) and (2) whether duration cues are perceived more categorically with increasing age. Production and perception data were obtained from eleven preschool, five school children and eleven adults. Results revealed that children produce the quantity contrast with temporal patterns that are similar to adults' patterns, although vowel duration was overall longer and variability slightly higher in faster speech and younger children. Apart from that, the two groups of children did not differ in production. In perception, however, school children's response patterns to a continuum from a long vowel to a short vowel word were in between those of adults and preschool children. Findings are discussed with respect to motor control and phonemic abstraction.

Index Terms: first language acquisition, quantity contrasts, production, perception, German

1. Introduction

It has been proposed that children acquiring English or German as their first language initially focus more on spectral than on durational cues when perceiving phonemic contrasts that are at least partially based on quantity [1]. This may be due to the fact that typologically these languages are not quantity languages, i. e. they heavily rely on non-durational cues in the phonetic distinction of the phonemic vowel length and other contrasts such as the voicing opposition ([2] for English; [3] for German). Duration, nevertheless, plays an important role which is why we refer to them here as quantity contrasts. In German, phonologically long vowels are longer in duration than phonemically short vowels [4] and voiced or lenis consonants, too, generally have a shorter duration compared to voiceless or fortis consonants.

From acquisition studies we know that adult listeners judge three-year-olds' productions of the German vowel length contrast as correctly produced [5, 6]. However, most acquisition studies are based on transcriptions of child speech, i. e. on auditory judgements, which make it difficult to assess whether these judgements come about due to a greater category separation along the spectral dimension or the durational dimension. Moreover, it remains unclear whether and, if so, to what extent children's productions differ from adults' productions. That is, children may produce the phonemic vowel length contrast in the minimal pair *lag* (/la:k/, 'sb./sth. lay') vs. *Lack* (/lak/, 'varnish') so that each word is recognizable as such for adult listeners but children may differ from adults in the phonetic implementation of the contrast [7].

There is ample evidence showing that children until the age of twelve vary from adult speakers in that their speech is characterized by more temporal variability ([8], [9]). For example, [10] found greater temporal variability in the speech of five- and eight-year-olds, although both age groups did not differ from adults in the general temporal patterns signaling the vowel length contrast. They interpreted their finding as supporting the so called *dissociation hypothesis* according to which children have acquired the representation of the contrast but the execution of articulatory timing is not yet fully developed, the former being reflected by a difference in vowel duration between long and short vowels, the latter by children's more variable duration.

Eventually, however, children need to fine-tune the language-specific temporal patterns [11] used to convey phonemic contrasts – in particular when the duration of a single segment can cue two phonemic contrasts. This is the case with vowel duration in English and German, which can signal both the vowel length and the postvocalic voicing contrast: that is, phonemically short vowels are not only shorter than phonemically long vowels, but even more so when they precede voiceless consonants; likewise, phonemically long vowels are lengthend when they precede voiced stops. Languages differ with respect to when these pattern are acquired: for example, children with Finnish as their first language acquire the geminate-singleton contrast in consonants earlier than children with Japanese as their first language [9].

The current study is embedded in a larger project investigating the acquisition of the vowel length contrast and the voicing contrast in German vowel (V) plus consonant (C) sequences. The specific aims of the current study were to test (1) whether temporal variability in the production of the vowel length contrast decreases with increasing age (in general and more so when the variability is speech rate induced) and (2) whether duration cues are perceived more categorically with increasing age as the phonemic contrast stabilizes during phonological development.

2. Production

2.1. Method

2.1.1. Speakers and Experimental Design

Eleven preschool and five school children took part in a picture naming task. The age range of the preschool children was from 5 years, one month (= 5;1) to 6;7 with a mean age of 5;7. School children's age range was from 7;7 to 9;5 with a mean age of 8;5. All children were born and raised in Munich, Germany, acquiring Southern Standard German.

Recordings were made using the SpeechRecorder software [12] (version 3.4.2a; with a 44.1 kHz sampling rate and a 16-bit

http://dx.doi.org/10.21437/Interspeech.2017-1282

resolution), a laptop computer, and mobile recording equipment (BeyerDynamic headset microphone, M-Audio audio interface) in a quiet room of a kindergarten and a day care center that the preschool and school children, respectively, attended.

All children were presented with the same 26 pictures of mono and disyllabic words which they had to name upon presentation. Each picture was repeated six times and shown in randomized order in six blocks each consisting of the 26 items (i. e. no repetitions per block). Additionally, children were asked to say the words in every second block as loudly as possible. This condition was introduced to investigate within speaker variation. Here loud speech refers to slower and more hyperarticulated speech [13]. We did not vary speech rate directly because for children it was easier to vary loudness. Thus, the database contains 20 (children) x 26 (words) x two conditions (normal vs. loud voice) x three repetitions = 3120 recordings.

A subset of the children's data was compared to the production data of a group of ten young adults (8 female) from the same regional area (Munich and surrounding areas) that were obtained for a previous study [14]. In this study speakers were asked to read out loud a total of 10 times 46 different words each presented in isolation and in randomized order on a computer screen in a moderate voice and tempo. The age range of the young adults was from 20 to 30 years.

2.1.2. Speech Materials

From the database described in 2.1.1 above we selected the following three word pairs with phonologically long and short vowels for the current analysis: *rote-Motte* (/roːtə/, 'red'; /mɔtə/ 'moth'), *Lupe-Suppe* (/luːpə/, 'magnifying glass'; /sʊpə/, 'soup'), *hacken-Haken* (/hakən/, 'to chop'; /haːkən/, 'hook').

From the database containing adult speech we chose the minimal pair *Hüte-Hütte* (/hyːtə/, 'hats'; /hʏtə/, 'hut'). We are currently recording more adult speakers producing exactly the same words the children produced to allow for more comparable data sets.

2.1.3. Analysis

Children's recordings were automatically segmented using WebMAuS [15] and saved as an EMU speech database [16]. In a second step, each recording file was checked by two labelers and segment boundaries were corrected whenever necessary, whereby a vowel's on and offset had to coincide with a clearly visible second formant (F2). In addition, the labelers marked each file with respect to the loudness condition, i. e. whether or not the child produced the utterance with a loud or normal voice. These labeling steps were carried out in the EMU Speech Database Management System [16]. Adults' data were already available as an EMU speech database.

Acoustic analyses from both databases were caried out in R [17] (version 3.3.2) using the emuR-package. Tokens were excluded from the analysis (1) when the condition (i. e. repetitions 1, 3, and 5 = normal voice vs. repetitions 2, 4, and 6 = loud) did not match the labelers' judgment of loudness (binary decision: loud vs. normal) and (2) when word duration exceeded 1100 ms (a word duration above this value was taken as indicating a disfluent utterance, cf. [10]).

The dependent measures analyzed in this study were absolute vowel duration and the so called *scaled covar* – a measure of temporal variability. Following the method described in [10] and [18], the *scaled covar* was calculated by dividing the standard deviation of the absolute vowel duration by the mean which was then multiplied by 100. Higher *scaled covar* values indicate greater temporal variability.

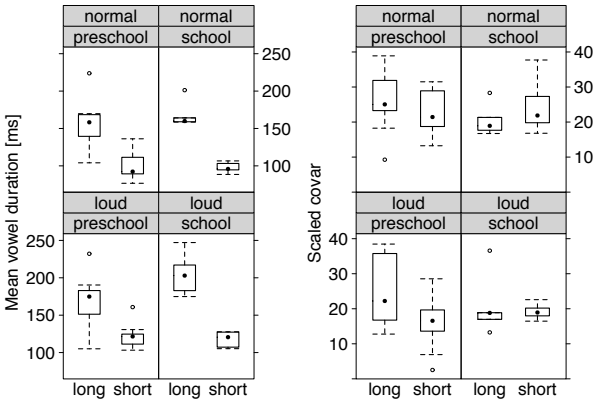

Figure 1: *Vowel duration in milliseconds (left) and temporal variability (right) shown separately for preschool and school children.*

Repeated measures ANOVAs were chosen for statistical analyses because some of the linear mixed models we ran with word and speaker as random factor did not converge. Prior to the actual analyses we therefore checked in two separate analyses with long-vowel-words and short-vowel-words, respectively, whether word (i. e. *rote* vs. *Lupe* vs. *Haken* and *Motte* vs. *Suppe* vs. *hacken*) had an effect on vowel duration, which was the case (long vowel words: $F[2, 24] = 16.2, p < .000$; short vowel words: $F[2, 24] = 23.4, p < .000$). However, since word pair did neither interact with style (loud vs. normal) nor age, we aggregated over word pair in the subsequent final analyses. Thus, in the actual analyses the only fixed factors were phonemic vowel length (within-subject factor with two levels), speaking style (within-subject factor with two levels) and age group (between-subject factor with two or three levels); speaker was entered as a random factor.

2.2. Results

2.2.1. Effect of speaking style in younger vs. older children

Commensurate with the data in the left panel of Figure 1, phonemic vowel length ($F[1, 12] = 112.1, p < .000$) and speaking style ($F[1, 12] = 60.8, p < .000$) significantly affected absolute vowel duration. There were no significant main effect for age or significant interaction effects between any of the fixed factors. These results suggest that preschool and school children realized the vowel length contrast in terms of absolute vowel duration and that they did so to the same degree. When speaking loudly, vowel duration of both long and short vowels increased compared to the normal speaking condition.

Speaking style was the only factor significantly affecting temporal variability ($F[1, 12] = 5.0, p < .05$) with overall more variability in the normal (i. e. faster) condition. Neither vowel length nor age affected temporal variability, although a trend towards slightly less variability in school children's normal voice condition can be observed in the right panel in Figure 1.

2.2.2. Comparison with adults

A subset of children's data containing words with alveolar stops (i. e. *rote-Motte*) produced in a normal voice was compared

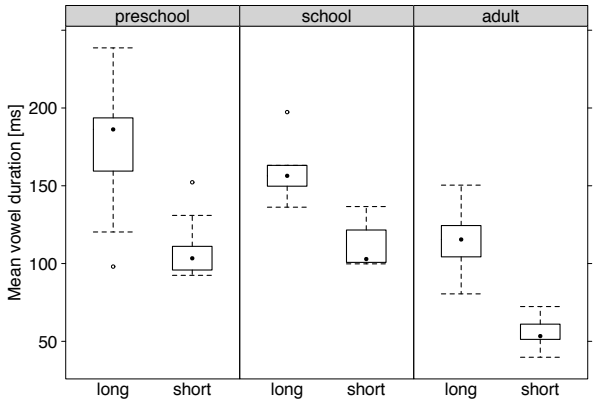

Figure 2: *Vowel duration in milliseconds shown separately for long and short vowels and for preschool children, school children, and adults.*

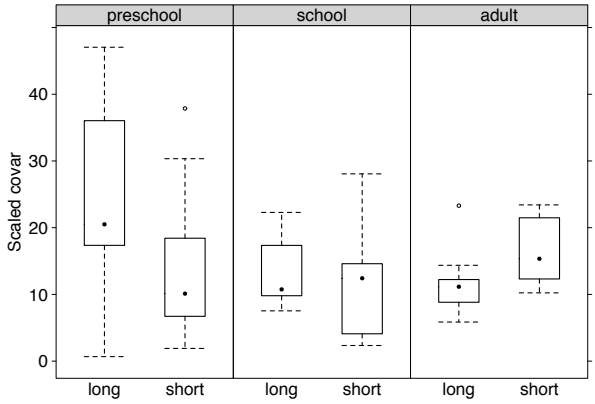

Figure 3: *Temporal variability shown separately for long and short vowels and for preschool children, school children, and adults.*

to the *Hüte-Hütte* tokens produced by adults. Commensurate with the absolute vowel duration shown in Figure 2, the repeated measures ANOVA revealed significant main effects for phonemic vowel length ($F[1, 22] = 119.8, p < .000$) and age ($F[2, 22] = 21.2, p < .000$) but no significant interaction between phonemic length and age. Phonologically long vowels had longer vowel durations than phonologically short vowels and overall vowel duration decreased with age. The difference between long and short vowels was about the same for all age groups, that is the temporal patterns with which the contrast was realized were similar across age groups. Note that some of the variance between both groups of children, on the one hand, and adults, on the other hand, may also be explained by the different words compared (i. e. words containing either back or front vowels), but the intermediate position of school children suggests that at least some of the variance is due to developmental differences.

Figure 3 shows again the temporal variability which was greater in the youngest age group but only for long vowels, as the significant interaction effect between phonemic length and age ($F[2, 22] = 3.8, p < .05$) suggests. None of the two fixed factors independently reached significance.

3. Perception

3.1. Method

3.1.1. Listeners

The same child participants who took part in the production study described in 2 above also completed a perception experiment. In addition, ten adult listeners, who did not partake in the production study described in 2, participated in this perception experiment. The age range of the adult listeners was from 18 to 27 years.

3.1.2. Stimuli and Experiment

The continnuum used for this perception experiment was taken from a previous study on the perception of the combined vowel length and voicing contrast in German [14] and slightly modified for the current study. Originally, the continuum spanned the words *Hagen* (/haːgən/, the name of a German city and a male name) to *Haken* to *hacken*. The parameter modified in this continuum was the proportional vowel duration (here and hereafter *VCratio*) as defined in (1)

$$VCratio = \frac{V_{dur}}{(V_{dur} + Clos_{dur})} \quad (1)$$

where V_{dur} is the duration of the vowel and $Clos_{dur}$ the duration of the postvocalic velar stop's closure phase. The stimulus specific *VCratios* are given in Table 1 (see [14] for further details on stimulus creation). *VCratio* has been shown to be a relevant perceptual cue both to the voicing [19] and the vowel length contrast [14] with *VCratios* above and below 0.5 being indicative of long and short vowels, respectively.

Table 1: *Grouping of stimuli along the continuum and VCratio values per stimulus number.*

Part of continuum	Stimulus number	VCratio
left	1	0.80
left	2	0.75
left	3	0.65
middle	4	0.60
middle	5	0.55
right	6	0.50
right	7	0.45
right	8	0.40

The aforementioned modifications of the stimuli were such that the first segment /h/ was cut out in each stimulus in praat ([20], version 5.0.27) leaving disyllbic /VCən/ stimuli. This was done based on the assumption that children neither know the city nor the less common male name *Hagen*. Prior to the experiment, participants were told that the speaker they will hear throughout the listening test does not speak very clearly and that the word's initial sound is hard to understand. Upon auditory presentation of each stimulus, participants were asked to decide whether the stimulus sounded more like *Wagen* (/vaːgən/, 'trolley'), *Haken*, or *hacken*. While children were presented with three repetitions of each stimulus, adults judged five repetitions. All participants responded by clicking on one of three different pictures presented on a computer screen, each corresponding to one response option. Stimuli were presented in randomized order. The experiment was conducted using praat [20] (version 5.0.27).

3.1.3. Analysis

For each participant individual response curves of long-vowel-word (i. e. *Wagen, Haken*) and short-vowel-word (i. e. *hacken*) judgments were plotted. On the basis of visual inspection of these speaker-specific response curves, we excluded the results from one adult participant because s/he was not able to classify the endpoint stimuli unambiguously as containing a long and a short vowel, respectively. Although the same was true for some child participants, none of the children's data were excluded because for these two age groups we did expect less categorical perception.

Since many response curves were not clearly s-shaped and therefor did not converge when fitting sigmoid functions to the responses using binary logistic regression, we instead grouped together all responses to stimuli one to three (i. e. the left part of the continuum), to stimuli four and five (i. e. the middle part) and to stimuli six to eight (i. e. the right part). This procedure was based on the aggregated response curves of long-vowel-word and short-vowel-word judgments averaged across all adults showing roughly 80 % long and short vowel responses to the first and last three stimuli of the continuum, respectively, and chance level responses to the ambiguous stimuli four and five. The data was statistically analyzed using general linear mixed effects models (GLMM with family is binomial) with response as the dependent variable, age group (preschool vs. school vs. adult) and stimulus group (left vs. middle vs. right part of the continuum) as fixed factors, and listener as random factor. The hypothesis to be tested was whether younger and older children responded differently to these stimulus groups.

3.2. Results

Adult and child participants judged only 10 % and 12 %, respectively, of the stimuli to sound like *Wagen*. This finding differs from the perception results in [14] where the first two stimuli (i. e. *VCratios* above 0.65) were predominantly rated as *Hagen*. The few *Wagen*-responses in this experiment are likely due to the absence of a steep F2-transition into the initial vowel which listeners could have expected in a /va:/ sequence. Interestingly, children and adults did not differ significantly in the number of *Wagen*-responses, suggesting that all age groups focus to the same extent on spectral cues. In the following analysis, we therefore collapsed all *Wagen* and *Haken* responses and refer to these as long-vowel-word responses.

Figure 4 shows the aggregated long- and short-vowel-word responses to the three stimulus groups (i. e. the left, middle, and right part of the continuum) separately for the two groups of children and the adult group. The GLMM revealed a significant interaction between age group and stimulus group ($\chi^2_{10} = 113.9, p < .000$) suggesting that the three groups of participants differed in the proportion of long-vowel-word-responses as a function of stimulus group. Adults' categorical separation into the two categories is not surprising given that the continuum was subdivided based on adults' response curves. Older children showed a similar response pattern as adults with respect to the two ambiguous stimuli from the middle of the continuum, although, overall, the response curve was generally flatter compared to adults. Younger children's judgments of stimuli four to eight were, on the other hand, at chance level while they perceived the first three stimuli predominantly as containing a long vowel.

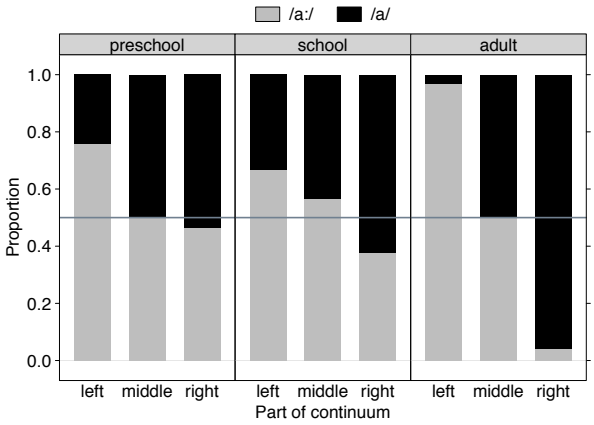

Figure 4: *Proportion of long-vowel-word (/a:/, grey) and short-vowel-word (/a/, black) responses split by age group and continuum part. The horizontal grey line signals the 50 % boundary.*

4. Discussion and Conclusion

The findings of the present study are threefold: Children aged 5 to 9 years produce the vowel length contrast in terms of vowel duration and they exploit the duration cue to the same extent as adults. Temporal variability was greater in words that were spoken at a faster speech rate, in this case at a normal voice, but the increased variability in faster speech did not vary across the two groups of children. These two findings suggest that preschool and school children do not differ in the execution of articulatory timing. In comparison to adults, however, in particular preschool children show greater temporal variability indicating that overall children's motor control abilities are not fully developed at this age. This result is in line with previous findings showing greater variability in children ([8], [9], [10]). However, it deviates from the finding of overall greater variability that is independent of phonemic vowel length which was described in [10]. The greater variability in preschoolers' realization of long vowels suggests that representations of phonemic categories and their phonetic implementations are not entirely dissociated. In addition, the effect of quantity on variability may reflect a particular stage in the acquisiton of a language-specific pattern showing more variation in long vs. short vowels in German ([21]).

Our third result was a small age effect between the two groups of children in perception where school children's response pattern to a continuum from a long- to a short-vowel-word were in between those of adults and preschool children. This finding indicates that the abstract representation of the phonemic vowel length contrast, too, is not fully stabilized in children of both age groups, even though the realization of the contrast in production suggests that it is. More analyses of the link between production and perception during acquistion are thus needed to better understand how children "climb[...] the ladder of abstraction" [22] while learning to obtain control over language-specific temporal patterns in production.

5. Acknowledgements

This research was supported by DFG grant number KL 2697/1-1 "Typology of Vowel and Consonant Quantity in Southern German varieties: acoustic, perception, and articulatory analyses of adult and child speakers" to the author.

6. References

[1] S. Nittrouer, "The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults," *The Journal of the Acoustical Society of America*, vol. 115, no. 4, pp. 1777–1790, 2004.

[2] M. Chen, "Vowel length variation as a function of the voicing of the consonant environment," *Phonetica*, vol. 22, no. 3, pp. 129–159, 1970.

[3] K. H. Ramers, *Vokalquantität und -qualität im Deutschen*. Tübingen: Niemeyer, 1988.

[4] R. Wiese, *The phonology of German*. Oxford: Oxford University Press, 2000.

[5] A. V. Fox-Boyer, *Kindliche Aussprachestörungen: phonologischer Erwerb, Differenzialdiagnostik, Therapie*. Schulz-Kirchner, 2009.

[6] M. Kehoe and C. Lleó, "A phonological analysis of schwa in german first language acquisition," *Canadian Journal of Linguistics/Revue canadienne de linguistique*, vol. 48, no. 3-4, pp. 289–327, 2003.

[7] M. A. Redford and C. E. Gildersleeve-Neumann, "The development of distinct speaking styles in preschool children," *Journal of Speech, Language, and Hearing Research*, vol. 52, no. 6, pp. 1434–1448, 2009.

[8] B. L. Smith, M. K. Kenney, and S. Hussain, "A longitudinal investigation of duration and temporal variability in children's speech production," *The Journal of the Acoustical Society of America*, vol. 99, no. 4, pp. 2344–2349, 1996.

[9] S. Kunnari, S. Nakai, and M. M. Vihman, "Cross-linguistic evidence for the acquisition of geminates," *Psychology of Language and Communication*, vol. 5, no. 2, pp. 13–24, 2001.

[10] M. A. Redford and G. E. Oh, "The representation and execution of articulatory timing in first and second language acquisition," *Journal of Phonetics*, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.wocn.2017.01.004

[11] I. Yuen, F. Cox, and K. Demuth, "Three-year-olds' production of Australian English phonemic vowel length as a function of prosodic context," *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1469–1479, 2014.

[12] C. Draxler and K. Jänsch, "Speechrecorder-a universal platform independent multi-channel audio recording software." in *LREC*, 2004.

[13] C. Dromey and L. O. Ramig, "Intentional Changes in Sound Pressure Level and RateTheir Impact on Measures of Respiration, Phonation, and Articulation," *Journal of Speech, Language, and Hearing Research*, vol. 41, no. 5, pp. 1003–1018, 1998.

[14] F. Kleber, "Complementary length in vowel-consonant sequences: acoustic and perceptual evidence for a sound change in progress in bavarian german," *Journal of the International Phonetics Association*, In Press.

[15] T. Kisler, U. D. Reichel, F. Schiel, C. Draxler, B. Jackl, and N. Pörner, "Bas speech science web services-an update of current developments," in *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, 2016.

[16] R. Winkelmann, J. Harrington, and K. Jänsch, "Emu-sdms: Advanced speech database management and analysis in r," *Computer Speech & Language*, 2017.

[17] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: http://www.R-project.org/

[18] N. Zharkova, "Voiceless alveolar stop coarticulation in typically developing 5-year-olds and 13-year-olds," *Clinical Linguistics & Phonetics*, 2016.

[19] K. J. Kohler, "Dimensions in the perception of fortis and lenis plosives," *Phonetica*, vol. 36, no. 4-5, pp. 332–343, 1979.

[20] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," [Computer program].

[21] C. Mooshammer and C. Geng, "Acoustic and articulatory manifestations of vowel reduction in german," *Journal of the International Phonetic Association*, vol. 38, no. 02, pp. 117–136, 2008.

[22] B. Munson, J. Edwards, M. E. Beckman, A. Cohn, C. Fougeron, and M. Huffman, "Phonological representations in language acquisition: Climbing the ladder of abstraction," *Handbook of laboratory phonology*, pp. 288–309, 2011.