Studies on phonetic convergence between speakers in interactional speech

Noël Nguyen

Hong-Kong, 7 January 2016

Outline of talk

- Convergence effects at a short-time scale (within a conversational interaction).
- Convergence effects at a longer time scale (within an individual’s lifetime).
- Convergence effects at a still longer time scale (that of language itself).

Phonetic convergence — the tendency for two speakers engaged in a conversational exchange to sound more like each other — is one important aspect of between-speaker adaptation.

It is a recurrent phenomenon in mature speakers.

It relies on sensory-motor abilities that infants may already possess at birth.

It may have long-lasting consequences on how we speak.

It may be a driving mechanism in the acquisition of a second language.

It may have played an central role in the emergence of phonological systems.

Convergence in conversation: Pardo, 2006

Pairs of speakers produced sets of words

1. Individually in a reading task (baseline);
2. In a conversational interaction (Map Task).
3. Individually and in a reading task again.

Convergence effects were perceptually assessed by an independent group of listeners (AXB task).

Perceived similarity between participants in words’ acoustic forms was found to

- Increase over the course of the Map Task;
- Persist beyond the end of the interaction.
Alignment to visual speech information

RACHEL M. MILLER, KAYUMARI SANCHEZ, AND LAWRENCE D. ROSENBAUM
University of California, Riverside, California

Convergence effects extend to visual speech.

Miller et al. (2010) asked participants to lipread and shadow out loud a talker uttering words in a silent video.

The participants’ shadowed word forms were perceived to converge towards the talker’s auditory word forms by listeners (with respect to a baseline condition).

Simply seeing the talker’s moving face allows the participant to imitate the talker’s idiosyncratic way of producing words.

Phonetic convergence and social desirability

Natale, 1975: Participants in a dialogue tend to converge towards each other in mean vocal intensity.

Convergence in mean vocal intensity occurs to a greater extent in participants with a higher degree of social desirability.

Social desirability refers to a need for social approval and acceptance.

Phonetic convergence and gender

Namy et al., 2002: In a shadowing task, female participants tend to phonetically converge towards the recorded speaker to a greater extent than male participants.

Convergence in female participants is also greater when the recorded speaker is a male compared with a female speaker.

Gender differences in accommodation emerged even when immediate social motives were minimized.
Phonetic convergence and social status

Gregory & Webster, 1996: speakers with a lower social status converge phonetically to a greater extent towards speakers with a higher social status.

Introduction

Phonetic convergence between speakers

Short-term effects

Long-term effects

Very long-term effects

Phonetic convergence was characterized by measuring the amount of acoustic energy in the low-frequency range (0-500 Hz) in the speech signal.

Two current central issues

Pardo et al. (2013): the available acoustic studies “present a bewildering array of patterns of convergence and divergence, with a great deal of variability and inconsistencies across measures”.

The Why question: phonetic convergence has been attributed to at least two possible mechanisms

- A low-level, automatically-triggered, perception-action link between the perceived speech signal and the listener’s own repertoire of speech motor commands;
- A high-level strategy deliberately employed by speakers to modulate the position they occupy with respect to each other in a social space, as their interaction unfolds.

Convergence effects in f0

Sato and colleagues (2013) examined between-speaker convergence effects in f0 frequency in a non-interactive vowel production task.

Participants were asked to repeat or overtly imitate instances of /i/, /e/ and /e/ previously recorded by 3 female and 3 male speakers.

We asked to what extent participants tend to converge towards the perceived speakers in vowel f0.

To establish the size of convergence effects, we compared

- a) the difference in the perceived speaker’s f0 relative to the participant’s baseline (y axis, in Barks), with
- b) the difference in the participant’s f0 in the repetition or imitation task relative to the participant’s baseline (y axis, in Barks).

Correlations between these two measures are indicative of a convergence effect.

Correlations significantly differed from 0 in both the production and imitation tasks.
Phonetic convergence can occur in response to hearing a recorded speaker in a laboratory setting. Convergence does not, or not only, depend on mutual adjustment and social attunement. It may also be triggered by low-level sensory and motor adaptive processes.

In a recent work with Vincent Aubanel (now at the MARCS Institute, Sydney), we asked to what extent do speakers converge towards each other in f0 in an interactive setting.

We asked pairs of speakers to perform an alternate reading task.

The story was divided into a number of “reading turns” (from 6- to 13-word long). Speaker A’s turns in black, Speaker B’s turns in grey.

The task was aimed to mirror the turn-taking organization of a conversation between two people.

Points correspond to reading turns.

f0 in both speakers was manipulated using a sine function with a period about equal to the duration of the reading task (~4 min).

The amplitude of the sinewave transformation was ±200 cents (~±1 tone).

The reading task was performed twice.

In one reading, the phase angle of the sine function was 0 (i.e. f0 was shifted upwards, then downwards).

In the other reading, the phase angle was π (i.e. f0 was shifted downwards, than upwards).

This was done to avoid artifacts potentially associated with the reading task (e.g. speakers could systematically move their f0 down across the first part of the reading irrespective of what the other speaker does).

The order of the two readings was counterbalanced across dyads.

The speakers were in different rooms and communicated with each other using microphones / headphones. Unbeknownst to both of them, we manipulated f0 frequency in the two speakers in real time.
Convergence effects between speakers were measured by computing, for each reading turn, the difference in median F0 value between the phase=$\pi$ reading and the phase=0 reading.

The expected response pattern is shown in this figure.

The task was performed by 32 dyads of female speakers. In each dyad, the two speakers regarded themselves as friends. None of the participants noticed that the other participant’s voice had been altered.

**Long-term effects of phonetic convergence**

Perceived similarity in pronunciation between speakers increases over the course of the interaction, and maintains itself beyond the end of that interaction. It therefore has long-term effects on the pronounced form of words.

It has been taken as suggesting that words are stored in memory in a form that is highly dependent on their context of occurrence (Goldinger, 1998; Johnson, 2006).

More specifically, the representations associated with words in the mental lexicon for each talker, may dynamically evolve during conversation, under the influence of the other talker’s speech patterns, and retain the traces of that influence once the conversation has ended.

Between-speaker phonetic convergence may make it easier for each speaker to understand the other speaker. It may contribute to making each speaker more attuned to the phonetic characteristics of words produced by the other speaker, via a perception-action resonance phenomenon.

In some theoretical frameworks (e.g., Pickering and Garrod, 2007), imitating an individual’s actions makes it easier to predict what that individual will do next, particularly in the case of ambiguous or distorted input.

In a recent behavioral study, Adank et al. (2010) showed that imitation of a novel accent improves language comprehension in adverse listening conditions. In another study, however, phonetic imitation did not have a significant impact on how listeners later recognized words in a non-native regional accent.
Very long-term effects of phonetic convergence

To what extent does between-speaker phonetic convergence play a role in the emergence of phonological systems?

Moulin-Frier, Diard, Schwartz & Bessière have recently developed a new computational model that simulates the emergence of phonological systems: COSMO.

COSMO: Communicating about Objects using Sensory-Motor Operations.

In COSMO, convergence does not directly contribute to the building up of a phonological system.

By contrast, in other theoretical frameworks, imitation is a driving force in the emergence of language and its phonological component.


In human evolution, it is one of the factors that made our brains “ready for language” (Arbib, 2002).

It is much more than a behavioral echo. It is selective and compositional.

While it allows sets of actions to be passed on from one individual to another through observation, it can be a factor of innovation and change.
Articulatory Phonology also assumes that gestural structures emerge as the by-product of imitation between speakers.

For example, Browman & Goldstein (2000) did a computational experiment in which “speakers” must conform to an “accommodation condition” (to act like each other).

As these dyadic interactions developed, phase relationships between gestures arose that were both stable and shared across speakers.

Goldstein (2003, see also Goldstein & Fowler, 2003, and Goldstein et al., 2006) also showed that the establishment of gestural contrasts within a single vocal organ, can be modeled as resulting from imitation games.

Imitation is a central mechanism in de Boer’s (2000) computational model of the emergence of vowel systems.

Agents engage in a sequence of interactions called “imitation games”.

One of two agents (the imitator) has to imitate the vowel produced by the other agent (the initiator).

The initiator’s and imitator’s repertoires of vowel prototypes are then updated in a way which depends on whether the game is found successful or unsuccessful by both agents.

Recent experimental work has attempted to reproduce the conditions that may have presided over the emergence of language in the lab.

This work concerns graphical communication systems mostly, much less acoustic communication systems.

Verhoef et al. (2014) used a task involving learning an artificial whistle language.

Chains of participants had to memorize and reproduce whistle sounds of a variety of forms.

The sounds did not refer to anything. Participants were simply instructed to imitate them as best they could.

As the sounds were transmitted from one participant to the next one, they came to be broken down into smaller components that were then reused in combination with each other.
Fragment of the whistles plotted as pitch tracks in the last set of a chain. Basic elements (falling–rising slides and short level notes) can be identified that are systematically recombined. From Verhoef et al., 2014.

Thus, in an experimental setting where each individual is asked to learn through imitation a repertoire of sounds produced by another individual, combinatorial properties emerge in that sound repertoire that resemble those of phonological systems.

In a linear transmission chain, information is transmitted in a one-way fashion, from the sender to the receiver.

In other studies (e.g., Galantucci, 2005; Garrod et al., 2010), participants engaged in pairwise interactions.

These studies showed that pairwise interactions have a significant influence on how a (graphical) communication system takes shape.

The experimental evidence now available suggests that both learning by imitation and pairwise interactions have an important role to play in the process that allow human participants to jointly build up a repertoire of acoustic or graphical forms endowed with combinatorial properties.

A few words of conclusion

Convergence in speech may be dependent on both low-level and high-level processes.

- Low-level, automatically triggered, sensori-motor integration processes;
- High-level processes that operate under the control of a variety of linguistic factors (phonetic convergence can occur inasmuch as phonological contrasts are preserved), individual, situational and social factors.

The strategic, socially-driven component of imitation develops during ontogenesis, on top of the sensory-motor component, as children acquire the skills that are required for them to engage in social interactions.

Contributors

Vincent Aubanel (MARCS, Sydney)
Véronique Delvaux (Mons, Belgium)
Robert Espesser (LPL)
Amandine Michelas (LPL)
Marc Sato (LPL)