A Perceptual Correlate of the Labial-Coronal Effect

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: The authors 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, they 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 (I)LCLC(I) 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 online speech processing. This result is interpreted in relation with theories assuming a link between perception and action in the human speech processing system.

KEY WORDS: labial-coronal effect, phonological tendencies, verbal transformation effect, perceptuo-motor interactions

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: 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 the University of California Los Angeles Phonological Segment Inventory Database (UPSID; Maddieson, 1984). Preferences for certain sounds over others have been described and analyzed in many studies (e.g., Crothers, 1978; Ladefoged & Maddieson, 1990; Maddieson, 1984; Schwartz, Boë, Vallée, & Abry, 1997; Vallée, 1994, for vowels; Ladefoged & Maddieson, 1996; Lindblom & Maddieson, 1988; Maddieson, 1984; Stefanuto & Vallée, 1999; Vallée, Boë, Schwartz, Badin, & 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, see MacNeilage & Davis, 2000b). Ingram (1974) first reported that during childhood, there is a strong preference to produce words beginning with a prevocalic consonant articulated with a more anterior place of articulation than the postvocalic one. This “fronting” tendency was further observed by
MacNeilage, Davis, Matyear, and Kinney (1999) in 9 out of 10 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 (/LaVCoV/) to coronal–vowel–labial (/CoVLaV/) sequences was 2.55. This labial-coronal (LC) effect (henceforth abbreviated as LC effect) is so strong that it sometimes occurs even if the word that the infant is 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 10 languages revealed a strong preference for /LaVCoV/ over /CoVLaV/ sequences at the beginning of words (/LaVCoV/ to /CoVLaV/ ratio: 2.23; MacNeilage et al., 1999). Similar results were obtained in our laboratory by analyzing a 15-language syllabified lexicon database (Rousset, 2004), partly available from Maddieson and Precoda’s (1992) data. The lexicons were representative of the main language families and were fairly well distributed geographically. They contained from approximately 2,000 words (Ngizim) to 12,200 words (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 in intersyllabic patterns not only at word onsets (/LaVCoV/ to /CoVLaV/ ratio: 2.40) but also throughout the entire words between two consecutive syllables (ratio: 1.75). The LC effect also occurs in intrasyllabic patterns, between onset and coda, without restriction on the position in the words (ratios: 1.73 for /LaVCoV/ vs. /CoVLaV/ 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 regardless of their position within the word, the strongest effect was found for disyllabic words (ratio: 2.77) and, regardless of the word length, was found 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, 2000b, 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, consonant–vowel (CV) 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 & Davis, 2000b, p. 529).

**The Labial-Coronal Chunking Hypothesis**

The developmental explanation proposed by MacNeilage and Davis (2000a, 2000b, 2001) seems however a bit ad hoc, especially 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. Rather, an alternative hypothesis may come from the gestural overlap asymmetry previously observed between LC and coronal-vowel–labial-vowel (CL) sequences in consonantal clusters, possibly reflecting a preferential coordinative articulatory mode for LC compared with CL sequences.

In C1C2 sequences in which 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 /LaC0/ sequence) than when it is posterior (e.g., in a /CaLa/ sequence). This fact has been described for several languages (examples in English include Byrd, 1992, 1996; Hardcastle & Roach, 1979; Surprenant & Goldstein, 1998; and Zsiga, 1994; examples in Georgian include Chitoran, Goldstein, & Byrd, 2002) and various manners of articulation (extending the many findings on stops, provided in the previously cited studies, to liquids; Chitoran & Goldstein, 2006). The fact that the overlap is larger when C1 is anterior to C2 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 realized in a vocal tract that is 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 they are displaying a larger overlap) in /LaC0/ sequences than in /CaLa/ ones.

In recent work, Rochet-Capellan and Schwartz (2005a, 2005b) 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 /LaVCoV/ sequences would be better “in phase” than /CoVLaV/ ones, thanks to the possibility of preparing the coronal gesture in the course of the labial realization, whereas 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 & 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, & Tuller, 1986), and the accelerated repetition of a given speech sequence might induce a reorganization of the coordination between the effectors toward a preferential coordinative mode. In a series of acoustic and articulatory measurements on speeded /LaVCoV/ and /CoVLaV/ sequences involving both plosives and fricatives (e.g., /pata/ and /tapa/, /pasa/ and /sapa/, /fata/ and /tafa/), Rochet-Capellan and Schwartz (2005a, 2005b) obtained two main results. First, acoustical measurements (Rochet-Capellan & 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 /LaVCoV/ sequences are actually more stable than /CoVLaV/ ones, with a progression of both kinds of sequences toward a LaCo attractor (e.g., /pa’ta/ → /pa’sa/ → /fa’ta/). Second, articulatory measurements (Rochet-Capellan & 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 the jaw closing phase, while the coronal release occurs around the jaw’s 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 release rather than the inverse. 

A Possible Perceptual Correlate of the Labial-Coronal Chunking Hypothesis

The foregoing studies (Rochet-Capellan & Schwartz, 2005a, 2005b) indicate that speakers tend to chunk /LaVCoV/ 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 to 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 involving a link between perception and action (e.g., in the motor theory of speech perception [Liberman & Mattingly, 1985; Liberman & Whalen, 2000]; in the direct realist theory of speech perception [Fowler, 1986]; or in the perception for action control theory [Schwartz, Abry, Boë, & Cathiard, 2002; Schwartz, Boë, & Abry, 2006]. With this aim, we exploited the verbal transformation effect (Warren, 1961; Warren & Gregory, 1958). 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 pseudowords 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 & Meyers, 1987), auditory streaming/perceptual grouping (the repeated stimulus being separated into different audio streams, giving rise to verbal transformations involving only part of the presented material; Pitt & Shoaf, 2001, 2002), and lexical and semantic transformations (Kaminska, Pool, & Mayer, 2000; Shoaf & Pitt, 2002; Warren, 1961). In a recent study, Sato, Schwartz, Abry, Cathiard, and Luvenbruck (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., /psa/ or /sapa/) 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 /sapa/ (with the coronal on the syllable onset and the labial on the coda), toward a better-phased labial-coronal cluster on syllable onset in /psa/ (with a tight synchronization of /p/ and /s/ thanks to gestural overlap asymmetry). Verbal transformation data displayed a large preference for
In a verbal transformation paradigm, would LC sequences provide a segmentation into LaCo chunks than into CoLa 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 online 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 on various lexical factors related to the repeating stimulus. 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 /LaVCoV/ sequence to the associated /CoVLaV/ 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, & Abrams, 1993; Shoaf & Pitt, 2002), and considering the general preference for LaVCoV sequences in French lexicons according to syllabic inventories (Rousset, 2004), various stimuli were selected to control for possible lexical interferences and to attempt to focus on “pure” sensorimotor mechanisms. As a matter of fact, the manifestations of the LC effect in French are quite diverse. For instance, although /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, although 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 /LaVCov/ and /CoVLov/ disyllabic sequences, Experiment 1 involving unvoiced plosives with /pVtV/ and /tVpV/ sequences and Experiment 2 involving 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 fewer 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 24 students from Grenoble University participated in the study. In Experiment 1, 17 men and 7 women were tested ($M = 24$ years of age; range = 18–48 years). In Experiment 2, 11 women and 13 men were tested ($M = 28$ years of age; range = 19–52 years). The number of participants was chosen on the basis of previous studies that tested a similar range of participants (e.g., Pitt & Shoaf, 2002; Shoaf & Pitt, 2002). All participants were native French speakers and reported no history of hearing or speaking disorders. None of them were aware of the purpose of the experiment.

**Materials**

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—that is, /pata/, /tapal/, /piti/, /tipi/, /poto/, /topo/ in Experiment 1 and /bada/, /daba/, /bidi/, /bido/, /bodo/ 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 and native French speaker (third author [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-bit digitized on the hard disk of a PC 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).

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 interstimulus 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, none of whom participated in 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 (13 women, 2 men; M = 26 years of age; range = 19–42 years). 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 10 times in a randomized sequence for a total of 120 trials. Each trial started with a fixation cue presented for 500 ms, immediately followed by the audio stimulus. Once the response was written, participants had to press a computer key 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 (approximately 105,000 words; Dufour, Peerman, Pallier, & 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 = 2.4 per million; range = 0–9), whereas 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, & Goldinger, 1990) and the disyllabic frequencies (defined as the number of lexical entries incorporating a CVCV structure identical to that of the stimulus 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 positions 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 8,486, 4,998, 5,568 and 8,249, respectively.

Finally, global analyses were conducted on /LaVCoV/ and /CoVLaV/ syllabic structures (La being any labial

| Table 1. Acoustic properties of the CV sequences (duration, vowel formants, intensity, and pitch) used in (A) Experiment 1 and (B) Experiment 2. |
| --- | --- | --- | --- | --- | --- | --- |
| (A) Experiment 1 | CV sequence | /pa/ | /ta/ | /pi/ | /ti/ | /po/ | /to/ |
| Duration (ms) | 134 | 120 | 114 | 157 | 139 | 112 |
| Consonant | 131 | 118 | 122 | 133 | 140 | 141 |
| Vowel | 265 | 238 | 236 | 290 | 279 | 253 |
| CV | 6 | 15 | 17 | 15 | 13 | 12 |
| Burst | 72 | 66 | 77 | 64 | 58 | 65 |
| Intensity (dB) | 78 | 78 | 83 | 83 | 77 | 77 |
| Vowel | 753 | 747 | 255 | 258 | 372 | 373 |
| F1 | 1357 | 1329 | 2267 | 2282 | 889 | 868 |
| F2 | 2867 | 2835 | 3404 | 3313 | 2960 | 2881 |
| F3 | 110 | 111 | 112 | 114 | 113 | 113 |
| Pitch (Hz) |
| Vowel | 108 | 109 | 112 | 113 | 117 | 116 |
| Burst | 68 | 68 | 69 | 70 | 71 | 72 |
| Vowel | 696 | 711 | 244 | 251 | 328 | 328 |
| F1 | 1446 | 1335 | 2278 | 2357 | 916 | 988 |
| F2 | 2821 | 2796 | 3309 | 3234 | 2874 | 2804 |
| F3 | 108 | 109 | 112 | 113 | 117 | 116 |

<table>
<thead>
<tr>
<th>(B) Experiment 2</th>
<th>CV sequence</th>
<th>/ba/</th>
<th>/da/</th>
<th>/bi/</th>
<th>/di/</th>
<th>/bo/</th>
<th>/do/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration (ms)</td>
<td>88</td>
<td>93</td>
<td>102</td>
<td>116</td>
<td>99</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Consonant</td>
<td>138</td>
<td>138</td>
<td>115</td>
<td>114</td>
<td>127</td>
<td>126</td>
<td></td>
</tr>
<tr>
<td>Vowel</td>
<td>226</td>
<td>231</td>
<td>217</td>
<td>230</td>
<td>226</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td>CV</td>
<td>5</td>
<td>12</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Burst</td>
<td>65</td>
<td>60</td>
<td>61</td>
<td>59</td>
<td>62</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>Intensity (dB)</td>
<td>68</td>
<td>68</td>
<td>69</td>
<td>70</td>
<td>71</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Vowel</td>
<td>696</td>
<td>711</td>
<td>244</td>
<td>251</td>
<td>328</td>
<td>328</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>1446</td>
<td>1335</td>
<td>2278</td>
<td>2357</td>
<td>916</td>
<td>988</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>2821</td>
<td>2796</td>
<td>3309</td>
<td>3234</td>
<td>2874</td>
<td>2804</td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td>108</td>
<td>109</td>
<td>112</td>
<td>113</td>
<td>117</td>
<td>116</td>
<td></td>
</tr>
</tbody>
</table>
consonant and C₀ being any coronal consonant in French). The numbers of lexical entries incorporating /L₀VC₀V/ syllabic structures, either at the word onset or at other positions in the word, were 2,570 and 3,238, respectively. The numbers of lexical entries incorporating /C₀VL₀V/ syllabic structures were 1,751 and 1,162, respectively. The /L₀VC₀V/ to /C₀VL₀V/ 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-language syllabified lexicon database (including French).

**Apparatus**

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

---

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

<table>
<thead>
<tr>
<th>(A) Experiment 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical factor</td>
</tr>
<tr>
<td>Disyllabic frequency</td>
</tr>
<tr>
<td>Disyllabic frequency (initial)</td>
</tr>
<tr>
<td>Neighborhood density</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lexical factor</th>
<th>/pVW/</th>
<th>/WpV/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disyllabic frequency</td>
<td>874 (447)</td>
<td>376 (184)</td>
</tr>
<tr>
<td>Disyllabic frequency (initial)</td>
<td>383 (207)</td>
<td>210 (105)</td>
</tr>
<tr>
<td>Neighborhood density</td>
<td>1887 (645)</td>
<td>916 (340)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lexical factor</th>
<th>/po/</th>
<th>/ta/</th>
<th>/pi/</th>
<th>/ti/</th>
<th>/po/</th>
<th>/ta/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigram frequency</td>
<td>2975 (1515)</td>
<td>3408 (1926)</td>
<td>1784 (912)</td>
<td>4743 (2490)</td>
<td>2135 (1050)</td>
<td>1897 (1076)</td>
</tr>
<tr>
<td>Bigram frequency (initial)</td>
<td>1648 (823)</td>
<td>767 (376)</td>
<td>541 (277)</td>
<td>336 (178)</td>
<td>696 (345)</td>
<td>466 (218)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lexical factor</th>
<th>/pVV/</th>
<th>/WpV/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigram frequency</td>
<td>12132 (5877)</td>
<td>21818 (10679)</td>
</tr>
<tr>
<td>Bigram frequency (initial)</td>
<td>5177 (2554)</td>
<td>3037 (1507)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(B) Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical factor</td>
</tr>
<tr>
<td>Disyllabic frequency</td>
</tr>
<tr>
<td>Disyllabic frequency (initial)</td>
</tr>
<tr>
<td>Neighborhood density</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lexical factor</th>
<th>/ba/</th>
<th>/da/</th>
<th>/bi/</th>
<th>/di/</th>
<th>/ba/</th>
<th>/da/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigram frequency</td>
<td>2193 (1100)</td>
<td>875 (551)</td>
<td>1347 (698)</td>
<td>2966 (1524)</td>
<td>1072 (554)</td>
<td>1147 (592)</td>
</tr>
<tr>
<td>Bigram frequency (initial)</td>
<td>1300 (618)</td>
<td>174 (99)</td>
<td>459 (243)</td>
<td>1128 (531)</td>
<td>420 (217)</td>
<td>358 (176)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lexical factor</th>
<th>/bVdV/</th>
<th>/dVbV/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigram frequency</td>
<td>221 (109)</td>
<td>389 (182)</td>
</tr>
<tr>
<td>Bigram frequency (initial)</td>
<td>137 (67)</td>
<td>380 (176)</td>
</tr>
<tr>
<td>Neighborhood density</td>
<td>1100 (397)</td>
<td>615 (261)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lexical factor</th>
<th>/bV/</th>
<th>/dV/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigram frequency</td>
<td>8108 (4058)</td>
<td>14349 (6913)</td>
</tr>
<tr>
<td>Bigram frequency (initial)</td>
<td>3872 (1926)</td>
<td>7712 (3602)</td>
</tr>
</tbody>
</table>

Note. 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. 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 (approximately 105,000 words; Dufour, Peerman, Pallier, & Radeau, 2002), and were computed for token frequency and type frequency (indicated in parentheses).
**Procedure**

The participants were individually tested in a quiet room. The experiment began with a lengthy briefing during which they were introduced to the verbal transformation task. The participants listened to the experimenter repeating the word “life” at a rate of 2 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 in French, aimed at displaying the verbal transformation effect on materials that all participants 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 to participants that these changes could be subtle or very noticeable and could correspond to a word as well as a pseudoword. 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 approximately 30 min.

**Data Analysis**

For each participant and each stimulus, transformations were phonetically labelled by a trained French phonetician (fourth author [I. R.] in Experiment 1 and second author [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 us 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 according 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/ vs. /tapa/ in Experiment 1, and /bada/ vs. /daba/ or {/bada/, /bata/, /pada/, /pata/} vs. /daba/, /taba/, /dapa/, /tapa/) in Experiment 2 were calculated for each participant (delta score). This allows us 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 the between-subject variable.

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

**Results**

**Experiment 1**

The mean relative stability durations (MSDs) 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 ± SD: 69% ± 3). The reported transformations differing from the repeated sequence by the presence of a consonantal voicing were marginal (e.g., /bata/ for /pata/; MSD ± SD: 2% ± 2). The other transformations included *auditory streaming*—that is, discarding part of the stimulus embedded in another auditory flow (e.g., /ata/ for...
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.

The MSDs of reported transformations observed in Experiment 2 for the three pairs of disyllabic reversible CVCV sequences are displayed in Figure 2. Although the perceptual stability of the two possible forms of the reversible CVCV structure was lower than in Experiment 1 (MSD ± SD: 34% ± 16), the reported transformations differing from the repeated sequence by the absence of consonantal voicing were very frequent (MSD ± SD: 37% ± 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 /dodo/ 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 were taken into account, the mean perceptual stability was still dominated by two forms, always incorporating the repeated stimulus (MSD ± SD: 52% ± 4).

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

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 -4%), $F(1, 23) = 8.86, p < .05$. There was no reliable effect of the vowel, $F(2, 46) = 1.16$. Nor was there an interaction between the two factors, $F(2, 46) = 0.92$.

Once the transformations differing from the repeated sequence by the lack of consonantal voicing were 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 was there an 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% vs. 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% vs. 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 Experiment × Vowel interaction, $F(2, 92) = 5.55$, $p < .005$, 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, 9% versus -6%), $F(1, 46) = 19.5$, $p < .005$. There was no reliable effect of the experiment, $F(2, 46) = 0.15$, nor of the vowel, $F(2, 92) = 0.45$, nor

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 differed from it by the lack of consonantal voicing (e.g., /pata/, /bata/, /pada/, /bada/ for /pata/). Error bars represent standard errors of the mean.
were there 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 were 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 Experiment × Vowel interaction, $F(2, 92) = 7.20, p < .005$, 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 among 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 regardless of 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 this section, we first discuss possible psycholinguistic interpretations of the LC asymmetries found in both experiments. Then, we explore some possible explanations of the two supplementary results observed in Experiment 2. Finally, we comment on the consequences of the observed asymmetries for the LC chunking hypothesis and, more generally, we discuss its implications in relation to 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 (MacKay et al., 1993; Yin & MacKay, 1992). 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 for words (as words have strong and well-formed representations) and for phonologically illegal pseudowords than for permissible pseudowords (Natsoulas, 1965; Shoaf & 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 & 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, whereas 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 (MacKay et al., 1993; Yin & MacKay, 1992). A similar competitive effect between lexical candidates was also observed during auditory word recognition tasks (e.g., Vitevitch & Luce, 1998, 1999)—in which words with few competitors were 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 (Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, McQueen, & Cutler, 2000). On the other hand, probabilistic phontactics, which refers to the frequency with which a particular sequential arrangement of phonetic 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, & 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 to more quickly than were low-probability/density nonwords, whereas the reverse effect was observed for words (see also Vitevitch & 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/ compared with /piti/, the neighborhood density effect is not likely to provide an explanation for the greater stability observed for LC compared with 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, compared with 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 increased the reliability of these lexical analyses, none of the psycholinguistic interpretations that are based on the lexical status, the lexical frequency, the neighborhood density, or the probabilistic phonotactics of the repeated stimuli appear to satisfactorily explain the present results. Alternatively, 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, although 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 with the LC effect.

The Potential Role of Voicing and Vocalic Context

There appeared to be 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), whereas 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 & Smits, 2004). However, in these experiments the perceptual loss of voicing was related to a production problem 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” (van Alphen & Smits, 2004, p. 485). Almost all tokens produced with prevoicing were identified as “voiced” (van Alphen & Smits, 2004, p. 485).

In the present case, the voicing contrast for French is indeed mainly carried by the presence versus absence of prevoicing (e.g., Serniclaes & 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. Rather, the explanation seems to rely on the transformation phenomenon itself. Actually, a number of transformations reported in the literature involve voicing (e.g., Pitt & Shoaf, 2002; Warren, 1961). 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 participants while listening to repetitions of /tu/ and /du/ (“to” and “do”) presented in regular alternation. It appeared that over the 30 participants, 19 reported the “to to” form, and only 6 reported “do do.” Warren and Meyers (1987) 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-lived components” (Warren & Meyers, 1987, p. 179).

Pitt and Shoaf (2001, 2002) later developed this idea with the introduction of the streaming process. In this process, as previously mentioned, the repeated audio content can be perceptually separated into different audio streams, giving rise 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 voice bar apart would remove 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 because the listeners are then focused on changes between various forms of the LC–CL sequences. Actually, a majority of the participants 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 participants 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 in which the participant 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 semivowels between the stop and the vowel, both for /b/ (typically /bo/ becoming /bro/ with the introduction of a uvular liquid/fricative) and for /d/ (typically /di/ becoming /dzi/ with the introduction of an alveolar or postalveolar fricative). This is in good agreement with a classical type of transformation of speech occurrences both in speech perception and in language shifts (e.g., Chang, Plauché, & 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 with 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 behavior 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, 2000b, 2001) and then be captured in the phonology by a generalization process. The fact that there can be exceptions to this process (as in the case of voiced /CvLCnV/ sequences, more frequent than /LnVCvV/ ones in French) would not prevent the listeners from using this LC principle in the verbal transformation paradigm and from showing a systematic preference for LC sequences. There are, however, several problems with this assumption. Apart from the weaknesses of the developmental scenario, previously mentioned in The Labial-Coronal Effect in Human Languages section, 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, 2005b). 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 The Labial-Coronal Effect in Human Languages section. According to this hypothesis, the speaker tends to chunk LC sequences on a single jaw gesture, with a neighboring 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 be audible in spite of the closing gesture being 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 & 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 supports the proposition of an articulatory basis to this perceptual process. Such a perceptuo-motor link is compatible with both motor (Fowler, 1986; Liberman & Mattingly, 1985; Liberman & Whalen, 2000) and perceptuo-motor (Schwartz et al., 2002, 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 online analysis of articulatory speech forms that support communicative or interpretative speech, and particularly of syllable parsing in the verbal transformation task. This neuroanatomical 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 & Goldstein, 2006) and the speeding production paradigm (Rochet-Capellan & 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 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, which is 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 (Liberman & Mattingly, 1985; Liberman & Whalen, 2000; Schwartz et al., 2002, 2006).

**Acknowledgments**

This work was supported by Centre National de la Recherche Scientifique and Ministero Italiano dell’Istruzione, dell’Università e della Ricerca as well as by the French project “Patiipapa: Modélisation de l’émergence d’un langage articulé dans une société d’agents sensori-moteurs en interaction [Patiipapa: Modeling the emergence of an articulated language in a society of sensorimotor agents in interaction]” (ACI Systèmes Complexes en SHS).

We thank Christian Abry, Barbara Davis, Judith Gierut, and Peter MacNeilage 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.

**References**


Received March 8, 2006
Revised September 19, 2006
Accepted April 4, 2007
DOI: 10.1044/1092-4388(2007/101)

Contact author: Nathalie Vallée, Institut de la Communication Parlée, UMR CNRS N° 5009, Institut National Polytechnique de Grenoble, Université Stendhal, BP25, 38040 Grenoble Cedex 9, France. E-mail: nathalie.vallee@icp.inpg.fr.
A Perceptual Correlate of the Labial-Coronal Effect

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

*J Speech Lang Hear Res* 2007;50;1466-1480
DOI: 10.1044/1092-4388(2007/101)

The references for this article include 3 HighWire-hosted articles which you can access for free at: http://jslhr.asha.org/cgi/content/full/50/6/1466#BIBL

This article has been cited by 2 HighWire-hosted article(s) which you can access for free at:

http://jslhr.asha.org/cgi/content/full/50/6/1466#otherarticles

This information is current as of February 12, 2013

This article, along with updated information and services, is located on the World Wide Web at:

http://jslhr.asha.org/cgi/content/full/50/6/1466