Study of the phenomenon of phonetic convergence thanks to speech dominoes

Amélie Lelong, Gérard Bailly

Study of the Phenomenon of Phonetic Convergence thanks to Speech Dominoes

Amélie Lelong & Gérard Bailly

GIPSA-Lab, Speech & Cognition dpt., UMR 5216 CNRS/Grenoble INP/UJF/U. Stendhal,
38402 Grenoble Cedex, France
{amelie.lelong, gerard.bailly}@gipsa-lab.grenoble-inp.fr

Abstract. During an interaction people are known to mutually adapt. Phonetic adaptation has been studied notably for prosodic parameters such as loudness, speech rate or fundamental frequency. In most of the cases, results are contradictory and the effectiveness of phonetic convergence during an interaction remains an open issue. This paper describes an experiment based on a children game known as speech dominoes that enabled us to collect several hundreds of syllables uttered by different speakers in different conditions: alone before any interaction vs. after it, in a mediated interaction vs. in a face-to-face interaction. Speech recognition techniques were then applied to globally characterize a possible phonetic convergence.

Keywords: face-to-face interaction phonetic convergence, mutual adaptation

1 Introduction

The Communication Adaptation Theory (CAT), introduced by Giles et al [1], postulates that individuals accommodate their communication behavior either by becoming much closer of their interlocutor (convergence) or on the contrary by increasing their differences (divergence). People can adapt to each other in different ways. For example, conversational partners notably adapt to each other's choice of words and references [2] and also converge on certain syntactic choices [3]. Zoltan-Ford [4] has shown that users of dialog systems converge lexically and syntactically to the spoken responses of the system. Ward et al [5] demonstrated that adaptive systems mimicking this behavior facilitate learning. This alignment [6] may have several benefits such as easing comprehension [7], facilitating the exchange of messages of which the meaning is highly context-dependent [8], disclosing ability and willingness to perceive, understanding or accepting new information [9] and maintaining social glue or resonance [10].

Researchers have examined also adaptation of phonetic dimensions such as pitch [11], speech rate [12], loudness [13], dispersions of vocalic targets [14] as well as more global alignment such as turn-taking [15]. But the results of these different studies show a weak convergence and even in some cases no convergence at all. In the perceptual study conducted by Pardo [16], disparities between talkers have been

attributed to various dimensions such as social settings, communication goals and varying roles in the conversation. Sex differences have also been put forward: female interlocutors show more convergence than males.

The emerging field of research is crucial to the comprehension of adaptive behavior during unconstrained conversation on one hand and to versatile speech technologies that aim at substituting one partner with an artificial conversational agent on the other hand. Literature shows that two main challenges persist: (a) the need of original experiments that allow us to collect sufficient phonetic material to study and isolate the impact of the numerous factors influencing adaptation; (b) the use of automatic techniques for characterizing the degree of convergence if any.

2 State of the art

In the following section, several influential articles will be presented. These papers thoroughly summarize research about phonetic adaptation.

2.1 Convergence and social role

There are only a few studies that explain the role of convergence in a social interaction. Different interpretations have been given.

First of all, convergence could be a consequence of the episodic memory system [17]. People keep a trace of all their multimodal experiences during social interaction. An exemplar-based retrieval of previous behavior given similar social context is triggered so that the current interaction benefits from previous attunement.

Adaptation can also be used in a community to let a more stable form emerge across those present in the community [18] or to help people to define their identity by categorizing others and themselves into groups constantly compared and evaluated [19].

Other studies have shown that convergence may help to accomplish mutual goal [20], align representations [18], increase the quality of an interaction [21], and furthermore contribute to mutual comprehension by decreasing social distance [21].

According to Labov [22], convergence could be due to the need to add emphasis to expression and persist for the next interaction.

Finally, adaptation could be interpreted as a behavioral strategy to achieve particular social goals such as approval [11, 23] or desirability [24].

2.2 Description of key studies on phonetic convergence

Pardo [16] examined whether pairs of talkers converged in their phonetic repertoire during a single conversational interaction called a map-task. Six same-sex pairs were recruited to solve a series of 5 map tasks where their role – instruction giver or receiver – were exchanged. The advantage of the map task is to collect landmark names that are uttered several times during the interaction by each interlocutor in order to have the receiver replicate the itinerary described by the giver. One or two weeks before any interaction, talkers read out the set of map task landmark labels in order to obtain reference pronunciations. Just after interaction, the same procedure was also performed to test the persistence of convergence, i.e. to distinguish stimulus-dependent mimicry from mimesis which is supposed to originate from a deeper change of phonetic representations [25]. To measure convergence, 30 listeners were asked to judge similarity between pronunciations of pre-, map- and post-task landmark labels in a AXB test, X being a map-task utterance and (A,B) pre-, map- or

post-task of the same utterance pronounced by the corresponding partner. Results of this forced choice showed significant main effects of expose and persistence but there was also dependence of role and sex: givers' instructions converged more than receivers' instructions and particularly for female givers. It is in agreement with the results found by Namy [26].

Delvaux and Soquet [14] questioned the influence of ambient speech on the pronunciations of some keywords. These keywords were chosen in order to collect representatives of two sounds (the mid-open vowels [ɔ] and [ɛ]) the allophonic variations of which are typical of the two dialects of French spoken in Belgium. During these non interactive experiments, subjects were asked to describe a simple scene: "C'est dans X qu'il y a N Y" (*It's in X that there are N Y*), where X were locations, N numbers and Y objects. This description was either uttered by the speaker or by recorded speakers using the same or the other dialect. Pre- and post-tasks were also performed for the same reasons enounced previously. The phonetic analysis focused on the production of the two sounds that were used in the two possible labels X. The authors sought for unintentional imitation. To characterize the amplitude of that change, they compared durations and spectral characteristics of target sounds. In most cases, small but significant displacements towards the prototypes of the other ambient dialect were observed for both sounds (see the lowering of the canonical values in Test 1 and 2 in Fig. 1). Similar unconscious imitation of characteristics of ambient speech has also been observed by Gentilucci et al [27] for audiovisual stimulations.

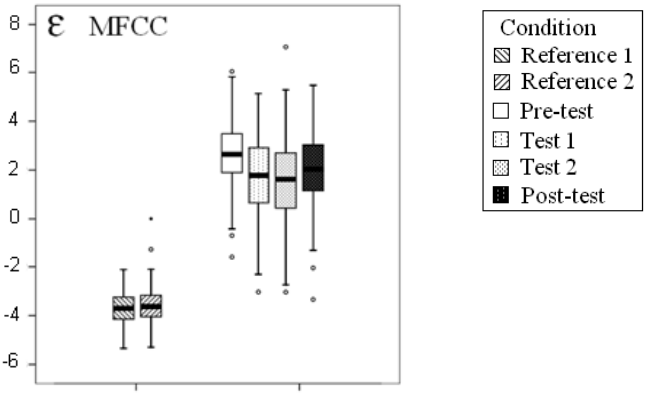

Fig. 1. Results on spectral distance calculated by Delvaux and Soquet [14]. It can be seen that, during tests (Tests 1 & 2), subjects are getting away from their own reference (Pre-test) and closer to the other dialect (References 1 & 2).

Aubanel and Nguyen [28] also conducted experiments to study the mutual influence between French accents, i.e. northern versus southern, that could be part of the subjects' experience. They have proposed an original paradigm to collect dense interactive corpora made up of uncommon proper nouns. They defined some criteria in order to discriminate the two accents, i.e. schwa, back mid vowels, mid vowels in word-final syllables, coronal stops, and nasal vowels. Uncommon proper nouns containing these segments are chosen so as to maximize coverage of alternative spellings. They chose their subjects in a major high school and grouped them

according to their sex and to a similar score on the Crowne-Marlowe [29] social desirability scale. One week before any interaction, subjects read out three sets of 16 names to get reference pronunciations. This session was repeated just after the interactions to measure mimesis. During the interaction, dyads were asked to associate names with photographs and the corresponding characters' statements. Aubanel and Nguyen used a Bayes classifier to automatically assign subjects to a group and test different levels of convergence in the dyads (towards the interlocutor, the interlocutor's group and accent) using linear discriminant analysis performed on spectral targets. They found very few instances of convergence. Additionally convergence was quite dependent of the critical segments analyzed, the sessions and the pairs.

2.3 Comments

These studies show that phonological and phonetic convergence is very weak. The experimental paradigms used so far either collect few instances (typically a dozen in Aubanel and Nguyen) of few key segments or many instances of a very small set of key segments (two in Delvaux and Soquet). These segments are always produced in a controlled context within key words.

Both studies have focused on inter-dialectal convergence and segments that carry most of the dialectal variation. This a priori choice is questionable since it remains to be shown that subjects at first negotiate these critical segments before or more easily than others. Since the convergence is segment-dependent, it is interesting to study the speakers' alignment on the common repertoire of their mother tongue. In our experiments, we will examine the convergence of the eight French peripheral oral vowels.

In most studies, interlocutors or ambient speech are not known a priori by the subjects. The authors were certainly expecting to observe on-line convergence as the dialog proceeds. The hypothesis that adaptation and alignment is immediate and fast is questionable: in the following we will compare convergence of unknowns with those of good friends.

Table 1. First speech dominoes used in the interactive scenario. Interlocutors have to choose and utter alternatively the rhyming words. Correct chainings of rhymes are high lightened with a dark background.

spk 1	spk 2	spk 1	spk 2	spk 1	spk 2	spk 1	
rotɔr	tordy	ʃimi	ʃema	ʒilɛ	leto	ɡeri	…
	bɛrly	dyre	repi	pile	kepi	todi	

3 Material and Protocol

During our experiments, speakers were instructed to choose between two words displayed on a computer screen.

3.1 Speech Dominoes

The rule of the game is quite simple. Speakers have to choose between two words the one that begins with the same syllable as the final syllable of the word previously uttered by the interlocutor (see Table 1). Such rhyme games - here speech dominoes –

are part of the children's folklore and widely used in primary school, for example for language learning. We decided to chain simple disyllabic words such as:

bateau [bato], taudis [todi], diffus [dify], furie [fyri], etc.

We used only two disyllabic words in order to limit the cognitive load and ease the running of successive sessions.

The words were chosen to uniformly collect allophonic variations of the eight peripheral oral vowels of French: [a], [ɛ], [e], [i], [y], [u], [o], [ɔ].

To force mutual attention during the interaction, the word list has been built so that the speaker could not guess the next domino given the sole history of the dialog. In fact, he has to pay attention to the word uttered by his interlocutor to decide which "domino" he would have to utter next. For instance, spk 2, after having chosen [tordy] in Table 1, will be presented with the following two alternatives, namely [ʃema] and [repi]. Since [dyʃe] and [dyre] are two valid French common words with almost the same word frequency, spk 2 will have to wait until spk1 chooses the right rhyme to decide about his own.

A chain of 350 dominoes was thus established that permitted us to collect almost 40 exemplars of each peripheral oral vowel (see Table 2).

Table 2. Number of phones collected for each speaker during the dominoes' game. 350 CV or CVC syllables are pronounced in total.

phones	a	ɛ	e	i	y	u	o	ɔ	others
#items	47	48	45	43	44	40	43	31	9

3.2 Conditions

The speakers pronounced dominoes under different conditions. First of all, we needed to get references for each speaker, we called this condition *pre-test*. To do this, they uttered a list of 350 words before any dialog with their interlocutor. The pre-test words were the same as those pronounced by the two speakers during the dominoes' game. It allowed us to characterize each speaker's phonetic space and to measure the amplitude of adaptation if any.

Fig. 2. Face-to-face interaction

3.3 Experiments

In this paper, we only contrast the pre-test condition and the interactive game played during three experiments:

- Experiment I: speakers were in two different rooms and communicated through microphones and headphones. This setup was easy to realize thanks to the MICAL platform of our laboratory (two rooms separated with a tinted mirror). Speakers were unknown to each other.
- Experiment II: same as Experiment I but with a reduced set of good friends, people that know each other or work together since a long time (mean of 15 years from 10 years to 25 years).
- Experiment III: speakers were in a face-to-face interaction. We studied also dyads of good friends (mean of from 6 months to 3 year and 6 months).

In the two cases, they were instructed to avoid speech overlaps and repairs so as to ease automatic segmentation and alignment.

3.4 Experimental settings

For Experiments I and II, people played through sets of microphones and headphones. Signals were digitized at 16 kHz thanks to a high-quality stereo sound card. Dominoes were displayed on two computer screens displaying a pdf file.
For Experiment III, the setting is quite different. Speakers sat on each side of a table facing two back-to-back computer screens. They were recorded with a camera - a mirror allowed us to capture both interactants (see Figure 2) - their head movements were monitored using four infrared cameras (Qualisys system®).

We used two keyboards connected to the same computer to forward turns: when a speaker finishes uttering his domino, he presses a key on his keyboard to display the two choices for his next turn on his own screen.

3.5 Characterization

Delvaux and Soquet [14] noticed that global automatic analysis of spectral distributions by MFCC (Mel Frequency Cepstral Coefficients) lead to quasi-identical but more robust characterization of convergence than a more detailed semi-automatic phonetic analysis such as formant tracking. Aubanel and Nguyen [28] similarly used automatic recognition techniques to recognize idiolects.

Here, we trained phone-sized context-independent HMMs with 5 states trained using HTK on the pre-test data. The input parameters are the first 12 MFCC + energy + deltas of theses parameters. After various forced alignments, we compared the distributions of normalized self vs. other's recognition scores of central states of each vowel (see Fig. 4 and Fig. 5). Paired t-tests were also performed to compare changes of distributions of scores of vowels produced in the same words (175 words for each speaker).

We align signals with a network of pronunciation variants for each word to semi-automatically segment the signals. This segmentation was then checked by hand (by three different annotators). Devoiced or creaky vowels - often high vowels [i], [y] in unvoiced context - were discarded. Dialectal variations were also considered: allophonic variations of mid-vowels (open or closed) are determined according to the speaker-dependent partition of the range of the first formant.

4 Results

4.1 Phonological variations

Despite the fact that our corpus was not designed to enhance dialectal variations (unlike Aubanel and Nguyen [28]), we observed some dialectal variations mainly concerned with allophonic variations of mid-vowels. Most participants came from North of France and thus used exclusively open vowels in closed syllables (e.g. *sabord* /sabɔʀ/ vs. *sabot* /sabo/). Other interlocutors spectrally contrasted minimal pairs such as *vallée* vs. *valais* (/vale/ vs. /valɛ/), *miné* vs. *minet* (/mine/ vs. /minɛ/), etc. We observed few cases of phonological adaptation, i.e. subjects adopting a pronunciation different from the one chosen in their pretest to get closer to the pronunciation of their interlocutor. Most interactions resulted in convergence of allophonic choices (see Fig. 3) but this is not significant due to limited data. For example, for the vowel [e], a mutual adaptation can be seen during the interaction between ALa and MGB and also between ALa and MSM.

We should however mention that the labeling of allophonic variations of mid-vowels is very difficult, since French speakers have now the tendency to front mid-closed vowels [30-31]. The labeling is particularly difficult in non accented positions where vowel undershoot or co articulation may override perceptual intuition. We always privileged labeling based on objective measurements that tend to favor mid-closed options.

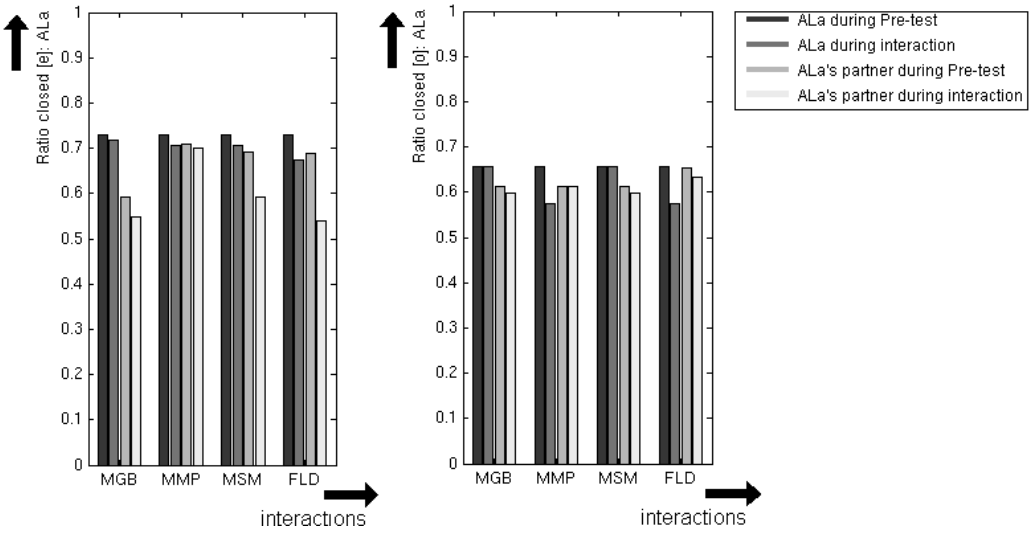

Fig. 3. Proportion of peripheral mid vowels pronounced as closed by 4 pairs (left: the vowel [e]; right: vowel [o]). The initiator was the same female ALa interacting with 3 males (MGB, MMP, MSM) and one female (FLD). For each pair, bars represent the proportion uttered during respectively the ALa pretest, ALa interacting with her interlocutor, her interlocutor interacting with ALa and the interlocutor's pretest.

4.2 Sub-phonemic convergence

Given the assumption that each phone was properly labeled, we compared the distributions of normalized recognition scores of the pre-test and interactive utterances, as explained previously. These utterances were recognized by the speaker's own HMMs in a first time and in a second time by the HMMs of his interlocutor. For pre-test data, we expected high scores for HMMs tested on their own training data by construction and lower scores for HMMs of the interlocutor. The recognition score of each vowel is the average log likelihood per frame for the central state of the corresponding HMM. The difference between the scores somehow reflects the inter-speaker distance. Convergence would be characterized by a decrease of scores by self HMMs and an increase of scores by the other's HMMs.

The recognition is thus performed by HMM models of each speaker and of his/her interlocutor. Fig. 4 and Fig. 5 compare the distributions of normalized recognition scores for the pre-test (left) versus the interaction (right). Scores are typically higher for phones uttered by one speaker and recognized by his own HMMs. In case of an interaction between two unknowns (cf. Fig. 4), the distributions computed for the interactive speech do not evolve so much. We observe stronger convergence in case of good friends (cf. Fig. 5).

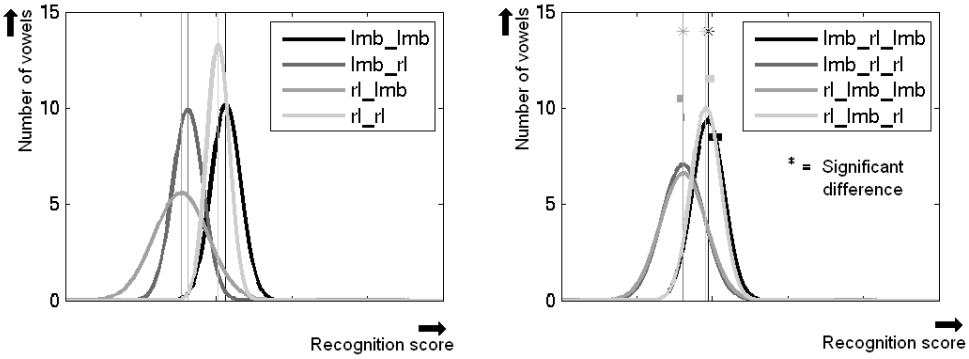

Fig. 4. Distribution of recognition scores for the vowels of disyllabic words produced by two unknowns. The recognition is performed by their own HMM models and by the HMM models of their interlocutor. Scores are expected to be higher by using their own HMM models. Left: scores for word lists read aloud in isolation; this speech data is used to train the speaker-specific HMM models. It can be seen that, by using the own model of each interlocutor (lmb_lmb and rl_rl), higher recognition scores are obtained than for cross recognition (lmb_rl and rl_lmb) Right: same words pronounced in a verbal domino game. In this case, we expect a decrease of recognition scores by using the own model of each interlocutor (lmb_rl_lmb and rl_lmb_rl) and an increase of recognition scores by using cross recognition (lmb_rl_rl and rl_lmb_lmb). Here, only small adjustments are observable (weak shift on the left for lmb_rl_lmb and rl_lmb_rl and on the right for lmb_rl_rl and rl_lmb_lmb).

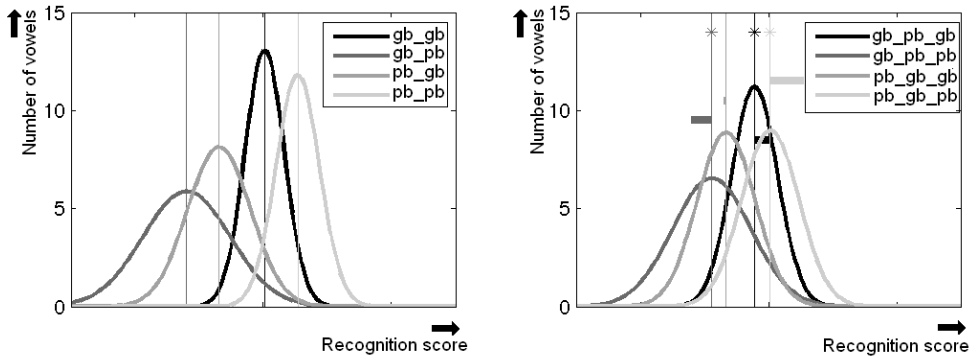

Fig. 5. Same as Fig. 4 but for disyllabic words produced by two old friends. Stronger convergence is observed here since larger shifts are observed.

4.3 Distributions of recognition scores

Fig. 6 shows the average convergence rate for all dyads recorded in the two experiments. This is computed as the relative distance of vocalic targets produced in pretest vs. interaction (central state of the HMM alignment).

Similarly to Delvaux & Soquet [14], linear discriminant analysis is performed on the target MFCC parameters for each vowel to categorize each interlocutor's vocalic space into two distinct groups. For each pair, pre-test and interactive vocalic targets are projected onto the first discriminant axis. A normalized convergence rate is then computed by dividing the distance between targets produced during the interaction with that produced during the pre-test for the same word.

Convergence is not systematic. We can see that the phenomenon is amplified with pairs of the same sex and particularly women (it can be due to a priori more similar speakers). We used this statement for the seven last interactions by selecting only women for the experiment and the results confirm this statement. An ANOVA has been done to assess significance of adaptation. Distributions with significant convergence rates are drawn in bold.

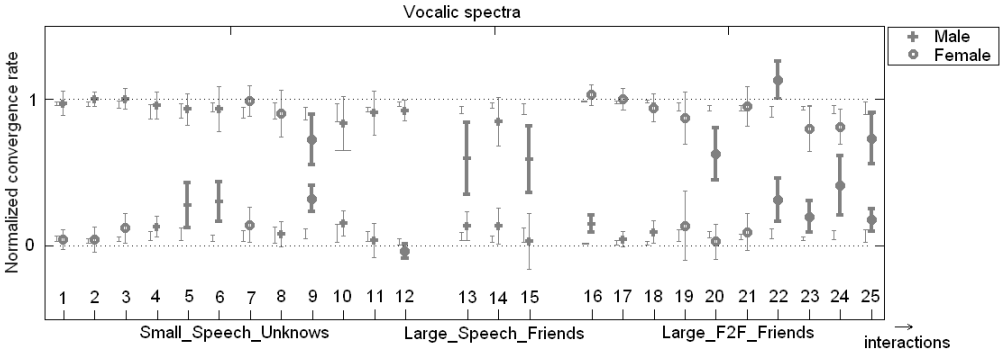

Fig. 6. Average convergence rate (calculated on 100 iterations) of vocalic targets of interlocutors for all conditions. A linear discriminant analysis on one random half of the pre-test data has been used to separate each interlocutor's vocalic space. Thus, a reference discriminant distance between interlocutors is obtained. It is used to calculate normalized convergence rates. First, it is used to calculate the convergence rate on the other half of the pre-test to have our reference departure for each interlocutor. The two dotted lines represent the mean of pre-test of the tested subjects (line 0) and of the reference subjects (line 1). Distributions displayed with bold lines are significantly different ($p<0.05$) from the corresponding pre-test (reference departure). Note that only two significant divergences are found (one speaker in the pair number 12 and one on the pair 22). Most convergence cases are observed with pairs with same sex.

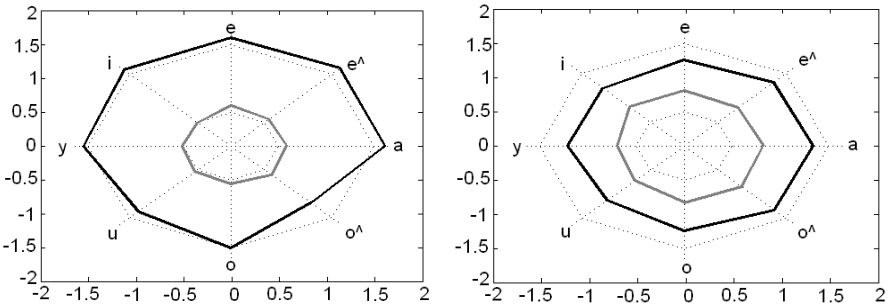

Fig. 7. Detailing convergence rates for two different pairs and for each vowel. Pair (a) seems not convergence at all except for the mid-open vowel [ɔ] while pair (b) exhibits complete mutual adaptation. The rates are calculated the same way as Figure 6. For each figure (or interaction), the dotted line on the outside corresponds to the reference subject and the other dotted line to the tested subject. The darker grey corresponds to the reference subject's convergence rates and the lighter one to those of the tested subject

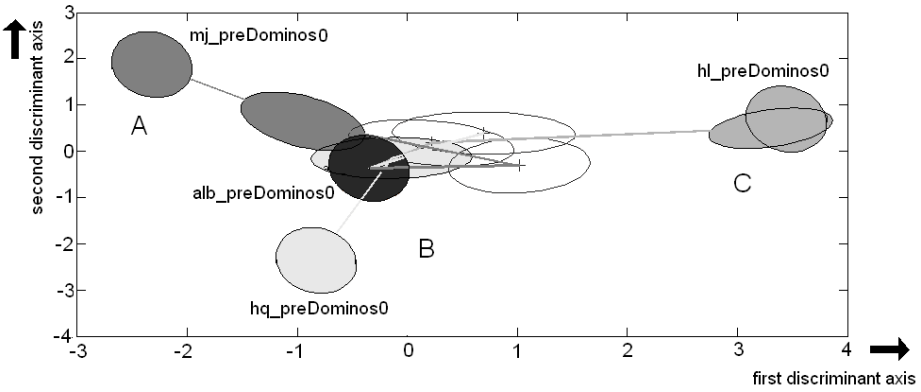

Fig. 8. First discriminant space projection of MFCC targets for [ɔ] produced by speaker alb (dark dispersion ellipsis for the pre-test drawn at the center) interacting successively with three interlocutors A, B, C (pre-test ellipses located at the periphery). Realizations for interactions are displayed with unfilled ellipsis for alb and filled ellipsis with same color as pre-test for interlocutors. While A and B converge to alb, alb and C do not adapt.

4.4 Convergence of vocalic targets

Convergence is a vowel- and interlocutor- dependent phenomenon. Fig. 7 shows that some pairs do not adapt at all while others show a significant mutual adaptation.

To study interlocutor-specific adaptation strategies, our game initiators interacted with 2 to 5 different interlocutors. Fig. 8 illustrates examples of target- and interlocutor-specific behaviors. When considering each vowel separately (40 occurrences in average, see Table 1), we do observe cases of full convergence (speakers A and B in Fig. 8).

Note however that our analysis is based on relative distances and should take into account the whole structure of the vocalic space of the interlocutors. Speakers notably fill differently their available acoustic space, especially between mid-vowels [32].

An evolution of convergence rates with time was expected. Convergence rates as a function of time as been plotted for each interlocutor but nothing relevant has been observed. Maybe the proposed task was too short to observe this phenomenon.

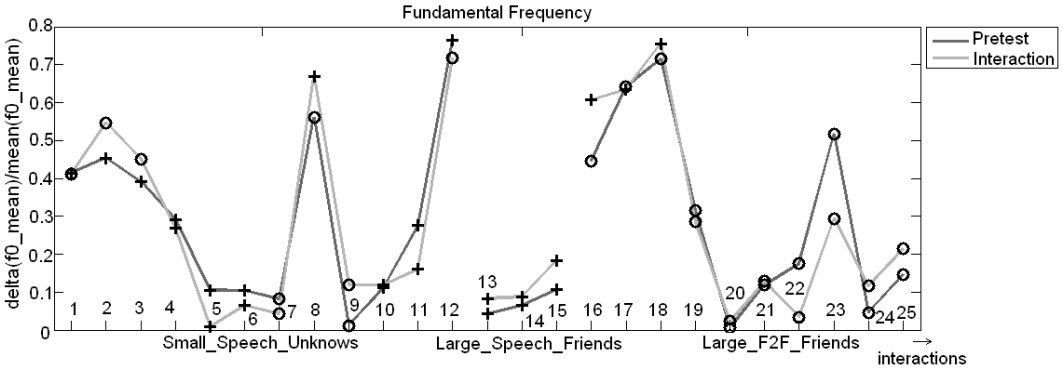

Fig. 9. Mean changes of relative difference between F0 registers. As expected the values are higher for pairs with different sex. No significant narrowing of this difference is induced by interaction. This can be due to the task that imposes short utterances.

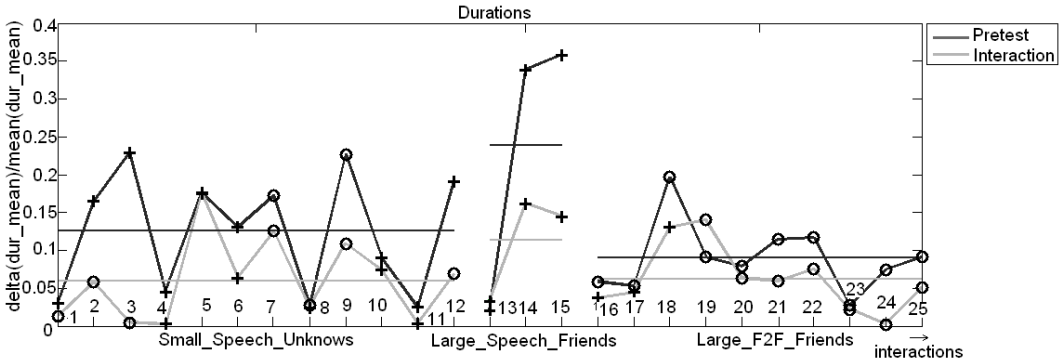

Fig. 10. Mean changes of relative difference between syllabic durations.

4.5 Prosody

Fig. 10 shows that fundamental frequency register was relatively unaffected by the interaction. The exchange of simple words does not favor attunement of melody.

Convergence of speech rhythm is clearly observed, certainly due to the 'ping-pong' task. This can also be due to the fact that speech rhythm was also much quicker in interactive speech compared to isolated word reading (cf. Fig. 10) with a notable shortening of final syllables. This is probably due to the task focusing on rhyme matching. Delvaux and Soquet [14] advise to discard final syllables for studying phonetic convergence. In our case, we did not find any difference between global and partial statistics except stronger convergence for the durations of vocalic nuclei of initial syllables.

5 Conclusions and perspectives

We proposed here an original speech game that quickly collects many instances of target sounds with a mutual influence that force interlocutors to engage into active action-perception loops. Distribution of target sounds can be explicitly controlled to observe convergence in action if any.

We found occurrences of strong phonetic convergence with only one instance of small divergence. This convergence strongly depended on the dyads – with strongest convergence observed for pairs of the same sex – and seemed to be phoneme-dependent. We used this observation to select our last subjects and the results confirmed a strongest convergence for dyads composed of women.

These objective measurements should be confirmed by subjective assessments such as promoted by Pardo [16]. We are also planning to conduct a series of subjective tests to determine if adapted stimuli offer a clearer perceptual benefit for listeners compared as to non adapted stimuli. Perception of degraded stimuli such as used by Adank et al [7] is an interesting option.

This gaming paradigm will now be used to select subjects and dyads who exhibit the strongest adaptation abilities and study more complex conversational situations. This data will be used to train speech synthesis engines that will implement these adaptation strategies. Such interlocutor-aware components are certainly crucial for creating social rapport between humans and virtual conversational agents [33].

Acknowledgments

This work has been financed by ANR Amorces and by the Cluster RA ISLE. We thank Frederic Elisei, Sascha Fagel and Loïc Martin for their help.

References

1. Giles, H., et al., *Speech accommodation theory: The first decade and beyond*, in *Communication Yearbook*, M.L. McLaughlin, Editor. 1987, Sage Publishers: London, UK. p. 13-48.
2. Brennan, S.E. and H.H. Clark, *Lexical choice and conceptual pacts in conversation.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 1996. **22**: p. 1482-1493.
3. Lockridge, C.B. and S.E. Brennan, *Addressees needs influence speakers early syntactic choices.* Psychonomic Bulletin and Review, 2002. **9**: p. 550-557.
4. Zoltan-Ford, E., *How to get people to say and type what computers can understand.* International Journal of Man-Machine Studies, 1991. **34**: p. 527-547.
5. Ward, A. and D. Litman. *Dialog convergence and learning*. in *International Conference on Artificial Intelligence in Education (AIED)*. 2007. Los Angeles, CA.
6. Pickering, M., et al., *Activation of syntactic priming during language production.* Journal of Psycholinguistic Research, 2000. **29**(2): p. 205–216.
7. Adank, P., P. Hagoort, and H. Bekkering, *Imitation improves language comprehension.* Psychological Science, 2010. **21**: p. 1903-1909.
8. Lakin, J., et al., *The chameleon effect as social glue: evidence for the evolutionary significance of nonconscious mimicry.* Nonverbal Behavior, 2003. **27**(3): p. 145–162.

9. Allwood, J., *Bodily communication - dimensions of expression and content*, in *Multimodality in Language and Speech Systems*, B. Granström, D. House, and I. Karlsson, Editors. 2002, Kluwer Academic Publishers: Dordrecht. p. 7-26.
10. Kopp, S., *Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors.* Speech Communication, 2010. **52**(6): p. 587-597.
11. Gregory, S.W. and S. Webster, *A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions.* Journal of Personality and Social Psychology, 1996. **70**: p. 1231-1240.
12. Edlund, J., M. Heldner, and J. Hirschberg. *Pause and gap length in face-to-face interaction*. in *Interspeech*. 2009. Brighton.
13. Kousidis, S., et al. *Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues*. in *Interspeech*. 2008. Brisbane.
14. Delvaux, V. and A. Soquet, *The influence of ambient speech on adult speech productions through unintentional imitation.* Phonetica, 2007. **64**: p. 145-173.
15. Benus, S. *Are we 'in sync': Turn-taking in collaborative dialogues*. in *Interspeech*. 2009. Brighton.
16. Pardo, J.S., *On phonetic convergence during conversational interaction.* Journal of the Acoustical Association of America, 2006. **119**(4): p. 2382–2393.
17. Dijksterhuis, A. and J.A. Bargh, *The perception-behavior expressway: automatic effects of social perception on social behavior.* Advances in Experimental Social Psychology, 2001. **33**: p. 1-40.
18. Garrod, S. and G. Doherty, *Conversation, co-ordination, and convention: An empirical investigation of how groups establish linguistic conventions.* Cognition & Emotion, 1994. **53**: p. 181-215.
19. Tajfel, H. and J. Turner, *An integrative theory of intergroup conflict*, in *The Social Psychology of Intergroup Relations*, W.G. Austin and S. Worchel, Editors. 1979, Brooks-Cole: Monterey, CA. p. 94-109.
20. Clark, H.H., *Using Language*. 1996, Cambridge, UK: Cambridge University Press.
21. Babel, M.E., *Phonetic and social selectivity in speech accommodation*, in *Department of Linguistics* 2009, University of California: Berkeley, CA. p. 181.
22. Labov, W., *The anatomy of style-shifting*, in *Style and Sociolinguistic Variation*, P. Eckert and J.R. Rickford, Editors. 2001, Cambridge University Press: Cambridge, UK. p. 85-108.
23. Giles, H. and R. Clair, *Language and Social Psychology*. 1979, Oxford: Blackwell.
24. Natale, M., *Social desirability as related to convergence of temporal speech patterns.* Perceptual Motor Skills, 1975. **40**: p. 827-830.
25. Donald, M., *Origins of the Modern Mind: three stages in the evolution of culture and cognition*. 1991, Cambridge, MA: Harvard University Press.
26. Namy, L.L., L.C. Nygaard, and D. Sauerteig, *Gender differences in vocal accommodation: The role of perception.* Journal of Language and Social Psychology, 2002. **21**: p. 422–432.
27. Gentilucci, M. and P. Bernardis, *Imitation during phoneme production.* Neuropsychologia, 2007. **45**(3): p. 608-615.

28. Aubanel, V. and N. Nguyen, *Automatic recognition of regional phonological variation in conversational interaction.* Speech Communication, 2010. **52**: p. 577-586.
29. Crowne, D.P. and D. Marlowe, *A new scale of social desirability independent of psychopathology.* Journal of Consulting Psychology, 1960. **24**: p. 349-354
30. Coveney, A., *The Sounds of Contemporary French : Articulation and Diversity*. 2001, Exeter, UK: Elm Bank Publications.
31. Boula de Mareüil, P., et al., *Accents étrangers et régionaux en français : Caractérisation et identification.* Traitement Automatique des Langues, 2008. **49**(3): p. 135-163.
32. Ménard, L., J.-L. Schwartz, and J. Aubin, *Invariance and variability in the production of the height feature in French vowels.* Speech Communication, 2008. **50**(1): p. 14-28.
33. Gratch, J., et al. *Creating rapport with virtual agents*. in *Intelligent Virtual Agents (IVA)*. 2007. Paris, France.