Articulation rate in child-directed speech increases as a function of child age

Johan Sjons & Thomas Hörberg
Department of Linguistics, Stockholm University, Sweden
johan@ling.su.se, thomas_h@ling.su.se

Abstract

It has been shown that articulation rate (AR), the number of produced linguistic units per time unit with pauses excluded, is lower in child-directed speech (CDS) than in adult-directed speech (ADS). The present study is the first corpus-based longitudinal study to investigate AR in Swedish CDS as a function of child age while also controlling for utterance length in terms of number of syllables and for individual differences between speakers. AR in transcribed utterances of 7 parents directed at their respective child during different ages was analyzed with mixed effects modeling. Results show a significantly higher AR in longer than in shorter utterances and a significant increase in AR as a function of infant age. Future studies include comparison with entropy-based measures.

Introduction

There are differences between child-directed speech (CDS) and adult-directed speech (ADS). For instance, CDS has overall higher pitch (Fernald and Simon, 1984), more varied pitch contour (Garnica, 1977), shorter utterances with fewer words (Phillips, 1973), and lower speech rate (Broen, 1972).

Liu, Tsao & Kuhl (2009) suggest that some features of CDS might change in relation to the development of the child, whereas other features might not. This is an area in which "there is little data" (p. 910). They mention on the one hand Stern et al. (1983) who found that CDS had a wider pitch range when directed at 4-month-old children than at newborns, 1-, and 2-year olds; and on the other Newport, Gleitman and Gleitman (1977) who found that the syntactic features displayed in CDS do not become more complex when the child is between 1-2 years old.

The reasons for why speakers employ the features of CDS are not entirely clear. A principled explanation for this might arise in accordance with the H&H-theory (Lindblom, 1990). [1] This theory predicts that speakers make online updates about how much information their recipients need, and move along a continuum of more or less well-articulated speech accordingly. Applied to CDS, adults could be sensible to what is suitable for the child as input during different stages of its development. To this end, the field needs quantitative measures that can be applied to CDS longitudinally.

One such measure, which has been somewhat less studied as applied to CDS longitudinally, is articulation rate (AR). The present study investigates whether AR in Swedish CDS changes as a function of child age. AR is the number of produced linguistic units (e.g., phonemes.) per time unit (e.g., seconds) excluding pauses. It is distinguished from the related measure speech rate. The distinction lies in whether pauses are excluded (articulation rate) or included (speech rate) (e.g., Goldman-Eisler, 1961; Trouvain et al., 2001).

It has been shown that AR is lower in CDS than in ADS. Fernald and Simon (1984) found the AR of German-speaking mothers (measured in terms of syllables per second, excluding pauses

[1] Hyper- and hypospeech.

>300ms) was significantly lower in mothers speech directed at their newborns than in their speech directed at an adult interviewer (ADS: 5.8; CDS: 4.2).

Van de Weijer (1997) investigated AR of caretakers speaking Dutch and German to each other as well as to a child of 6-9 months of age. Results showed AR in ADS to be higher (5.64 syllables/s) than in CDS (3.32 syllables/s). The author refers to speech rate, but this corresponds to what we call AR: "the duration of the pauses was subtracted from the total utterance duration." (p. 291). No definition of a pause in terms of length is provided.

Since AR has been shown to be higher in ADS than in CDS, a natural place to look for an increase is in the early life of the child, which a couple of studies have done. Narayan and McDermott (2016) showed that AR in CDS in Sri Lankan Tamil, Tagalog, and Korean increases during the first 4-16 months of the child. Utterances had a duration around 5s excluding pauses >300ms.

However, Thanavisuth and Luksaneeyanawin (1998) failed to find an increase in AR as a function of infant age in Thai CDS.[2] They studied the speech of 6 mothers talking to their children during the first year of the child (newborn, 3, 6, 9 and 12 months). It is unclear what duration defined a pause: "[Recordings] were transcribed into utterances by using auditory pauses as a marker to delimit them" (p. 2).

For Swedish, Fuchs et al. (2014) did a corpus-based longitudinal study of AR in CDS. The same recordings as in the present study were used (see Material). They transcribed 9 parent-child dyads from when children were 8-9, 16-17 and 26-27 months old. AR was quantified as the number of transcribed vowels per utterance duration. Utterances were between pauses of >300ms. Results showed a non-significant increase in relation to child age.

The present study is the first corpus-based longitudinal study to investigate AR in Swedish CDS as a function of child age, while also controlling for utterance length in terms of number of syllables. Since AR varies as a function of utterance length, where longer utterances tend to have a higher AR than shorter utterances (e.g., Lindblom and Rapp, 1973; Quené, 2008), a potential increase in AR could depend on an increase in utterance length *per se*, rather than reflecting an adaptation to the age or development of the child. We therefore control for utterance length in terms of number of syllables, as well as individual differences in AR between speakers. The study is thus an attempt to bring further clarity to the question of whether AR in CDS increases during the early years in the life of the child.

Method and Data

Material

The data in the present study consist of the transcribed utterances of 4 mothers and 3 fathers directed at their respective child (i.e., 4 in total). These utterances are derived from the orthographical transcriptions in the MINGLE-corpus (Nilsson Björkenstam and Wirén, 2014; Nilsson Björkenstam, 2012), but have in some cases been modified (see next section). The MINGLE-corpus is a multi-modally annotated[3] corpus, consisting of several recordings of Swedish-speaking subjects (alternately mothers and fathers) in interaction with their child. Recordings were collected at Stockholm Babylab, part of the Phonetics Laboratory at Stockholm University. Parents and children were equipped with a wireless lavalier microphone, allowed them to move around unhindered. Also, two cameras, which recorded each session, were placed in the laboratory room. Various toys were available for the parents and children in the room. The parents were instructed

[2] But was lower than in ADS.

[3] Using the transcription software ELAN, available for free at https://tla.mpi.nl/tools/tla-tools/elan/

to refer to and use these. Parents were asked to act freely as if being at home. During the second part of some of the recordings, the experiment leader entered the room and guided the parent through the Swedish Early Communicative Development Inventory (SECDI). The sessions lasted on average around half an hour (Lacerda, 2009; Nilsson Björkenstam and Wirén, 2014).

Data processing

For the measure of AR, we used number of syllables per second within an utterance, defined as a string of sounds in between pauses of length >200ms. A syllable was approximated to each vowel in the transcribed material, something which in Swedish gives a very good approximation. The transcriptions were, when needed, changed from the somewhat formal Swedish orthographic representation employed in MINGLE to a representation closer to the actual pronunciation. For instance, "någonting" was changed to "nånting" or "nåt" (something) and "ramlade" to "ramla" (fell), etc. Each utterance was also, where needed, split, merged with another utterance, or re-segmented, in accordance with the definition of a pause. The start and end of the segments were moved as closely to the start and end of the vocalizations as possible, while allowing for a little bit of silence before and after, in order for the utterance not to sound truncated. However, the duration of this silence was never longer than 40ms. Laughter, non-words and singing were not included. The following was extracted with a Python (v. 3.2.2) script for each session: the age of the child in days (approximated to each month containing 30 days); AR (number of syllables per second) for each utterance; and the number of syllables per utterance. The procedure resulted in 8458 utterances and 30703 tokens.

Results

Figure 1 shows the mean ARs in CDS for each recorded session as a function of child age, with error bars illustrating the 95% confidence interval for each mean.

The relationship between AR and the age of the child – controlling for utterance length and individual ARs between speakers – is illustrated in Figure 2A. Figure 2B shows AR as a func-

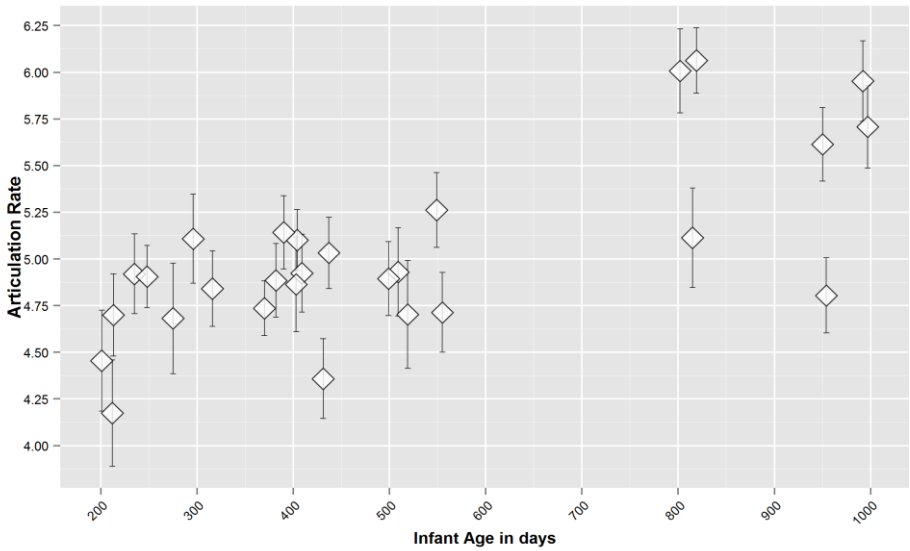

Figure 1. Mean articulation rate per session as a function of age of the infant. Error bars illustrate 95% confidence intervals of these means.

tion of utterance length in terms of the log of the number of syllables the utterance contains. The figure indicates that AR increases as a function of the age of the child. It also shows that longer utterances tend to have a higher AR.

In order to evaluate these effects, the data were analyzed with mixed effects modeling (cf., Gelman & Hill 2006). The mixed effects model is a special instance of the general linear model (e.g., Howell 2010). It allows for the inclusion of random effects, such as differences in AR between speakers. It is therefore possible to rule out that any observed effect between AR and age is not an artifact of a bias in the distribution of individual speakers across ages – for example, that faster speakers produce more utterances directed at children in the higher age range.

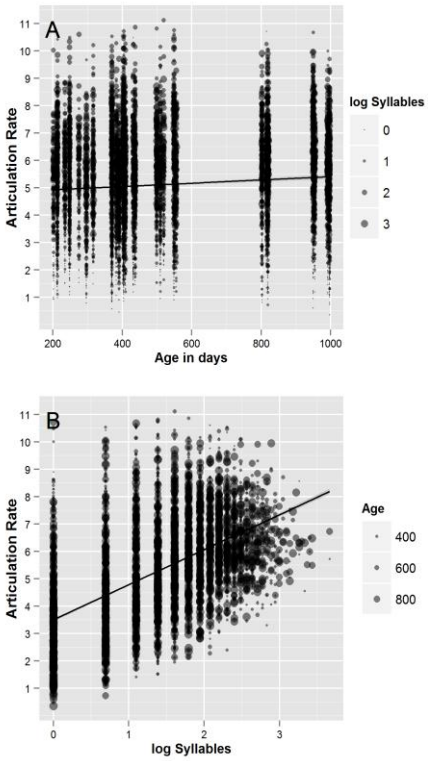

Figure 2. Articulation rate of individual syllables in CDS as a function of the age of the child (Panel A) and utterance length in terms of log number of syllables (Panel B). Dot size indicates either utterance length (Panel A) or child age (Panel B).

In the present analysis, the model predicts AR in CDS (i.e., the outcome variable, Articulation rate) as a function of the age of the child (i.e., the predictor variable Age), utterance length in terms of the log of the number of syllables [4] (i.e., predictor variable logSyllables) and the interplay between child age and log number of syllables (i.e., the Age × logSyllables interaction). As such, the model tests whether AR in CDS increases as the child gets older, while controlling for differences in AR between utterances of different length.

Analyses were conducted in the statistical language R (v. 3.1.2). Degrees of freedom for the calculation of *p*-values were estimated using Welch-Satterthwaite approximation, as implemented in the lmerTest() package (v. 2.0-20) (Kuznetsova, Brockhoff & Christensen 2014). AR that deviated by more than 3 standard deviations from the overall mean AR were excluded. Both the Age and the logSyllables predictor were standardized[5]. Apart from the fixed effects for Age, logSyllables, and the Age × logSyllables interaction, the model included a random intercept for speaker and a by-speaker random slope for logSyllables. In other words, the model controls for AR differences between speakers as well as speaker differences in the influence of utterance length on AR.

The model found a significant effect of logSyllables, $\beta = 1.12$, $t(6) = 17.55$, $p < .0001$, showing that longer utterances are articulated significantly faster than shorter utterances. Importantly, the model also found a significant effect of Age, $\beta = 0.14$, $t(4278) = 6.49$, $p < .0001$. This shows that AR in CDS increases with the age of the child. The

[4] The log was used because the distribution of the number of syllables is highly positively skewed.

[5] For each data point, the mean is subtracted and it is divided by the standard deviation so that the resulting distribution has a mean of 0 and a standard deviation of 1.

Age × logSyllables interaction was not significant, β = -0.01, $t(960) = -0.57$, p=.57. This shows that the effects of child age and utterance length on AR in CDS are independent of each other.

Discussion

This is the first study showing that articulation rate (AR) in Swedish child-directed speech (CDS) increases as a function of child age, even when the number of syllables per utterance and AR between subjects are controlled for.

The results are consistent with earlier findings of Narayan & McDermott (2016), where, however, the method of controlling for utterance length differed, since they chose utterances of duration around 5s.

Furthermore, the results give a hint about that parents adjust their speech in order to fit their approximation of the linguistic capacity of the child. The child's linguistic capacity is a more interesting predictor than child age, but harder to quantify. We plan to use SECDI as an additional approximation in a future study.

A drawback of the study is the lack of data-points for when children were under 7 months, approximately 20-27 months, and over 33 months of age. This means the linear function might not give a complete picture of the change in AR.

Future research includes investigating the difference in AR between CDS and ADS in the same corpus, to see if the results in the literature concluding that AR in CDS is lower than AR in ADS hold for Swedish. The reason for why this was not included was that it would require too much work of annotation of when the parent was concentrating on the SECDI-forms, that is, having a lower AR than in normal conversation, due to disfluencies, etc.

It would also be interesting to see to what extent the parent's AR correlates with AR in child production, since it is known that children gradually increase their AR (Pettinato, Tuomainen, Granlund, & Hazan, 2016).

Finally, the results of AR in CDS should be compared with, for instance, some entropy-based measure, in order to get a clearer picture of the language-related input data the child receives during its early years.

Acknowledgements

Thanks to Mats Wirén, Anders Eriksson and Ellen Marklund for useful remarks, Stockholm Babylab for sharing the recorded material, and Kristina Nilsson Björkenstam & Mats Wirén, for sharing the MINGLE corpus.

References

Broen, P. A. (1972). The Verbal Environment of the Language-Learning Child. *ASHA Monographs*, No. 17.

Fernald, A. (1985). Four-month-old children prefer to listen to motherese. *Child behavior and development*, *8*(2), 181-195.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of child preference for motherese speech. *Child behavior and development*, *10*(3), 279-293.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental psychology*, *20*(1), 104.

Fuchs, S., Hasselqvist, H., Hjärpe, A., Jonsson, L., Kos, D. & Påhls, M. (2014). Articulation rate in child-directed speech – changes related to age. *Unpublished.*

Garnica, O. K. (1977). Some prosodic and paralinguistic features of speech to young children. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition.* Cambridge, England: Cambridge University Press.

Goldman-Eisler, F. (1961). The significance of changes in the rate of articulation. *Language and Speech*, *4*(3), 171-174.

Howell, D. C., (2012) *Statistical methods for psychology*. 7th ed. Belmont, CA: Thomson Wadsworth.

Kuhl, P., Williams K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. (1992). Linguistic experience alters phonetic perception in children by 6 months of age. *Science*, *255*(5044), 606-608.

Lacerda, F. (2009). On the emergence of early linguistic functions: A biologic and interactional perspective. In K. Alter, M. Horn, M. Lindgren, M. Roll, & J. von Koss TorklCDSen (Eds.), *Brain talk: Discourse with and in the brain* (pp. 207-230).

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403-439). Springer Netherlands.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Auditory Analysis and Speech Perception (London, 1975)*, 387-96.

Liu, H-M., Tsao, F-M., & Kuhl, P. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to proverbal children and five-year-old children: a longitudinal study. *Journal of Child Language*, *36*(04), 909-922.

Logan, K. J., Byrd, C. T., Mazzocchi, E. M., & Gillam, R. B. (2011). Speaking rate characteristics of elementary-school-aged children who do and do not stutter. *Journal of Communication Disorders*, *44*(1), 130-147.

Narayan, C. R. & McDermott, L. C. (2016). Speech rate and pitch characteristics of child-directed speech: Longitudinal and cross-linguistic observations. *The Journal of the Acoustical Society of America*, *139*(3), 1272-1281.

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. *Talking to children: Language input and acquisition*, 109-149.

Nilsson Björkenstam, K. (2012). The MINGLE annotation scheme: Multimodal annotation of parent-child interaction in a free play setting (version 1.0).

Nilsson Björkenstam, K. & Wirén, M. (2014). Multimodal Annotation of Syncrony in Longitudinal Parent-Child Interaction. *MMC 2014 Multimodal Corpora: Combining applied and basic research targets: Workshop at LREC 2014*.

Pettinato, M., Tuomainen, O., Granlund, S., & Hazan, V. (2016). Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication. *Journal of Phonetics*, *54*, 1-14.

Phillips, J. R. (1973). Syntax and vocabulary of mothers' speech to young children: Age and sex comparisons. *Child development (44)*, 182-185.

Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America*, *123*(2), 1104-1113.

Stern, D. N., Spieker S., Barnett, R. & MacKain, K. (1983). The prosody of maternal speech: Child age and context related changes. *Journal of child language*, *10*(1), 1-15.

Thanavisuth, C. & Luksaneeyanawin, S. (1998). Acoustic qualities of CDS and ADS in Thai. In *International Conference on Lexicography and Semantic Prosody* , (pp. 445-448).

Trouvain J., Koreman, J., Erriquez, A. & Braun, B. (2001). Articulation rate measures and their relation to phone classification in spontaneous and read german speech. *ISCA Tutorial and Research Workshop (ITRW) on Adaptation Methods for Speech Recognition*.

Van de Weijer, J. (1997). Language input to a prelingual child. *The*

GALA '97 Conference on Language Acquisition, Edinburgh University Press, (pp. 290-293).

Van de Weijer, J. (1999). Language input for word discovery. Ph.D. thesis, Radboud University Nijmegen.