Relations between Opinion Convergence, Acoustic Convergence and Movement Convergence in Interlocutors

Charlize Ma, Effie Kao, Raechel Kitamura, Stephanie Wang, Jahurul Islam, Gillian De Boer, Bryan Gick

Department of Linguistics, University of British Columbia (Canada)

sy.charlize@gmail.com

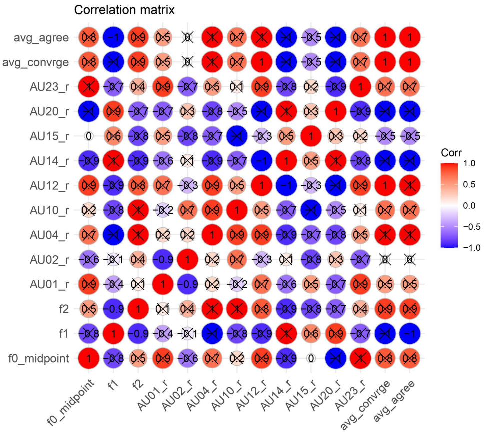
Background: Speakers in a conversation (interlocutors) can exhibit convergent behaviours in a variety of ways, including influencing one another's speech acoustics, movements, and opinions. Past research shows that interlocutors appear to converge in a descending F_0 pattern nearing the end of a conversation [1]. Additional research has also shown that speakers tended to imitate each other's changes in F_0 across turns during a turn-taking reading task [2]. Notably, individuals who perceived a Voice User Interface (VUI) as having the same opinion and characteristics as themselves had an increased likelihood of convergence [3]. Furthermore, the degree of closeness in the relationship between interlocutors appeared to be a factor in the polarization of their opinions [4]. Most of the research into speech convergence has been focused on acoustics, but there have been few attempts to assess if the same applies to visual cues, like lip and eyebrow movement. Past studies have found that our facial movements change during speech depending on our interlocutor. Lip movements were observed to increase significantly during infant-directed speech [5] and in congenitally blind speakers [6]. We sought to discover whether facial movement and speech convergence could be linked to the convergence of opinions.

Methods: 36 participants (M:9, F:27) above the age of 18 were recruited. Each participant was randomly paired with another participant to have a short conversation (3-5 mins) in a Zoom meeting where they discussed their views of online vs. in-person schooling. At the end of the conversation, they completed a questionnaire asking how much they thought their opinion converged with their partners' (convergence), and how much they agreed with each others' ideas (agreement), on a 7-point Likert scale. The whole conversation process was videotaped and recorded using Zoom's recording system.

OpenFace 2.0 [7] software was used to extract lip and eyebrow movement information from the video data. The first and last minutes of the conversation were selected to generate differences in action units (AUs) in each dyad. 9 AUs were targeted (brows: 1, 2, 4; lips: 10, 12, 14, 15, 20, and 23). Audio data was transcribed and force-aligned using Montreal Forced Aligner [8]. Acoustics values (F_0, F1 values etc.) were extracted from the vowel midpoints using Praat [9]. Acoustic data was synchronized with the facial movement data from OpenFace 2.0 using timestamps.

From this acoustic data, plots for seven vowels (ɪ, i, ɛ, ɑ, ɔ, ʊ, u) were examined to aid in visualizing the relationship between specific vowels and dependent variable values from the experiment, namely the Likert scale data taken from the questionnaire after the discussion and facial movement differences. The average agreement and convergence values from the Likert scale after the conversation for each dyad was then calculated.

Results: A correlation matrix was run on the acoustics values from Praat, the AU values from OpenFace, and the average Likert scale data. In the matrix in Figure 1, circles that are crossed out denote non-significance. A significant positive correlation was found between lip corner pull (AU12_r) and average convergence (avg_converge) ($r = 1$, $p < .001$), and a significant negative correlation was found between F1 values and average agreement (avg_agree) ($r = -1$, $p = .018$). However, there was no overall difference observed between F_0 values and AUs within participants in a conversation from their first to last minutes of conversation. A box plot was generated to display the differences between the first and last minutes of conversation for each AU as well as F_0 (Figure 2). Additionally, a U-Test was run using R [10], that indicated no significant difference between F_0 values and AUs ($p > .05$ for all comparisons).



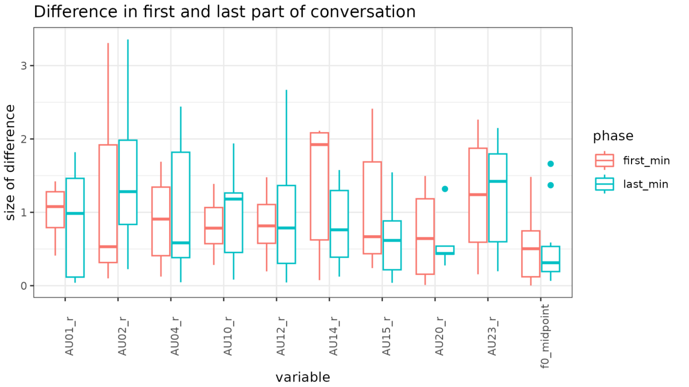


Fig. 1 Correlation Matrix of all variables

Fig. 2 Box plot of AU differences in first vs. last min

Discussion: Our initial analysis shows a correlation between a consensus of agreement among participants and increased lip corner pulling (AU 12). This could possibly demonstrate a relationship between opinion convergence and facial movements (in this case, smiling). Additionally, the correlation between participants who agreed more and those who exhibited higher vowel height (through acoustic analysis) could indicate that participants expended more effort in trying to converge with their interlocutor. However, the vast majority of facial action units analyzed did not appear to be affected by opinion convergence, suggesting that speech convergence and opinion convergence appear to work largely independently. The lack of significant F_0 convergence shows different results from that of previous literature [1], but there is room for further investigation with regard to interactions between facial movements and opinion convergence.

References

[1] Yang, Li-Chiung. (2013). Prosodic convergence, divergence, and feedback: Coherence and meaning in conversation. 27th Pacific Asia Conference on Language, Information, and Computation, PACLIC 27. 85-91.

[2] Aubanel V, Nguyen N. Speaking to a common tune: Between-speaker convergence in voice fundamental frequency in a joint speech production task. PLoS One. 2020 May 4;15(5):e0232209. doi: 10.1371/journal.pone.0232209. PMID: 32365075; PMCID: PMC7197779.

[3] Farr, C., Purnomo, G., Cardoso, A., Shamei, A. & Gick, B. (2021). Speaker accommodations and VUI voices: Does human-likeness of a voice matter? In Proceedings of the XVIIth Conference of Associazione Italiana di Scienze della Voce, 63-64.

[4] Balietti, S., Getoor, L., Goldstein, D. G., & Watts, D. J. (2021). Reducing opinion polarization: Effects of exposure to similar people with differing political views. Proceedings of the National Academy of Sciences, 118(52), e2112552118. https://doi.org /10.1073/pnas.2112552118

[5] Green, J. R., Nip, I. S. B., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. Journal of Speech, Language, and Hearing Research, 53(6), 1529-1542. https://doi.org/10.1044/1092-4388(2010/09-0005)

[6] Ménard, L., Leclerc, A., & Tiede, M. (2014). Articulatory and acoustic correlates of contrastive focus in congenitally blind adults and sighted adults. Journal of Speech, Language, and Hearing Research, 57(3), 793-804. https://doi.org/10.1044/2014_JSLHR-S-12-03

[7] OpenFace: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.

[8] MFA: McAuliffe, Michael, Michaela Socolof, Elias Stengel-Eskin, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger (2017). Montreal Forced Aligner [Computer program].

[9] Boersma, Paul & Weenink, David (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.07, retrieved 6 February 2023 from http://www.praat.org/

[10] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.