Effects of Caregiver Prosody on Child Language Acquisition

Soroush Vosoughi[1], *Brandon C. Roy*[1], *Michael C. Frank*[2], *Deb Roy*[1]

[1]The Media Laboratory
[2]Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology, Cambridge, MA USA
soroush@media.mit.edu, bcroy@media.mit.edu, mcfrank@mit.edu, dkroy@media.mit.edu

Abstract

This paper investigates the role of prosody in one child's lexical acquisition using an ecologically valid, high-density, longitudinal corpus. The corpus consists of high fidelity recordings collected from microphones embedded throughout the home of a family with a young child. We analyze data collected continuously from ages 9 – 24 months, including the child's first productive use of language at about 11 months and ending at the child's active use of more than 500 words. We found significant correlations between prosody of caregivers' speech and age of acquisition for individual words.

Index Terms: prosody, child language acquisition, word learning, corpus data

1. Introduction

What role does the linguistic environment play in child language acquisition? Why do children learn some words earlier than other words? Does the caregivers' use of language in the presence of the child have an effect on the child's language development? This paper will try to answer these questions by investigating the role of the prosody of caregiver speech in one child's lexical acquisition using the corpus of Human Speechome Project [1].

Previous studies have shown that the prosody of child-directed speech is different from the prosody of adult-directed speech [2, 3]. More recent studies have shown that infants are sensitive to the prosodic aspects of speech [4, 5, 6]. Furthermore, it has been suggested that one of the roles that prosody plays in child language acquisition is in word segmentation [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17].

Despite these studies, our overall understanding of the role of prosody in child language development is limited. The Human Speechome Project was launched to address this and other questions in language development by collecting a dense, longitudinal corpus in a naturalistic setting that is orders of magnitude larger than prior studies.

In this study we use the Human Speechome Project's corpus to investigate the influence of prosody in caregiver speech on the lexical development of the child. The caregiver speech used in our analysis is speech the child was exposed to, which we call "child-available speech" (CAS). Child-available speech includes child-directed speech (CDS) as well as other speech when the child was present. More specifically, we create a predictive model of the child's word acquisition based on prosodic aspects of child-available speech.

2. Method

2.1. Corpus

This study uses the corpus collected for the Human Speechome Project (HSP). The HSP corpus is high-density, longitudinal and naturalistic. The corpus consists of high fidelity recordings collected from microphones embedded throughout the home of a family with a young child [1]. For this study we look at data collected continuously from ages 9 to 24 months, including the child's first productive use of language at about 11 months and ending at the child's active use of more than 500 words.

Although the corpus as a whole contains more than 70% of the child's total language input (an estimated 12 million words for the 9-24 month age range), we analyze an evenly-sampled 400,000 word portion that has been hand-transcribed using new, semi-automatic methods and for which the speaker has been automatically identified with high confidence [18]. The corpus contains both the child's productions and child-available speech. Using this subset of the data, we tracked the child's vocabulary development over time by defining a "word birth" as the child's first productive use of a word captured in our transcripts. This serves as a conservative estimate for the child's age of acquisition (AoA) for each word. A more detailed account of our methodology and first results on the child's language development can be found in [19].

2.2. Measuring prosody

Our analysis investigated the relationship between prosody and age of acquisition. As a proxy for prosodic emphasis, we used a standardized measure of mean word duration, relative fundamental frequency (F0) and relative intensity. In order to do our analysis, we regressed the word birth date for each word in the child's productive vocabulary (517 words by 24 months.) against the three prosodic variables mentioned above (duration, f0, and intensity). In order to esure reliable estimates for all three prosodic variables, we excluded those words from the child's vocabulary for which there fewer then six caregiver utternaces. In total 56 words were excluded leaving 461 total words included in the current analysis.

Below is our definition of the three prosodic variables used in our analysis. All three variables are computed using caregiver speech up to the AoA for a particular word.

2.2.1. Duration

The duration predictor is a standardized measure of word duration for each word. We first extracted duration for all vowel tokens in the corpus. We next converted these to normalized units for each vowel separately (via z-score), and then measured the mean standardized vowel duration for the tokens of a particular word type. For example, a high score on this measure for the word “dog” would reflect that the vowel that occurred in tokens of “dog” was often long relative to comparable vowel sounds that appeared in other words. We grouped similar vowels by converting transcripts to phonemes via the CMU pronunciation dictionary.

2.2.2. Fundamental frequency

The fundamental frequency predictor is the measure of a word’s change in fundamental frequency (F0) relative to the utterance in which it occurred. We first extracted the F0 contour for each utterance in the corpus using the PRAAT system [20]. We then calculated the change in F0 as a sum of two terms shown in the equation below. The first term captures the change in F0 for the word relative to the utterance in which it’s embedded. $\overline{\mathrm{F0}}_{\mathrm{w}}$ is the mean F0 value of the word, and $\overline{\mathrm{F0}}_{\mathrm{utt}}$ is the mean F0 of the whole utterance. The second term captures the maximum change in F0 within the word. $t_{\max}$ and $t_{\min}$ are the times at which the max and min F0 values occur within the word. α_0 and α_1 are constants set using a brute force optimization technique.

$$\alpha_0 * \left|\overline{\mathrm{F0}}_{\mathrm{w}} - \overline{\mathrm{F0}}_{\mathrm{utt}}\right| + \alpha_1 * \left|\frac{\max(\mathrm{F0}_{\mathrm{w}}) - \min(\mathrm{F0}_{\mathrm{w}})}{t_{\max} - t_{\min}}\right|$$

2.2.3. Intensity

Relative word intensity was calculated in the same manner as F0 using the intensity contour in place of the F0 contour. The intensity contour was extracted using the PRAAT system.

3. Results

The result of the linear regression analysis of AoA against mean standardized duration, relative F0 and relative intensity are shown in Figures 1, 2, and 3 respectively. We found significant correlations between age of acquisition and standardized duration ($r = -0.29, p < .001$), relative F0 ($r = -0.19, p < .001$) and relative intensity ($r = -0.35, p < .0001$) across all words. This indicates that words that were often spoken with relatively greater emphasis were acquired earlier by the child. These correlations were mediated by the syntactic category of the words being examined, similar to the findings in [21]. In all cases, adjectives showed strong correlations between their prosodic variables and AoA, while verbs showed considerably weaker correlation (see Table 1), likely due to other factors mediating acquisition (eg. [22]).

The combination of duration, F0 and intensity resulted in overall predictions that were more accurate than those produced by either alone ($r = -0.44, p < .0001$). Correlations between the three prosodic variables are shown in Table 2. The largest and most significant correlation is between duration and intensity. Though the reason behind this relatively high correlation is not completely clear, it could mean that words uttered by the caregivers are more likely to have longer duration vowels when they are being accented in a sentence with greater change in intensity. The result of linear regression analysis of AoA against the best linear combination of duration, F0 and intensity is shown in Figure 4. These results suggest that the three prosodic variables (duration, f0, and intensity) are complementary sources of information and that children may be able to leverage all three in the service of early word learning.

The equation below is the result of regressing AoA against the combination of the three prosodic variables, complete with the three coefficient values and the intercept. Here, AoA is measured in *months*. Moreover, this equation shows the predictive relationship between duration, F0 and intensity and the AoA for any given word in the child’s vocabulary.

$$\mathrm{AoA}_w = -2.66 * \mathrm{Dur}_w - 3.42 * \mathrm{F0}_w - 4.78 * \mathrm{Int}_w + 25.57$$

Table 1: Pearson’s r values measuring the correlation between age of acquisition and standardized mean duration, fundamental frequency, intensity and their best linear combination for each category in the child’s speech. Note: $' = p < .1$, $* = p < .05$, and $** = p < .001$.

	Duration	F0	Intensity	Dur+F0+Int
Nouns	-.13*	-.17*	-.37***	-.39***
Adjectives	-.44***	-.27*	-.43***	-.63***
Verbs	-.19’	-.09	-.20*	-.25’
All	**-.29*****	**-.19*****	**-.35*****	**-.44*****

Table 2: Pearson’s r values measuring the correlation between duration, F0 and intensity. Note: $' = p < .1$, $* = p < .05$, and $** = p < .001$.

	Duration	F0	Intensity
Duration	1.0	.12*	.22**
F0	.12*	1.0	.10*
Intensity	.22**	.10*	1.0

Figure 4 can also be used to identify outliers in our model. These outliers are words whose age of acquisition is not very strongly correlated with any of the three prosodic variables. By identifying these outliers we can further study the hidden signals in caregiver speech that our current model is not capturing. These could be non-prosodic speech signals, visual signals or even social signals. For example, according to our current model, the prosodic values of the word ‘dad’ are not very highly correlated with age of acquisition. That is, the age of acquisition of the word ’dad’ is much earlier than predicted by our model. This could be because prosody alone can not account for the social implications of learning the word ‘dad’ early on.

4. Discussion and Conclusions

Investigating the effects of prosody in child-available speech on child language acquisition is not only of scientific interest, it may also have clinical importance. The results of this and related studies may help in developing techniques for addressing developmental language disorders. For example, these results may be useful for evaluating caregiver speaking styles, or in guiding caregivers toward better communication with their children.

We found evidence that the prosody of child-available caregiver speech influences the lexical development of the child. Specifically we found that relative intensity has the most significant effect on the age of acquisition of words, followed by duration and lastly fundamental frequency. These results agree

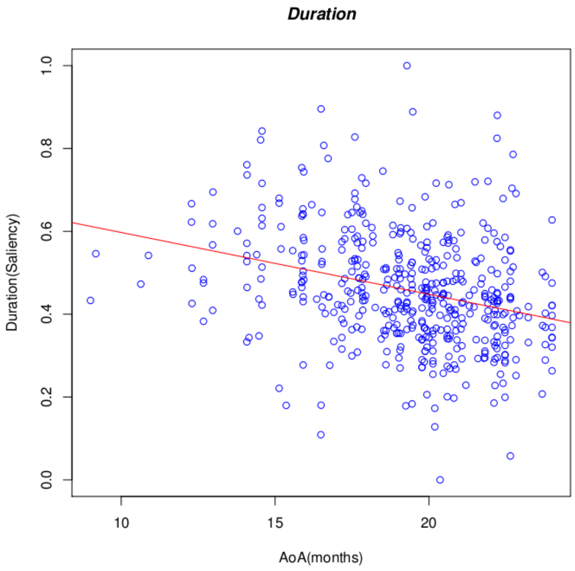

Figure 1: *Words plotted by the age they were first produced and their normalized mean duration, along with the best linear fit.*

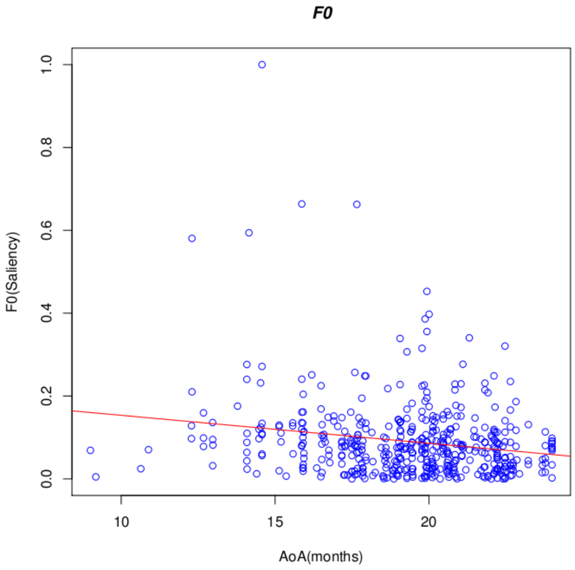

Figure 2: *Words plotted by the age they were first produced and their normalized relative mean F0, along with the best linear fit.*

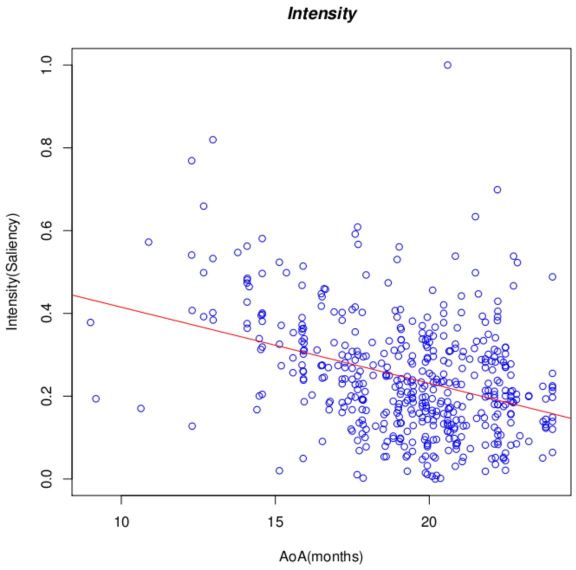

Figure 3: *Words plotted by the age they were first produced and their normalized mean relative intensity, along with the best linear fit.*

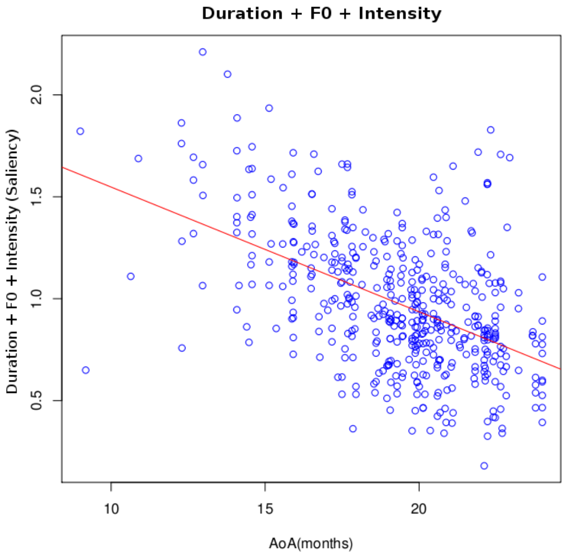

Figure 4: *Words plotted by the age they were first produced and the best linear combination of duration, F0 and intensity, along with the best linear fit.*

with a previous study by [23], where the authors found that fundamental frequency plays a minor role in distinguishing prominent syllables from the rest of the utterance and that instead, speakers primarily marked prominence with patterns of intensity and duration. We also found that the three prosodic variables, duration, F0 and intensity are complementary sources of information.

However, the limitations of this study must be acknowledged since we use a linear input-output model where the child is treated as the only developing agent. In reality, the caregivers may also be developing and adapting along with the child. In future analyses, we hope to use a circular model where the caregiver(s) and the child are treated as two adapting players.

Finally, though we only look at prosodic variables in this study, we have started looking at other interesting variables such as time of day, word recurrence and mean length of caregiver utterance that could be used in conjunction with prosody to better predict the lexical development of the child.

5. References

[1] D. Roy, R. Patel, P. DeCamp, R. Kubat, M. Fleischman, B. Roy, N. Mavridis, S. Tellex, A. Salata, J. Guinness, M. Levit, and P. Gorniak, "The Human Speechome Project," in *Proceedings of the 28th Annual Cognitive Science Conference*. Mahwah, NJ: Lawrence Earlbaum, 2006, pp. 2059–2064.

[2] A. Fernald, T. Taeschner, J. Dunn, M. Papousek, B. de Boysson-Bardies, and I. Fukui, "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants," *Journal of Child Language*, vol. 16, no. 3, pp. 477–501, 1989.

[3] O. Garnica, "Some prosodic and paralinguistic features of speech to young children," *Talking to children: Language input and acquisition*, pp. 63–88, 1977.

[4] R. Cooper and R. Aslin, "Developmental differences in infant attention to the spectral properties of infant-directed speech," *Child Development*, vol. 65, no. 6, pp. 1663–1677, 1994.

[5] A. DeCasper and W. Fifer, "Of Human Bonding: Newborns Prefer Their Mothers' Voices," in *Readings on the Development of Children*, M. Gauvain and M. Cole, Eds. Worth Publishers, 2004, ch. 8, p. 56.

[6] J. Mehler, P. Jusczyk, G. Lambertz, H. Nilofar, J. Bertoncini, and C. Amiel-Tison, "A precursor of language acquisition in young infants," *Cognition*, vol. 29, pp. 143–178, 1988.

[7] L. Gleitman, H. Gleitman, B. Landau, and E. Wanner, "Where learning begins: initial representations for language learning," in *Linguistics: The Cambridge Survey*, F. J. Newmeyer, Ed. Cambridge University Press, 1988, ch. 6, pp. 150–192.

[8] L. Gleitman and E. Wanner, "The state of the state of the art," in *Language acquisition: The state of the art*. Cambridge University Press, 1982, pp. 3–48.

[9] K. Hirsh-Pasek, K. Nelson, G. Deborah, P. Jusczyk, K. Cassidy *et al.*, "Clauses are perceptual units for young infants." *Cognition*, vol. 26, no. 3, pp. 269–286, 1987.

[10] P. Jusczyk, K. Hirsch-Pasek, D. Kemler Nelson, L. Kennedy *et al.*, "Perception of acoustic correlates of major phrasal units by young infants." *Cognitive Psychology*, vol. 24, no. 2, pp. 252–293, 1992.

[11] N. Kemler, K. Hirsh-Pasek, P. Jusczyk, and K. Cassidy, "How the prosodic cues in motherese might assist language learning." *Journal of Child Language*, vol. 16, no. 1, pp. 55–68, 1989.

[12] J. Morgan, *From simple input to complex grammar*. MIT Press, 1986.

[13] J. Morgan, R. Meier, and E. Newport, "Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language," *Cognitive Psychology*, vol. 19, no. 4, pp. 498–550, 1987.

[14] J. Morgan and E. Newport, "The role of constituent structure in the induction of an artificial language." *Journal of Verbal Learning & Verbal Behavior. Vol*, vol. 20, no. 1, pp. 67–85, 1981.

[15] A. Peters, "Language segmentation: Operating principles for the perception and analysis of language," in *The crosslinguistic study of language acquisition*, D. I. Slobin, Ed. Lawrence Erlbaum, 1985, vol. 2, pp. 1029–1067.

[16] ——, "Language typology, individual differences and the acquisition of grammatical morphemes," in *The cross-linguistic study of language acquisition*, D. I. Slobin, Ed. Lawrence Earlbaum, 1992, vol. 4.

[17] A. M. Peters, *The Units of Language Acquisition*. Cambridge University Press, 1983.

[18] B. C. Roy and D. Roy, "Fast transcription of unstructured audio recordings," in *Proceedings of Interspeech*, Brighton, England, 2009.

[19] B. C. Roy, M. C. Frank, and D. Roy, "Exploring word learning in a high-density longitudinal corpus," in *Proceedings of the 31st Annual Cognitive Science Conference*, 2009.

[20] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 5.1.01)," http://www.praat.org/, 2009.

[21] J. Goodman, P. Dale, and P. Li, "Does frequency count? Parental input and the acquisition of vocabulary," *Journal of Child Language*, vol. 35, pp. 515–531, 2008.

[22] L. Gleitman, "The structural sources of verb meanings," *Language acquisition*, vol. 1, pp. 3–55, 1990.

[23] G. Kochanski, E. Grabe, J. Coleman, and B. Rosner, "Loudness predicts prominence: Fundamental frequency lends little," *The Journal of the Acoustical Society of America*, vol. 118, p. 1038, 2005.