Multi-scale analysis of vocal coordination in infant-caregiver daily interaction

Jiarui Li
International Research Center for Neurointelligence (WPI-IRCN)
The University of Tokyo
Tokyo, Japan
ORCID: 0000-0003-3306-7810

Marisa Casillas
Comparative Human Development
University of Chicago
Chicago, USA
mcasillas@uchicago.edu

Sho Tsuji and Yukie Nagai
International Research Center for Neurointelligence (WPI-IRCN)
Institute for AI and Beyond
The University of Tokyo
Tokyo, Japan
shotsuji@ircn.jp, nagai.yukie@mail.u-tokyo.ac.jp

***Abstract*—Infants participate in vocal coordination with others early in their life, even before they can rely on linguistic cues. They react sensitively to caregiver vocalizations, for instance, by imitating the caregiver and/or modulating their own vocalizations. When talking to an infant, caregivers also modulate their vocalizations, e.g., talk more slowly or with exaggerated prosody, which might attract infants' attention and increase the clarity of vocal information. However, it is still unclear to what extent both parties' vocal modifications dynamically influence each other. In this study, we investigate infants' and caregivers' vocal coordination in their daily interactions by applying multi-scale analysis on a global scale (i.e., a day), a middle scale (i.e., a conversational block), and a local scale (i.e., a turn). The day-long auditory recording data of nine infants, ages two months to three years, and their caregivers were analyzed. The results revealed that infants' and caregivers' vocalizations are differently coordinated on each timescale. On a global scale, infants and mothers react sensitively to each other's vocalizations. Their conversation length varies across a day with a decreasing tendency. On a middle scale, infant-caregivers' prosodic alignments increase over multiple turns in a conversation, indicating a continuous influence between them. Finally, more fine-grained analyses found that pitch-related features and pitch contours are aligned in each turn. The multi-scale analysis reveals the complexity of infant-caregiver interaction in the natural social environment, which inspires us to investigate the benefits of alignment in infants' language learning at different timescales.**

Keywords—infant-caregiver interaction, vocal exchanges, vocal coordination, multi-scale analysis

I. Introduction

Infants' ability to take turns while exchanging vocalizations with other people presents early in development [1]. Gratier et al. [2] found that infants already take turns in proto-conversations at two months of age, preceding the ability to understand the linguistic content of a conversation or to produce meaningful language. Thus, young infants start actively participating in vocal exchanges, raising the question of what kinds of cues they use to do so.

One way to look into this question is to ask to what extent infants and caregivers coordinate to each other's vocalizations. During vocal exchanges, speakers are found to modulate the characteristics of their productions to match their interlocutor's characteristics, which is often referred to as alignment of vocalizations. Many studies have researched the alignment during adult conversation from different perspectives according to different levels, such as matching contents, speech rate, pauses, and other acoustic features [3, 4]. As mentioned above, infants start to take turns from a very young age. Do they also align vocalizations with others? Several studies have investigated infant-caregiver prosodic alignment. For example, in [5], an alignment in pitch-related features was investigated while infants and caregivers were playing picture naming games. Other studies [6, 7] conducted similar analyses but used day-long recordings. These studies reveal that infants and caregivers do align with the other's vocalization.

However, previous studies statically analyzed alignments on a limited timescale, i.e., a turn, which makes it difficult to picture the overall perspective of vocal coordination between infants and caregivers. In natural daily life, infants and caregivers exchange varieties of vocalizations. Their vocal coordination changes *dynamically* and is impacted by many factors, e.g., the environmental factors, including activities they are doing and the existence of other people, and the internal factors, including genetic factors and personalities. More specifically, infants' and caregivers' vocal coordination timeline varies during the day when doing different activities, while its prosodic features vary in a conversation because of the continuous alignment. Even in a single turn, the utterances are not static but with complex patterns, which match dynamically across interlocutors. Furthermore, the tendency of such *dynamic changes* may be identical/differ on timescales by considering different factors. Thus, it is interesting to investigate infant-caregiver's vocal coordination on multiple timescales to identify the hierarchical structure of the vocal coordination in varieties of perspectives.

In this study, we introduce a multi-scale analysis to investigate infants' and caregivers' vocal coordination.

Different methods are used at different timescales to reveal different temporal dynamics. In the global timescale, i.e., a day, Cross Recurrence Plot (CRP) [8] is used to analyze the timeline of vocal coordination. Dynamic Prosodic Alignment Analysis is utilized in the middle timescale, i.e., a conversational block, to observe the ongoing influence between infants and caregivers. Finally, in the local timescale, i.e., a turn, we use Dynamic Time Warping (DTW) [9] technology to measure the dynamic alignment of the prosodic features. The vocal coordination between infants and caregivers is comprehensively pictured by using this multi-scale analysis. The results revealed multiple dynamics about their vocal exchanges progress. Moreover, the vocal coordination procedures are found to be distributed hierarchically with different tendencies, which is helpful to investigate further the benefits of vocal exchanges to infants' learning of vocal interactions from different aspects.

The remains of this paper are organized as follows. Section II describes the corpus and the preprocessing procedures of the speech signals used in this study. Section III illustrates the analysis methods in multi-scales. The results are presented in Section IV. Finally, Section V discusses the results and gives future perspectives from the current study.

II. Data Collection and Preprocessing

A. Data collection

The data used for the analysis come from the Casillas HomeBank Corpus [10]. Speech signals of ten infants ranging in age 0;2-3;0 and those of their mothers speaking native Tseltal were collected and annotated with transcripts. A lightweight stereo audio recorder (Olympus WS-832, Olympus Corporation, Tokyo, Japan), which was put in a vest worn by the infants, was used to record their natural vocalization at home. Over 9-11 hours of data were collected while no intervention by an experimenter. The recording started at 8-10 AM and finished at 6-7 PM. During the infants' waking hours, they are usually tied to their mothers' backs while the mothers do daily activities. If the mother goes to work in the field, the infant is sometimes left with other family members and sometimes taken along. Infants' vocal activities were all recorded, including the interactions with their mothers (i.e., the target of our analysis), other family members, neighbors, and soliloquize. The basic information of the participants is shown in TABLE I.

TABLE I. Basic Information about the Infants and their Mothers

No.	*Infants' Age*	*Infants' Sex*	*Mothers' Age*	*Number of the selected utterances (Infant, Mother)*
1	0;01.25	M	26	(177,148)
2	0;03.18	M	22	(441,93)
3	0;05.29	F	17	-
4	0;07.15	F	24	(206,199)
5	0;10.21	M	24	(269,215)
6	1;02.10	M	21	(488,177)
7	1;10.03	F	31	(204,231)
8	2;02.25	F	17	(570,319)
9	2;08.05	F	28	(406,134)
10	3;00.02	M	28	(175,162)

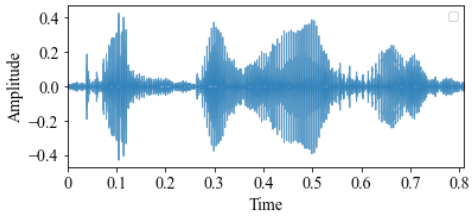
B. Preprocessing and prosodic features extraction

The speech of every speaker was manually annotated by trained native annotators. The raw audio data were first cut into utterances, and then only the utterances of the target infant (4669 clips) and mother (3507 clips) were selected for the analysis. Among them, 1504 infants' clips and 1593 mothers' clips with overlapping speech or with much noise were removed from the dataset to guarantee the quality of the prosodic feature extraction described next.

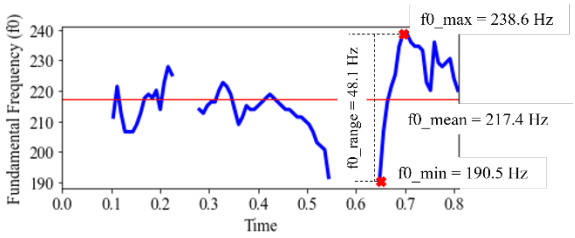
The Librosa package with version 0.8.0 of python was used to extract prosodic features of the audio clips. This study used the fundamental frequency (f0) to represent the pitch. The pitch-related features, including the maximum value (f0_max), minimum value (f0_min), mean value (f0_mean), and the range of f0 (f0_range), measured as the difference between the max and min, were calculated using the customer-written scripts in python. Figs.1 (a) and (b) show an example of the speech signal and the identified pitch of the mother (No.1).

The boundary of [120Hz, 600Hz] was used to filter the mothers' vocalization, and [120Hz, 1000Hz] was given to the infants' ones [11]. 122 infants' clips and 149 mothers' clips that failed to extract pitches in the given boundaries were further removed from the dataset.

After obtaining f0, linear interpolation was conducted to fill the missing values in each utterance. The pitch signals were normalized for each person to resolve individual and age-related differences.



(a). The original signal of an utterance



(b). The pitch of the utterance

Fig. 1. The original signal and the identified pitch of an utterance from the mother (No.1)

C. *Definitions of a turn and a conversation block*

The relationship between three timescales we analyzed are illustrated in Fig. 2: a day (top), a conversational block (middle), and a turn (bottom). A conversational block was first identified in a day-long recording if the infant's and mother's speech were produced consecutively with a pause shorter than 3 seconds [12]. Next, a turn was identified if the pause between their speech was shorter than one second. A turn initiated by the infant was named as I-M turn, whereas a turn led by the mother was named as M-I turn. Other utterances with no reply were defined as Non-turn.

Young infants do not always produce meaningful words according to the contents as observed in adults' conversations. Therefore, turns and conversational blocks were defined based on their pauses. One second, which is the average length of infant-caregiver turn-taking time found in previous studies [1], was chosen to define a single turn. In contrast, the longest pause around three seconds, no matter intra-interlocuter or inter-interlocuters, was chosen to define a conversational block.

Under the limitations of pitch extraction and turn definitions, there were no valid turns and conversation blocks in the data of No.3. Thus, the other nine infants and their mother's data were finally used for the analysis. In total, 1678 utterances from the mothers and 2936 utterances from the infants were used for the analysis. The numbers of the data used in each dyad are shown in TABLE I.

III. Methods

During the daily vocal exchange, infants would practice different interaction manners. In order to make a suitable response to the interlocutor, they need to understand the contents and give a response at an appropriate time. In addition, they sometimes mimic the vocalizations they hear to learn the language. In order to comprehensively investigate infants' learning progress through interactions, we propose a multi-scale analysis that applies different methods to different timescales. On the global scale (i.e., a day), Cross Recurrence Plot (CRP) was introduced to evaluate the infant's sensitivity to the mother's vocalizations. The alignment of their speech onset was examined to reveal the global coordination. In the middle scale (i.e., a conversational block), we analyzed the prosodic similarity between their vocalizations. This analysis aimed to measure the dynamic influence between the infant's and mother's speech. Finally, in the local scale, the temporal alignment of the pitch contours was measured in each turn. How similarly/differently the infant and mother modified their speech was investigated. The following sections explain the technologies used to analyze each timescale.

A. *Cross Recurrence Plot – A Day Scale*

A CRP is a graph that shows all those times at which a state in one dynamical system occurs simultaneously in a second dynamical system [8]. CRP has been used to analyze infant-caregiver's interaction activities. For instance, it is used in [13] to investigate infants' and mothers' motion coordination while playing with different toys, which inspires us to extend CRP in the analysis of vocal coordination. CRP-based Quantification Analysis (CRQA) is the nonlinear method that provides the quantification of dynamic systems and their trajectories. In [14], the authors utilized CRQA to quantify the overlaps between the onsets of infants and their mothers' vocalizations. Our study applies CRQA to reveal the global dynamics of vocal exchanges between an infant and a mother, that is, how sensitive the infant and mother are to the partner's speech. We further analyze the length of the conversational blocks in a day to see whether there is a change in infants' sensitivity to the mother's vocalization.

A CRP is a $M \times N$ matrix, where N is the number of data points of the temporal stream on the x-axis and M is the one on the y-axis (here, $M = N$). Our study aims to analyze the onset time of infants' and mothers' vocalizations. For any time when the mother's vocalization onset at time point T_i and infant's vocalization onset at time point T_j, value 1 is given to R_{ij} on CRP presented as a black dot on the graph. If the two events onset at the same time, a dot occurs on the diagonal. However, infants' and mothers' vocalizations start usually in turn with a latency. Thus, we are interested in the value of R_{ij} nearby the diagonal. According to the criteria defined in Section II.C, the

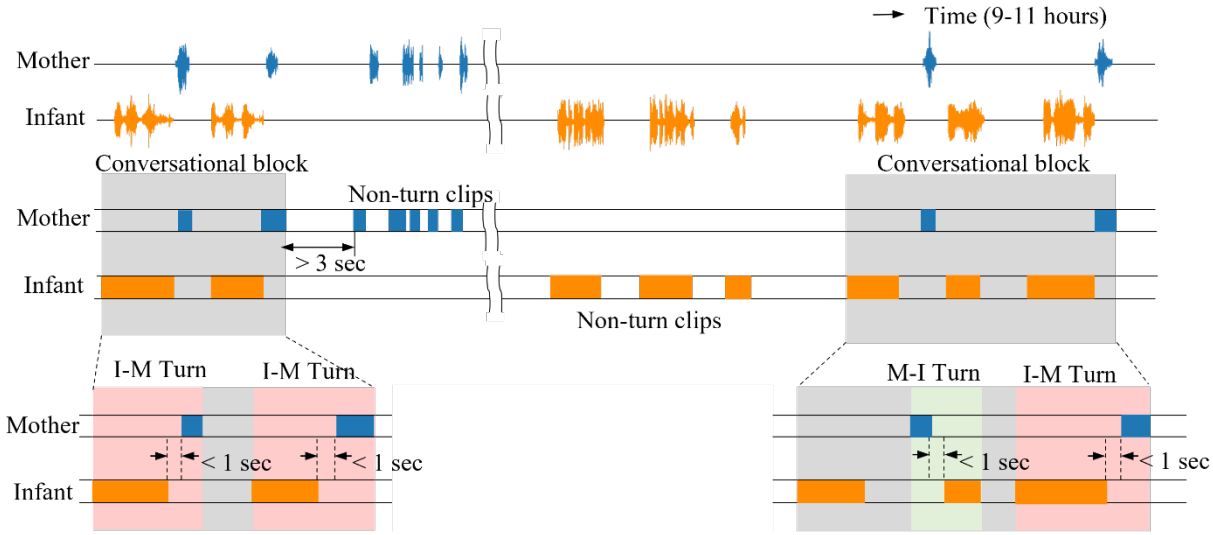


Fig. 2 Multi-scale analysis for a day (top), conversational blocks (middle), and turns (bottom). Data are from infant No.10 and his mother. A conversational block and a turn were identified when the pause between consecutive speeches was shorter than 3 secs and 1 sec, respectively

longest conversational block found in the presented data is 58.68 seconds. Thus, the Region of Interest (ROI) is set as a squared window lasting 60 seconds across the diagonal. We investigate the sensitivity of infants' and mothers' vocalizations to each other using the following equation,

$$Sensitivity\_k = \sum R_{ij}\ ,60 * (k-1)\ \leq\ i,\ j \leq 60 * k,\ k = 1, \dots, Int(\ Total\ time/60) \quad (1)$$

where the $Total\ time$ is the total recording time in seconds of each day long audio date. k is the number of the window. A higher value of the sensitivity presents a higher amount of infants' and their mothers' responses to each other's vocalizations than randomly making a vocalization or interacting with other people. A sum of the sensitivities in every minute is calculated and correlated with the time to investigate its change through a day.

B. Entrainment Analysis – Conversational Block Scale

To investigate the prosodic alignment in a conversational block, the similarities of the pitch-related features (i.e., f0_max, f0_min, f0_mean, and f0_range) between an infant's and a mother's speech are analyzed. We use Equation (2) to present the prosodic similarity, then correlate it with the turn numbers in a conversational block.

$$Similarity = -abs(Pitch\_M - Pitch\_I) \quad (2)$$

where $Pitch\_M$ and $Pitch\_I$ are pitch-related features of the mother's and her infant's vocalizations. Positive correlations ought to be found when the alignment is enhanced through a conversation. However, each turn in a conversational block sometimes contains more than one utterance from the same interlocutor . For example, as shown in Fig.2, there are two utterances from the infant in the first turn in the latter block. Since there is no reference to the content in infants' speech, it is hard to identify their vocalizations directing to which utterances in a conversation. Thus, in such a situation, the features of the utterances from the same interlocutor in the same turn will be averaged. The differences of the averaged values between the two interlocutors will be used in correlation.

C. Dynamic Time Warping – Turn scale

We focused on pitch patterns in the turn-level analysis to examine whether the infant and caregiver align their vocal rhythm. Dynamic Time Warping (DTW) is an algorithm to measure the similarity between two temporal series with varying lengths [9]. It calculates the optimal match between two given sequences. The key idea is to build one-to-many and many-to-one matches to minimize the total distance between two signals. Since the vocal length of mother and infant varies, we calculate the DTW distance between their pitch contours to present the similarity of pitch patterns. Fig. 3 shows an example of the DTW distance between the mother's and infant's pitch contours in an M-I turn.

There is no golden standard to evaluate how short a DTW distance is, and it depends on the data used for analysis. According to [15], however, the DTW distances of the frequency waves are around 12-15 if the same person produces the same sentences, whereas they become higher than 200 if the sentences are from different conversations. In this study, we assume that a DTW distance under 15 presents that the two vocalizations are similar. Furthermore, we compared the DTW distances in true turns with pseudo turns, made up of the same utterance from the leading person with a randomly selected utterance from the non-turn clips.

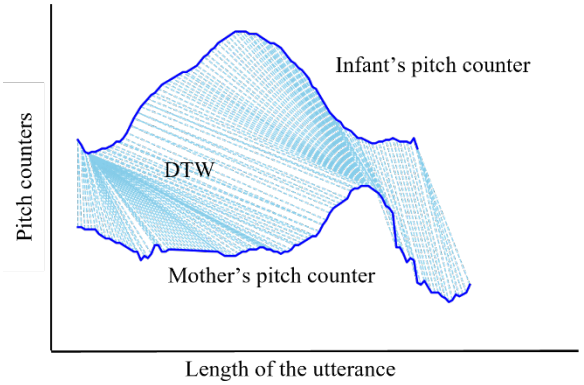


Fig. 3. An example of the DTW distances between the mother's and infant's pitch contours from infant-caregiver dyads No.7. In a conversational turn, the blue line represents an infant's and mother's pitch contour (normalized f0). The dotted lines are the optimal matches between the pitches, and the sum of the matches is the DTW distance between two pitch contours.

IV. Results

A. Cross Reccurrence Plot - Vocal Coordination Decreases on A Day Scale

Fig. 4 shows four examples of the CRP of infant-mother dyads (No. 1, 4, 5, and 10). Around the diagonal of the CRP, several blocks exist in the whole day-long recording data, showing that the infant and mother exchange vocalizations frequently on a global timescale. The infants' and mothers'

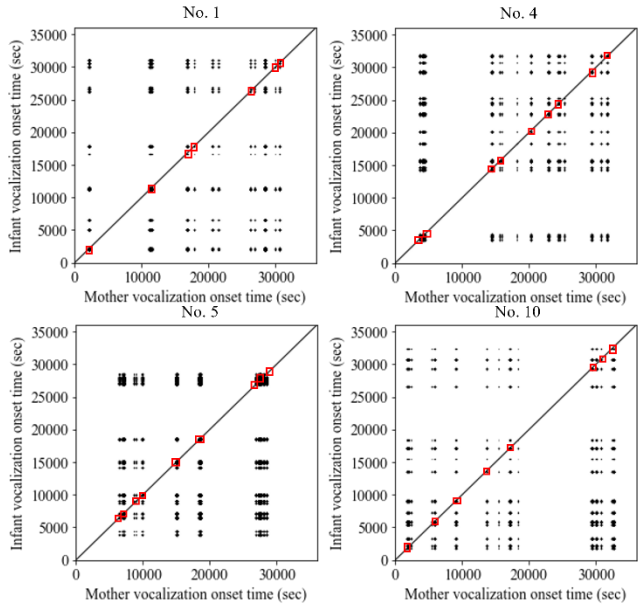


Fig. 4. The CRPs of 4 infant-mother. The black dots show the recurrence of the vocalizations. The red rectangle shows the continuous vocal exchanges between the two interlocutors.

sensitivities to each other's vocalizations are quantified by Eq. (1). The individual differences of the vocal onsets are large across dyads. We did not find a common patterns of infants' and mothers' sensitivities to each other's vocalizations.

Next, the temporal change of the sensitivity on the day scale was investigated using Eq. (1). As shown in Fig.5, a higher value of $\sum R_{ij}$ was found at the beginning of the recordings (i.e., in the morning) while it becomes lower in the evening. This result suggests that the amount of continuous vocal exchanges between infants and mothers significantly decreases across a day. (r = - 0.178).

B. *Entrainment Analysis – Similarities of Prosodic Features Increase on Conversational Block Scale*

Fig. 6 shows the correlation of the negative pitch feature differences with the number of turns in a conversational block. Although the results are not significant, positive correlation coefficients are found in all the pitch-related features with the increase of the turn numbers (f0_max: r=0.030, f0_min: r=0.030, f0_mean: r=0.033, f0_range: r=0.044). These results suggest that mothers and infants influence each other continuously in a conversational block. With the increase of the turn numbers, their pitches become more similar to each other by modifying their own vocalizations.

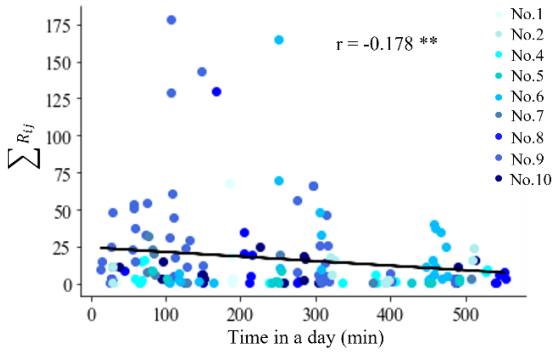


Fig. 5. The sensitivity changing of the infants toward mothers' during a day

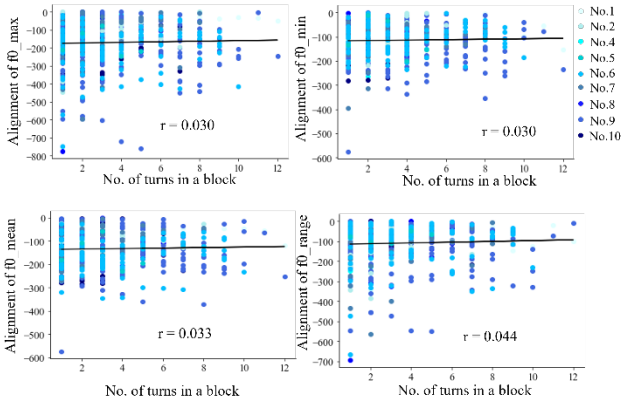


Fig. 6. Infants' and mothers' vocalizations become more similar across conversational blocks

C. *Dynamic Time Warping – Pitch patterns align on Turn scale*

We first calculated the correlations between infants' and mothers' pitch-related features in M-I turns and I-M turns separately. Tables II summarizes the results. In M-I turns, weak but significant correlations are found in the maximum, minimum, and mean values of f0. In I-M turns, weak but significant correlations are found in the maximum, minimum, mean values, and the range of f0. Thus, in the turn-level alignment analysis, infants and mothers' pitches are similar to each other, which shows that the alignments occur when they are exchanging vocalizations. This result is consistent with the previous research [5]. However, in M-I turns, no significant alignment was found in F0_range, indicating that young infants still do not capable to control the range of pitch changes in an utterance to coordinate mothers' speeches.

Next, the DTW distances evaluating the pitch pattern's similarity were analyzed. Fig. 7 shows the comparisons of DTWdistances in the M-I turns and I-M turns. The mean values of the M-I and I-M turns are 12.68 and 14.32, respectively, which shows that mothers' and infants' vocalizations are similar to each other in every single turn. Furthermore, no matter M-I or I-M turns' DTW distances are significantly shorter than pseudo turns, indicating that infants and mothers align each other's vocalizations when they are involved in vocal exchanges. However, although it is not significant, the DTW distances in M-I turns are slightly shorter than in I-M turns, indicating that infants tend to mimic their mothers' vocal patterns more strongly than mothers do.

TABLE II. CORRELATIONS IN THE TURNS

	M-I Turns		*I-M Turns*	
Features	***r***	***significance***	***r***	***significance***
F0_range	0.075	-	0.148	***
F0_max	0.235	***	0.264	***
F0_min	0.242	***	0.223	***
F0_mean	0.312	***	0.247	***

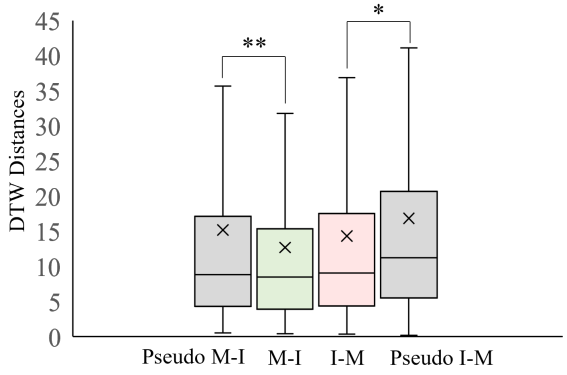


Fig. 7. DTW distances between mother's and infant's vocalizations in M-I turns and I-M turns

V. Discussions

Infants start to take turns from others' vocalization from two months old. Through vocal exchanges, they practice responding to their partner frequently in a short time. Investigating infants' vocal exchanges with others helps to understand the development of language learning procedures. Through exchanging vocalizations with people around them, infants learn to identify the vocalizations needed to give a reply compared with others. Also, they are trying to imitate others' vocalizations to give a speech-like response. Furthermore, except for mimicking voice as immediate feedback to others' vocalization, infants modify their vocalizations in the conversation to improve their language ability. It is impossible to investigate all the aspects of vocal coordination only by analyzing the alignment on a local timescale. Moreover, the dynamic change in such complex vocal coordination behavior between infants and caregivers varied on different timescales. Thus, we proposed a multi-scale analysis to evaluate infants' and caregivers' vocal alignment using different technologies in this study.

We found several interesting results about infants' and caregivers' vocal alignment in different timescales. In the global timescale, i.e., a day level, we found that mothers' and infants' vocal exchanges last longer at the beginning and middle of a day compared with the end. To explain this result, environmental factors, e.g., the activities conducted by the infants and whether other people exist in the environment, should be taken into account. Then, the influence of these factors on the vocal exchangeability of infants can be further considered. The entrainment between mothers and infants can be investigated on the conversational block scale. In the presented study, we found that mothers' and infants' vocalizations become more similar at the end of the blocks than at the beginning. The results show that more information can be obtained when investigating infants' interactions with their caregivers dynamically, which cannot be revealed by simply analyzing each local turn. Furthermore, in the local timescale, i.e., turn level, except for the usually used summarized pitch-related features, we introduced DTW distance to measure the similarity of pitch patterns. Comparing the summarized values, DTW distance is able to present the detailed change in pitch contour, which measures the alignment of the vocalizations more accurately.

Thus, our results show that the vocal coordination between infants and caregivers is complex on multi-scales. Difference dynamic changing tendencies are found on different timescales, i.e., decreasing sensitivity in a day, increasing similarity in a conversation, and better coordination from infants than caregivers. The different changing tendencies may relate to varieties of factors and can be connected to multiple perspectives of infants' language learning. It is also vital to use different technologies in different timescales and comprehensively picture the interactions.

Additionally, a contrast tendency of alignments were found between the block and day scales analysis, enlightening us that infants' learning procedures may not always following a growing trend, e.g., coordination get better across a conversation vs get worse at the end of a day. Noted here, that the effects of "better" or "worse" vocal coordination on infants' development have not been verified in this study. A decreasing in alignment on a day scale does not mean it works negatively on language learning. Also, the corpus used in this study has its specialty, e.g., Tseltal caregivers produce less infant directed speeches comparing with other language speakers [16]. In our future work, the vocal coordination on different timescales will be further investigated across cultures so that their contributions on language learning and infants' development can be further revealed.

Acknowledgment

This research was supported by JSPS KAKENHI (Grant Number: 21H04981), Japan, by Institute for AI and Beyond, The University of Tokyo, Japan, and by the World Premier International Research Center Initiative (WPI), MEXT, Japan.

References

[1] Casillas, M. (2014). Turn-taking. *Pragmatic development in first language acquisition*, pp. 53-70.

[2] Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in psychology*, Vol.6, No. 1167, pp. 1-10.

[3] Ostrand, R., & Chodroff, E. (2021). It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. *Journal of phonetics*, Vol.88, pp.101074.

[4] Truong, K. P., & Heylen, D. (2012). Measuring prosodic alignment in cooperative task-based conversations. *In Thirteenth Annual Conference of the International Speech Communication Association*.

[5] St. Pierre, T., Cooper, A., & Johnson, E. K. (2021). Cross-generational Phonetic Alignment between Mothers and Their Children. *Language Learning and Development*, pp. 1-22.

[6] Ko, E. S., Seidl, A., Cristia, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of child language*, Vol. 43 No.2, pp.284-309.

[7] Seidl, A., Cristia, A., Soderstrom, M., Ko, E. S., Abel, E. A., Kellerman, A., & Schwichtenberg, A. J. (2018). Infant–mother acoustic–prosodic alignment and developmental risk. *Journal of Speech, Language, and Hearing Research*, Vol. 61 No. 6, pp.1369-1380.

[8] Marwan, N., Thiel, M., & Nowaczyk, N. R. (2002). Cross recurrence plot based synchronization of time series. *Nonlinear processes in Geophysics*, Vol. 9 No. 3, pp. 325-331.

[9] Müller, M. (2007). Dynamic time warping. I*nformation retrieval for music and motion*, pp. 69-84.

[10] Casillas, M., Brown, P., & Levinson, S. C. (2017). Casillas HomeBank corpus. https://doi.org/10.21415/T51X12

[11] Seidl, A., Cristia, A., Soderstrom, M., Ko, E. S., Abel, E. A., Kellerman, A., & Schwichtenberg, A. J. (2018). Infant–mother acoustic–prosodic alignment and developmental risk. *Journal of Speech, Language, and Hearing Research*, Vol.61 No.6, pp. 1369-1380.

[12] Stern, D. N., Beebe, B., Jaffe, J., & Bennett, S. L. (1977). The infant's stimulus world during social interaction: A study of caregiver behaviors with particular reference to repetition and timing. *Studies in mother-infant interaction*, pp.177-202.

[13] Xu, T., & Yu, C. (2016). Quantifying Joint Activities using Cross-Recurrence Block Representation. *In 2016 Cognitive Science Society*.

[14] Leonardi, G., Nomikou, I., Rohlfing, K. J., & Rączaszek-Leonardi, J. (2016, September). Vocal interactions at the dawn of communication: the emergence of mutuality and complementarity in mother-infant interaction. *In 2016 joint IEEE international conference on development and learning and epigenetic robotics (ICDL-EpiRob)* , pp. 288-293. IEEE.

[15] Permanasari, Y., Harahap, E. H., & Ali, E. P. (2019, November). Speech recognition using dynamic time warping (DTW). *In Journal of Physics: Conference Series*, Vol. 1366, No. 1, pp. 012091. IOP Publishing.

[16] Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, Vol. 91. No.5, pp. 1819-1835.