Testing the effects of inter-speaker coordination on the stability of speech production patterns via the Virtual Parrot, a modelling framework for inter-speaker coordination

Leonardo Lancia¹, Noël Nguyen¹, Thierry Chaminade², Laurent Prévol¹

¹ Laboratoire Parole et Langage, Aix-en-Provence, France.
² Institut de Neurosciences de La Timone, Marseille, France.

Corresponding author: Leonardo Lancia, leonardo.lancia@blri.fr

Introduction. A growing body of research supports the hypothesis that inter-speaker coordination during linguistic interactions emerges from between-speakers dynamical coupling at several levels of physiological and cognitive activity (cf. Fowler, 2013). Supporting evidence comes mainly from studies focusing on the similarity between the behaviours of speakers involved in conversational tasks. However, although surface similarity can be interpreted as the consequence of dynamical coupling, other explanations are also possible (cf. for example Pickering and Garrod, 2013). The present work has a double aim. First, we introduce a modelling framework to study inter-speaker coordination in speech repetition tasks as resulting from dynamical coupling and beyond surface similarity. This is done by porting to speech the Human Dynamical Clamp paradigm (HDC, cf. Dumas et al. 2014). Second, we test if the dynamical coupling between a human and a virtual partner affects the stabilization of the repeated production of a tongue twister by the human speaker.

Method. In our adaptation of the HDC, the virtual partner is renamed as Virtual Parrot (VP) because it is designed to coordinate with a real speaker during the repetition of simple speech utterances. The VP is built to repeat an utterance while adapting in real time its local speech rate to that of a human speaker to achieve a predefined coordinative relation. The target coordinative relation is defined as a relative phase angle between the amplitude modulations of the audio signals produced by the speaker and by the VP and it corresponds to the lag between the two signals relative to the duration of a syllabic cycle. To achieve the target relative phase, the speech signal of the VP is windowed (dur.: 25ms), ad its time-scale is deformed through WSOLA synthesis. The amount of temporal deformation is determined by submitting to the Kuramoto equation (Kuramoto, 1984) the instantaneous relative phase measured between the low-pass filtered (freq. cut-off: 8Hz) amplitude modulations of the acoustic signals produced by the VP and by the speaker. The outcome of the Kuramoto equation is corrected in order to maintain the speech rate of the VP inside pre-established ranges.

Procedure. 8 speakers participated in the experiment. In each experimental trial, participants were instructed to repeat the tongue twister /top kop/ simultaneously (i.e. in-phase) with the VP, without interruptions, during 16s. At the beginning of each trial, the VP was parameterized to cooperate with the speaker (i.e. to produce in-phase coordination). At a random point in the experimental trial, one of the following perturbations could occur: 1) the target relative phase of the VP switched suddenly to π (anti-phase coordination) and was kept at that value up to the end of the trial. 2) The target relative phase of the VP changed gradually from 0 to π during 2s and was kept at that value up to the end of the trial. 3) The target relative phase of the VP was randomly perturbed during 2s and then was set to its initial value of 0 during the remaining part of the trial. We tested four different speech rates (corresponding to utterance durations of 1.12s, 0.7s, 0.6s and 0.5s). We also tested three different values of the parameter k that in the Kuramoto equation modulates the coupling strength between the VP and the speaker (0.05, practically corresponding to no coupling, 0.2 and 0.5). Each speaker had to complete one experimental trial per combination of perturbation type, speech rate and coupling strength.

Analyses. Portions of experimental trials in which the VP was parameterized to cooperate with the speaker were analysed separately from portions in which the target relative phase of the VP was in
competition with that of the speaker (i.e. the VP was parameterized to produce anti-phase coordination). To study the coordination regimes over the parameters space explored, we computed the relative phase ($\Phi_r$) between the amplitude modulations of the VP and that of the speaker. We obtained an index of the dispersion of the observed $\Phi_r$ values by computing the first Fourier moment of their distribution in each portion of the trials. In order to measure the regularity of the speech patterns produced by the speakers and the asymmetry of the coupling between the VP and the speakers, we computed mel spectra of the acoustic signals (window length: 25ms, time-step: 1ms, freq. range: 80 - 8000 Hz, bands num.: 40) and submitted the corresponding multivariate time-series to a version of recurrence analysis adapted to non-stationary signals (Lancia et al. 2016). Recurrences are defined here as productions of similar spectral slices at different and non-consecutive points in time. The recurrences produced by the speakers and by the VP in each portion of the experimental trials were detected separately through the computations of recurrence plots. A regularity index of the recorded speech patterns was determined by following the method proposed in Lancia et al. (2016). In each portion of experimental trial, we estimated the asymmetry of the coupling between the VP and the speaker by computing the mean conditional probability of recurrences (MCR, cf. Romano et al., 2007) of the speaker’s time-varying spectrum given the probability of recurrences of the VP’s spectrum and vice-versa. The difference between the two MCRs is an estimate of the relative strengths of the dependencies between the VP and the speaker.

**Results and discussion.** Linear mixed models regressions show that: 1) When the VP is parameterized to cooperate with the speakers, the distributions of $\Phi_r$ show strong peaks at in-phase values ($0$ and $2\pi$). When the VP is in competition with the speakers, the distributions become significantly flatter, indicating a continuously changing relative phase. However the evolution of $\Phi_r$ over the duration of each experimental trial reveals that this value changes more slowly around $0$ and $\pi$ (i.e. around the two target relative phase values). 2) The analysis of the asymmetry of the coupling between the VP and the speakers reveals that for small values of the coupling strength parameter $k$ the behaviour of the speakers depends on the behaviour of the VP more than the reverse. However, as the value of $k$ increases, the behaviour of the VP becomes more dependent on the behaviour of the speakers. 3) The regularity of the speakers’ behaviours decreases as speech rate increases. However, when the VP cooperates with the speakers, for a given speech rate, the regularity increases as the value of $k$ increases. 4) When the value of $k$ increases the distribution of $\Phi_r$ becomes flatter both in the cooperative and in the competitive conditions. Results 1 and 2 demonstrate the presence of dynamical coordination between the VP and the speakers, thus showing that the proposed modelling framework is appropriate to the coordination task under study. Result 3 shows that the coupling between a speaker and a surrogate of himself has a beneficial effect on the stabilization of an inherently unstable coordination pattern, such as that underlying the production of the tongue twister /top kop/. This result cannot be explained as due to smaller delays between the speech signal of the human speaker and that of the VP. Indeed when the strength of the coupling increases, the relative phase between the productions of the two partners becomes more variable (cf. result 4).

**References**


