Constructing the Corpus of Infant-Directed Speech and Infant-Like Robot-Directed Speech

Ryuji Nakamura
Waseda University
Tokorozawa, Saitama,
359-1164, Japan
ryuji.n@suou.waseda.jp

Kouki Miyazawa
RIKEN Brain Science Institute
Wako-shi, Saitama, 351-0198,
Japan
kouki.miyazawa@riken.jp

Hisashi Ishihara
Osaka University
Suita, Osaka, 565-0871, Japan
ishihara@ams.eng.osaka-u.ac.jp

Ken'ya Nishikawa
RIKEN Brain Science Institute
Wako-shi, Saitama, 351-0198,
Japan
nisi@brain.riken.jp

Hideaki Kikuchi
Waseda University
Tokorozawa, Saitama,
359-1164, Japan
kikuchi@waseda.jp

Minoru Asada
Osaka University
Suita, Osaka, 565-0871, Japan
asada@ams.eng.osaka-u.ac.jp

Reiko Mazuka
RIKEN Brain Science Institute
Wako-shi, Saitama, 351-0198,
Japan
mazuka@brain.riken.jp

ABSTRACT

The characteristics of the spoken language used to address infants have been eagerly studied as a part of the language acquisition research. Because of the uncontrollability factor with regard to the infants, the features and roles of infant-directed speech were tried to be revealed by the comparison of speech directed toward infants and that toward other listeners. However, they share few characteristics with infants, while infants have many characteristics which may derive the features of IDS. In this study, to solve this problem, we will introduce a new approach that replaces the infant with an infant-like robot which is designed to control its motions and to imitate its appearance very similar to a real infant. We have now recorded both infant- and infant-like robot-directed speech and are constructing both corpora. Analysis of these corpora is expected to contribute to the studies of infant-directed speech. In this paper, we discuss the contents of this approach and the outline of the corpora.

Author Keywords

speech register; infant-directed speech; language acquisition; human agent interaction.

ACM Classification Keywords

J.4 [Social and Behavioral Sciences]: Psychology.

INTRODUCTION

When talking to infants and young children, adults use a special speech register that is called infant-directed speech (IDS). Different features are observed in IDS relative to adult-directed speech (ADS). These features have been examined across various languages [1]. The features that IDS has shown are the exaggeration of pitch-range, the appearance of the characteristic boundary pitch movement (BPM), the expansion of the vowel space, and giving listeners a positive hearing impression [1–4]. The pitch-range is greater than in ADS and this specific character also appears in pet-directed speech [5]. Therefore, it has been hypothesized that these phenomena appear as the reflection of the speaker's affective expression and intention of "wanting to convey affect". The vowel space in IDS is larger than in ADS and this characteristic also appears in speech during lectures, speech toward foreigners, and that directed to speech recognition systems [4,6,7]. Thus, it has been hypothesized that these phenomena appear as the reflection of the speaker's pronunciation to be easy to understand and intention of "wanting to convey what I'm saying" if the listeners are seen to have a linguistic ability. These results indicate that the expansion of the vowel space has the role of "conveying meaning," the exaggeration of pitch has the role of "conveying affect," and the listener that requires both of these is an infant. Recently however, there have been a few different findings. The variations and variance of the vowel pronunciation in IDS are much greater than in ADS and read speech [8,9]. If there is a greater variance of the vowels, it becomes more difficult to hear. Because of this, evaluating ease in speech hearing by the measure of the vowel space is controversial. Due to an inability to control the appearance, behavior, and linguistic ability of infants, it has been difficult to examine why IDS has such an implication based on the experiments. There have been studies that observe the speech sound's changes of the mothers by controlling the situation to the infants. However, there was a problem that the degree of intervention with regard to the infants is limited. To resolve this problem, we developed the corpus of infant-like robot-directed speech (RDS) to be as similar to an infant as possible and also be controllable. With the infant-like robot we can control the degree of its design and contingency and are able to observe the dynamic transition on a series of dialog. In addition, this allows for recording the dialog in face-to-face, not remote, conditions. Therefore, how the speech is modified when the characteristics of the listeners are changed, can be investigated more directly by using the infant-like robot.

OUTLINE OF THE CORPUS

We recorded two types of corpora. The first was IDS and ADS; the second was RDS and ADS. Twenty-one mothers (all standard Japanese speakers) and their children (11 female and 10 male, aged 18 months ± 3 weeks) were asked to participate for the first corpus. Twenty-one Japanese mothers, different from the first corpus, were asked to participate for the second corpus. For both corpora the same post-graduate student (male, aged 21) listened to the ADS. We used an infant-like robot "Affetto" [10] that was developed by Minoru Asada laboratory, Osaka University, Japan. Affetto (Figure 1) is able to move the joints of its neck, arm, and trunk and its movement is quiet and smooth because of its air pressure controller. Affetto has only the upper half of the body and it was attached firmly to the table. Affetto can vocally express for the speech of the participant by producing monosyllabic synthesized voice /aH/[1] from the speaker attached to its neck.

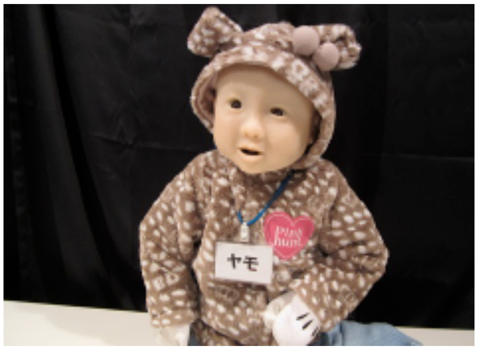

Figure 1. Affetto [10].

Recording of speech was conducted at the Laboratory for Language Development, Brain Science Institute, RIKEN. The recording environment comprises the experiment room and a control room. Participants are in the experiment room with the experimenter. The experiment organizer and the robot operator are in the control room. The participant sits on the small chair and faces the robot on the table. The experimenter is with the participant and explains the procedure. When the participant is talking to the robot, their speech is recorded through the participant's headset microphone. During the conversation, videos are also recorded by the three cameras, which are on the wall adjacent to the participant, upper side of the back, and front upper. The organizer and operator supervise the experiment and control the robot in real time while watching the participant through the camera. In all of the conditions—infant-directed, infant-like-robot-directed, and adult-directed—we recorded a conversation held while using toys. The names of toys were determined to nine kinds of nonsense words that consist of three moras (manarji, namjirji, manura, etc.) For analyzing the salience of the vowels, the words began with /i/, /u/, and /a/, their accents fall on the first syllable, and their second mora was a nasal sound that least affects the nearby vowels. We let the participant practice the pronunciation in advance. The total of the recorded time is shown in Table 1. To analyze the acoustic features of the corpus, we are tagging transcriptions, morphological information, and X-JToBI labels based on the schemes developed for the Corpus of Spontaneous Japanese (CSJ) .

	Per person time	Total time
RDS+ADS	50 min (RDS 25min)	Approx. 18 h
IDS+ADS	30 min (RDS 15min)	Approx. 11 h

Table 1. Recording time of the corpus.

IMPRESSION RATINGS

An evaluation experiment by a third party was conducted to investigate how the hearing impression of RDS corresponds with that of IDS. Raters were Japanese students with no difficulty in hearing (aged 18–24). Three spoken sounds that uttered a toy name of the corpus of IDS, RDS, and

[1] STRAIGHT[11], the package of MATLAB was used for developing the speech sound that is heard like a baby's cooing by stretching the spectral envelope of the adult's speech sound.

ADS were extracted and shuffled at random, and then were rated on seven points of linear measure (7: obviously IDS to 1: obviously ADS). We obtained 10 evaluation values for each token. Paired *t*-tests were conducted in ADS (corpus 1, ADS1) and IDS and in ADS (corpus 2, ADS2) and RDS using a Bonferroni correction, and unpaired *t*-tests were also conducted in the other combination with using. The result of the impression ratings is indicated in Figure 2. RDS that we recorded were rated, as well as IDS, as significantly more "infant-directed speech-like" than ADS (ADS1-IDS(paired): t(19) = -6.15, p < 0.001, ADS2-RDS(paired): t(19) = -7.95, p < 0.001, IDS-RDS(unpaired): t(34.02) = 0.06, p > 0.1, ADS1- ADS2(unpaired): t(37.91) = -0.85, p > 0.1).

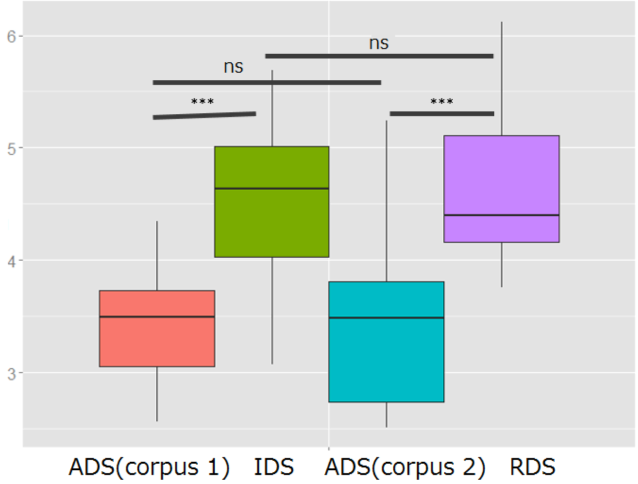

Figure 2. The result of the impression ratings.

DISCUSSION

The impression ratings has indicated that the hearing impression of infant-like-robot directed speech recorded by this study did not differ from IDS. Thus, we suggest that similar speech to that addressed to real infants can be induced by the proposed approach using Affetto [10]. This study showed the effectiveness of adding infant-like robot in infant-directed speech. We will conduct detailed phonetic analysis of the corpus that is being constructed. By revealing the similarity and difference of RDS and IDS it would be expected to enable a more accurate analysis of the acoustic features of IDS.

REFERENCES

1. Anne Fernald. 1989. Intonation and Communicative Intent in Mothers Speech to Infants – is the Melody the Message. *Child Development* 60, 1497-1510.
2. Yosuke Igarashi, Ken'ya Nishikawa, Kuniyoshi Tanaka, and Reiko Mazuka. 2013. Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech. *JASA* 134, 2:1283–1294.
3. Patricia K. Kuhl, Jean E. Andruski, Inna A. Chistovich, Ludmilla A. Chistovich, Elena V. Kozhevnikova, Viktoria L. Ryskina, Elvira I. Stolyarova, Ulla Sundberg, and Francisco Lacerda. 1997. Cross-Language Analysis of Phonetic Units in Language Addressed to Infants. *Science* 277, 5326:684–686.
4. Maria Uther, Monja A. Knoll, and Denis Burnham. 2007. Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication* 49, 1:2-7.
5. Denis Burnham. 2002. What's New, Pussycat? On Talking to Babies and Animals, *Science* 296, 1435.
6. Masanobu Nakamura, Koji Iwano, and Sadaoki Furui. 2008. Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech & Language* 22, 2:171–184.
7. Denis Burnham, Sebastian Joeffry, and Lauren Rice. 2010. Computer-and human-directed speech before and after correction. *In Proc. of SST* 2010, 13-17.
8. McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. 2013. Infant directed speech and the development of speech perception: enhancing development or an unintended consequence? *Cognition*, *129*(2), 362–78.
9. Kouki Miyazawa, Hideaki Kikuchi, Takahito Shinya, and Reiko Mazuka. 2009. The dynamic structure of vowels in Infant-directed speech -RIKEN Japanese Mother-Infant Conversation Corpus-. *IEICE Technical Report* 109, 308:67–72. (in Japanese)
10. Hisashi Ishihara and Minoru Asada. 2013. "Affetto": towards a design of robots who can physically interact with people, which biases the perception of affinity (beyond "uncanny"). *IEEE International Conference on Robotics and Automation*.
11. Hideki Kawahara. 2006. STRAIGHT, Exploration of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. *Acoustic Science and Technology* 27, 6:349-353.