Neural correlates of spoken language processing

Pablo Brusco [1,2]

[1] Departamento de Computación, FCEyN, Universidad de Buenos Aires, Argentina
[2] Instituto de Investigación en Ciencias de la Computación, CONICET-UBA, Buenos Aires, Argentina
pbrusco@dc.uba.ar

1. Motivation

In conversation, humans normally exchange turns in swift coordination, with few silences and overlaps, due partly to the existence of *turn-yielding cues*: acoustic, prosodic, lexical or syntactic events that signal an upcoming turn-taking transition. In the past few decades, much effort has been put into understanding the nature of turn-yielding cues. However, not much is known about what happens inside our brains upon the perception of these cues. We are interested in understanding *when* and *how* turn-yielding cues are produced and perceived, by looking directly at the **speaker's or listener's brain activity**. In other words, we want to understand how humans process the information produced by an interlocutor (whether human or computer) who currently holds the floor and signals an imminent point of possible turn completion. Specific questions we want to answer include:

- Can the perception and production of turn-yielding cues be detected in the brain activity?
- Can we understand which cues are projected over different cortical areas?
- When exactly does this take place, and under what conditions?

A second direction of research will consist in predicting the intention of a speaker to take the floor—i.e. anticipating the moment in which s/he will decide to start talking.

2. Aims of the research

- **Aim 1.** To develop machine-learning models capable of **detecting the occurrence of turn-yielding cues in the EEG signals** recorded from subjects engaged in conversation. We will use state-of-the-art techniques in order to achieve the best classifier performance. In addition to aiming at the best possible performance, we will analyze the trained models themselves, in an attempt to understand and *describe* the brain activity involved in the perception of turn-yielding cues.
- **Aim 2.** To implement machine-learning models using techniques similar to those in Aim 1 that allow to predict **the intention** of a speaker to start speaking even before the action is externalized. The ultimate goal is to build a BCI system that can anticipate the user's decision of taking the floor in conversation.

3. Related work

It has been shown that when holding a conversation, we generally alternate turns in a smooth way, that is, respecting the order, in a coordinated manner, without much silence or overlapping speech [1]. This is possible in part due to prosodic, lexical and syntactic cues produced by speakers at the end of each speech segment. These turn-yielding cues allow interlocutors to anticipate appropriate moments for starting their next contribution to the conversation, and include variations in pitch, intensity, speaking rate and voice quality over the last 500 milliseconds of every speech segment [2, 3]. Furthermore, it has been shown that these cues are indeed perceived by interlocutors in an incremental manner – the more cues present, the higher the perception of turn finality [4, 5].

The production of turn-yielding cues has been studied in the literature for years. Hypotheses such as Duncan's (1972), which claims that turn-yielding cues are linearly correlated with the occurrence of turn-taking attempts [6], have recently been aimed at from a computational point of view. For example, in [2] the authors show that the accumulation of seven turn-yielding cues, all of which can be extracted automatically, significantly increase the chances of a turn exchange attempt by the interlocutor. On the listener's side, perception studies have shown that these cues are actually perceived by the listener, and also that the number of turn-taking cues affects the reaction times for these decisions: the higher number of cues, the faster the reaction times [4]. In addition, in recent studies we have presented evidence supporting the claim that these cues are produced and perceived in suprisingly similar ways across very different languages, such as English, Slovak and Spanish [7, 8].

After years of research in linguistics, psychology, speech processing and related areas, turn-taking has started to be of interest in the neuroscientific community [9, 10]. There exists now plenty of literature regarding the neural networks involved in the linguistic processing of utterances [9, 11, 12, 13]. However, little research has been conducted yet on the less conscious system that monitors the activity of turn exchanges [9, 10], and in fact there is evidence suggesting both systems to be completely separate [14].

Regarding the second direction of research proposed here, detecting the intention of making a decision, Cerf and colleagues identified neurons that fire 0.2–1.5 seconds prior to the subject's movement, and even 0.1–1 seconds prior to the subject's reported 'will' to initiate a movement. This knowledge provided the basis for building an online classifier capable of anticipating the subject's urge to move their fingers to push a button [15, 16].[1] Further literature on intention prediction includes the work by Bai et al., who explore computational methods for predicting the production of self-paced right- and left-hand movements, and show that using a combination of Independent Component Analysis, Power Spectral Density, and Support Vector Machines, the discrimination accuracy was as high as 75% [17].

[1] A live demo of this experiment can be found at `https://youtu.be/lmI7NnMqwLQ?t=849`

4. Methodology

We will analyze the single-trial scalp electroencephalogram (EEG) signal of conversation participants who are currently listening to their interlocutor. In these data, we will apply state-of-the-art machine-learning techniques to look for patterns that allow us to explain and anticipate the moment in which the person will decide to start talking – whether taking the turn in a smooth manner, interrupting, or simply inviting the current speaker to continue (e.g. *uh-huh*).

A central issue regarding EEG signals is how to get the data. In our research group[2], we have been collecting a corpus of ten spontaneous, task-oriented, collaborative dyadic conversations in Argentine Spanish with simultaneous recordings of speech and EEG activity from each participant. We plan to use this corpus for our machine-learning experiments.

The EEG activity was recorded at 128 electrode positions using the Biosemi Active-Two system.[3] The experimental task consisted in two subjects playing a series of object-placing games (as described in [2]). Each subject used a separate laptop computer and could not see the screen of the other subject. Subjects sat facing each other in a booth, with an opaque curtain hanging between them, so that all communication was verbal. All recordings have been manually transcribed and annotated for type of turn-taking transition.

From this work, we expect to arrive at conclusions regarding which patterns emerge in the cortical activity of a person when turn-yielding cues are perceived. These patterns should allow us to build a temporal map of the human brain showing the dynamics of the perception of these cues. Specifically from aim 2, we expect to have an accurate classifier that can differentiate between moments in which the user wants to take the floor versus moments in which s/he does not. Furthermore, we expect to introduce new machine learning techniques tested not only in experimental setups but also using low-cost BCI devices by building a demo application as a proof of concept.

5. Results from completed work

Last year, we started working on this dataset by building a framework that would allow us to run machine learning experiments on this kind of data. We presented preliminary results at an Interspeech 2016 workshop, on the automatic detection of turn-taking events (such as keeping the floor in a conversation or yielding the turn) in continuous EEG data from spontaneous dialogue [18].

Also, in this year's Interspeech, we present a work where we use data from this corpus and an English version of the games corpus [2] for a cross-linguistic analysis of prosodic features automatically extracted from the conversations [8]. We found that, when signaling Holds, speakers of both languages tend to use roughly the same combination of cues. However, in speech preceding Smooth Switches or Backchannels, we observe the existence of the same set of prosodic turn-taking cues in both languages, although the ways in which these cues are combined together to form complex signals differ. Still, we find that these differences do not degrade below chance the performance of cross-linguistic systems for automatically detecting turn-taking signals. These results are relevant to this PhD plan, since understanding the differences in both languages on the way prosodic turn-taking cues are produced, should lead to better understanding how new findings in our Spanish dataset may generalize.

[2] `http://habla.dc.uba.ar/english.html`

[3] Biosemi, Amsterdam, Holland, http://www.biosemi.com

6. Future work

Next steps include to continue working on these ideas, now focusing specifically on the automatic detection of the *perception* of turn-yielding cues. We will also work on predicting the speaker's *intention* to continue talking or to yield the turn, and on the listener's side, the decision to take the floor or not, given the signals produced by the interlocutor.

Additionally, we plan to experiment with low-cost brain-computer interface (BCI) devices, such as EMOTIV EPOC 14-electrode EEG, which have opened many exciting opportunities for research [19]. We will use such devices aiming at **developing new BCI protocols**, similar to P3-speller (an effective assistive device for patients with severe motor diseases [20]), but in our case, using the findings of our investigations on turn-taking.

7. Main contributions

Both aims should pave the way for the medium-term goal of building BCI systems that use turn-taking information. That is, systems that can detect the user's perceptions and anticipate their intentions, with potential applications in gaming, medicine and communication. Our ultimate goal consists in building new BCI protocols using low-cost devices.

8. Acknowledgments

Work partially supported by ANPCYT PICT 2014-1561, Bilateral Cooperation Program CONICET-SAS, and the Air Force Office of Scientific Research, Air Force Material Command, USAF under Award No. FA9550-15-1-0055. This work is supervised by Prof. Agustín Gravano (thesis advisor) and by Juan Kamienkowski.

9. References

[1] H. Sacks, E. A. Schegloff, and G. Jefferson, "A simplest systematics for the organization of turn-taking for conversation," *Language*, vol. 50, pp. 696–735, 1974.

[2] A. Gravano and J. Hirschberg, "Turn-taking cues in task-oriented dialogue," *Computer Speech and Language*, pp. 601-634, 2011, 25(3).

[3] A. Raux and M. Eskenazi, "Optimizing the turn-taking behavior of task-oriented spoken dialog systems," *ACM Transactions on Speech and Language Processing (TSLP)*, vol. 9, no. 1, p. 1, 2012.

[4] A. Hjalmarsson, "The additive effect of turn-taking cues in human and synthetic voice," *Speech Communication*, vol. 53, no. 1, pp. 23–35, 2011.

[5] M. Zellers, "Duration and pitch in perception of turn transition by Swedish and English listeners," in *Proceedings of Fonetik*, 2014.

[6] S. Duncan, "On the structure of speaker-auditor interaction during speaking turns," *Language in Society*, vol. 3, no. 2, pp. 161–180, 1974.

[7] A. Gravano, P. Brusco, and Š. Beňuš, "Who do you think will speak next? perception of turn-taking cues in slovak and argentine spanish," *Interspeech 2016*, pp. 1265–1269, 2016.

[8] B. Pablo, J. M. Pérez, and A. Gravano, "Cross-linguistic study of the production of turn-taking cues in american english and argentine spanish," *Accepted in Interspeech 2017*, 2017.

[9] J. C. Hoeks and H. Brouwer, "Electrophysiological research on conversation and discourse," *Holtgraves, TM (Ed.)*, pp. 365–386, 2014.

[10] S. C. Levinson, "Turn-taking in human communication–origins and implications for language processing," *Trends in cognitive sciences*, vol. 20, no. 1, pp. 6–14, 2016.

[11] S. Bögels, L. Magyari, and S. C. Levinson, "Neural signatures of response planning occur midway through an incoming question in conversation," *Scientific reports*, vol. 5, 2015.

[12] A. D. Friederici, "Event-related brain potential studies in language," *Current neurology and neuroscience reports*, vol. 4, no. 6, pp. 466–470, 2004.

[13] M. Kutas and K. D. Federmeier, "Thirty years and counting: Finding meaning in the n400 component of the event related brain potential (erp)," *Annual review of psychology*, vol. 62, p. 621, 2011.

[14] D. Foti and F. Roberts, "The neural dynamics of speech perception: Dissociable networks for processing linguistic content and monitoring speaker turn-taking," *Brain and language*, vol. 157, pp. 63–71, 2016.

[15] M. Cerf, N. Thiruvengadam, F. Mormann, A. Kraskov, R. Q. Quiroga, C. Koch, and I. Fried, "On-line, voluntary control of human temporal lobe neurons," *Nature*, vol. 467, no. 7319, pp. 1104–1108, 2010.

[16] M. Cerf and M. Mackay, "Studying consciousness using direct recording from single neurons in the human brain," in *Characterizing Consciousness: From Cognition to the Clinic?* Springer, 2011, pp. 133–146.

[17] O. Bai, P. Lin, S. Vorbach, J. Li, S. Furlani, and M. Hallett, "Exploration of computational methods for classification of movement intention during human voluntary movement from single trial eeg," *Clinical Neurophysiology*, vol. 118, no. 12, pp. 2637–2655, 2007.

[18] P. Brusco, J. Kamienkowski, and A. Gravano, "Automatic detection of turn-taking events in continuous eeg data from spontaneous dialogue," in *1st Workshop on Speech Engineering and Computational Neuroscience*, San Francisco, CA, 2016. [Online]. Available: https://sites.google.com/site/secns16/

[19] N. A. Badcock, P. Mousikou, Y. Mahajan, P. de Lissa, J. Thie, and G. McArthur, "Validation of the emotiv epoc® eeg gaming system for measuring research quality auditory erps," *PeerJ*, vol. 1, p. e38, 2013.

[20] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.