A novel paradigm to investigate phonetic convergence in interaction

Martijn Wieling[1,2] Mark Tiede[2], Teja Rebernik[1], Lisanne de Jong[1], Anouck Braggaar[1], Martijn Bartelds[1], Masha Medvedeva[1], Penny Heisterkamp[1], Tom Freire Offrede[1], Hedwig Sekeres[1], Anna Pot[1], Mara van der Ploeg[1], Karin Volkers[3] & Gregory Mills[1]

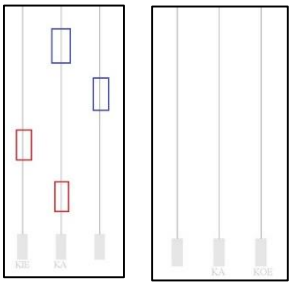
[1]University of Groningen, [2]Haskins Laboratories, [3]Philadelphia Care Foundation

Introduction

Many studies have shown that when two people converse with each other, they progressively adapt their linguistic resources to those of their partner (Pickering and Garrod, 2013). One form of adaptation is phonetic convergence, which is highly context-sensitive, variable and driven by the interactional goals of the participants (see e.g. Pardo et al., 2017). Consequently, experimental approaches to phonetic convergence are faced with a methodological trade-off between experimental control and validity: On the one hand, tasks which use shadowing allow high levels of control but block many interactive mechanisms that underpin convergence. On the other hand, more spontaneous tasks allow high levels of interaction, but remove the tight control over stimuli afforded by shadowing. To side-step this trade-off, we present data from a phonetic variant of the procedural coordination task (Mills, 2011) which presents pairs of participants with the recurrent coordination problem of jointly producing sequences of congruent and complementary vocalizations.

Paradigm

The setup (see: https://www.youtube.com/watch?v=99PC3a3Pscg) consists of a two-player music game in which in each round one player has the role of director, and the other the role of follower. The director sees the melody which needs to be played from bottom to top (see left-most figure; each line represents one distinct note, the color indicates the supposed player: self = red, other = blue), whereas the follower only sees three empty lines (see right-most figure).



The notes are played through pronouncing different consonant-vowel (CV) sequences. The left-most note is played by pronouncing /ki/, the middle by pronouncing /ka/, and the right-most by pronouncing /ku/. Each player can only play two of these CV sequences (/ka/ and /ki/, or /ka/ and /ku/). For the example melody shown, the required pronunciation sequence is /ka/ (director), /ki/ (director), /ku/ (follower) and /ka/ (follower). If a mistake is made in this sequence (i.e. a speaker pronounces a wrong sound, or the wrong speaker pronounces a correct sound), the sequence has to be played anew from the start. Correctly played notes can be visually identified by the director (but not the follower) as these will be filled (in red or blue).

Participants are not able to see each other, but can hear each other and are instructed to only communicate with each other via their two assigned CV sequences. (The experimenter fails the present trial if the participants use other sounds or words.) Effectively, this means that each dyad has to converge on some kind of shared communication system.

After an initial calibration phase in which a real-time vowel recognizer (implemented in Matlab) is trained to recognize the individual sounds for both speakers, the speakers first finish a few director-only melodies to become familiar with the game. Subsequently, the 15-minute experiment starts. Initially only simple random melodies are shown, but these increase in

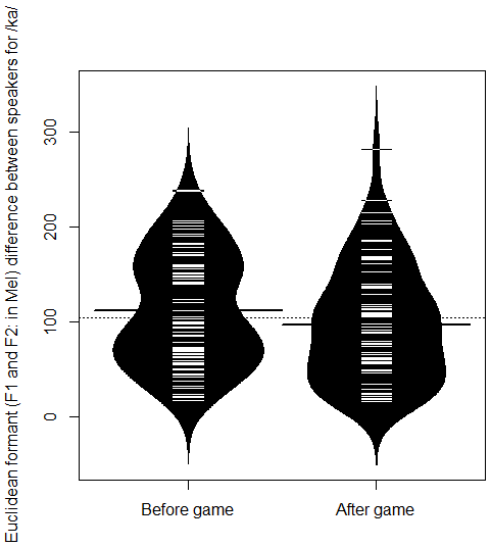
complexity when participants successfully complete the melody within the time limit (90 seconds). A higher level of complexity is realized through increasing the length of the melody (up to 5 notes), but also through requiring two notes to be played simultaneously.

Data collection

Data was collected at *Lowlands Science 2019*, a science event at the three-day Dutch music festival *Lowlands*, with over 50,000 visitors every year. After answering initial assessment questions (including information regarding musical ability and substance use; we also measured blood alcohol concentration using a professional breathalyzer), 77 pairs of (mostly Dutch) speakers – generally friends, partners or family (67 pairs) – played the music game. Before and after the experiment, participants produced a single sentence with three words for each of the vowels /a/, /i/ and /u/. In addition, during the training phase and at the end, both participants pronounced the sequence /ka/, /ka/, /ki/ (or /ku/ depending on the speaker), /ki/ (or /ku/), /ka/. The sentence and the sequence were used to assess phonetic convergence. All pronunciations were recorded with head-worn microphones (Shure WH20).

Results

To assess phonetic convergence in the pre- and post-game sentences and sequences, we calculated F1-F2 (Mel-scaled) based Euclidean distances between the two speakers in a pair for the shared vowel /a/, both before and after the experiment. Using mixed-effects regression analysis with the optimal random-effects structure, we observed no significant phonetic convergence for the sentences (β = -1.2, t = -0.2, p = .81). However, phonetic convergence was clearly present for the /a/-vowels in the sequence (β = -12.6, t = -2.5, p = .01; see the bean plot to the right). The effect appeared to be robust, as it remained significant when excluding 41 pairs who had used alcohol or drugs. No influence of personal characteristics (e.g., gender, age, musical ability) on convergence reached significance.



Discussion

In this study we have illustrated a new experimental paradigm which combines the merits of the (controlled) shadowing task with the advantage of incorporating communicative interaction. Our paradigm was shown to result in *task-specific* phonetic convergence after only 15 minutes of interaction. No more general convergence was shown, but this may have been caused by the large majority of the speakers already knowing each other well.

References

Mills, G. (2011). The emergence of procedural conventions in dialogue. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

Pardo et al. (2017). Phonetic Convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, 79(2), 637-659

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences*, *36*(4), 329-347.