Perceptually driven VOT lengthening in initial stops by French-L1 English L2-learners

Véronique Delvaux[+*], Kathy Huet[*], Myriam Piccaluga[*], Bernard Harmegnies[*]
[+] FNRS, [*]Laboratoire des Sciences de la Parole, AUWB, Université de Mons-Hainaut, Belgium.
E-mail: delvaux@umh.ac.be

Abstract

This paper reports on an exploratory study of the processes involved in the acquisition of new phonetic control regimes in L2 learning. We focus here on the acquisition of long VOT initial stops by French native speakers. Francophone speakers were asked to repeat /ta/ stimuli varying in VOT and burst intensity (BI). The productions were analysed in terms of VOT, BI and duration of the following vowel (Dur). Results show that the speakers are fairly accurate in the imitation task. They exhibit more fidelity to VOT than to BI variation, more sensitivity to BI than to VOT variation, and a good potentiality to respond in terms of Dur (even though vowel duration is maintained constant). These results are presented in relation with the high interspeaker variation.

1 Introduction

The framework of this paper is a research project on the processes involved in the acquisition of new phonetic control regimes in L2 learning [2].

Extensive phonetic training has been shown to improve the production and perception of non native segmental and suprasegmental contrasts in adult speakers (e.g. [1,5,6,8,9]). Introducing high variability in the phonetic input has revealed to be a successful method: improvements in the perception of the training material reasonably generalize to new words and talkers, improvements in perception partly transfer to production, and both are retained after the completion of the training. Typically, phonetic training procedures are extensive (at least 10 sessions of ½ hour) and explicit (identification task with feedback). The focus is on perception: the pre-test and post-test measurements, as well as the training material, consist in perceptual tasks. When present, post-training productions are also indexed by perceptual judgements (from native speakers) [1,4,9].

In this study, we focus on the *production* of speech by subjects faced with models that do not fit into their native system. More precisely, we report on the production of lengthened VOT in initial stops by French natives who are involved in a syllable repetition task. Their situation is similar, among others, to the one of French-L1 English-L2 beginners trying to produce the English word-initial aspirated /t/ → [t^h]. Although /t/-/d/ is a native contrast in French (realized as [t]-[d] in all word positions), those learners need to master the aspirated allophonic variant in order to avoid foreign accented speech in English. Acquiring aspiration involves mastering the appropriate motor control regimes to produce new patterns of inter-gestural timing, i.e. desynchronizing supra-laryngeal (oral) release and vocal fold adduction. Among the numerous acoustic cues signalling this desynchronization [7], long VOT is generally considered as the most reliable.

2 Material and method

2.1 Experimental paradigm

The aim of the paper is to provide a fine-grained acoustic analysis of the learners productions *during* training, in a first attempt to investigate *in progress* the acoustic effects of the motor control regimes reorganization process. 'Training' was limited to the exposure to acoustic signals (the stimuli, presented one syllable at a time). Speakers were asked to 'repeat the word as faithfully as possible, as if it was a word from a foreign language'. There were 10 participants in the experiments, all native French speakers from Belgium with very limited knowledge of English. An AB discrimination task was carried out on the stimuli before the production experiment itself. To complement the acoustic analysis, the productions were also indexed by similarity judgements from the same Belgian French speakers, as well as by similarity and typicality judgements from 15 American English listeners. In this paper, we present the results of the production experiment only.

2.2 Stimuli

In total, 36 stimuli (4*9) were built up by manipulating a natural production of /ta/ from a native American English speaker. The original [t^ha] was modified along two dimensions: (i) the intensity of the 20 ms burst plus friction phase, i.e. 4 series across which instantaneous amplitudes were multiplied by a factor ½, 1, 2 and 4[1] (ii) the duration of the following aspiration phase (from 0 to 80 ms by 10 ms steps), i.e. 9 stimuli ranging from 20 to 100 ms VOT. The following vowel remained unchanged. On both dimensions, the stimuli continuum ranges from a typical L1 [t] to an *enhanced* L2 [t^h]. Stimuli were presented in 3 blocks of 4 series of 45 trials.

2.3 Measurements

Measurements of the learners' productions involve: (i) the ratio between RMS amplitude over the burst duration, and the RMS amplitude in a similar time window centred at the maximum amplitude of the following vowel ('BI', for 'Burst Intensity'); (ii) the duration of the burst plus aspiration/friction phase ('VOT'); (iii) the duration of the following vowel ('Dur').

2.4. Processing

A three-way ANOVA procedure was carried out, using the properties of the speakers' responses as the three dependent variables: BI_{resp}, VOT_{resp}, Dur_{resp}. The properties of the stimuli (BI_{stim}: 4 levels and VOT_{stim}: 9 levels) were considered as independent variables, as well as the Speaker variable (fixed factor, 10 levels).

[1] Below we refer to the four series using the RMS ratios resulting from these factors, i.e. BI_{stim} = -3dB, -9dB, -15dB and -21dB.

3 Results

3.1 All speakers

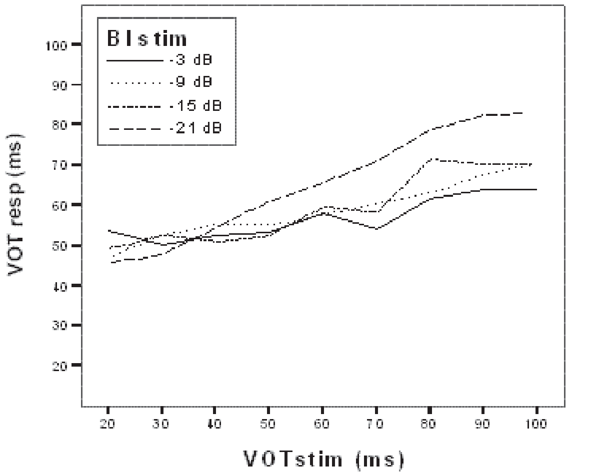

Figure 1 : *Mean VOT_{resp} as a function of VOT_{stim} across BI_{stim} levels (all subjects).*

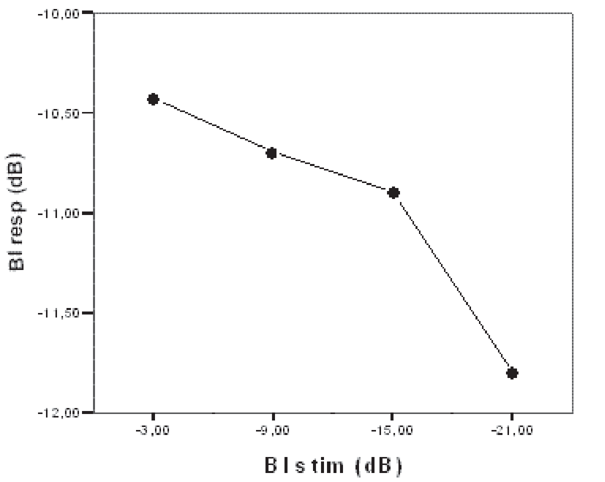

Figure 2 : *Mean BI_{resp} as a function of BI_{stim} (all subjects).*

Figure 1 and Figure 2 respectively illustrate the relation between VOT_{stim} and VOT_{resp}, and between BI_{stim} and BI_{resp}. They indicate that, generally speaking, the speakers were fairly accurate in the imitation task, in the sense that VOT_{resp} increases (from 49 to 72 ms on average) while VOT_{stim} increases (from 20 to 100 ms), and BI_{resp} decreases (from -10 dB to -12 dB on average) as BI_{stim} decreases (from -3 to -21 dB), even if the range of (mean) variation in the responses is lesser than in the stimuli.

The statistical analysis allows us to refine the observed variations. VOT_{stim} does have a significant effect on VOT_{resp} ($F_{(8,2874)}=46.7$; $p<.001$), so do $BI_{stim.}$ ($F_{(3,2874)}=18.6$; $p<.001$) and Speaker ($F_{(9,2874)}=470.7$; $p<.001$). BI_{stim} has a significant effect on BI_{resp}, ($F_{(3,2710)}=4.4$; $p=.005$) so does Speaker ($F_{(9,2710)}=438.1$; $p<.001$), but not $VOT_{stim.}$ ($F_{(8,2710)}=1.2$; $p=.284$).

In other words, when all subjects are considered together, a variation in stimulus VOT induces variations in response VOT that may be mediated by Burst intensity, whereas a variation in stimulus Burst intensity induces variations in response Burst intensity independently of VOT. There is also a significant Speaker effect (and an interaction between the Speaker and other main effects) that will be considered in the next section.

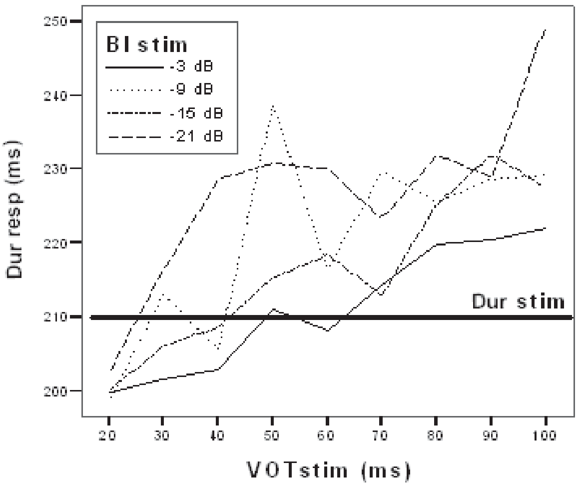

Figure 3 : *Mean Dur_{resp} as a function of VOT_{stim} across BI_{stim} levels (all subjects). The horizontal line corresponds to the vowel duration in the stimuli (210ms).*

Figure 3 shows that Dur_{resp} increases as long as VOT_{stim} increases and BI_{stim} decreases. Indeed, VOT_{stim} does have a significant effect on Dur_{resp} ($F_{(8,2873)}=9.7$; $p<.001$), so do Speaker ($F_{(9,2873)}=173.1$; $p<.001$) and $BI_{stim.}$ ($F_{(3,2873)}=9.6$; $p<.001$). Although vowel duration remained constant across stimuli (210 ms), vowel duration significantly varied across responses (from 200 to 232 ms on average). The variations in Dur_{resp} are related with the variation of stimulus parameters other than vowel duration, i.e. $BI_{stim.}$ and VOT_{stim}.

3.2 Inter-speaker variation

The above three-way ANOVA have revealed that the Speaker variable has a significant effect on all three dependent variables. In order to compare individual behaviours, two-way ANOVA procedures were carried out for each speaker separately. Table 1 presents a summary of the results. It mentions the number of subjects (out of 10) for whom each response parameter varies significantly (and appropriately) as a function of the stimulus parameters.

Table 1. *Number of subjects (out of 10) for whom the response parameters vary significantly depending on the stimulus parameters according to the ANOVA.*

		STIMULUS	
		VOT	BI
RESPONSE	VOT	8*(9)	7*(8)
	BI	1*	5*
	Dur	5*(6)	6*

* Significant at the .05 level
( ) Significant at the .06 level

The results given in Table 1 show that there are more speakers who produce VOT_{resp} variation in response to variation in VOT_{stim} (8/10) than speakers who produce variation in BI_{resp} in response to BI_{stim} (5/10). In this sense, the speakers seem to demonstrate more *fidelity* to VOT variations than to BI variations. Note, however, that 'unprovoked' variations in VOT_{resp} (in response to BI_{stim} variations) are also much more common (7/10) than unprovoked BI_{resp} variations (1/10).

Table 1 also indicates that BI_{stim} variations are more prone to elicit some variation in the response from the speakers, whatever the nature of this variation. In total, BI_{stim} variations induce significant variations in the response in 7+5+6=18 out of 30 cases (vs. 14/30 for VOT_{stim}). In other words, speakers seem to exhibit a higher *sensitivity* to variations in the burst intensity of the stimulus.

Finally, it appears from these results that some dimensions are more 'available' for the speakers to use in their performing of the task. In particular, speakers exhibit more *potentiality* to use VOT variations in their responses (8+7=15 out of 20 cases) than to use BI (1+5=6/20). They also show a good potentiality to handle vowel duration since it is varied in 5+6=11 out of 20 cases even though there were no such variations in the stimuli.

4 Discussion

In this paper, we examined the reproduction of /ta/→[tʰa] stimuli by French natives. Our main goal was to provide a detailed acoustic analysis of the speakers productions, in order to investigate what happens during the process of reorganization of the cognitive and/or motor control regimes that govern the mastering of an initial aspirated stop.

First, results show that there is a fair amount of correspondence between the acoustic properties of the stimuli and the responses, meaning that the speakers were quite successful in the imitation task. The general undershoot in the responses when compared to the stimuli suggests that stimulus enhancement might be a successful training strategy for improving speech production in L2.

Second, the specific responses to the manipulation of different cues in the stimuli suggest that these dimensions of articulatory/acoustic variation may be treated differently by the speakers. Results indicate that the French speakers are more sensitive to BI variation than to VOT variation, but that they achieve a higher degree of fidelity to the input when VOT vs. BI is manipulated. Generally speaking, the dimensions involving timing and/or duration (VOT, vowel duration) are more readily used by the French speakers than Burst intensity. A tentative hypothesis to account for these phenomena is that BI level is more easily detected (particularly the lowest level) but less easily reproduced than time-related VOT. Also recall that the correspondence between VOT_{resp} and VOT_{stim} is better when BI level is at its lowest (Fig.1), which may denote an inhibiting effect of BI on the perception of the duration of the aspiration phase.

Particularly interesting is the speakers' potentiality to respond to either VOT or BI variations in terms of duration of the following vowel. It may be considered as a valid 'replacement strategy' from French speakers in the process of reorganization of their motor control routines. However, results not presented here show that this strategy is poorly rewarded by both French and American listeners, and that it is less exploited by the speakers themselves as time goes by.

Finally, interindividual variation is very high, either in the performances or in the response patterns. The specific patterns of covariation between the different acoustic dimensions for different speakers suggest the existence of distinct individual strategies, which obviously need further research in order to be established.

References

[1] Bradlow, A., Pisoni, D., Akahane-Yamada, R., Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *JASA,* 101, 2299-2310.

[2] Delvaux, V., Piccaluga, M., Huet, K., Harmegnies, B. (2008). Modelage perceptuel du contrôle de la matière phonique en L2 : Variabilité inter-individuelle. *Actes des XXVIIe JEP*, 389-392.

[3] Harmegnies, B., Delvaux, V., Huet, K., Piccaluga, M. (2005). Oralité et cognition : pour une approche raisonnée de la pédagogie du traitement de la matière phonique. *PArole,* 37-39, 277-348.

[4] Hazan, V., Sennema, A., Iba, M., Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants in Japanese learners of English. *Spe. Comm.*, 47, 360-378.

[5] Iverson, P., Hazan, V, Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *JASA*, 118, 3267-3278.

[6] Lambacher, S., Martens, W., Kakeki, K. (2002). The influence of identification training on identification and production of the American English mid and low vowels by native talkers of Japanese. *Proceedings 7th ICSLP Denver*, 245-248.

[7] Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech,* 29, 3-11.

[8] Lively, S., Logan, J., Pisoni, D. (1994). Training Japanese listeners to identify English /r/ and /l/. III: long- term retention of new phonetic categories. *JASA,* 96, 2076-2087.

[9] Wang, Y., Jongman, A., Sereno, J.A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *JASA* 113, 1033-1043.