Is Turn-Shift Distinguishable with Synchrony?

Jieyeon Woo, Liu Yang, Catherine Pelachaud, Catherine Achard

Is turn-shift distinguishable with synchrony?*

Jieyeon Woo, Liu Yang, Catherine Pelachaud, and Catherine Achard

Institut des Systèmes Intelligents et de Robotique (CNRS-ISIR), Sorbonne University, Paris, France
{woo,yangl,pelachaud,achard}@isir.upmc.fr

Abstract. During an interaction, interlocutors emit multimodal social signals to communicate their intent by exchanging speaking turns smoothly or through interruptions, and adapting to their interacting partners which is referred to as interpersonal synchrony. We are interested in understanding whether the synchrony of multimodal signals could help to distinguish different types of turn-shifts. We consider three types of turn-shifts: smooth turn exchange, interruption and backchannel in this paper. We segmented each turn-shift into three phases: before, during and after, we calculated the synchrony measures of the three phases for multimodal signals (facial expression, head pose, and low-level acoustic features). In this paper, a brief analysis of synchronization during turn-shifts is presented, we also study the evolution of interpersonal synchrony before, during and after the turn-shifts. We proposed computational models for the turn-shift classification task only using synchrony measures. The best performance was obtained with an FNN model using the three phases' synchrony score of all features (accuracy of 0.75).

Keywords: Turn-shift · Synchrony · Neural network.

1 Introduction

During an interaction, people communicate information via verbal and nonverbal channels. Verbal communication transfers information through language containing explicit content. Nonverbal behavior conveys through "body language" including gestures, facial expressions, body movement, and gaze [8]. Intra-synergies are formed within one's own behavior [14].

While the intent is communicated in a direct manner, by emitting multimodal social signals, people also coordinate and adapt their behavior to that of their interlocutors [14] in a continuous manner. Being in sync enables a fluid exchange of information and increases the engagement level [21]. This coordination of behaviors, which may occur unintentionally [40], is also referred to as synchrony and we define it as in [19].

In conversations, speaking turns are exchanged between the interlocutors which is done smoothly or through interruptions. We call this change of turns as

* Supported by ANR-JST-CREST TAPAS (ANR-19-JSTS-0001) and IA ANR-DFG-JST Panorama (ANR-20-IADJ-0008) projects.

turn-shift in this paper. Beattie [3] and Schegloff and Sacks [39] classified turn-shift into three main categories based on simultaneous speech and willingness to yield the floor: smooth switch, interruption, and overlap. Overlap happens at the end of a speaking turn when the listener starts speaking and over-anticipating the end of the current speaker's turn [37]. On the other hand, interruption grabs the floor against the speaker's will when she/he is not finished. Here we also consider the backchannels which are produced by the listener without the intent to grab the speaking turn. Similar to interruption, backchannels always happen during a speaking floor. They may be mistakenly identified as an interruption when conducting real-time analysis of interlocutors' multimodal signals.

We are interested in understanding whether the synchrony of multimodal signals could help to distinguish different turn-shift types along with backchannel via analysis. A predictor (computational model) is built for the classification task using synchrony measures. We focus on dyadic interactions. To our knowledge, we are the first to build a computational model to classify turn-shift types using only synchrony measures.

In this paper, overlap and smooth switch are merged as smooth turn exchange since they are at the end of a turn. Thus, we analyze synchrony measures for the following three turn-shift types: smooth turn exchange, backchannel and interruption.

Related works of turn-shift and synchrony will be introduced in Section 2. In Section 3, the analyzed corpus and the studied features will be explained. The analysis will be shared in Section 4 and our turn-shift type predictors and their results will be presented in Section 5. The paper will be concluded with a brief discussion of the possible future applications and extensions of our work.

2 Related works

Turn-shift during interaction has been an interesting subject of research for a long time. Emanuel A. Schegloff [38] firstly defined conversation sequencing rules. During the course of a conversation, interlocutors dynamically collaborate with each other by yielding and taking the speaking turns based on rules in order to keep the flow of information exchange and maintain the communication [15, 9]. The idea of conversation analysis was then proposed by Harvey Sacks [37] which describes its most basic structure as turn-taking.

Various works analyzed turn-taking, to get a better understanding of the coordination taking place during turn-shifts by looking into multimodal features, such as eye-gaze [17, 26], respiration [25, 27], and head-direction [42]. Linguistic features such as syntactic structure, turn-ending markers, and language model were also investigated [29, 32, 28]. They highlighted the importance of prosodic feature variation (e.g. fundamental frequency $F0$ and intensity) during turn-shifts [22, 30, 43]. Interruptions were observed to be often combined with higher voice energy [41, 23, 24]. These differences in the three turn-shift types might lead to an increased or decreased interpersonal synchrony.

To study the interpersonal synchrony of whether the partners are in sync or not, a multitude of methods were introduced.

Pioneer works consists of manual assessments that rely on trained observers. The synchrony perception was done by directly observing the data on a local time scale using behavior coding methods [12, 16]. For larger time scales, judgment methods were employed [12, 6]. The rating was done using a Likert scale [12, 6].

Manual annotation is a laborious and time-consuming task. This tedious workload was relieved by the appearance of automatic measures. The measures capture relevant signals to detect synchrony. One of the most commonly used measures for interpersonal synchrony is the correlation [11, 18, 35] that calculates the synchrony during a same period. Interlocutors' social signals constantly react to those of the other which leads to behavior coordination. When conversing, the perception of the other interacting partner's behavior is delayed by a certain time period (2 to 4 seconds [13, 31]). Several works consider this time delay by employing the time-lagged cross-correlation [7, 1, 4]. As such behavior signals are shifted in time, but they can also vary in length. Dynamic Time Warping (DTW) [33], which measures the similarity between two temporal sequences while being invariant to speed and length, can address such problems. It is widely used to find common patterns [5]. Some other studies perform spectral analysis to capture the synchrony between signals. The evolution of the relative phase is measured to obtain information related to synchrony stability [34, 36].

For our study, we choose to employ frequently used synchrony measures of correlation (Pearson correlation coefficient), time-lagged cross-correlation, and DTW to study the synchrony of a turn-shift. As explained above, the three measures differ in the way how they measure synchrony. Correlation expresses the linear relation of signals within the same time window. Time-lagged cross-correlation takes into account the time swift between the signals and DTW maps the signals that are shifted in time and differ in length. This leads us to use all three of them.

Prior works analyzed the synchrony of interlocutors' behavior during the entire course of the interaction. They do not specifically look into them during the turn-shift moments.

We want to check if there is a visible link between synchrony and turn-shift which allows synchrony measures to serve as a potential feature for the characterization of turn-shift types. We also intend to verify the usefulness of synchrony measures in classifying the turn-shift types via computational models.

3 Corpus

The NoXi database [10], which contains screen-mediated face-to-face dyadic interactions, was used for this study. The database is made up of 3 parts depending on the recording location (France, Germany, and UK). For our study, we choose to use the French part that contains 21 dyadic interactions performed by 28 participants with a total duration of 7h22.

All turn-shift moments (1403 smooth turn exchanges, 1651 backchannels, and 929 interruptions) were manually annotated following Yang's annotation schema [44].

The turn-shift and backchannel moments were identified on the onset point of the listener's voice activity which we note as t_0. We segmented each moment into three phases:

- **Before:** $t_0 - 6\text{s} \sim t_0 - 2\text{s}$;
- **During:** $t_0 - 2\text{s} \sim t_0 + 2\text{s}$;
- **After:** $t_0 + 2\text{s} \sim t_0 + 6\text{s}$.

We define the three phases of turn-shift (before, during, and after) to refine the detection of different shifts. For each phase, multimodal features were extracted, and the synchrony scores between partners were calculated separately.

For our study, the features employed are the following:

- **Facial features:** AU1, AU2, AU4, AU12, and AU15;
- **Head features:** Head translation and rotation;
- **Acoustic features:** F0 and loudness.

Facial features were obtained using OpenFace [2] and acoustic features were extracted via openSMILE [20].

As the initial head position of the interlocutor may create a bias, instead of using the absolute position we applied the following equation for the head translation, called head motion activity:

$$v_{Head}(i) = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 + (z_i - z_{i-1})^2} \tag{1}$$

where x_i, y_i and z_i are the coordinates of the head position in the image at timestep i.

And for head rotation, we also calculate the head rotation activity:

$$r_{Head}(i) = |x_i - x_{i-1}| + |y_i - y_{i-1}| + |z_i - z_{i-1}| \tag{2}$$

where x_i, y_i and z_i are the head angles according to the 3 axes at timestep i.

A z-score normalization was applied to all features for them to be invariant to the quantity of behaviors of interlocutors.

4 Analysis

To understand the relationship between synchrony and turn-shift types, we analyzed the presented multimodal signals. The significance of the turn-shift type difference was checked via a two-tailed t-test.

We start our analysis by looking at the interpersonal synchrony scores (correlation, time-lagged cross-correlation, and DTW) during the turn-shift (*during* phase of $t_0 - 2\text{s} \sim t_0 + 2\text{s}$) for all features.

Significant differences can be seen for several signals of interruption (Int), smooth turn exchange (ST), and backchannel (BC) with t-test ($p < 0.01$) in

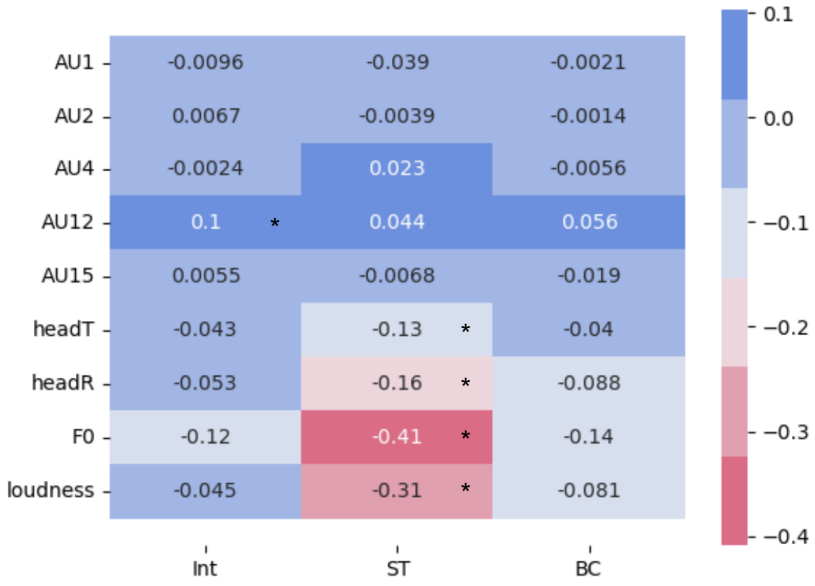

Fig. 1. Correlation of multimodal features *during* Interruption(Int), Backchannel(BC), Smooth turn exchange(ST) (*: $p < 0.01$)

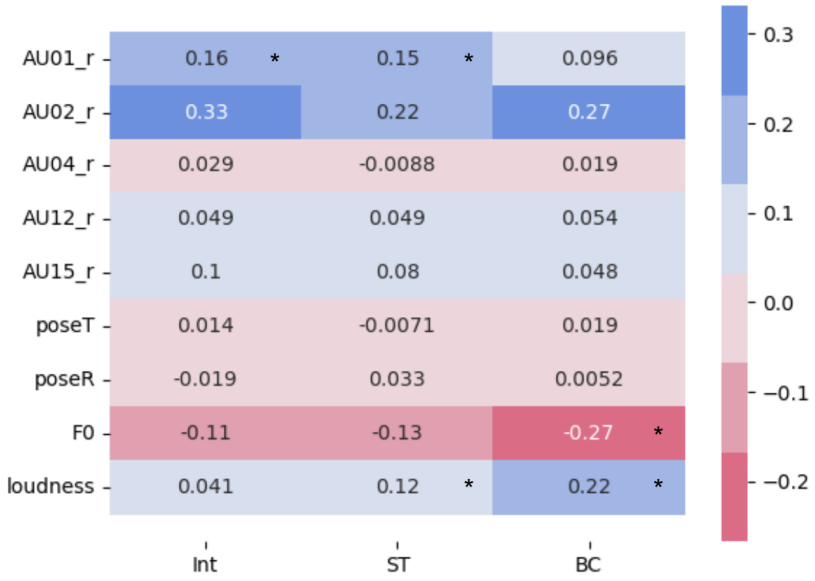

Fig. 2. Time-lagged cross-correlation of multimodal features *during* Interruption(Int), Backchannel(BC), Smooth turn exchange(ST) (*: $p < 0.01$)

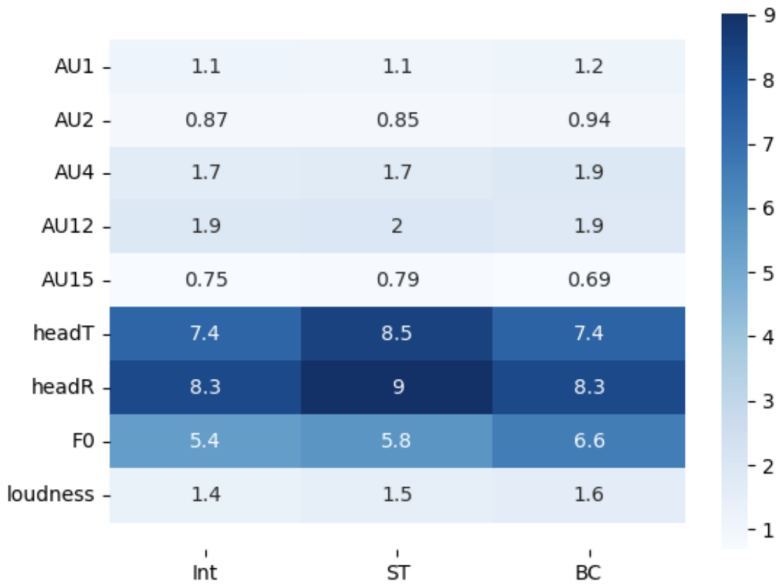

Fig. 3. DTW of multimodal features *during* Interruption(Int), Backchannel(BC), Smooth turn exchange(ST)

Figures 1, 2, and 3. To detail, with the synchrony score obtained through correlation, in Figures 1, smooth turn exchange gets higher negative correlation scores for acoustic and head features (showing opposite trends), these features allow smooth turn exchange to be differentiated from interruption and backchannel. The values of the other two are mostly uncorrelated (close to 0) or comparatively less correlated. Interruption gets a higher positive correlation score for AU12 while smooth turn exchange and backchannel shows no relation (close to 0). Thus, correlation measure can be used to distinguish smooth turn exchange and interruption.

In Figure 2, backchannel is significant for the acoustic features of F0 and loudness using time-lagged correlation. For all three types, a positive correlation can be observed for AU1 and loudness. For AU1, no correlation can be found for backchannel while the other two are positively correlated. An increasing trend of synchrony can be seen in the order of interruption, smooth turn exchange, and backchannel for loudness. F0 is negatively correlated for all three types. A noticeably higher correlation score can be noticed for backchannel compared to the other two. Backchannels can thus be identifiable among the others via time-lagged correlation scores of AU1, F0 and loudness.

Using DTW, in Figure 3, backchannel for F0 and smooth turn exchange for head features are significant. Via DTW, the distance between two signals can be measured, which can be interpreted to be more synchronized when the distance gets smaller. Here we can note a lower sync during smooth turn exchange via the

head translation and rotation compared to interruption and backchannel. In the same manner, a lower sync can be seen for backchannel with F0 compared to the other two. Therefore, DTW can be used to distinguish smooth turn exchange and backchannel.

We can thus identify the three types of smooth turn exchange, interruption, and backchannel using the synchrony measures at the *during* phase of $t_0 - 2\text{s} \sim t_0 + 2\text{s}$.

The usefulness of synchrony scores has been proved for the task of identifying turn-shift types. However, a clearer way of distinguishing them would be more desirable. To do so, we observe the variation of synchrony scores *before*, *during*, and *after* the turn-shifts.

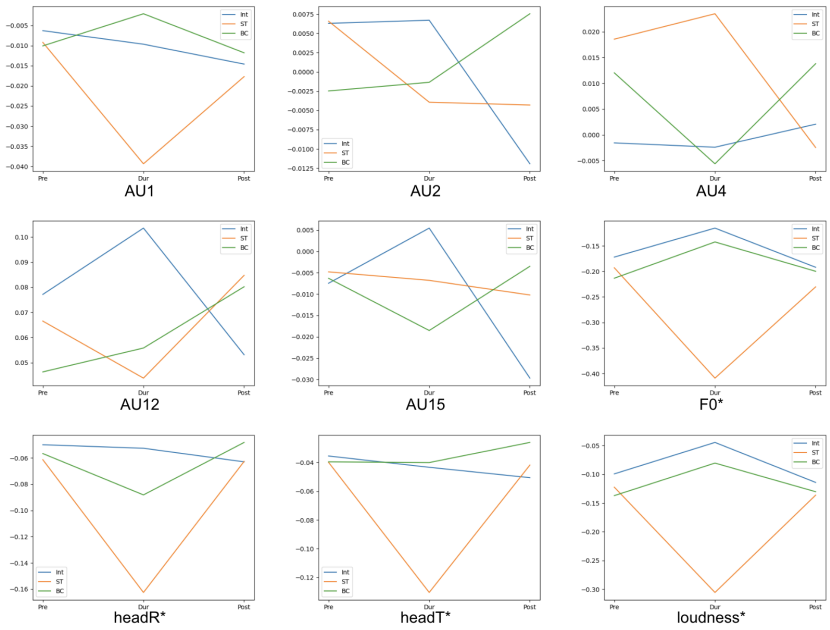

Fig. 4. Correlation of multimodal features *before*, *during*, and *after* Interruption(Int), Backchannel(BC), Smooth turn exchange(ST) (*: $p < 0.01$)

By evaluating the evolution of correlation measures *before*, *during*, and *after* turn-shifts, in Figure 4, for smooth turn exchange we can find a remarkable sudden increase in negative correlation in the features of F0, loudness, head rotation and translation. For these acoustic and head features, a stable trend or only a slight change in synchrony can be observed for backchannel and interruption.

In the same respect, in Figure 5 synchrony evolution trends of acoustic features obtained via time-lagged cross-correlation render significant information. The trends of backchannels are easily distinguishable compared to interruption

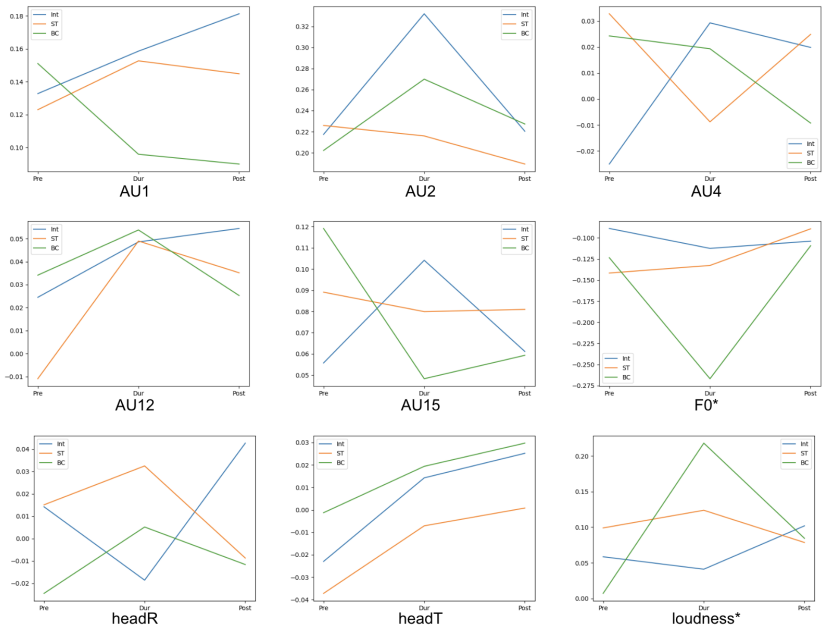

Fig. 5. Time-lagged cross-correlation of multimodal features *before*, *during*, and *after* Interruption(Int), Backchannel(BC), Smooth turn exchange(ST) (*: $p < 0.01$)

and smooth turn exchange. An increase in inverse correlation can be seen for F0, and for loudness, the synchrony score rises during backchannels while there is only a minor change in synchrony score for the other two types.

Figure 6 present the evolution of the DTW synchrony score. Head rotation, F0, and loudness show an increase in synchrony scores during the turn-shifts. This could be interpreted as the turn-shift event causing an effect on interpersonal synchrony.

The difference between the three phases was calculated to check the variation significance of the phase transitions (*before-during* and *during-after*) via the t-test ($p < 0.01$).

The evolution of synchrony scores of the three phases of *before*, *during*, and *after* provided additional information on distinguishing turn-shift types. As seen above, each turn-shift type has different synchrony evolution trends which have been proven to enable the identification of smooth turn exchange, interruption, and backchannel.

5 Turn-shift Classification Models

With the analysis of the relationship between synchrony and turn-shift types, we want to employ synchrony measures in identifying the different turn-shift types to verify their usefulness. For this, we built computational models only using the

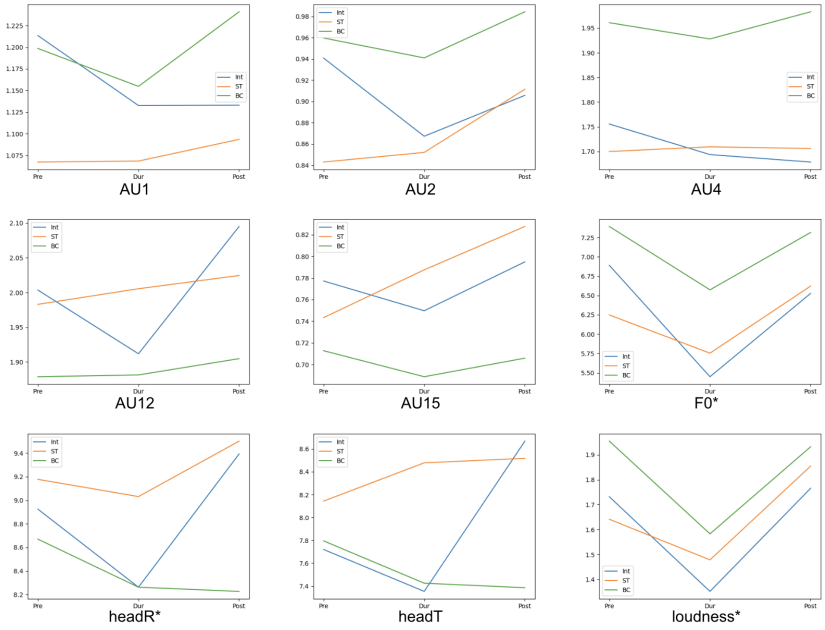

Fig. 6. DTW of multimodal features *before*, *during*, and *after* Interruption(Int), Backchannel(BC), Smooth turn exchange(ST) (*: $p < 0.01$)

synchrony measures of the three segmented phases in the turn-shift classification task.

We approach this task considering two aspects:

- choice of turn-shift phase(s)
- choice of features

As our Model 1, we start by looking into the synchrony measures of all features *during* turn-shifts. We use a feedforward neural network (FNN) to classify the turn-shift types.

In Section 4, we have identified several features by analysis which were significant in differentiating the turn-shift types. We selected these features to check if the same performance could be obtained only by using these features. The selected features for the *during* phase are:

- **Correlation:** AU12 and head translation and rotation;
- **Time-lagged cross-correlation:** AU1, F0, and loudness;
- **DTW:** Head translation and rotation, and F0.

Via Model 1, an accuracy of 0.627, in Table 1, is obtained. With this, we can see that the synchrony measure can be used to identify turn-shift types. Also, the same accuracy of 0.627 is obtained using only the selected features. This

Table 1. Accuracy scores of different turn-shift classification models.

Model	Accuracy
Model 2	**0.750**
Model 2 (selected features)	0.650

supports our analysis that these features indeed are significant in identifying turn-shift types.

We continue by using the synchrony measures obtained through all three phases of *before*, *during*, and *after*. We have tested structures of FNN, one-dimensional convolutional neural network (1D CNN), and long short-term memory network (LSTM) with all features. The best-performing model was the FNN which we have chosen as our Model 2.

We have also evaluated Model 2 using the selected features for all three phases are:

- **Correlation:** Head translation and rotation, F0, and loudness;
- **Time-lagged cross-correlation:** F0 and loudness;
- **DTW:** Head rotation, F0, and loudness.

An accuracy score of 0.750 is obtained for Model 2, in Table 1. However, the accuracy decreases to 0.636 when using only the selected features. This can be explained by the fact that the cross-modality information was missed in the analysis, as it is implicit and thus hard to visually capture them.

Model 2 renders a promising accuracy, however, its application is restricted as the future is required. To enable real-time turn-shift identification, we assess Model 2 by varying the moment time range. We studied the 3 time ranges of: $t_0 - 6\text{s} \sim t_0 + 2\text{s}$, $t_0 - 6\text{s} \sim t_0$, and $t_0 - 6\text{s} \sim t_0 - 2\text{s}$.

Table 2. Accuracy scores of model 2 using all features with different time ranges.

Moment time range	Accuracy
$t_0 - 6\text{s} \sim t_0 + 2\text{s}$	**0.727**
$t_0 - 6\text{s} \sim t_0$	0.478
$t_0 - 6\text{s} \sim t_0 - 2\text{s}$	0.530

We can remark that a similar accuracy score of 0.727 can be obtained using the moment time range of $t_0 - 6\text{s} \sim t_0 + 2\text{s}$. This implies that the *after* phase ($t_0 + 2\text{s} \sim t_0 + 6\text{s}$) does not play a critical role in identifying the turn-shift types, which might be too far from the turn-shift moment to provide useful information. For the identification to work in real-time, the moment time range must be restricted to before the turn-shift moment of t_0. However, the results of $t_0 - 6\text{s} \sim t_0$ and $t_0 - 6\text{s} \sim t_0 - 2\text{s}$ are not acceptable for real-time detection, this indicates that the period just after the turn-shift may carry the most important information to identify the turn-shift type.

Table 3. Accuracy scores of Model 2 using selected features with different time ranges.

Moment time range	Accuracy
$t_0 - 6\text{s} \sim t_0 + 2\text{s}$	**0.682**
$t_0 - 6\text{s} \sim t_0$	0.397
$t_0 - 6\text{s} \sim t_0 - 2\text{s}$	0.434

As seen above for Model 2, in Table 1, the same result of accuracy score falling (0.682, 0.397, and 0.434 respectively) when using the selected features can be observed.

Thus, the best is to use Model 2 with all features with the turn-shift moment time range of $t_0 - 6\text{s} \sim t_0 + 2$.

6 Conclusion and Discussion

Several works have been done studying turn-shift types of smooth turn exchange and interruption, backchannels by analyzing multimodal signals. However, the research on turn-shift types and synchrony is still to be done.

Through the analysis of multimodal signals (visual and acoustic features), we investigated the synchrony scores for three phases before, during, and after turn-shift. We were able to find a link between synchrony scores and turn-shift types and backchannel. This relationship was used to build computational models to automatically classify the turn-shift types. The modeling of all features of all three phases showed the most promising result which proved the usefulness of synchrony measures in turn-shift identification task. We also looked into whether the classification could be done in real-time by varying the moment time range. A lower accuracy was obtained compared to that using future information, although it is better than random chance and is a compromise to be considered.

Our turn-shift type identification model can be applied to various purposes. Manual annotation of turn-shifts is a lot of work, this is the problem we need to face every time a new corpus is generated, we are looking forward to integrating this model into automatic annotation systems that can help to detect and annotate different turn-shifts. The identified turn-shifts could also be used to analyze the personality or characteristic of people. We studied the synchrony of the same features of the interlocutors. We would also include cross-modality synchrony measures in the future to improve the performance of our classification model.

References

1. Ashenfelter, K.T., Boker, S.M., Waddell, J.R., Vitanov, N.: Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation. Journal of Experimental Psychology: Human Perception and Performance **35**(4), 1072 (2009)

2. Baltrušaitis, T., Robinson, P., Morency, L.P.: Openface: an open source facial behavior analysis toolkit. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1–10. IEEE (2016)
3. Beattie, G.W.: Interruption in conversational interaction, and its relation to the sex and status of the interactants (1981)
4. Beňuš, Š., Gravano, A., Hirschberg, J.: Pragmatic aspects of temporal accommodation in turn-taking. Journal of Pragmatics **43**(12), 3001–3027 (2011)
5. Berndt, D.J., Clifford, J.: Using dynamic time warping to find patterns in time series. In: KDD workshop. vol. 10, pp. 359–370. Seattle, WA, USA: (1994)
6. Bernieri, F.J., Reznick, J.S., Rosenthal, R.: Synchrony, pseudosynchrony, and dissynchrony: measuring the entrainment process in mother-infant interactions. Journal of personality and social psychology **54**(2), 243 (1988)
7. Boker, S.M., Rotondo, J.L., Xu, M., King, K.: Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. Psychological methods **7**(3), 338 (2002)
8. Burgoon, J.K., Guerrero, L.K., Manusov, V.: Nonverbal signals. The SAGE handbook of interpersonal communication pp. 239–280 (2011)
9. Burgoon, J.K., Stern, L.A., Dillman, L.: Interpersonal adaptation: Dyadic interaction patterns. Cambridge University Press (1995)
10. Cafaro, A., Wagner, J., Baur, T., Dermouche, S., Torres Torres, M., Pelachaud, C., Andre, E., Valstar, M.: The noxi database: multimodal recordings of mediated novice-expert interactions. pp. 350–359 (11 2017). https://doi.org/10.1145/3136755.3136780
11. Campbell, N.: Multimodal processing of discourse information; the effect of synchrony. In: 2008 Second International Symposium on Universal Communication. pp. 12–15. IEEE (2008)
12. Cappella, J.N.: Behavioral and judged coordination in adult informal social interactions: Vocal and kinesic indicators. Journal of personality and social psychology **72**(1), 119 (1997)
13. Chartrand, T.L., Bargh, J.A.: The chameleon effect: the perception–behavior link and social interaction. Journal of personality and social psychology **76**(6), 893 (1999)
14. Condon, W.S., Ogston, W.D.: Sound film analysis of normal and pathological behavior patterns. Journal of nervous and mental disease (1966)
15. Condon, W.S., Ogston, W.D.: A segmentation of behavior. Journal of psychiatric research **5**(3), 221–235 (1967)
16. Condon, W.S., Sander, L.W.: Neonate movement is synchronized with adult speech: Interactional participation and language acquisition. Science **183**(4120), 99–101 (1974)
17. De Kok, I., Heylen, D.: Multimodal end-of-turn prediction in multi-party meetings. In: Proceedings of the 2009 international conference on Multimodal interfaces. pp. 91–98 (2009)
18. Delaherche, E., Chetouani, M.: Multimodal coordination: exploring relevant features and measures. In: Proceedings of the 2nd international workshop on Social signal processing. pp. 47–52 (2010)
19. Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., Cohen, D.: Interpersonal synchrony: A survey of evaluation methods across disciplines. IEEE Transactions on Affective Computing **3**(3), 349–365 (2012)
20. Eyben, F., Wöllmer, M., Schuller, B.: Opensmile: the munich versatile and fast open-source audio feature extractor. In: Proceedings of the 18th ACM international conference on Multimedia. pp. 1459–1462 (2010)

21. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Robotics and autonomous systems **42**(3-4), 143–166 (2003)
22. French, P., Local, J.: Turn-competitive incomings. Journal of Pragmatics **7**(1), 17–38 (1983)
23. Gravano, A., Hirschberg, J.: A corpus-based study of interruptions in spoken dialogue. In: Thirteenth Annual Conference of the International Speech Communication Association (2012)
24. Hammarberg, B., Fritzell, B., Gaufin, J., Sundberg, J., Wedin, L.: Perceptual and acoustic correlates of abnormal voice qualities. Acta oto-laryngologica **90**(1-6), 441–451 (1980)
25. Heldner, M., Edlund, J.: Pauses, gaps and overlaps in conversations. Journal of Phonetics **38**(4), 555–568 (2010)
26. Ishii, R., Otsuka, K., Kumano, S., Matsuda, M., Yamato, J.: Predicting next speaker and timing from gaze transition patterns in multi-party meetings. In: Proceedings of the 15th ACM on International conference on multimodal interaction. pp. 79–86 (2013)
27. Ishii, R., Otsuka, K., Kumano, S., Yamato, J.: Using respiration to predict who will speak next and when in multiparty meetings. ACM Transactions on Interactive Intelligent Systems (TiiS) **6**(2), 1–20 (2016)
28. Ishii, R., Ren, X., Muszynski, M., Morency, L.P.: Multimodal and multitask approach to listener's backchannel prediction: Can prediction of turn-changing and turn-management willingness improve backchannel modeling? In: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents. pp. 131–138 (2021)
29. Ishimoto, Y., Teraoka, T., Enomoto, M.: End-of-utterance prediction by prosodic features and phrase-dependency structure in spontaneous japanese speech. In: Interspeech. pp. 1681–1685 (2017)
30. Kurtić, E., Brown, G.J., Wells, B.: Resources for turn competition in overlapping talk. Speech Communication **55**(5), 721–743 (2013)
31. Leander, N.P., Chartrand, T.L., Bargh, J.A.: You give me the chills: Embodied reactions to inappropriate amounts of behavioral mimicry. Psychological science **23**(7), 772–779 (2012)
32. Maier, A., Hough, J., Schlangen, D., et al.: Towards deep end-of-turn prediction for situated spoken dialogue systems (2017)
33. Müller, M.: Dynamic time warping. Information retrieval for music and motion pp. 69–84 (2007)
34. Oullier, O., De Guzman, G.C., Jantzen, K.J., Lagarde, J., Scott Kelso, J.: Social coordination dynamics: Measuring human bonding. Social neuroscience **3**(2), 178–192 (2008)
35. Reidsma, D., Nijholt, A., Tschacher, W., Ramseyer, F.: Measuring multimodal synchrony for human-computer interaction. In: 2010 international conference on cyberworlds. pp. 67–71. IEEE (2010)
36. Richardson, M.J., Marsh, K.L., Isenhower, R.W., Goodman, J.R., Schmidt, R.C.: Rocking together: Dynamics of intentional and unintentional interpersonal coordination. Human movement science **26**(6), 867–891 (2007)
37. Sacks, H., Schegloff, E.A., Jefferson, G.: A simplest systematics for the organization of turn taking for conversation. In: Studies in the organization of conversational interaction, pp. 7–55. Elsevier (1978)
38. Schegloff, E.A.: Sequencing in conversational openings 1. American anthropologist **70**(6), 1075–1095 (1968)

39. Schegloff, E.A., Sacks, H.: Opening up closings (1973)
40. Schmidt, R.C., Richardson, M.J.: Dynamics of interpersonal coordination. In: Coordination: Neural, behavioral and social dynamics, pp. 281–308. Springer (2008)
41. Shriberg, E., Stolcke, A., Baron, D.: Observations on overlap: Findings and implications for automatic processing of multi-party conversation. In: Seventh European Conference on Speech Communication and Technology (2001)
42. Skantze, G., Johansson, M., Beskow, J.: Exploring turn-taking cues in multi-party human-robot discussions about objects. In: Proceedings of the 2015 ACM on international conference on multimodal interaction. pp. 67–74 (2015)
43. Truong, K.P.: Classification of cooperative and competitive overlaps in speech using cues from the context, overlapper, and overlappee. In: Interspeech. pp. 1404–1408 (2013)
44. Yang, L., Achard, C., Pelachaud, C.: Annotating interruption in dyadic human interaction. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference. pp. 2292–2297 (2022)