



HAL
open science

A perceptual correlate of the labial-coronal effect

Marc Sato, Nathalie Vallée, Jean-Luc Schwartz, Isabelle Rousset

► **To cite this version:**

Marc Sato, Nathalie Vallée, Jean-Luc Schwartz, Isabelle Rousset. A perceptual correlate of the labial-coronal effect. *Journal of Speech, Language, and Hearing Research*, 2007, 50 (6), pp.1466-1480. 10.1044/1092-4388(2007/101) . hal-00194046

HAL Id: hal-00194046

<https://hal.science/hal-00194046>

Submitted on 5 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A perceptual correlate of the Labial-Coronal Effect

Marc Sato ^{1,2}, Nathalie Vallée ^{1,*}, Jean-Luc Schwartz ¹, Isabelle Rousset ¹

¹ *Institut de la Communication Parlée, CNRS UMR 5009,*

Institut National Polytechnique de Grenoble, Université Stendhal

² *Departimento di Neuroscienze, Sezione di Fisiologia, Università di Parma*

* Corresponding author.

Institut de la Communication Parlée - UMR CNRS N° 5009

INPG/Université Stendhal

BP25 - 38040 GRENOBLE Cedex 9 FRANCE

Phone: (+33) 476 827 789

Fax: (+33) 476 824 335

E-mail address: nathalie.vallee@icp.inpg.fr



ABSTRACT

Purpose: Statistical studies conducted in various languages on both infants and adults have revealed an intersyllabic preference for initiating words with a labial consonant–vowel–coronal consonant sequence. Speech motor constraints have been proposed to explain this so-called Labial-Coronal effect. This study was designed to test for a possible perceptual correlate of the Labial-Coronal effect in French adults.

Method: We examined the perceptual stabilities of repeatedly presented disyllabic sequences, involving either a labial-vowel-coronal-vowel (LC) or a coronal-vowel-labial-vowel (CL) phonological structure. With this aim, we exploited the Verbal Transformation Effect, which refers to the perceptual changes experienced while listening to a speech form cycled in rapid and continuous repetition. Two experiments were carried out, involving either voiced or unvoiced plosive consonants.

Results: In both experiments, a greater stability and attractiveness was observed for LC stimuli, which suggests that in a (...)LCLC(...) flow the listener could more naturally provide a segmentation into LC chunks.

Conclusion: This study demonstrates that the Labial-Coronal effect also occurs in the course of on-line speech processing. This result is interpreted in relation with theories assuming a link between perception and action in the human speech processing system.

Keywords: Labial-Coronal Effect; Phonological Tendencies; Verbal Transformation Effect; Perceptuo-motor Interactions.

INTRODUCTION

The Labial-Coronal effect in human languages

In spite of their large variability, sound systems of human languages share a number of common aspects, including the most obvious, that is the alternation between consonants and vowels. They also seem to follow a number of general trends, revealed by statistical analyses of phonological inventory databases such as UPSID (UCLA Phonological Segment Inventory Database, Maddieson, 1984). Preferences for certain sounds over others have been described and analysed in many studies (e.g., Crothers, 1978; Maddieson, 1984; Ladefoged and Maddieson, 1990; Vallée, 1994; Schwartz, Boë, Vallée and Abry, 1997, for vowels; Maddieson, 1984; Lindblom and Maddieson, 1988; Ladefoged and Maddieson, 1996; Stefanuto and Vallée, 1999; Vallée, Boë, Schwartz, Badin and Abry, 2002, for consonants).

Sound combinations also share many properties from one language to another. Statistical studies conducted in various languages on both infants and adults have revealed an intersyllabic preference for initiating words with a labial consonant–vowel–coronal consonant sequence (for a review, MacNeilage and Davis, 2000a). Ingram (1974) first reported that during childhood there is a strong preference to produce words beginning with a pre-vocalic consonant articulated with a more anterior place of articulation than the post-vocalic one. This ‘fronting’ tendency was further observed by MacNeilage, Davis, Matyear and Kinney (1999) in nine over ten infants during the

50-word stage (12-18 months) in an English-speaking environment. In their study, the mean ratio of numbers of labial-vowel-coronal (/L_aVC_o/) to coronal-vowel-labial (/C_oVL_a/) sequences was 2.55. This so-called labial-coronal effect (henceforth abbreviated LC effect) is so strong that it sometimes occurs even if the word infants are attempting to produce has the opposite sequence (as in 'pot' for 'top'; Macken, 1978). Interestingly, the LC effect was also found to be present in the inventories of human languages. Statistical analyses carried out on a sample of ten languages revealed a strong preference for /L_aVC_oV/ over /C_oVL_aV/ disyllabic sequences at the beginning of words (/L_aVC_oV/ to /C_oVL_aV/ ratio: 2.23; MacNeilage, Davis, Matyear and Kinney, 1999). Similar results were obtained in our laboratory by analysing a 15-languages syllabified lexicon database (Rousset, 2004), partly available from Maddieson and Precoda's data (1992). The lexicons were representative of the main language families and fairly well distributed geographically. They contained from near 2,000 words (Ngizim) to 12,200 (French). The mean number of lexical items was 5,908. Statistical analyses revealed the presence of the LC effect in 13 languages. The LC effect was observed not only in intersyllabic patterns at word onsets (/L_aVC_oV/ to /C_oVL_aV/ ratio: 2.40) but also all over the words between onsets of two consecutive syllables (ratio: 1.75). It also occurs in intrasyllabic patterns, between onset and coda, without restriction on the position in the words (ratios: 1.73 for /L_aVC_o/ versus /C_oVL_a/ syllables at the beginning of the words and 1.89 at other positions in the word). Although the LC

effect was observed for both CVC and CVCV structures whatever their position within the word, the strongest effect was found for disyllabic words (ratio: 2.77) and, whatever the word length, at the beginning of words between the onsets of two consecutive syllables.

The convergence between data on human languages and developmental observations led MacNeilage and Davis (2000a,b; 2001) to propose a developmental explanation for the LC effect, derived from speech motor constraints. For the authors, the Frame-Content Theory (FCT, MacNeilage, 1998), linking consonant and vowel alternations with the jaw cyclicities involved in mastication and babbling and providing a phylogenetic and ontogenetic framework for the birth of speech, may explain the preferred co-occurrences between consonant and vowel places of articulations in human languages. According to this theory, labial-vowel sequences, called ‘pure frames’, are produced in the course of babbling with no active movement except for jaw cyclicities (e.g., /ba/). In contrast, coronal-vowel sequences, called ‘fronted frames’, involve a fronting of the tongue together with jaw cyclicities, resulting in both a fronted consonant and a fronted vowel (e.g., /de/). During this period, the LC effect would result from a preference for ‘simple first’ utterances. A pure labial frame being supposedly easier to produce than a fronted coronal one, infants would have a ‘tendency to start a word in an easy way and then add a tongue movement’ (MacNeilage and Davis, 2000a, pp. 529).

The Labial-Coronal “chunking hypothesis”

The developmental explanation proposed by MacNeilage and Davis seems however a bit ad-hoc, the more so considering that Vilain and colleagues (1999) showed, in the comparison of various articulatory models, that pure jaw cyclicities could actually produce coronal as well as labial contacts, depending on individual morphology. Therefore, labials appear not so obviously more ‘simple’ to produce than coronals. An alternative hypothesis may rather come from the gestural overlap asymmetry previously observed between LC and CL sequences in consonantal clusters, possibly reflecting a preferential coordinative articulatory mode for LC compared with CL sequences.

In C1C2 sequences where C1 and C2 are two contiguous consonants, there is generally a “gestural overlap”, estimated for example as the delay between C1 and C2 articulatory onsets, divided by the whole duration of the C1 gesture from onset to offset. Gestural overlap occurs if C2 onset precedes C1 offset. Several studies show that this overlap is consistently larger when C1 is anterior to C2 (e.g., in a /L_aC_o/ sequence) than when it is posterior (e.g., in a /C_oL_a/ sequence). This fact has been described for several languages (e.g., in English, Hardcastle and Roach, 1979; Byrd, 1992, 1996; Zsiga, 1994; Surprenant and Goldstein, 1998; or in Georgian, Chitoran, Goldstein and Byrd, 2002) and various manners of articulation (extending the many findings on stops, provided in the previously cited papers, to liquids; Chitoran and Goldstein,

2006). It has sometimes been interpreted as the need to deal with the risk for the first consonant, if it is posterior to the second one, to be acoustically hidden by it. Indeed, in this case, the C1 burst is realised in a vocal tract more or less closed by the C2 closure if the overlap is too large, hence the burst is acoustically much less salient. The consequence of this asymmetry in gestural overlap is that the labial and coronal consonants are more synchronous (because displaying a larger overlap) in /L_aC_o/ sequences than in /C_oL_a/ ones.

In a recent work, Rochet-Capellan and Schwartz (2005a, b) capitalized on this asymmetry to propose a new explanation of the LC effect. Assuming that the gestural overlap asymmetry would extend to disyllabic utterances, their reasoning was that /L_aVC_oV/ sequences would be better ‘in phase’ than /C_oVL_aV/ ones, thanks to the possibility to prepare the coronal gesture in the course of the labial realization, while the inverse is unlikely, because of the gestural overlap asymmetry. This hypothesis has been tested in the framework of the speeding paradigm previously developed by Kelso and colleagues for displaying articulatory synergies (Tuller and Kelso, 1990, 1991). In this framework, the synergy between a given set of articulators according to segmental and suprasegmental characteristics of a task could be described by temporal phasing relationships between the articulators (Kelso, Saltzman and Tuller, 1986), and the accelerated repetition of a given speech sequence might induce a reorganization of the coordination between the effectors towards a preferential coordinative mode. In a series of

acoustic and articulatory measurements on speeded /L_aVC_oV/ and /C_oVL_aV/ sequences involving both plosives and fricatives (e.g., /pata/ and /tapa/, /pasa/ and /sapa/, /fata/ and /tafa/), Rochet-Capellan and Schwartz obtained two main results. First, acoustical measurements (Rochet-Capellan and Schwartz, 2005a) showed a favored trend for the accentuation of the vowel after the coronal consonant together with the reduction of the vowel after the labial consonant until its possible complete fading (e.g. /pata/-/tapa/ → /pa'ta/ → /p'ta/). This confirmed that /L_aVC_oV/ sequences are actually more stable than /C_oVL_aV/ ones, with a progression of both kinds of sequences towards a L_aC_o attractor (e.g., /pa'ta/, /pa'sa/, /fa'ta/). Second, articulatory measurements (Rochet-Capellan and Schwartz, 2005b) showed that the speeding process induces a shift from two jaw cycles per disyllable (one jaw cycle per syllable) to one in a way that tends to limit the jaw energy consumption. This shift, resulting in chunking the two syllables on a single jaw gesture, modifies the coordination between the jaw and the constrictors, that is the lower lip for the labial consonant and the tongue tip for the coronal one. Actually, the labial release occurs during jaw closing phase, while the coronal release occurs around jaw highest position or during the opening phase. Moreover, in agreement with acoustic measurements, there is a trend to anticipate the tongue constriction during the labial one rather than the inverse¹.

A possible perceptual correlate of the Labial-Coronal chunking hypothesis

In summary, the foregoing study indicates that speakers tend to chunk /L_aVC_oV/

sequences on a single jaw gesture, with the labial consonant on the closing phase and the coronal one on the opening phase. This led us wonder whether such a motoric chunk of the two syllables would not have a possible perceptual counterpart. This question logically arises from theories of speech perception invoking a link between perception and action (e.g., in the Motor Theory of Speech Perception, Liberman and Mattingly, 1985; Liberman and Whalen, 2000; in the Direct Realist Theory of Speech Perception, Fowler, 1986; or in the Perception for Action Control Theory, Schwartz, Abry, Boë and Cathiard, 2002; Schwartz, Boë and Abry, 2006). With this aim, we exploited the Verbal Transformation Effect (Warren and Gregory, 1958; Warren, 1961). This multistable perception phenomenon refers to perceptual changes experienced while listening to a speech form cycled in rapid and continuous repetition. Initially, a percept matching the original form is heard, but at some point illusory words or pseudo-words arise, corresponding to a change in the perceived stimulus. This transformation process goes on throughout the repetition procedure, leading to perceptual transitions from one speech form to another (or back to the original form). Previous studies have reported that perceptual changes relative to the auditory input could range from small phonetic deviations to strong semantic distortions. These transformations include substitution of a phoneme by a phonetically close one (Warren, 1961; Warren and Meyers, 1987), auditory streaming/perceptual grouping (the repeated stimulus being separated into different audio streams, giving birth to verbal transformations involving only part

of the presented material; Pitt and Shoaf, 2001, 2002), lexical and semantic transformations (Warren, 1961; Kaminska, Pool and Mayer, 2000; Shoaf and Pitt, 2002). In a recent study, Sato, Schwartz, Abry, Cathiard, and Lœvenbruck (2006) showed that verbal transformations were actually influenced by articulatory synergies, with more in-phase sequences demonstrating more stability and attracting transformations more often than less in-phase ones. This work involved the overt or covert repetition of CCV or CVC sequences with a labial and a coronal consonant (e.g., /ps↔/ versus /s↔p/), and the search for verbal transformations. A preliminary articulatory experiment showed there was a trend to resynchronize the labial and coronal gestures, out of phase in /s↔p/ (with the coronal on the syllable onset and the labial on the coda), towards a better-phased labial-coronal cluster on syllable onset in /ps↔/ (with a tight synchronization of /p/ and /s/ thanks to gestural overlap asymmetry). Verbal transformation data displayed a large preference for /ps↔/ over /s↔p/, with a greater stability for /ps↔/ and with much more transformations from /s↔p/ to /ps↔/ than the other way round. The fact that no psycholinguistic factors can fully account for these results argues in favour of the existence of articulatory control constraints acting on verbal transformations.

The question was therefore precisely the following. In a verbal transformation paradigm, would LC sequences be more stable than their CL counterpart, that is, in a (...)L_aC_oL_aC_o (...)

flow, would the listener more naturally provide a segmentation into L_aC_o chunks than into C_oL_a ones? Displaying such a perceptual correlate of the LC effect would have a number of interesting consequences. First, it would increase the understanding of the LC effect, showing that it does indeed exist not only in infants' first words and in adult languages, but also in the course of on-line speech processing and segmentation, providing LC sequences rather than CL ones. Second, it would provide a new piece of evidence for a link between speech perception and production. The aim of the present study was hence to realize verbal transformation experiments involving $/L_aVC_oV/$ and $/C_oVL_aV/$ sequences cycled in continuous repetitions. Ditzinger, Tuller and Kelso (1997) showed that verbal transformations are generally dominated by two forms incorporating the repeated stimulus. Our assumption was that the main organization of the reported transformations should be of a pairwise coupling between the two possible forms of each 'reversible' repeated stimulus (i.e., from a $/L_aVC_oV/$ sequence to the associated $/C_oVL_aV/$ sequence, and vice-versa), and that LC patterns would be both more stable (i.e., transform less) and more attractive (i.e., attract more transformations) than reverse CL ones in these verbal transformations. Because verbal transformations have been shown to depend on various lexical factors related to the repeating stimulus (MacKay, Wulf, Yin and Abrams, 1993; Shoaf and Pitt, 2002), and considering the general preference for L_aVC_oV sequences in French lexicons according to syllabic inventories (Rousset, 2004), various stimuli were selected in order to

control for possible lexical interferences and to attempt to focus on ‘pure’ sensori-motor mechanisms. As a matter of fact, the manifestations of the LC effect in French are quite diverse. For instance, though /pVtV/ sequences are more frequent than /tVpV/ ones, this is not systematically the case (e.g., /poto/ sequences are *less* frequent than /topo/ ones). Moreover, while the labial /p/ is more frequent than the coronal /t/ at the beginning of a word, the labial /b/ is less frequent than the coronal /d/ and the LC effect happens to be lacking in French with voiced plosives.

Accordingly, two experiments were designed to test the perceptual stability and attractiveness of /L_aVC_oV/ and /C_oVL_aV/ disyllabic sequences, Experiment 1 involving unvoiced plosives with /pVtV/ and /tVpV/ sequences, and Experiment 2, voiced plosives with /bVdV/ and /dVbV/ sequences. Furthermore, three different vowels were used in each experiment, that is /a/, /i/ and /o/. The working assumption was that LC sequences would be both more stable (i.e. displaying less transformations) and more attractive (i.e. capturing more transformations) than CL ones, for both experiments and independently of vocalic context.

METHOD

Participants

Two distinct groups of twenty-four students from Grenoble University participated in the study. In Experiment 1, seventeen males and seven females were tested (mean age: 24 years; range: 18-48). In Experiment 2, eleven females and thirteen males were tested (mean age: 28 years; range: 19-52). The number of subjects was chosen on the basis of previous studies, testing a similar range of participants (e.g., Shoaf and Pitt, 2002, Pitt and Shoaf, 2002). All subjects were native French speakers and reported no history of hearing or speaking disorders. None of them was aware of the purpose of the experiment.

Material

The stimuli consisted of a combination of the bilabial /p/ and coronal /t/ unvoiced consonants in Experiment 1, and of the bilabial /b/ and coronal /d/ voiced consonants in Experiment 2. For all sequences, the vowel was fixed, selected in the {/i/, /a/, /o/} set. For each experiment, three pairs of disyllabic reversible CVCV speech sequences were thus contrasted: i.e., /pata/, /tapa/, /piti/, /tipi/, /poto/, /topo/ in Experiment 1, and /bada/, /daba/, /bidi/, /dibi/, /bodo/, /dobo/ in Experiment 2.

Stimulus recordings

Multiple utterances of CV sequences - /pa/ and /ta/, /pi/ and /ti/, /po/ and /to/, /ba/ and

/da/, /bi/ and /di/, /bo/ and /do/ - were recorded in a soundproof room by a trained phonetician, native French speaker (J.-L.S.). The speaker was told to pronounce syllables naturally at a conversational rate, maintaining an even intonation and vocal intensity while producing the sequences. The items were 16-bits digitised on the hard disk of a PC computer at a 44.1-kHz sampling rate. One clearly articulated token was selected for each syllable, the syllables being matched as closely as possible for acoustic similarities (according to the duration, intensity, formant and pitch values) within each group of stimuli (as checked by a spectrogram analysis using the Praat software, Institute of Phonetic Sciences, University of Amsterdam, the Netherlands - see Table 1).

Insert Table 1 about here

The six pairs of disyllabic CVCV speech forms were then created by inserting each of the two appropriate CV sequences in an individual sound file. For unvoiced sequences in Experiment 1, a 100 ms silent period was added for each syllable between the offset of the vowel and the onset of the following consonantal burst. With this procedure, the mean duration of CVCV sequences (corresponding to the inter-stimulus interval) was 520 ms (range: 503-531 ms). The voiced stimuli in Experiment 2 were prepared by simply concatenating voiced syllables with no added portion of silence between the two syllables, to preserve the permanent voicing

characteristic of the stimuli. The resulting CVCV sequences happened to be slightly shorter, with a mean duration of 455 ms (range: 452- 457 ms).

Stimuli intelligibility

A control experiment was run in order to assess the intelligibility of the stimuli. Fifteen participants, who did not participate to the verbal transformation experiment, were individually tested in a quiet room. All were native French speakers and reported no history of hearing or speaking disorders (thirteen females, mean age: 26 years; range: 19-42). They were instructed to carefully listen to speech stimuli and to report what they heard. The six pairs of CVCV disyllabic stimuli were presented binaurally over headphones at a comfortable sound level. During the experiment, each stimulus was presented ten times in a randomized sequence for a total of 120 trials. Each trial started with a fixation cue presented during 500 ms, immediately followed by the audio stimulus. Once written the response, participants had to press a computer key placed in front of them to begin a new trial. The stimuli were correctly spelled 99% of the time, which validates the reliability of the recording procedure.

Lexical analyses

Lexical analyses were extracted from VoCoLex, a lexical database for the French language (~105,000 words; Dufour, Peerman, Pallier and Radeau, 2002). All the stimuli but one (/pata/) used in Experiment 1 are present in the French Lexicon (with a similar low word

frequency - mean frequency of 2.4 per million; range: 0-9), while all the stimuli used in Experiment 2 are nonwords. Because verbal transformations have been shown to vary as a function of distinct lexical factors related to the repeating stimulus, different lexical factors were calculated for each CVCV stimulus (computed for both token and type frequency - see Table 2): the neighborhood density (defined as the number of phonologically similar words that differed from the stimulus only by a single substitution, insertion or deletion at any position in the target stimulus; Luce, Pisoni and Goldinger, 1990) and the disyllabic frequencies (defined as the number of lexical entries incorporating a CVCV structure identical to that of the stimulus either at the word onset or at any positions in the word). Finally, the bigram frequencies (defined as the number of lexical entries incorporating a CV structure identical to that of the stimulus either at the word onset or at any position in the word) were also computed for each CV sequence.

Similar analyses were also performed on /pVtV/, /tVpV/, /bVdV/ or /dVbV/ stimuli (V being any French vowel). In addition, when considering the onset of the words, additional analyses showed that the number of lexical entries beginning with /p/, /t/, /b/ or /d/ was 8486, 4998, 5568 and 8249, respectively.

Finally, global analyses were conducted on /L_aVC_oV/ and /C_oVL_aV/ syllabic structures (L_a being any labial consonant and C_o being any coronal consonant in French). The numbers of lexical entries incorporating /L_aVC_oV/ syllabic structures, either at the word onset or at other

positions in the word, was respectively 2570 and 3238. The numbers of lexical entries incorporating /C_oVL_aV/ syllabic structures was respectively 1751 and 1162. The /L_aVC_oV/ to /C_oVL_aV/ ratios were 1.47 at word onset and 2.79 at other positions in the word.

These results thus confirm the LC effect in French (in spite of certain local discrepancies such as /bVdV/ versus /dVbV/ - see Table 2) and appear quite similar to those observed by Rousset (2004) using a 15-languages syllabified lexicon database (including French).

Insert Table 2 about here

Apparatus

The stimuli were presented binaurally over headphones at a comfortable sound level. Transformations were collected via a microphone and directly recorded as individual sound files onto the hard disk of the computer.

Procedure

The participants were individually tested in a quiet room. The experiment began with a lengthy briefing during which they were introduced with the verbal transformation task. The participants listened to the experimenter repeating the word "life" at a rate of two repetitions per second and were asked to listen carefully for any changes in the repeating utterance. The experimenter then asked the participants if they had perceived another sequence and, if not,

explained the possibility of hearing the word "fly". This “bootstrap” example, presented in English rather than French, aimed at displaying the Verbal Transformation Effect on a material that all subjects understood, while letting them experience the phenomenon later in their own language. Then, they were told that they would hear an utterance being played repeatedly and were asked to listen carefully for any changes in the repeated utterance. If the stimulus changed into another form, even if it changed into one they had heard previously, they were asked to immediately report the transformation by repeating it as it was heard. It was indicated that these changes could be subtle or very noticeable and could correspond to a word as well as a pseudo-word. Finally, the participants were assured that there were no correct or incorrect responses and that if they did not hear any transformation, they were to say nothing. For both experiments, the six stimuli were randomly presented for 300 repetitions each. The total duration of each experiment was around thirty minutes.

Data analysis

For each participant and each stimulus, transformations were phonetically labelled by a trained French phonetician (I.R in Experiment 1 and N.V. in Experiment 2) according to both the response sound file and the corresponding spectrogram (using the Praat software, Institute of Phonetic Sciences, University of Amsterdam, the Netherlands). This procedure ensured a careful and detailed phonetic transcription of the participant responses and allowed to determine the

precise timing at which these transformations were reported. Furthermore, although no interjudge transcription reliability was carried out, it is worthwhile to note that analyses were conducted relatively to each presented sequence and its reversible form, sequences for which possible errors in transcription are quite unlikely.

The perceptual stability duration of each phonetic form was computed by summing the time spent perceiving the given form before switching to the next one. For both experiments, a repeated measures analysis of variance (ANOVA) was performed in order to study the perceptual stability of the presented stimuli. The considered within-subject variables were related to the vowel type (i.e., /i/, /a/, /o/) and to the consonantal pattern (i.e., /p/-/t/ and /t/-/p/ in Experiment 1, /b/-/d/ and /d/-/b/ in Experiment 2). The dependent variable was the mean relative stability duration for the reported transformations strictly corresponding to the repeated sequence (i.e., the sum of the time spent perceiving the repeated sequence divided by the task duration). In addition, a similar analysis was conducted in Experiment 2 with the dependent variable related to the mean relative stability duration for the reported transformations strictly corresponding to the repeated sequence or differing from it by the absence of consonantal voicing (e.g., /bada/, /pata/, /bata/, /pada/ for /bada/). This analysis was not performed for Experiment 1 because of the very low number of transformations incorporating voiced instead of unvoiced consonants.

The difference between the mean relative stability durations observed for the presented

sequence and the associated one (e.g., /pata/ versus /tapa/ in Experiment 1, and /bada/ versus /daba/ or {/bada/, /bata/, /pada/, /pata/} versus {/daba/, /taba/, /dapa/, /tapa/} in Experiment 2) were calculated for each participant (*delta-score*). This allows to further explore the nature of the perceptual grouping across the two possible forms of the disyllabic reversible CVCV speech sequences. ANOVAs were performed on these data with the same factors as used in previous analyses. Finally, a comparison of both stability and delta scores was performed between Experiments 1 and 2 with the experiment as between-subject variable.

Because delta-scores are not independent of stability durations, a Bonferroni correction was applied on each pair of statistical analysis. Besides the ANOVAs, pairwise comparisons were conducted with Bonferroni method whenever appropriate. The significance level was always set at $p < .05$. Prior to perform these statistical analyses, Mauchly's tests showed that the sphericity assumption was not violated ($p > .05$).

RESULTS

Experiment 1

The mean relative stability durations (MSD) of reported transformations observed in Experiment 1 for the three pairs of disyllabic reversible CVCV sequences are displayed in Figure 1. For each stimulus, the main organization of the transformations was that of a pairwise coupling between the two possible forms of the reversible CVCV structure (e.g., /pata/ and /tapa/ for /pata/; MSD \pm SD: 69% \pm 3). The reported transformations differing from the repeated sequence by the presence of a consonantal voicing were marginal (e.g., /bata/ for /pata/; MSD \pm SD: 2% \pm 2). The other transformations included auditory streaming, that is discarding part of the stimulus embedded in another auditory flow (e.g., /ata/ for /pata/, /p/ being perceived in a separate stream; MSD \pm SD: 3% \pm 3), phonetic deviations (e.g. /potõ/ for /poto/; MSD \pm SD: 4% \pm 5), lexical transformations (e.g., ‘pitié’/’pity’ for /piti/; MSD \pm SD: 11% \pm 4), and more complex sequences (e.g. /pipit/ or /pitipiti/ for /tipi/; MSD \pm SD: 11% \pm 6).

Insert Figure 1 about here

Analysis of stability scores showed a significant effect of the consonantal pattern, with a greater perceptual stability of LC as compared with CL stimuli (on average 40% versus 30%, respectively; $F(1,23) = 12.43, p < .01$). There was no reliable effect of the vowel ($F(2,46) = 0.79$)

nor interaction between the two factors ($F(2,46) = 0.91$). Similarly, analysis of delta scores revealed a significant effect of the consonantal pattern. The delta score was greater for LC stimuli (on average 10%) than for CL ones (-07%; $F(1,23) = 10.53$, $p < .01$). There was no reliable effect of the vowel ($F(2,46) = 0.45$) nor interaction between the two factors ($F(2,46) = 1.81$).

Experiment 2

The mean relative stability durations of reported transformations observed in Experiment 2 for the three pairs of disyllabic reversible CVCV sequences are displayed in Figure 2. While the perceptual stability of the two possible forms of the reversible CVCV structure was lower than in Experiment 1 (MSD \pm SD: 34% \pm 16), the reported transformations differing from the repeated sequence by the absence of consonantal voicing were very frequent (MSD \pm SD: 37% \pm 31). Furthermore, the perceptual stability of these transformations strongly differed according to the vocalic context (representing on average 77% of the total duration for /bada/ and /daba/, 16% for /bidi/ and /dibi/, and 17% for /bodo/ and /dobo/ stimuli). Hence, contrary to Experiment 1, no pairwise coupling between the two possible forms of the reversible CVCV stimuli was observed. However, it is worthwhile to note that, once these ‘voicing’ transformations taken into account, the mean perceptual stability was still dominated by two forms, always incorporating the repeated stimulus (MSD \pm SD: 52% \pm 4)

The other transformations included auditory streaming (e.g., /odo/ for /bodo/, /b/ being perceived in a separate stream; MPS \pm SD: 3% \pm 5), phonetic deviations (e.g. /bedi/ for /bidi/; MPS \pm SD: 1% \pm 2), lexical transformations, although with a lower perceptual stability than in Experiment 1 (e.g., ‘débit’/’flow’ for /dibi/; MPS \pm SD: 2% \pm 1), and more complex sequences (e.g. /bRodo/ for /bodo/; MPS \pm SD: 22% \pm 15).

Insert Figure 2 about here

Analysis of stability scores indicated a significant effect of the consonantal pattern with a greater perceptual stability of LC as compared with CL stimuli (on average 21% versus 15%, respectively; $F(1,23) = 7.77, p < .05$). There was also a significant effect of the vowel ($F(2,46) = 9.04, p < .002$). On average, the perceptual stabilities for the stimuli related to /a/, /i/ and /o/ vowels were 9%, 22% and 23%, respectively. Post-hoc analyses revealed a significantly lower perceptual stability for speech stimuli related to /a/ as compared with /i/ and /o/ vowels. The interaction between the two factors was not significant ($F(2,46) = 0.79$). Analysis of delta scores revealed a significant effect of the consonantal pattern, with a greater delta score for LC than for CL stimuli (on average 15% versus -04%; $F(1,23) = 8.86, p < .05$). There was no reliable effect of the vowel ($F(2,46) = 1.16$) nor interaction between the two factors ($F(2,46) = 0.92$).

Once the transformations differing from the repeated sequence by the lack of consonantal

voicing taken into account, analysis of stability scores indicated a significant effect of the consonantal pattern, with a greater perceptual stability for LC than for CL stimuli (on average 41% versus 30%, respectively; $F(1,23) = 13.64, p < .005$). There was also a significant effect of the vowel ($F(2,46) = 15.27, p < .005$). On average, the perceptual stabilities for the stimuli related to /a/, /i/ and /o/ vowels were 47%, 29% and 32%, respectively. Post-hoc analyses revealed a significantly higher perceptual stability for speech stimuli related to /a/ as compared with /i/ and /o/ vowels. The interaction between the two factors was not significant ($F(2,46) = 1.51$). Analysis of delta scores revealed a significant effect of the consonantal pattern, with a greater delta score for LC than for CL stimuli (on average 12% versus -11%; $F(1,23) = 24.27, p < .005$). There was no reliable effect of the vowel ($F(2,46) = 0.33$) nor interaction between the two factors ($F(2,46) = 0.26$).

Experiment 1 versus Experiment 2

Analysis of stability scores indicated a significant effect of the experiment, with a greater perceptual stability in Experiment 1 than in Experiment 2 (on average 35% versus 18%, respectively; $F(1,46) = 33.94, p < .005$). This effect is related to the low stability of all stimuli in Experiment 2, mainly due to the loss of the voicing feature. There was a significant effect of the consonantal pattern, coherent with what occurred in both experiments, with a greater perceptual stability of LC stimuli as compared with CL stimuli (on average 31% versus 23%, respectively;

$F(1,46) = 20.46, p < .001$). The effect of the vowel was significant ($F(2,92) = 5.64, p < .01$). On average, the perceptual stabilities for the stimuli related to /a/, /i/ and /o/ vowels were 22%, 28% and 30%, respectively. Post-hoc analyses revealed a significantly lower perceptual stability for speech stimuli related to /a/ as compared with /i/ and /o/ vowels. This effect was largely due to Experiment 2. There was also a significant interaction between the experiment and the vowel ($F(2,92) = 5.55, p < .005$), coherent with the low stability in the /a/ context in Experiment 2, compared with the lack of differences between vowel contexts in Experiment 1. Analysis of delta scores revealed a significant effect of the consonantal pattern, with a greater delta score for LC stimuli (on average 09% versus -06%; $F(1,46) = 19.5, p < .005$). There was no reliable effect of the experiment ($F(2,46) = 0.15$), nor of the vowel ($F(2,92) = 0.45$), nor two-way or three-way interactions between the factors.

Once the transformations differing from the repeated sequence by the presence (Experiment 1) or absence (Experiment 2) of consonantal voicing taken into account, analysis of stability scores indicated a significant effect of the consonantal pattern, coherent with what occurred in both experiments, with a greater perceptual stability for LC than for CL stimuli (on average 41% versus 30%, respectively; $F(1,46) = 27.71, p < .005$). The effect of the vowel was significant ($F(2,92) = 11.02, p < .005$). On average, the perceptual stabilities for the stimuli related to /a/, /i/ and /o/ vowels were 47%, 29% and 32%, respectively. Post-hoc analyses

revealed a significantly greater stability for speech stimuli related to /a/ as compared with /i/ and /o/ vowels. There was also a significant interaction between the experiment and the vowel ($F(2,92) = 7.20, p < .005$), due to the larger stability in the /a/ context, and the smaller stability in the /i/ and /o/ contexts, in Experiment 2 compared with Experiment 1. Analysis of delta scores revealed a significant effect of the consonantal pattern with greater delta scores observed for LC than for CL stimuli (on average 11% versus -10%; $F(1,46) = 34.43, p < .005$). The only other significant effect was a three-way interaction between the factors ($F(2,92) = 4.27, p < .05$). This slightly significant interaction was due to the differences observed for the stimuli related to /i/ between the two experiments, with delta scores close to zero whatever the consonantal pattern in Experiment 1.

DISCUSSION

The major result of these verbal transformation experiments is the greater stability of LC sequences observed in both experiments and their larger attractiveness. This result, which was the focus of the present study, could be interpreted as an online perceptual correlate of the LC effect in adult French speakers. Two other results emerged from Experiment 2: the importance of devoicing and the dependence of stability on vocalic context. In the following, we shall first discuss possible psycholinguistic interpretations of the LC asymmetries found in both experiments. Then, we shall explore some possible explanations of the two supplementary results observed in Experiment 2. Finally, we shall comment on the consequences of the observed asymmetries for the “LC chunking hypothesis”, and more generally, we shall discuss its implications in relation with perceptuo-motor interactions in speech communication.

Psycholinguistic interpretations of the observed LC asymmetries

The two experiments provide a clear perceptual correlate of the LC effect, with greater stability and attractiveness observed for /pVtV/ and /bVdV/ stimuli over /tVpV/ and /dVbV/ ones. Given that one important source of influence in the verbal transformation paradigm comes from a set of lexical and sublexical factors, it is therefore important to check for possible psycholinguistic interpretations of the observed perceptual asymmetries between LC and CL reported sequences.

First, lexical status and word frequency of the presented stimulus may act on the reported transformations. It has been argued that low frequency words should suffer a competitive disadvantage as compared with higher frequency words (Yin and MacKay, 1992; MacKay, Wulf, Yin, and Abrams, 1993). As a consequence of the amount of lexical priming, high frequency words should be more stable than low frequency words. Similarly, less stability has been observed for pseudowords than words, which have strong and well-formed representations, and for phonologically illegal pseudowords than for permissible pseudowords (Natsoulas, 1965; Shoaf and Pitt, 2002). In addition, it has been shown that the proportion of nonword transformations also varied with the lexical status of the repeated stimulus, with words producing the fewest transformations (Shoaf and Pitt, 2002). These data appear in line with the contrastive results observed between the two experiments, with a lower perceptual stability of both the repeated stimuli and lexical transformations observed in Experiment 2. However, possible interpretations based on the lexical status and/or the frequency of the repeated stimuli cannot satisfactorily explain the perceptual preference observed for LC sequences within each experiment. Indeed, all the stimuli but one (/pata/) used in Experiment 1 are present in the French lexicon and have similar (low) word frequency, while all the stimuli used in Experiment 2 are phonologically acceptable nonwords.

Second, it has been argued that a greater number of lexical neighbors (i.e., the number of

lexical entries that are phonologically similar to the repeated stimulus) should increase the number of possible primed word candidates, resulting in a greater number and wider range of transformations and a lower stability of the repeated stimulus (Yin and MacKay, 1992; MacKay, Wulf, Yin and Abrams, 1993). A similar competitive effect between lexical candidates was also observed during auditory word recognition tasks (e.g., Vitevitch and Luce, 1998, 1999), with words with few competitors processed more quickly and accurately than words with many competitors, and is introduced in several models of speech perception that emphasize lexical competition in word recognition (McClelland and Elman, 1986; Luce and Pisoni, 1998; Norris, McQueen and Cutler, 2000). On the other hand, probabilistic phonotactics, which refers to the frequency that a particular sequential arrangement of phonetics segments or sequences of segments will occur in a given position in words, might also influence verbal transformations. Previous studies on auditory word recognition have shown that high phonotactic probability nonwords are processed more quickly and accurately than low probability nonwords (Vitevitch, Luce, Charles-Luce and Kemmerer, 1997). By contrasting high phonotactic probability/high neighborhood density stimuli and low phonotactic probability/low neighborhood density stimuli in a speeded auditory shadowing task, Vitevitch and Luce (1998) further showed that high probability/density nonwords were responded more quickly than low probability/density nonwords, whereas the reverse effect was observed for words (see also Vitevitch and Luce,

1999). From these results, it has been suggested that two distinct levels of representation (i.e., lexical and sublexical) are responsible for differential effects of phonotactics and neighborhoods. More specifically, facilitatory effects of probabilistic phonotactics are assumed to reflect differences among activation of sublexical units, whereas effects of similarity neighborhood arise from competitive interactions among multiple word forms.

Considering the greater number of lexical competitors for /poto/ than for /topo/ and for /tipi/ as compared to /piti/, the neighborhood density effect is not likely to provide an explanation for the greater stability observed for LC as compared to CL word stimuli in Experiment 1. In Experiment 2, the greater number of words incorporating a disyllabic sequence identical to that of the LC nonword stimuli, as compared to the CL stimuli, might indeed reflect a facilitatory effect of probabilistic phonotactics. However, given the rather small numbers of related words (on average 2 and 0 for LC and CL stimuli, respectively), such an interpretation does not, in our opinion, make much sense. Furthermore, the higher disyllabic frequency for /dVbV/ stimuli than for /bVdV/ stimuli, as well as the higher bigram frequency for /dV/ than for /bV/ does not fit well with an interpretation based on probabilistic phonotactics.

In sum, although a greater range of stimuli would have increase the reliability of these lexical analyses, none of the psycholinguistic interpretations based on the lexical status, the lexical frequency, the neighborhood density or the probabilistic phonotactics of the repeated

stimuli appears to satisfactorily explain the present results. On the other hand, marginal lexical effects might also have contributed to some local perceptual discrepancies. This might be the case for the lack of perceptual stability difference observed between /piti/ and /tipi/. Actually, though these stimuli involve two possible French words, the word corresponding to /piti/ (i.e., “Pythie”, the Greek oracle) is probably much less salient than the Amerindian word “tippee” corresponding to /tipi/, which could raise the stability and attractiveness of /tipi/ and decrease the difference with /piti/ associated to the LC effect.

The potential role of voicing and vocalic context

There appeared a strong asymmetry between Experiments 1 and 2 in terms of stability. This was due to the very frequent loss of the consonant voicing feature in transformations in Experiment 2 (representing on average 37% of the total duration), while acquisition of voicing in the unvoiced sequences of Experiment 1 was very rare (representing on average 1% of the total duration). This kind of asymmetry has already been displayed in perceptual experiments (e.g., van Alphen and Smits, 2004). However, in these experiments the perceptual loss of voicing was related to a production rather than to a perception problem, namely a lack of sufficient prevoicing in the produced voiced stimuli. Indeed, “most mismatches appeared when the plosive was intended to be voiced but was produced without prevoicing (...). Almost all tokens produced with prevoicing were identified as voiced” (van Alphen and Smits, 2004, pp. 485).

In the present case, the voicing contrast for French is indeed mainly carried by the presence versus absence of prevoicing (e.g. Serniclaes and Bejster, 1979). In the stimuli used in Experiment 2, prevoicing is strong and stable all along the closure period, for all voiced plosives and all stimuli, and of course it is stable during the repetition procedure. Therefore, it seems that the explanation is not simply contained in the stimuli nor in the pattern of confusions between voiced and unvoiced stimuli in classical perceptual experiments. The explanation rather seems to rely in the transformation phenomenon itself. Actually, a number of transformations reported in the literature involve voicing (e.g. Warren, 1961; Pitt and Shoaf, 2002). Furthermore, exactly the same asymmetry in voicing transformations was observed by Warren and Meyers (1987) with stimuli quite similar to ours. In their study dealing with perstimulatory and poststimulatory effects of listening to repeated syllables in English, the authors examined the verbal transformations heard by 30 subjects while listening to repetitions of /tu/ and /du/ ('to' and 'do') presented in regular alternation. It appeared that over the 30 subjects, 19 reported the 'to to' form and only 6 reported 'do do'. Warren and Meyers did not provide an explanation of this asymmetry, apart from a general 'disruption caused by bombardment of the listener's phonological organization with continuing repetitions of the same clear verbal signal' with 'a fragmentation into short-live components' (Warren and Meyers, 1987, pp. 179).

Pitt and Shoaf (2001, 2002) later developed this idea with the introduction of 'streaming'

processes. This process, as previously mentioned, consists in the fact that the repeated audio content can be perceptually separated into different audio streams, giving birth to verbal transformations involving only part of the presented material. We suggest that the voicing loss could be due to a 'streaming' of the consonant voice bar into a low-frequency flow, hence parsed apart from the presented material. Streaming the voicing bar apart would exactly play the role of removing prevoicing, hence turning the voiced plosives into perceived voiceless ones. This is coherent with the fact that the voicing loss was more frequent with /bada/ and /daba/ stimuli than with other vocalic contexts. Actually, the voicing bar is spectrally much more separated from the vowel components in /a/ than in /i/ and /o/, because F1 for /a/ is much higher. This would produce the lower stability in the /a/ context in Experiment 2. In contrast, once voiced and unvoiced transformations are grouped, /a/ sequences happen to be more stable. This is likely due to the fact that the listeners are then focussed on changes between various forms of the LC-CL sequences. Actually, a majority of the subjects in Experiment 2 switch between two dominant patterns, one of the LC type and one of the CL type (involving a loss of voicing in a majority of cases). However, in a number of cases the subjects also switch from one LC or CL pattern to another, only varying the value of the voicing feature. Globally, this results in increasing the percentage of the time where the subject hears an LC or CL sequence, hence increasing the sequence stability if voiced and unvoiced transformations are grouped. On the contrary, in the /i/

and /o/ contexts, devoicing occurs much less. In these contexts, there is a very large trend to incorporate liquids, fricatives or semi-vowels between the stop and the vowel, both for /b/ (typically /bo/ becoming /bro/ with the introduction of an uvular liquid/fricative) and /d/ (typically /di/ becoming /dzi/ with the introduction of an alveolar or post-alveolar fricative). This is in good agreement with a classical type of transformations of speech occurrences both in speech perception and in language shifts (e.g. Chang, Plauché and Ohala, 2001).

Perceptuo-motor interactions and the LC effect

Having better understood the stability differences associated with vowel context and consonant voicing in Experiments 1 and 2, we are left with a basic effect: presented with a sequence of CV syllables alternating labial and coronal stops, French listeners display a clear trend to segment the sequence into disyllables beginning by a labial onset.

This trend cannot be explained simply by regularities in the French lexicon, as shown previously. However, it could be argued that this perceptual behaviour is due to a phonological “Labial-Coronal principle”, which would be implicit in all languages, and would simply be translated into a perceptual preference in the course of the verbal transformation paradigm. This LC principle could for example be due to the developmental mechanism proposed by MacNeilage and Davis (2000a,b; 2001), and then captured in the phonology by a generalisation process. The fact that there can be exceptions to this process (as in the case of voiced /C_oVL_aV/

sequences, more frequent than /L_aVC_oV/ ones in French) would not prevent the listeners to use this LC principle in the verbal transformation paradigm and to show a systematic preference for LC sequences. There are however several problems with this assumption. Apart from the weaknesses of the developmental scenario, already mentioned in Introduction, this explanation cannot adequately explain the gestural overlap asymmetry, nor its extension into the LC chunking phenomenon in speeded production displayed by Rochet-Capellan and Schwartz (2005a, b). Moreover, it is unlikely that a pure psycholinguistic explanation linked with phonological rules would not lead to strong differences between voiced and unvoiced stimuli, and between vowel contexts, considering the large differences in lexical distributions from one case to the other. However, in spite of various differences in stability already discussed, there is almost no difference in delta scores in the whole set of results of Experiments 1 and 2: the preference for LC transformations is clear and stable in the whole study.

Rather, the present data are therefore much more likely to support the LC chunking hypothesis presented in Introduction. According to this hypothesis, the speaker tends to chunk LC sequences on a single jaw gesture, with a neighbouring of the labial and the coronal consonants. This chunking would be due to articulatory economy (minimizing jaw displacements) together with possible perceptual reasons (letting the labial consonant audible in spite of the closing gesture posterior in the vocal tract for preparing the coronal). It is possible

that other biomechanical reasons could also be at work for explaining the jaw-lips-tongue coordination (Chitoran and Goldstein, 2006). This articulatory chunking would result in a perceptual chunking in verbal transformations. It is noticeable that no obvious auditory process can explain the LC preference in verbal transformations, which gives support to the proposition of an articulatory basis to this perceptual process. Such a perceptuo-motor link is compatible with both motor (Liberman and Mattingly, 1985; Fowler, 1986; Liberman and Whalen, 2000) and perceptuo-motor (Schwartz, Abry, Boë and Cathiard, 2002; Schwartz, Boë and Abry, 2006) theories of speech perception. In this framework, it is interesting to note that in their fMRI study of the neuroanatomy of verbal transformations, Sato and colleagues (2004) showed that the verbal transformation task mainly activated a left-lateralized network of cerebral regions, including the left inferior frontal gyrus (Broca's area), the left supramarginal gyrus, the left superior temporal gyrus, the anterior part of the right cingulate cortex, and the cerebellar cortex, bilaterally. This suggests that the verbal transformation effect, which requires percept analysis, form interpretation, and attentional maintenance of verbal material, relies on a circuit sharing common components of speech perception and speech production systems. In this circuit, Sato and colleagues (2006) propose that Broca's area, classically involved in speech production tasks, would be in charge of the on-line analysis of articulatory speech forms that support communicative or interpretative speech, and particularly of syllable parsing in the verbal

transformation task. This neuro-anatomical portrait gives support to the present perceptuo-motor scenario.

It remains that a number of further experiments should be done to enhance the understanding of this scenario. First of all, it will be important to test the extension of the present data to other languages, particularly English, mostly studied in the verbal transformation paradigm. Extension to other consonantal modes of articulation would also be important, in order to confirm that there is really a coherent LC portrait, as it has been displayed in both the gestural overlap asymmetry (Chitoran and Goldstein, 2006) and the speeding production paradigm (Rochet-Capellan and Schwartz, 2005a, 2005b). Variations in the stress pattern of disyllabic sequences would be another dimension to be tested. A preliminary set of experiments involving stress, together with multisensory interactions (associating the sound and the image of the speaker's LC or CL gestures) indicate that the LC effect seems to resist to both stress variations and audiovisual interactions (Basirat, 2005). However, further experiments are required in order to test whether variations in the stress pattern of disyllabic sequences may influence the LC/CL perceptual asymmetry.

CONCLUSION

The present set of experiments provides a clear and robust online perceptual correlate of the LC effect in adult French speakers. The set of perceptual asymmetries observed during a verbal transformation paradigm mirror and complement the set of articulatory-acoustic asymmetries observed in parallel by Rochet-Capellan and Schwartz (2005a, 2005b). Altogether, these data enable us to propose a coherent explanation of the LC effect observed in both adult languages and infant first words, discovered by MacNeilage and colleagues (1999). This explanation based on the LC chunking hypothesis, assumes that speakers tend to chunk /LaVCoV/ sequences on a single jaw gesture, both consonants being tightened up rather closely inside the sequence. Conversely, listeners would rely on this articulatory chunking to perceptually parse (...)LCLC(...)utterances into LC units. This scenario fits well with theories assuming a link between perception and action in the human speech processing system (Lieberman and Mattingly, 1985; Liberman and Whalen, 2000; Schwartz, Abry, Boë and Cathiard, 2002; Schwartz, Boë and Abry, 2006).

ACKNOWLEDGMENTS

We thank Christian Abry, Barbara Davis, Peter MacNeilage, Judith Gierut and three anonymous reviewers for helpful discussions on this study. We also thank Alain Arnal, Anahita Bassirat and Christophe Savariaux for their help in preparing stimuli and scoring data. This work was supported by CNRS (Centre National de la Recherche Scientifique) and MIUR (Ministero Italiano dell'Istruzione, dell'Università e della Ricerca), and by the French project 'Patipapa: Modélisation de l'émergence d'un langage articulé dans une société d'agents sensori-moteurs en interaction' (ACI Systèmes Complexes en SHS).

REFERENCES

- Basirat, A. (2005). *Le traitement de percepts phonétiques multistables: Apport de la lecture labiale et étude neurophysiologique préliminaire*. Master 2A, Sciences Cognitives, INPG Grenoble.
- Byrd, D. (1992). Perception of assimilation in consonant clusters: A gestural model. *Phonetica*, 49: 1–24.
- Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24: 209-244.
- Chang, S.S., Plauché, M.C., & Ohala, J.J. (2001). Markedness and consonant confusion asymmetries. In: E. Hume & K. Johnson (eds.), *The role of speech perception in phonology*, pp. 80-101. New-York: Academic Press.
- Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In: C. Gussenhoven, T. Rietveld & N. Warner (eds.), *Laboratory Phonology VII*, pp. 419–447. Berlin/New York: Mouton de Gruyter.
- Chitoran, I., & Goldstein, L. (2006). Testing the phonological status of perceptual recoverability: Articulatory evidence form Georgian. *Proc. Labphon 10*, Paris.
- Crothers, J. (1978). Typology and universal of vowel systems in phonology. In: Greenberg, H.J (Editor), *Universal of Human Language (2)*, Standford University Press, Standford, pp.

95-152.

Ditzinger, T., Tuller, B. & Kelso, J.A.S. (1997). Temporal patterning in an auditory illusion: The verbal transformation effect. *Biological Cybernetics*, 77: 23-30.

Dufour, S., Peereman, R., Pallier, C. & Radeau, M. (2002). VoCoLex: A lexical database on phonological similarity between French words. *L'Année Psychologique*, 102: 725-746.

Fowler, C.A. (1986). An event approach to the study of speech perception from a direct-realistic perspective. *Journal of Phonetics*, 14: 3-28.

Hardcastle, W., & Roach, P. (1979). An instrumental investigation of coarticulation in stop consonant sequences. In: Harry Hollien and Patricia Ann Hollien (eds.), *Current Issues in the Phonetic Sciences*, 531–540. Amsterdam: John Benjamins.

Ingram, D. (1974). Fronting in child phonology. *Journal of Child Language*, 1: 49-64.

Kaminska, Z., Pool, M. & Mayer, P. (2000). Verbal transformation: Habituation or spreading activation? *Brain and Language*, 71: 285-298.

Kelso, J.A.S., Saltzman, E.L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, 14, 29-59.

Ladefoged, P. & Maddieson, I. (1990). Vowels of the world's languages. *Journal of Phonetics*, 18: 93-122.

Ladefoged, P. & Maddieson, I. (1996). *The Sounds of World Language*. Blackwell Publishers,

Oxford.

Lieberman, A.M. & Mattingly, I.G. (1985). The motor theory of speech perception revised.

Cognition, 21: 1-36.

Lieberman, A.M. & Whalen, D.H. (2000). On the relation of speech to language. *Trends in*

Cognitive Science, 3(7): 254-264.

Lindblom, B. & Maddieson, I. (1988). Phonetic universals in consonant systems. In: Hyman,

L.M. & Li, C.N. (Eds), *Language, Speech and Mind. Studies in honour of Victoria A.*

Fromkin, Routledge, London and New-York, pp. 62-78.

Luce, P.A., Pisoni, D.B. & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words.

In: G.T.M. Altman (Editor), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MIT Press, pp. 122-147.

Luce, P.A. & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation

model. *Ear and Hearing*, 19: 1-36.

MacKay, D.G., Wulf, G., Yin, C. & Abrams, L. (1993). Relations between word perception and

production: New theory and data on the verbal transformation effect. *Journal of Memory and Language*, 32: 624-646.

Macken, M. (1978). Permitted complexity in phonological development: One child's

acquisition of Spanish consonants. *Lingua*, 44: 219-253.

- MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21: 499-511.
- MacNeilage, P.F., Davis, B.L., Matyear, C.M. & Kinney, A. (1999). Origin of speech output complexity in infants and in languages. *Psychological Science*, 310(5): 459-460.
- MacNeilage, P.F. & Davis, B.L. (2000a). Origin of the internal structure of words. *Science*, 288: 527-531.
- MacNeilage, P.F. & Davis, B.L. (2000b). Deriving speech from nonspeech: a view from ontogeny. *Phonetica*, 57(2-4): 284-96.
- MacNeilage, P.F. & Davis, B.L. (2001). Motor mechanisms in speech ontogeny: Phylogenetic, neurobiological and linguistic implications. *Current Opinion in Neurobiology*, 11: 696-700.
- Maddieson, I. (1984). *Patterns of Sounds*. Cambridge University Press, Cambridge.
- Maddieson, I. & Precoda, K. (1992). Syllable structure and phonetic models. *Phonology*, 9: 45-60.
- McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18: 1-86.
- Natsoulas, T. (1965). A study of the verbal transformation effect. *American Journal of Psychology*, 78: 257-263.

- Norris, D., McQueen, J. & Cutler, A. (2000). Merging information in speech recognition. *Behavioral and Brain Sciences*, 23: 299-370.
- Pitt, M. & Shoaf, L. (2001). The source of a lexical bias in the verbal transformation effect. *Language and Cognitive Processes*, 16(5/6): 715-721.
- Pitt, M. & Shoaf, L. (2002). Linking verbal transformations to their causes. *Journal of Experimental Psychology: Human, Perception and Performance*, 28(1): 150-162.
- Rochet-Capellan, A. & Schwartz, J.-L. (2005a). The labial-coronal effect and CVCV stability during reiterant speech production: an acoustic analysis. *Proceedings of INTERSPEECH'2005*, Lisbon, 1009-1012.
- Rochet-Capellan, A. & Schwartz, J.-L. (2005b). The labial-coronal effect and CVCV stability during reiterant speech production: an articulatory analysis. *Proceedings of INTERSPEECH'2005*, Lisbon, 1013-1016.
- Rousset, I. (2004). *Structures Syllabiques et Lexicales des Langues du Monde: Typologies, Tendances Universelles et Contraintes Substancielles*. Unpublished PhD thesis, Université Stendhal, Grenoble.
- Sato, M., Baciú, M., Løevenbruck, H., Schwartz, J.-L., Cathiard, M.-A., Segebarth, C. & Abry, C. (2004). Multistable representation of speech forms: An fMRI study of verbal transformations. *NeuroImage*, 23(3): 1143-1151.

- Sato, M., Schwartz, J.-L., Abry, C., Cathiard, M.-A. & Lœvenbruck, H. (2006). Multistable syllables as enacted percept: A source of an asymmetric bias in the verbal transformation effect. *Perception & Psychophysics*, 68(3): 458-474.
- Schwartz, J.-L., Boë, L.J, Vallée, N. & Abry, C. (1997). Major trends in vowel system inventories. *Journal of Phonetics*, 25: 233-254.
- Schwartz, J.-L., Abry, C., Boë, L.-J. & Cathiard, M.A. (2002). Phonology in a theory of perception-for-action-control. In: Durand, J., Lacks, B. (Eds), *Phonology: From Phonetics to Cognition*. Oxford University Press, Oxford, pp. 240-280.
- Schwartz, J.L., Boë, L.J., & Abry, C. (2006). Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAF) principle in a Perception-for-Action-Control Theory (PACT). In M.J. Solé, P. Beddor & M. Ohala (eds.) *Experimental Approaches to Phonology*. Oxford University Press (to appear).
- Serniclaes, W, & Bejster, P. (1979). Cross-language differences in the perceptual use of voicing cues. In H. & P. Hollien (eds). *Amsterdam Studies in the Theory and History of Linguistic Science*, IV, 755-764. Amsterdam: Benjamins.
- Shoaf, L. & Pitt, M. (2002). Does node stability underlie the verbal transformation effect? A test of node structure theory. *Perception & Psychophysics*, 64(5): 795-803.
- Stefanuto, M. & Vallée, N. (1999). Consonant systems: From universal trends to ontogenesis.

- Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco, Vol. 3, pp. 1973-1976.
- Surprenant, A. M. & Goldstein, L. (1998). The perception of speech gestures. *Journal of the Acoustical Society of America*, 104, 518–529.
- Tuller, B. & Kelso, J.A.S. (1990). Phase transitions in speech production and their perceptual consequences. In: M. Jeannerod (Editor), *Attention and Performance XIII*, Hillsdale, NJ: Erlbaum, pp. 429-452.
- Tuller, B. & Kelso, J.A.S. (1991). The production and perception of syllable structure. *Journal of Speech and Hearing Research*, 34: 501-508.
- Vallée, N. (1994). *Systèmes Vocaliques : De la Typologie aux Prédications*. Unpublished PhD thesis, Université Stendhal, Grenoble.
- Vallée, N., Boë, L.-J., Schwartz, J.-L., Badin, P. & Abry, C. (2002). The weight of phonetic substance in the structure of sound inventories. *ZAS Papers in Linguistics*, 28: 145-168.
- van Alphen, P.M., & Smits, R. (200). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics*, 32, 455-491.
- Vilain, A., Abry, C., Brosda, S. & Badin, P. (1999). From idiosyncratic pure frames to variegated babbling: Evidence from articulatory modelling. *Proceedings of the XIVth International*

Congress of Phonetic Sciences, San Francisco, Vol. 4, pp. 1973-1976.

Vitevitch, M.S., Luce, P.A., Charles-Luce, J. & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, 40: 47-62.

Vitevitch, M.S. & Luce, P.A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9: 325-329.

Vitevitch, M.S. & Luce, P.A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40: 374-408.

Warren, M.R. & Gregory, R.L (1958). An auditory analogue of the visual reversible figure. *American Journal of Psychology*, 71: 612-613.

Warren, M.R. (1961). Illusory changes of distinct speech upon repetition – The verbal transformation effect. *British Journal of Psychology*, 52: 249-258.

Warren, M.R. & Meyers, D.M. (1987). Effects of listening to repeated syllables: Category boundary shifts versus verbal transformation. *Journal of Phonetics*, 15: 169-181.

Yin, C. & MacKay, D.G. (1992). Auditory illusions and aging: Transmission of priming in the verbal transformation paradigm. Paper presented to the *IVth Biennial Cognitive Aging Conference*, Atlanta.

Zsiga, Elisabeth C. 1994 Acoustic evidence for gestural overlap in consonant sequences. *Journal*

of Phonetics 22: 121–140.

Note.

¹ From this view, it is worthwhile to note that labio-velar consonants (e.g., /mɾ/, /ɪd/, / o /) occur principally in West and Central Africa in different language families and also in Papua New Guinea languages. The labio-velar consonants are the most common double-stop consonants in these languages. The closure at the soft palate slightly leads up to the closure at the lips, and the lip release immediately follows the velar one. The labio-velar consonants differ from consonant clusters in that the two occlusions significantly overlap. Accordingly, labio-velar consonants are considered as a complex consonant truly co-articulated at two simultaneous places of articulation. We believe therefore that the fact that these consonants occur in some languages is not in contradiction with the proposed explanation for the Labial-Coronal effect.

Table 1. Acoustic properties of the CV sequences (duration, vowel formants, intensity and pitch)

used in (A) Experiment 1 and (B) Experiment 2.

(A)

		/pa/	/ta/	/pi/	/ti/	/po/	/to/
Duration (<i>ms</i>)	Consonant	134	120	114	157	139	112
	Vowel	131	118	122	133	140	141
	CV	265	238	236	290	279	253
	Burst	6	15	17	15	13	12
Intensity (<i>dB</i>)	Burst	72	66	77	64	58	65
	Vowel	78	78	83	83	77	77
Formant (<i>Hz</i>)	F ₁	753	747	255	258	372	373
	F ₂	1357	1329	2267	2282	889	868
	F ₃	2867	2835	3404	3313	2960	2881
Pitch (<i>Hz</i>)	Vowel	110	111	112	114	113	113

(B)

		/ba/	/da/	/bi/	/di/	/bo/	/do/
Duration (<i>ms</i>)	Consonant	88	93	102	116	99	100
	Vowel	138	138	115	114	127	126
	CV	226	231	217	230	226	226
	Burst	5	12	7	8	5	12
Intensity (<i>dB</i>)	Burst	65	60	61	59	62	65
	Vowel	68	68	69	70	71	72
Formant (<i>Hz</i>)	F ₁	696	711	244	251	328	328
	F ₂	1446	1335	2278	2357	916	988
	F ₃	2821	2796	3309	3234	2874	2804
Pitch (<i>Hz</i>)	Vowel	108	109	112	113	117	116

Table 2. Lexical factors related to the stimuli used in (A) Experiment 1 and (B) Experiment 2.

The disyllabic and bigram frequencies are defined as the number of lexical entries incorporating a CVCV or a CV sequence identical to that of the stimulus either at the word onset or at any position in the word. The neighborhood density is defined as the number of phonologically similar words that differ from the stimulus only by a single substitution, insertion or deletion at any position in the stimulus. All lexical analyses were extracted from VoCoLex, a lexical database for the French language (~105000 words; Dufour, Peerman, Pallier and Radeau, 2002), and computed for token frequency and type frequency (indicated in coma).

(A)

	/pata/	/tapa/	/piti/	/tipi/	/poto/	/topo/
Disyllabic frequency	19 (15)	10 (7)	7 (4)	11 (7)	6 (5)	11 (7)
Disyllabic frequency (initial)	15 (11)	9 (6)	5 (5)	7 (4)	3 (2)	9 (5)
Neighborhood density	71 (35)	55 (25)	22 (10)	29 (13)	39 (19)	30 (13)

	/pa/	/ta/	/pi/	/ti/	/po/	/to/
Bigram frequency	2975 (1515)	3408 (1926)	1784 (912)	4743 (2490)	2135 (1050)	1897 (1076)
Bigram frequency (initial)	1648 (823)	767 (376)	541 (277)	336 (178)	696 (345)	466 (218)

	/pVtV/	/tVpV/
Disyllabic frequency	874 (447)	376 (184)
Disyllabic frequency (initial)	383 (207)	210 (105)
Neighborhood density	1887 (645)	916 (340)

	/pV/	/tV/
Bigram frequency	12132 (5877)	21818 (10679)

Bigram frequency (initial)	5177 (2554)	3037 (1507)
----------------------------	-------------	-------------

(B)

	/bada/	/daba/	/bidi/	/dibi/	/bodo/	/dobo/
Disyllabic frequency	6 (5)	0	2 (2)	0	9 (4)	0
Disyllabic frequency (initial)	3 (3)	0	0	0	0	0
Neighborhood density	25 (14)	25 (14)	19 (9)	4 (2)	22 (14)	12 (6)

	/ba/	/da/	/bi/	/di/	/bo/	/do/
Bigram frequency	2193 (1100)	875 (551)	1347 (698)	2966 (1524)	1072 (554)	1147 (592)
Bigram frequency (initial)	1300 (618)	174 (99)	459 (243)	1128 (531)	420 (217)	358 (176)

	/bVdV/	/dVbV/
Disyllabic frequency	221 (109)	389 (182)
Disyllabic frequency (initial)	137 (67)	380 (176)
Neighborhood density	1100 (397)	615 (261)

	/bV/	/dV/
Bigram frequency	8108 (4058)	14349 (6913)
Bigram frequency (initial)	3872 (1926)	7712 (3602)

Figure 1. Mean relative stability durations observed in Experiment 1 for the three pairs of stimuli and according to the reported transformations. Error bars represent standard errors of the mean.

Figure 2. Mean relative stability durations observed in Experiment 2 for the three pairs of stimuli and according to the reported transformations. Labial-coronal (LC) and coronal-labial (CL) transformations strictly corresponded to the repeated sequence or differing from it by the lack of consonantal voicing (e.g., /pata/, /bata/, /pada/, /bada/ for /pata/). Error bars represent standard errors of the mean.