

Toddlers are sensitive to prosodic correlates of disfluency in spontaneous speech

Melanie Soderstrom & James L. Morgan

Department of Cognitive & Linguistic Sciences
Brown University, Providence, RI, USA
{melsod; James_Morgan}@brown.edu

Abstract

The ability to distinguish fluent from disfluent speech could play an important role in infants' acquisition of their first language. Across two experiments using a Headturn Preference Procedure, we show that infants are able to distinguish fluent from disfluent speech based on its prosodic characteristics, and show a preference for listening to fluent English. In the first experiment, 22-month-old, but not 10-month-old, infants preferred to listen to fluent adult-directed speech samples over disfluent matched speech samples. In the second experiment, lexical and grammatical information were removed. Older infants still discriminated fluent from disfluent speech, but showed the reverse preference, for disfluent speech.

1. Introduction

When Chomsky described what was later termed the "poverty of the stimulus" in accounting for human language development, an important part of this argument was the inaccuracy and ill-formedness of the input to which infants are exposed. He noted the "interrupted fragments, false starts, lapses, slurring, and other phenomena that can only be understood as distortions of the underlying pattern" [1]. While this description of every day speech may be accurate with respect to adult speech, psychologists in the 70s and 80s noted that child-directed speech is much more fluent and grammatical [e.g. 4]. While some generative grammarians still hold the view that the input is highly degenerate, psychologists generally maintain that child-directed speech provides an input that is tailored to the needs of the language learner, and that the distortions found in adult-directed speech may be unimportant.

However, there are at least two reasons to think that adult-directed speech does constitute part of the input to the child. For one thing, children hear much more than just the child-directed speech that is aimed at them. One analysis of all the speech input heard by an infant between 6 and 9 months found that considerably less than half of the speech was directed at the infant or his older sibling [11]. It is likely that a significant portion of the speech that older infants and toddlers hear is adult-directed as well. Secondly, the well-formedness of child-directed speech is in large part due to its simplicity. Because child-directed utterances tend to be very short (with a mean length of utterance around 4 morphemes), there is less computational complexity, and hence a smaller likelihood of production problems which might cause disfluency. Yet it is longer utterances that might ultimately be the most informative to infants, once they have mastered the basic word order characteristics of their language, which arguably occurs by about 18 months [e.g. 2]. Such longer

utterances may be more likely to contain disfluencies. Hence, the problem of a "distorted" input may develop as the infant begins to access more complex utterances.

This study takes seriously the idea that infants might indeed pay attention to the adult-directed speech in their environment, particularly infants who are beginning to form a grammatical picture of their language. It asks whether infants might bring certain perceptual skills to bear in determining what constitutes fluent, well-formed speech input, based on prosodic and paralinguistic characteristics of the speech. In Experiment 1, prelingual 10 month old and more mature 20-23 month old infants heard both well-formed and ill-formed utterances in normal English speech. In Experiment 2, the lexical content was removed to isolate infants' sensitivity to the prosodic characteristics of the speech.

2. Experiment 1

Experiment 1 was designed to determine whether infants distinguish between fluent and disfluent speech. If so, infants should show a difference in their preference for listening to the two speech types. Additionally, it was predicted that infants might prefer to listen to fluent speech over speech that is less well-formed. If infants are using prosodic information to determine what constitutes well-formed input, they may be less interested in input they find to be ill-formed. We chose to examine the behavior of older infants because these infants already have some basic knowledge of the structural characteristics of their language [2, 7, 10], and may be on the cusp of benefiting from the longer and more complex input of adult-directed speech. We also examined younger infants whose knowledge of the grammatical properties of their language is limited at best, but who demonstrate sophisticated knowledge of the prosodic characteristics of speech [6, 9].

2.1. Stimuli

The disfluent utterances were culled from an audio transcript that was part of a larger corpus [8]. This recording took place in a busy airport lobby. The speaker was the mother of a 2.5 month old infant, who was present during the recording. However, many of the utterances in the transcript, and all but one of the disfluent utterances, were directed at fellow passengers who were strangers to the speaker. Disfluent utterances were culled from this transcript. Because of the ambient noise in the recording, and in order to produce fluent matched controls, the original recordings were not used. Instead, a trained speaker listened carefully to each utterance, and mimicked as closely as possible the prosodic characteristics of the original recordings. Fluent versions were produced at the same time, and matched as closely as possible the prosodic characteristics other than those of the disfluency itself. Below is an example of a disfluent utterance and the

fluent control version (# indicates pause, [///] indicates restart, +//. indicates self-interruption).

- (1) yeah , so we # ended up [///] we drove through Boston for like an hour , literally , with all the traffic +//.
- (2) yeah, so we ended up driving through Boston for like an hour, literally, with all the traffic.

The fluent and disfluent utterances were then rated by 8 adult listeners, using a 7 point rating scale, with 1 most fluent, and 7 most disfluent. The raters were able to discriminate the fluent and disfluent samples with a high degree of accuracy ($t(7)=7.27, p < .001$). The average rating for disfluent utterances was 4.52 and for fluent utterances 3.26. A subset of 20 utterance pairs with an average rating difference of at least 1 point were chosen and grouped into 4 disfluent and 4 fluent passages with 5 utterances each. Using the PSOLA method in Praat, the lengths of the fluent and disfluent versions of each utterance pair were made equal by slowing down the shorter version and speeding up the longer by an equivalent amount.

Table 1 provides information about the properties of these disfluent utterances and their fluent matched controls.

Table 1: Average properties of disfluent and fluent utterances (length is pre-adjustment)

	Length (ms)	Syllables	Repeated Words	Prosodic boundaries
Fluent	3565	17.6	0.85	1.85
Disfluent	3601	17.25	0.95	2.8

There were no significant differences between utterance types in length, number of syllables, or number of repeated words. The disfluent passages contained significantly more utterance-internal prosodic boundaries than the fluent passages ($t(19)=4.50, p < .001$). Along with prosodic disruptions, the four disfluent passages contained a total of 5 um/uhs and 9 part-words.

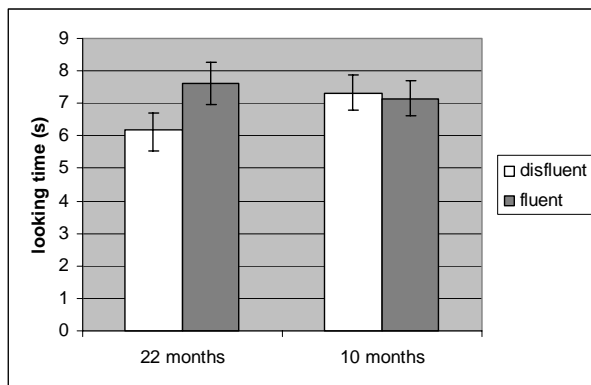


Figure 1: Mean looking times to fluent and disfluent passages, Experiment 1

2.2. Method

Participants were 24 infants between 20 and 23 months (range: 607-727, mean: 673 days, 12 male and 12 female), and 24 infants 10 months of age (range: 309-339 days, mean: 328 days, 11 male, 13 female). Five additional infants from the older age group participated in the study but their data were discarded due to fussiness or squirminess. No 10-month-

old's data were discarded. All participants were normally developing infants with normal hearing from Providence, RI, USA, and had parents and caregivers who were native speakers of American English.

Infants were tested using a version of the Headturn Preference Procedure [3]. In this method, speech stimuli are paired with a visual display (usually a flashing light). When the infant looks toward the visual display, the auditory stimulus begins to play. When the infant looks away for more than 2 consecutive seconds, the trial ends. Infants' level of interest in the speech stimuli is measured based on how much time they spend oriented toward the visual display. Because of the difficulty in maintaining the interest of older infants in this paradigm, a more engaging video display of a black and white geographic form was used for the older age group rather than the flashing light. Otherwise the procedure for the two age groups was identical. The dependent measure was average looking time toward the video/flashing light across trials.

Testing began with two pre-test trials to orient the infant to the task. This consisted of two trials of speech stimuli, one of which was a 10 second repetition of a fluent, prosodically well-formed phrase; the other was a 10 second repetition of a prosodically ill-formed phrase-like word sequence, taken from the familiarization stimuli for a previous study [9]. These were chosen so as not to bias the infant regarding prosodic well-formedness prior to the experiment. Infants were then tested on the 8 test passages – 4 fluent and 4 disfluent – presented in random order.

2.3. Results and Discussion

The 10-month-olds showed no preference for either passage type. Mean looking time to the disfluent passages was 7.3 s and to the fluent passages 7.2 s ($t(22) < 1, p > .5$). Thirteen out of twenty-four infants preferred the disfluent passages. By contrast, the 22-month-olds preferred the fluent passages, looking to the fluent passages for an average of 7.6 s compared with 6.2 s for the disfluent passages. This was significant by 2-tailed paired t-test ($t(23) = 2.3, p < .05$). Seventeen out of 24 infants preferred the fluent passages.

This finding suggests that 20-23 month old infants not only distinguish between fluent and disfluent passages, but also show a preference for the fluent passages. On the other hand, 10-month-olds do not show this preference. There are a number of possible explanations for this difference. Ten-month-olds might fail to detect the disfluency because they pay less attention to adult directed speech than older infants (although the similarities in looking times do not support this possibility) or because they are less able to detect the prosodic characteristics of disfluent speech in adult directed speech. Previous work in our lab using infant-directed speech (in which prosodic characteristics of speech tend to be exaggerated) demonstrated that 10-month-olds might be sensitive to this “infant-directed” disfluency, although the effect was not reliable across conditions [8]. Another possibility is that the older infants are using non-prosodic cues to disfluency. Although the utterances were similar in length and number of repeated words, the disfluent passages contained lexical cues (like “um/uh” and part-words), as well as grammatical cues such as the repair in example (1) above. Experiment 2 was designed to test this third possibility.

3. Experiment 2

3.1. Stimuli

We manipulated the stimuli used in Experiment 1 to remove grammatical and lexical cues to disfluency. This method has been used in previous studies in order to isolate certain phonological/prosodic properties of the speech stimuli [e.g. 6]. According to this method, the phonemes were reduced to a set of manner changes – all fricatives became /s/, vowels became /a/, liquids became /l/, glides became /y/, stops became /t/, and nasals /n/. The example stimuli above became:

- (3) ya, sa ya # antat at [///] ya tlas sla tastan sal lat an ayl, latalala, yas al sa tlasat +//.
- (4) ya, sa ya antat at tlasan sla tastan sal lat an ayl, latalala, ya al sa tlasat.

These stimuli were produced in a manner similar to that of Experiment 1 – the same trained speaker listened to the stimuli she had recorded for Experiment 1, and after careful practice produced the modified versions with prosodic properties as close as possible to the originals. These stimuli were then rated by a group of 11 adult listeners with the same rating scale as before. These raters were again able to discriminate the fluent from the disfluent stimuli with a high degree of accuracy ($t(10) = 7.2, p < .001$). The average rating for disfluent utterances was 5.22 and for fluent utterances 3.19. For 17 out of 20 utterances pairs, the fluent version was rated more than 1 point lower (more fluent) than the disfluent version.

3.2. Method

Participants were 20 infants between 20 and 23 months (range: 608-726, mean: 679 days, 10 male and 10 female). Seven infants participated in the study but their data were discarded due to fussiness (1) or lack of interest in the auditory stimuli (6). All participants were normally developing infants with normal hearing from Providence, RI, USA, and had parents and caregivers who were native speakers of American English. The testing method was identical to that of the older age group in Experiment 1.

3.3. Results and Discussion

With the modified stimuli, the infants preferred the disfluent passages, looking to the disfluent passages for an average of 7.7 s compared with 6.1 s for the fluent passages. This was significant by 2-tailed paired t-test, $p < .05$. Eight out of 20 infants preferred the fluent passages. This is the opposite pattern to that of Experiment 1, in which the infant preferred the fluent passages.

When lexical and grammatical information were removed from the stimuli, the infants still discriminated between the fluent and disfluent passages. However, their preference was reversed – they preferred the disfluent passages. This finding makes sense when one considers that the 20-month-olds have both lexical and grammatical knowledge of their language. While the modified stimuli preserve some of the phonological characteristics of English, they would not sound like English to the infants. Since the prediction of a preference in favor of the fluent passages in Experiment 1 was based on the infants recognizing them as useful input, it is unsurprising that infants might show a preference for the disfluent, more prosodically

complex, passages in Experiment 2. However, the infants' preference pattern does establish that they *discriminate* the fluent and disfluent passages even when lexical and grammatical information about disfluency are unavailable. This suggests that infants in both experiments were responding to the prosodic characteristics of the stimuli. Thus, infants by the end of the second year of life are sensitive to the prosodic characteristics of disfluencies in adult-directed speech.

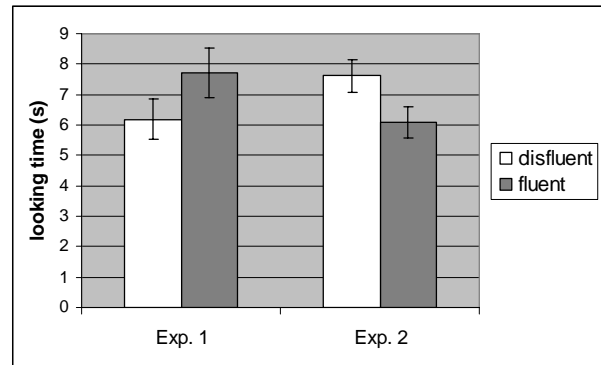


Figure 1: Mean looking times to fluent and disfluent passages, 20-23 month olds, Experiments 1 and 2.

4. Conclusions

Across two experiments, we have demonstrated that older infants are sensitive to the prosodic correlates of disfluency, and that they show a reliable preference for fluent adult English over disfluent English. By contrast, prelingual infants do not show a reliable ability to discriminate fluent and disfluent adult-directed speech.

Our results suggest that ambient adult-directed speech, despite the presence of disfluencies, may play a role in the linguistic development of infants. Younger, prelingual infants may not be able to detect disfluency, at least in adult-directed speech. They may rely primarily or exclusively on the highly fluent, prosodically exaggerated, and simple utterances that characterize infant-directed speech. To the extent that this younger age group makes use of prosodic cues to disfluency, it appears to be limited to infant-directed speech [8]. Perceptual experiments showing preferences for infant- over adult-directed speech have been geared toward these younger ages [e.g. 12]. However, older infants may indeed attend to adult-directed speech. If so, the ability to detect the presence of disfluencies might allow these infants to discard those utterances which might be less reliable as a source of information about the grammar. There is some evidence that infants at this age are able to use linguistic information from indirect input [5]. The current findings suggest that infants may come equipped to differentiate “good” input from “bad” input – and they may be hearing more than we think.

5. References

- [1] Chomsky, N., 1962. Explanatory Models in Linguistics. In *Logic, Methodology and Philosophy of Science*, Ernest Nagel, Patrick Suppes, and Alfred Tarski (Eds.), pp. 528-550, Stanford, CA: Stanford University Press. (cited from Thomas, 2004)

- [2] Hirsh-Pasek, K., & Golinkoff, R. M. 1996. *The Origins of Grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press.
- [3] Kemler Nelson, D.G., Jusczyk, P.W., Mandel, D.R., Myers, J., Turk, A., & Gerken, L.A., 1995. The headturn preference procedure for testing auditory perception. *Infant Behavior and Development*, 18, 111-116.
- [4] Newport, E.L., Gleitman, H., & Gleitman, L.R., 1977. Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C.E. Snow & C.A. Ferguson (Eds.), *Talking to children: language input and acquisition*, pp. 109-149. New York: C.U.P.
- [5] Oshima-Takane, Y., Goodz, E., & Derevensky, J., 1996. Birth order effects on early language development: Do secondborn children learn from overheard speech? *Child Development*, 67, 621-634.
- [6] Ramus, F., Hauser, M.D., Miller, C., Morris, D., & Mehler, J., 2000. Language discrimination by human newborns and cotton-top tamarin monkeys. *Science*, 288, 349-351.
- [7] Santelmann, L., & Jusczyk, P.W., 1998. *Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space*. *Cognition*, 69, 105-134.
- [8] Soderstrom, M., & Morgan, J. L., 2005. Disfluency in speech input to infants? The interaction of mother and child to create error-free speech input for language acquisition. In *Proceedings of DiSS'05*. Aix-en-Provence, France.
- [9] Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W., 2003. The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49, 249-267.
- [10] Soderstrom, M., White, K.S., Conwell, E. & Morgan, J.L., (submitted) Receptive grammatical knowledge of familiar content words and inflection in 16-month-olds. Under review: *Infancy*.
- [11] Weijer, J. van de, 2002. How much does an infant hear in a day? In J. Costa and M. João Freitas (Eds.). *Proceedings of the GALA2001 Conference on Language Acquisition*, pp. 279-282. Lisboa: Associação Portuguesa de Linguística.
- [12] Werker, J. F., & McLeod, P. J., 1989. Infant preference for both male and female infant-directed talk: a developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, 43(2), 230-246.