Inducing Imitative Phonetic Variation in the Laboratory

Véronique Delvaux[$*], *Alain Soquet*[*]

[$]FNRS, [*]Université de Mons-Hainaut
delvaux@umh.ac.be

ABSTRACT

Vocal imitation governs speech acquisition. But the role of imitation in the routine phonetic behavior of adult speakers still needs to be investigated. The experiment reported here is an attempt to induce imitative phonetic variation in the laboratory. Results show that the speakers' productions get closer to the productions they are exposed to (coming from a recorded speaker of another dialect), although the instructions are not to imitate or even to listen to the recorded speaker. We discuss the implications of this finding for the study of the emergence and the propagation of sound changes within a speech community.

Keywords: Imitation, mimesis, sound change.

1. INTRODUCTION

This paper is about the role of imitation in the routine phonetic behavior.

Vocal imitation is one of the basic mechanisms governing the acquisition of speech by children [1]. Adults use imitation of many aspects of their interlocutor's behavior (gestures, postures) in order to smooth social interactions [2]. Social psychology and sociolinguistics have long been interested in imitative behaviour. In particular, speech accommodation theory (SPA) [3] studies the ways in which individuals adjust their communication behaviour to one another, either to become more alike (convergence) or to exaggerate their differences (divergence). SPA states that individual short-term accommodation may become a long-term accommodation, which may in turn spread throughout the community at large. But the mechanisms that would be responsible for such developments (from short-term to long-term accommodation, and from the individual to the community level) still need to be specified. In this paper, we specifically focus on the phonetic and cognitive aspects of the accommodation process.

Indeed, imitation of speech by adult speakers has been little studied as such. Phoneticians have rather used imitation as a tool to study speech perception, in particular through the shadowing task [4]. In the present study direct measurements have been made on adult speech productions to investigate their ability to spontaneously imitate the speech of others. A distinction is made between *imitation*, the acquisition of a new phonetic variant as a result of the exposure to a matching model, and *mimesis*, a higher-order cognitive ability that involves a modification in the representations associated with the motor task to perform [5]. The hypothesis to be tested is that interacting speakers imitate each other's speech, so that not only are individual phonetic realizations modified, but phonetic representations are also. Then, the mimetic mechanism, occuring at the (inter)individual level, may account for two apparently paradoxical features of the system: (i) the stability or at least the convergence of phonetic realizations within a speech community, i.e. the existence of different dialects, and (ii) in accordance with SPA, the flexibility of the system through sound change. In this view, sound change could be the product of mimetic interactions between speakers.

In the following experimental study, speakers have been exposed to phonetic realizations that come from another dialect. The primary goal is to induce imitative phonetic behavior in the laboratory, and to design a way of assessing the magnitude of the effect. The second goal is to differentiate between imitation and mimesis. Particular attention has been paid to the behavior of the speakers after the period of exposure: if part of the imitative change is maintained in the absence of the stimulus, it means that it has been kept in long-term memory, i.e. that it has affected phonetic representations.

2. MATERIAL AND METHOD

2.1. Subjects and dialects

Eight female speakers (S1 to S8) participated in the experiment. They are all French native speakers who were born and have always lived in the area of Mons, Belgium. The 'reference speaker' (REF) is a female speaker who has spent her whole life in the area of Liège, Belgium. The Mons and Liège dialects were selected because they differ in the phonetic realization of the two target words of this study: "caisse" /kɛs/ and "frigo" /fʀigo/. Typically, the

vowel in /kɛs/ is realized as [ɛ] in Mons, but with a long vowel [ɛ:] in Liège. The vowel in /fʀigo/ is realized as [o] in Mons, but as a low-mid back vowel [ɔ] in Liège.

2.2. Experimental paradigm

The experimental setting is the following. The speaker is seated at a table, in front of a computer screen, with loudspeakers at both sides. The task is to pronounce a sentence based on the information given on the screen. All sentences follow the same pattern: "C'est dans X qu'il y a N Y" (*In X there are N Y*). The computer screen displays ideograms for X and Y, and a number for N. X may be {caisse, frigo}, N {1, 2, 3} and Y {stylo, bouquin, tortue, cageot, caisse enregistreuse, fleur, bombe, gourmand, main, mur}. For example, when a box, the number '1' and a book are displayed on the screen, the speaker must say the French sentence for '*In the box there is one book'*.

Prior to the experiment, a short training session consisting in 60 trials (#X * #N * #Y) allows the speaker to become familiar with the experimental setting and material. A written sample of the expected sentence pattern is provided on paper, and words appear on the screen under the corresponding ideograms. Speakers are instructed that they must learn both the sentence pattern and the associations between words and ideograms. Indeed, during the experimental sessions, only the ideograms are displayed on the screen to avoid read speech. Speakers are told that the purpose of the experiment is to assess their abilities of memory and attention. Actually, a phonetic analysis of the realizations of X ("caisse"/"frigo") was performed.

Speakers are tested under three experimental conditions. During the pre-test (PRE), the speaker performs the task on her own. During the test (TEST), the sentences are produced alternatively, either by the speaker or by the pre-recorded 'reference speaker' (REF) via one of the loudspeakers. For a period of 500 ms preceding each trial, an arrow points at the one that has been selected to produce the sentence (either the speaker or one of the loudspeakers). The order is pseudo-random. Finally, during the post-test (POST), the speaker performs the task on her own again.

Both PRE and POST consist of 100 trials (50 "frigo" and 50 "caisse"). TEST consists in two recording sessions, $TEST_1$ and $TEST_2$, each comprising 300 trials from which 100 are performed by the speaker (50 "frigo" and 50 "caisse"). The data analyzed in this paper consist in 200 occurrences of "caisse" and 200 occurrences of "frigo" per speaker, i.e. 3200 vowels as a whole.

2.3. Measurements

Speech signals were recorded and digitalized by the computer used for the presentation of the stimuli (22050Hz, 16 bits). Segmentation of the corpus was performed using a semi-automatic procedure [6].

The analysis of the phonetic realizations of the vowels /ɛ, o/, in /kɛs/ and /fʀigo/ respectively, is based on the following measurements: (i) the duration of the vowel (D); (ii) its spectral properties, as expressed by the 24 -frequency cepstra coefficients (MFCC) and by the values for the first three formants (F1, F2, F3). Spectral properties were measured at the first third of the vowel.

2.4. Data processing

The purpose is to measure the extent to which the phonetic realizations of the vowels /ɛ, o/ vary across experimental conditions. Discriminant analysis was used to obtain a space in which the productions of REF can be distinguished from those of the Mons speakers in PRE using a given set of acoustic cues, either D, or MFCC, or F1, F2, F3. Data for TEST and POST were then transferred in the same discriminant space. Fig. 1 illustrates a stylized case, in which the phonetic realizations of the speakers are modified during the experiment: (1) from PRE to TEST, the statistical distribution moves on the discriminant axis towards that of the reference speaker (REF); (2) from TEST to POST, the statistical distribution moves back towards the speakers' typical pronunciation as measured in PRE.

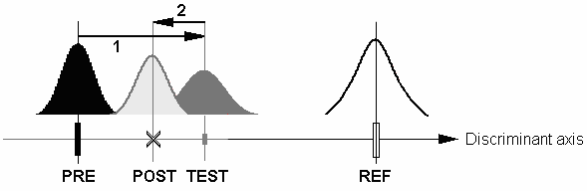

Figure 1. Data processing

Finally, the significance of the potential changes across conditions has been tested using a repeated measures ANOVA (one per discriminant space). The dependent variables are the discriminant scores. Experimental condition is a four-level within-subjects variable (PRE, TEST1, TEST2 and POST). The eight speakers form the between-subjects variable. Running the ANOVA on the discriminant scores rather than on the measures themselves ensures that any significant change is towards the reference.

3. RESULTS

The analysis of variance shows that both the within-subjects and the between-subjects factors, as well as the interaction between them, determine a significant variation in the discriminant scores at the .001 level. This is true of both vowels /ɛ, o/ for duration as well as for spectral properties.

2.5. The Experimental condition effect

Fig. 2 gives the results for /fʀigo/ and /kɛs/ (all subjects). Each horizontal line in the figure corresponds to a different discriminant analysis, based on the cues that were used for discrimination, either D, or the 24 MFCC, or F1, F2, F3. The left part of the figure represents the performances of the discriminant analysis, i.e. the rates of correct classification of the PRE and REF data into the categories of PRE and REF (in %). The right part of the figure gives the means obtained for the discriminant scores on the normalized discriminant axis in the four experimental conditions.

Fig. 2 shows that, when data for all subjects are considered together, there is a strong imitative effect. During the TEST condition, the realizations of the vowels /ɛ, o/ by the speakers from Mons get closer to the realizations they are exposed to (REF). During the post-test, when the speakers perform the task on their own again (POST), they only partly come back to their typical realizations, as measured in the pre-test (PRE). Thus, part of the imitative change is maintained in the absence of the stimulus, especially in the case of /ɛ/ (see Fig. 2). It may be that mimesis, more than imitation, is at work here.

Data obtained for /o/ duration are non typical. This is probably due to the lack of power of this parameter for discriminating between the two dialects (PRE vs. REF): the rates of correct classification are indeed around 70 % in this case.

2.6. The Speaker effect

Both the Speaker variable and its interaction with Experimental condition yield significant variation in the data. Fig. 3 illustrates this result in the case of /ɛ/ duration. The average discriminant scores and the corresponding duration values (in ms) are plotted for each speaker in each experimental condition. As shown in Fig. 3, each speaker starts from her specific phonetic habits/duration level in PRE. Then, there are significant changes across conditions: the phonetic realizations of /ɛ/ get longer when the speakers are exposed to the long /ɛ/ pronounced by REF. But each speaker remains in her own duration range, which accounts for the Speaker effect. However, part of the Experimental condition effect depends on the speaker itself, e.g. whether the imitative change is larger in TEST 2 than in TEST 1, or to what extent the change is undone in POST. This accounts for the significant interaction between the main factors.

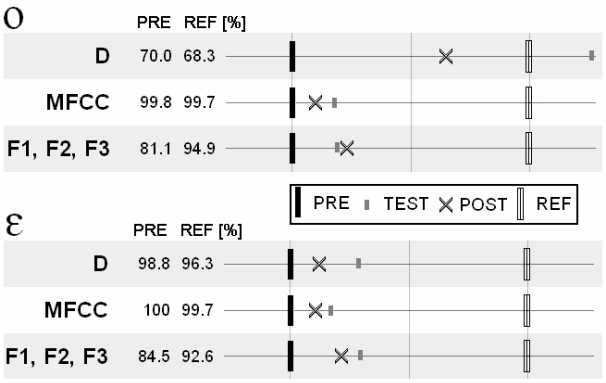

Figure 2. Results for /fʀigo/ (above) and /kɛs/ (below) (all subjects). Rates of correct classification by discriminant analysis (in %). Discriminant scores distribution means across experimental conditions.

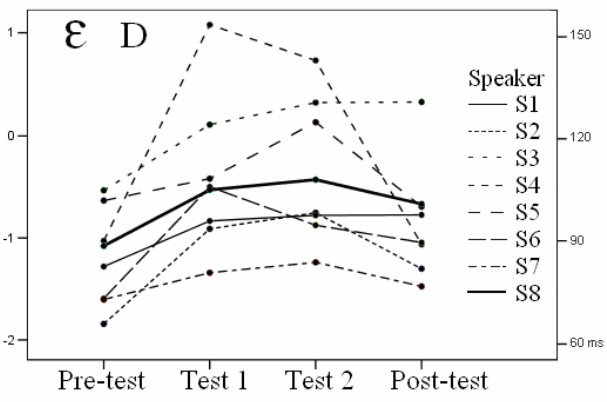

Figure 3. Results for /ɛ/ duration: average discriminant scores and corresponding duration values for each speaker in each experimental condition.

2.7. Time evolution

Fig. 4 gives the time evolution of the discriminant scores in the case of S7 (/o/, MFCC) across experimental conditions, starting from the first 50 trials of PRE and ending by the last 50 ones of POST. Data for REF are also included as a reference. The six horizontal lines on the figure represent the six medians, one per condition by this speaker. Fig. 4 has been selected to illustrate some of the most typical time evolution patterns found in the data. First, the imitative change usually occurs quickly, i.e. quite early in the TEST 1 session. In some cases, the new target seems to be achieved after a couple of trials only. Second, the imitative change is only gradually abandoned during the POST session. At the

beginning of POST, S7 pronounces an imitative /o/, i.e. a lower vowel than her typical (PRE) [o]. Then, she gradually comes back to her initial pronunciation, as measured in PRE.

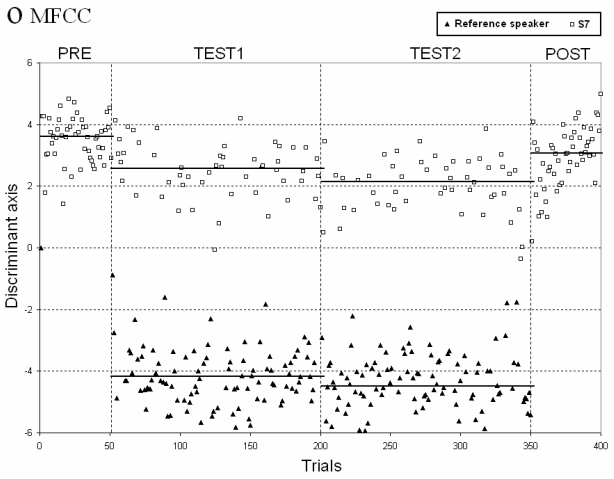

Figure 4. Time evolution of the discriminant scores for MFCC (in [o], /frigo/) along each experimental condition comparing Reference speaker and S7.

3. DISCUSSION

In this study, we have devised an experimental setup to induce imitative changes in the phonetic realizations of speakers and a method to measure the magnitude of the change. The imitative effect has been observed for all subjects, although there were some interesting inter-subject differences in the modalities of the change. Both the duration and the spectral characteristics of /ɛ, o/ were consistently modified. The speakers got closer to the phonetic realizations of the reference speaker although they were not instructed to imitate what they heard, or even to listen to it. After the experiment, they were not aware of having imitated REF. This means that the imitative process occurs at a low level of processing, not directly accessible to awareness or control; it is unintentional. Moreover, the phenomenon has been observed although the situation of communication was not interactive. We expect the effect to be stronger when the speakers are engaged in inter-individual spoken exchanges, since vocal imitation may increase the perceived smoothness of social interactions [2].

Turning to the potential role of this process as one of the regulators governing speech, we need to know if mimesis, more than imitation, is at work here. The results of the study show that a significant part of the imitative change is maintained in the POST condition. At the beginning of the POST session, the speakers still use the modified pronunciation, and then at some point during the session, they start to get back to their typical realization. This process may span over 50 trials (350 sec), which indicates that the new target has been kept somehow in long-term memory. Thus, not only the phonetic realization, but also the representation of the motor task to perform and/or the acoustic target to reach have been modified during the process. This is compatible with the claim that phonetic representations are flexible, graded, and constantly updated through a general mechanism of adaptation to experience [7]. Mimesis is a good candidate for describing such a mechanism.

Future work will have to confirm the present results on a larger scale, in particular with male speakers. Only then may mimesis be considered as a serious candidate in the search for the mechanisms that lead to sound change without threatening the stability of sound systems. Convergence and stability of the phonetic representations among a speech community may be achieved through constant re-actualizations of these representations via mimetic interactions between speakers. Constant re-actualizations also imply intra-individual variability, which can lead to sound change in the long run, in line with speech accommodation theory.

The potential link between imitation, representations updating, and the emergence and collective structuring of phonological systems may be studied using computer simulations [8]. The present study defines a methodological framework to gather the experimental data necessary to feed computer models, and as such provides a way to study self-organization in phonological systems.

4. REFERENCES

[1] Kuhl, P.K., Meltzoff, A.N. 1996. Infant vocalizations in response to speech: vocal imitation and developmental change. *J. Acous. Soc. Am*, 100(4), 2425–2438.

[2] Chartrand, T.L. and Bargh, J.A. 1999. The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910.

[3] Giles, H., Powesland, P. 1997. Accommodation theory. *Sociolinguistics: A reader*, N. Coupland et al. (eds.) Basingstoke: Macmillan.

[4] Nye, P. W., Fowler C.A. 2003. Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. *J. Phon.* 31, 63–79.

[5] Donald, M. 1991. *Origins of the Modern Mind: three stages in the evolution of culture and cognition.* HUPress.

[6] Delvaux, V., Demolin, D., Soquet, A. 2004. Interactions mimétiques entre locuteurs : une étude expérimentale, *XXVe JEP*, Fès, 153-156.

[7] Pierrehumbert, J. 2003. Probabilistic Phonology: Discrimination and Robustness. Bod et al. (eds.) *Probabilistic Linguistics.* MIT Press, 177-228.

[8] de Boer, B. 2000. Self-organization in vowel systems. *J. Phon.* 27, 1-25.