Perception & Psychophysics
2006, 68 (3), 458-474

Multistable syllables as enacted percepts: A source of an asymmetric bias in the verbal transformation effect

MARC SATO, JEAN-LUC SCHWARTZ, CHRISTIAN ABRY, MARIE-AGNÈS CATHIARD,
and HÉLÈNE LŒVENBRUCK
CNRS UMR 5009, Institut National Polytechnique de Grenoble,
and Université Stendhal, Grenoble, France

Perceptual changes are experienced during rapid and continuous repetition of a speech form, leading to an auditory illusion known as the *verbal transformation effect*. Although verbal transformations are considered to reflect mainly the perceptual organization and interpretation of speech, the present study was designed to test whether or not speech production constraints may participate in the emergence of verbal representations. With this goal in mind, we examined whether variations in the articulatory cohesion of repeated nonsense words—specifically, temporal relationships between articulatory events—could lead to perceptual asymmetries in verbal transformations. The first experiment displayed variations in timing relations between two consonantal gestures embedded in various nonsense syllables in a repetitive speech production task. In the second experiment, French participants repeatedly uttered these syllables while searching for verbal transformation. Syllable transformation frequencies followed the temporal clustering between consonantal gestures: The more synchronized the gestures, the more stable and attractive the syllable. In the third experiment, which involved a covert repetition mode, the pattern was maintained without external speech movements. However, when a purely perceptual condition was used in a fourth experiment, the previously observed perceptual asymmetries of verbal transformations disappeared. These experiments demonstrate the existence of an asymmetric bias in the verbal transformation effect linked to articulatory control constraints. The persistence of this effect from an overt to a covert repetition procedure provides evidence that articulatory stability constraints originating from the action system may be involved in auditory imagery. The absence of the asymmetric bias during a purely auditory procedure rules out perceptual mechanisms as a possible explanation of the observed asymmetries.

In a study on visual imagery, Chambers and Reisberg (1985) defended the view that

> mental images in any modality have no existence outside our understanding of them, making the image and its comprehension inseparable. In perception, there is a physical stimulus, existing independently of the perceiver, which needs interpretation. However, in imagery there is no free-standing icon waiting to be interpreted, and no interpretation is needed to learn what the image depicts. (cited in Reisberg, Smith, Baxter, & Sonenshine, 1989, p. 620)

To verify this hypothesis, the authors carried out a set of experiments in which participants were asked to imagine standard ambiguous figures, such as the Necker cube. They showed that the participants were uniformly unable to mentally discover shapes other than the one provided by the experimenter. Following the experiment, however, the participants were able to draw the figure and to discover interpretations different from the first. Chambers and Reisberg concluded that visual mental images are inherently unambiguous.

Considering that the stimuli and perceptual properties of auditory imagery could differ from those of visual imagery, Reisberg et al. (1989) attempted to examine ambiguous images in the auditory modality. For this purpose, they made use of the verbal transformation effect (Warren & Gregory, 1958). This "word game" (Treiman, 1983), which bears an analogy with the depth perceptual rivalry present in the Necker cube, relies on the fact that certain words, if repeated over and over, yield a soundstream compatible with more than one segmentation. For example, rapid repetitions of the word *life* may perceptually switch into sequences of the word *fly*. Reisberg et al. used this verbal transformation paradigm to test the unambiguity assumption on mental images in the auditory domain by examining whether or not imagined repetitions could produce verbal transformations just as heard repetitions do. In their experiments, the authors asked participants to

We thank Kenneth de Jong and three anonymous reviewers for a helpful review of the manuscript. We also thank Alain Content, Pascal Perrier, Daniel Reisberg, and Rudolph Sock for helpful discussions, and Anahita Basirat, Anne Vilain, Christophe Savariaux, and Willy Serniclaes for their assistance in scoring and analyzing data. Aspects of this article were presented at the 15th International Congress of Phonetic Sciences, Barcelona, August 2003. Correspondence concerning this article should be addressed to M. Sato, Institut National Polytechnique de Grenoble, 46, avenue Félix Viallet, 38031 Grenoble Cedex 01, France (e-mail: sato@icp.inpg.fr).

imagine the repetition of a word and to report any transformation of the auditory image. In order to test their assumption more thoroughly, they asked other participants to detect possible transformations during overt repetition of the same word produced either by the experimenter or by themselves.

However, Reisberg et al. (1989) noted that, in the imagined condition, participants might supplement the auditory image with a subvocalized *enactment*. According to the authors, enactment could provide real physical cues that might encourage auditory transformations, whereas "pure" auditory imagery would not. To control for this potential shortcoming in the paradigm, the authors tested various conditions that differed in the degree of enactment (whispering, silent mouthing, imaging with no mouthing) or even eliminated enactment (with a concurrent articulatory task, by having participants chew candy, or by having participants clamp the articulators). The experimental data showed that transformations were largely eliminated when subarticulation was blocked. Eliminating enactment prevented participants from detecting a transformation. Therefore, transformations in auditory imagery seem to require subvocalization. Moreover, the authors found that the transformation probability gradually decreases from a condition of complete externalization to one of complete internalization, through a condition of partial externalization (whispering, mouthing).

Reisberg et al. (1989) concluded that subvocalized enactment enables refreshment and, thus, elaboration of verbal auditory images, whereas "pure," unenacted auditory images remain unambiguous, just as visual images do. They thereby provided the first demonstration that speech production constraints, specifically speech enactment, may intervene in the emergence of verbal transformations.

Verbal Transformations: A Window Into Speech Representations

In past research, verbal transformations have been studied mainly as purely perceptual effects. The classic paradigm consists in presenting participants with an auditory speech stimulus looped on a tape (see, e.g., Kaminska, Pool, & Mayer, 2000; Pitt & Shoaf, 2001, 2002; Shoaf & Pitt, 2002; Warren, 1961, 1982; Warren & Meyers, 1987). It has been shown that the number of transformations heard by listeners depends on stimulus length, interstimulus interval, and listening duration (Warren, 1961). Previous studies have reported that perceptual changes to auditory input could range from small phonetic deviations to strong semantic distortions, including substitution of a phoneme by a phonetically close one (Warren, 1961; Warren & Meyers, 1987), auditory streaming/perceptual grouping (Pitt & Shoaf, 2001, 2002), and lexical and semantic transformations (Kaminska et al., 2000; Shoaf & Pitt, 2002; Warren, 1961). Lexical and sublexical levels of representation have been suggested as the loci of such effects. Accordingly, verbal transformations should vary as a function of distinct factors related to the repeated stimulus: its neighborhood density (i.e., the number of lexical entries that are phonologically similar to it), its frequency in the language, and whether or not it is a word (MacKay, Wulf, Yin, & Abrams, 1993; Natsoulas, 1965; Shoaf & Pitt, 2002; Yin & MacKay, 1992).

Although the processes implied in verbal transformations have challenged unified theoretical explanation for more than four decades, these transformations are considered to reflect mainly the operation of processes devoted to the perceptual organization and interpretation of speech (Warren, 1982). Two functions seem to be involved in the reinterpretation of the signal when it no longer makes sense: satiation and criterion shift (Kaminska et al., 2000; MacKay et al., 1993; Warren & Meyers, 1987). Repeatedly listening to a stimulus causes its memory representation to satiate. Simultaneously, the criteria used to categorize it abruptly shift, and a new representation is then built. These processes would repeat themselves throughout the presentation of the stimulus.

The Articulatory Synchrony Hypothesis

Viewed as temporary fluctuations of the online linguistic information processing that arise during veridical perception, auditory illusions provide a useful framework for enhancing our understanding of the language system by revealing otherwise hidden mechanisms (Warren, 1982). From this point of view, the verbal transformation effect appears to be well suited to exploring the organization of lexical and sublexical representations by examination of variations in the perceptual stability of reported transformations. Furthermore, the facts that (1) the effect occurs not only during a purely auditory procedure but also during an overt or covert repetition condition and (2) speech production constraints, specifically speech enactment, could also intervene in the transformation process (Reisberg et al., 1989; Smith, Wilson, & Reisberg, 1995) make the verbal transformation effect a nice pivot point from which to examine whether speech production and perception constraints act on verbal auditory images. In this framework, the aims of the present study were to further test how specific articulatory constraints may contribute to verbal transformations and to examine the functional equivalence, in terms of transformation mechanisms, between the classic perception procedure and the production variant procedure.

In this regard, an important issue considered neither in Reisberg et al. (1989) nor in other verbal transformation studies concerns the existence of possible verbal transformation asymmetries. For instance, the reverse transformation from *fly* to *life* seems less likely than a reverse transformation from *life* to *fly*. This could be related to differences in the articulatory cohesion of speech stimuli (Browman & Goldstein, 1989): The two consonantal gestures in the syllable onset of *fly* are temporally very compact, and the speaker can produce the three gestures of /flai/ almost in synchrony. In the *life* sequence, the synchronization of [l] (syllable onset) and [f] (syllable coda) is of course impossible. Hence, the temporal clustering in /flai/ could explain the facilitation for the transformation from *life* to *fly*.

More generally, we suggest that transformations from less to more temporally clustered stimuli should be more frequent, since they are associated with more compact and tightly synchronized sequences of articulatory gestures. From this point of view, since the pioneering work of Stetson (1951), several studies have reported evidence of variations in the articulatory phase relationship during speech production (e.g., de Jong, 2001; de Jong, Nagao, & Lim, 2002; Gleason, Tuller, & Kelso, 1996; Tuller & Kelso, 1990, 1991). By using a repetitive speech production task, Stetson first observed that fast repetition rates could induce specific syllabic parsing (i.e., the resyllabification of codas in vowel–consonant [VC] sequences into onsets in consonant–vowel [CV] sequences). Tuller and Kelso (1990, 1991) further replicated this finding by introducing the concept of relative phasing of articulatory events. They showed that as speaking rate increased, the sequence /pi/ remained stable throughout the task, whereas they observed a switch from /ip/ to /pi/ at a critical rate (see also de Jong, 2001; de Jong et al., 2002; Gleason et al., 1996). The greater stability of the CV phonetic structure in comparison with that of the VC structure is in accordance with studies of intrasyllabic tendencies during the babbling and single-word periods in early vocal acquisition (MacNeilage, 1998; MacNeilage & Davis, 2000) and is also supported by the cross-linguistic typology literature showing a clear predominance of CV syllables (see, e.g., Jakobson, 1966; Maddieson, 1984). Hence, the finding of variations in phasing patterns has provided a useful framework for rationalizing a number of typological facts.

Since the verbal transformation effect is well suited for assessing perceptual stability and phonological consciousness, the present study was first designed to test whether or not variations in the articulatory cohesion of repeated nonsense syllables could lead to perceptual asymmetries in verbal transformations. With this purpose in mind, the first experiment was designed to help us select a convenient phonetic material and to display variations in the temporal clustering of articulatory-acoustic events within this material. The second experiment, in which an overt repetition mode was used, was designed to test whether or not these variations could lead to perceptual asymmetries in the reported verbal transformations. The goal of Experiment 3 was to further explore such possible perceptual asymmetries using a covert repetition mode and then to examine whether synchrony constraints originating from the action system may participate in the elaboration of verbal auditory images. Finally, the goal of Experiment 4, involving a purely perceptual condition, was to disentangle the roles of perception mechanisms and production constraints in the pattern of results provided by the previous experiments.

EXPERIMENT 1

In order to explore possible perceptual asymmetries in the verbal transformation effect, a convenient set of speech stimuli that were likely to display various degrees of articulatory cohesion was first selected. The goal of the first experiment was to study variations in timing relations of articulatory-acoustic events within this material during a repetitive speech production task performed at various rates. This is in line with a large body of literature on speech timing, in which reorganizations in timing at high rates provide an implicit attractor to the speech production system and shed light on the respective cohesion of various competing sequences.

Method

Phonetic material. Six monosyllabic nonsense words (/psə/, /səp/, /əps/, /spə/, /pəs/, and /əsp/) were selected. Each sequence consisted of a combination of the bilabial [p] and coronal [s] consonants and the neutral vowel [ə]. The neutral vowel [ə] was selected because it imposes minimal constraints on the vocal tract shape and hence leaves the articulators for /p/ and /s/ free to accomplish their consonantal tasks.

None of these syllables occurs as a word in the French lexicon, minimizing lexical interferences in the verbal transformation task (Shoaf & Pitt, 2002). However, once the neutral vowel is removed, the phonological types of the speech sequences (i.e., /psV/, /sVp/, /Vps/, /spV/, /pVs/, and /Vsp/, where V is any spoken French vowel) are all phonotactically attested in French. From this point of view, it is also important to note that [ə] is generally realized as a mid-open front rounded vowel, making its articulatory realization close to [œ], which can occur in both closed and open syllables in French.

Articulatory-acoustic events. The consonants and the vowel were characterized by onset acoustic events with clear articulatory interpretations. Schematic unfolding of onset events for the six sequences are shown in Figure 1. For /səp/, the frication onset for [s] is followed by the voicing onset for [ə] and the [p] release after the bilabial closure. For /psə/, two of these events more or less cohere into a single event, defined by the onset of friction at the beginning of the logatome and corresponding to tightly synchronous gestures for [p] (lip opening) and [s] (tongue tip placing). For /əps/, these two events still cohere, but this happens long after the voicing onset for [ə] and after the lip closure for [p]. The three individual onset events for [s], [p], and [ə] are all separated in isolated /spə/, /pəs/, and /əsp/ sequences, respectively. For /spə/ and /pəs/ cycles, the two onset events for [s] and [p] remain separated by the lip closure period; hence they never cohere. For /əsp/, the frication onset and the bilabial release events occur only after the voicing onset for [ə].

In order to test the stability and cohesion of sequences during a repetitive speech production task, we focused on the relative timing of the onset events for the two consonantal gestures. As was described above, these events should remain always separated in cycles for /spə/ and /pəs/. For /psə/ and /səp/ sequences, an important question is whether or not the onset events might cohere in cycles for /səp/, the release of the final [p] becoming synchronous with the tongue-driven onset of the initial [s] in the next /səp/ utterance. This would result in resyllabification from /səp/ to /psə/.

Apparatus and Procedure. Examination of /səp/ cycles showed that [p] becomes implosive almost at once. This means that the [p] burst disappears, and hence it becomes impossible to track the /ps/ coherence on the acoustic signal, since the [p] onset event is no longer noticeable. To avoid this difficulty and to test the reality of

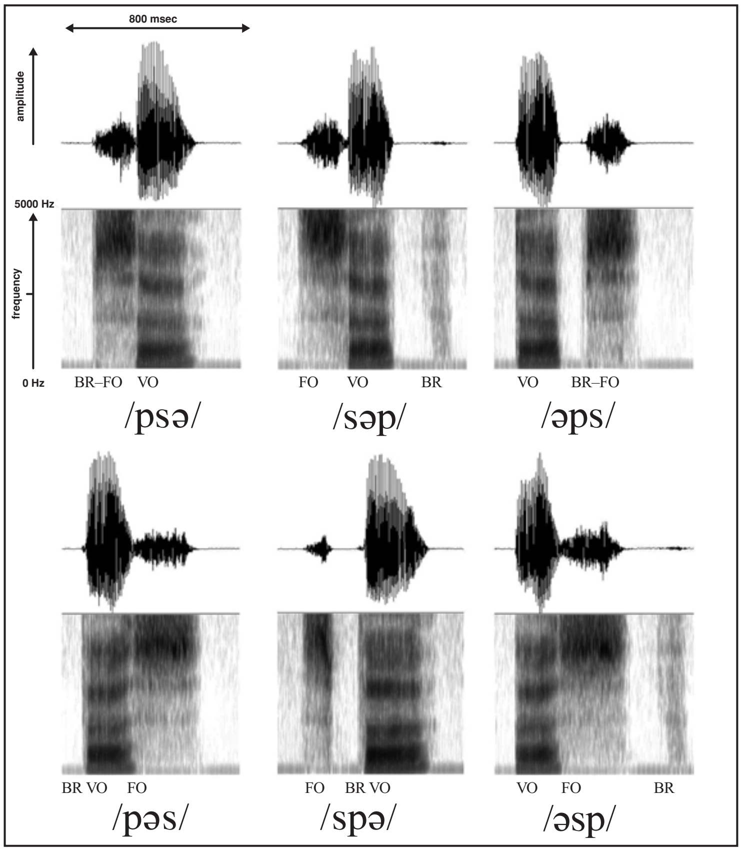

Figure 1. Unfolding of acoustic events for the /psə/, /səp/, /əps/, /pəs/, /spə/, and /əsp/ speech sequences. For each sequence, the acoustic signal (top of each panel) is displayed in synchrony with the corresponding spectrogram (temporal frequency representation, bottom of each panel). BR, bilabial release for [p] after the bilabial closure; FO, frication onset for [s]; VO, voicing onset for [ə].

coordinative patterns of onset events, we performed audiovisual recordings of /psə/, /səp/, /pəs/, and /spə/ sequences. We suspected that [p] is actually released on the lips, even with no acoustic noise. We used the setup designed at the Institut de la Communication Parlée of Université Stendhal for such phonetic analyses, in which the speaker's lips are colored with blue makeup to allow precise video analyses using a Chroma-Key process (Lallouache, 1990).

The four sequences were individually recorded by a trained phonetician who is a native speaker of French (J.-L.S.). Two rate manipulations were examined. In the increased rate condition, the speaker progressively speeded up the rhythm from low (around 1 cps) to high (around 6 cps). In the fixed rate condition, the speaker uttered the cycles at two stable rhythms: low (around 1 cps) and moderate (around 2 cps). Although previous studies of articulatory phasings always used a speeding-up speech production paradigm (de Jong, 2001; de Jong et al., 2002; Gleason et al., 1996; Tuller & Kelso, 1990, 1991), the fixed rate condition was designed to test whether variations in timing relationships between articulatory-acoustic events would occur at a sufficient rate but without any acceleration.

All the stimuli were then analyzed in the following way. The initiation of the high-frequency noise characteristic of the [s] onset event was detected on the stimulus spectrogram (using Praat software, Institute of Phonetic Sciences, University of Amsterdam). The [p] onset event was localized by analyzing the variations of the lip area and by detecting the first video frame with a nonzero area after the closure period. (The system can detect lip areas as small as 1 mm^2, with a temporal precision of 20 msec; Abry, Cathiard, Vilain, Laboissière, & Schwartz, 2004). Then, we defined the /ps/ coherence index as the time separating the onset events for [p] and [s] divided by the cycle duration, defined as the time between two consecutive [p] events.

Results

Plotted in Figure 2A are the variations of the coherence index as a function of cycle duration (in seconds) for the /psə/ and /səp/ speech sequences in the increased rate condition. Whatever the speech rate, the index was stable, at around zero, for /psə/. Although the index for /səp/ was around 0.50 at low speed (around 1 cps), it abruptly decreased toward zero from 2 cps. We observed a similar pattern for the two speech sequences in the fixed rate condition (see Figure 2B). Whether the speech rate was low or moderate, the index was stable at around zero for /psə/ (*M* = 0.02, *SD* = 0.01 in the low speech rate condition; *M* = 0.03, *SD* = 0.01 in the moderate speech rate condition). At a low speech rate, the index for /səp/ was 0.40–0.50 (*M* = 0.43, *SD* = 0.02), whereas it was 0.10–0.20 during the moderate speech rate condition (*M* = 0.15, *SD* = 0.06). A two-way ANOVA with stimulus condition (type of speech sequence) and speech rate condition (low vs. moderate) as independent variables and coherence index as the dependent variable yielded a significant effect of stimulus condition [$F(1,32) = 620.79$, $MS_e = 0.65$, $p < .001$], a significant effect of speech rate condition [$F(1,32) = 159.76$, $MS_e = 0.17$, $p < .001$], and a reliable interaction between the two conditions [$F(1,32) = 186.32$, $MS_e = 0.19$, $p < .001$].

The pattern for /pəs/ and /spə/ is different. In the increased rate condition (see Figure 3A), the indexes of these sequences got gradually closer starting at a speech rate of around 2 cps. In the fixed rate condition (see Figure 3B), for the low speech rate condition the index was 0.70–0.80 for /pəs/ ($M = 0.76$, $SD = 0.01$) and 0.20–0.30 for /spə/ ($M = 0.25$, $SD = 0.02$). When cycle duration decreased and rate increased (around 2 cps), we observed a nearly identical index of 0.45–0.60 for both speech sequences ($M = 0.58$, $SD = 0.02$ for /pəs/; $M = 0.47$, $SD = 0.01$ for /spə/). Hence, regardless of repetition rate, [s] and [p] always remain separated in cycles for /spə/ and /pəs/. Furthermore, for a rhythm of around 2 cps, /pəs/ and /spə/ sequences are barely distinguishable in terms of acoustic events. A two-way ANOVA yielded a significant effect of stimulus condition [$F(1,32) = 3{,}046.85$, $MS_e = 0.85$, $p < .001$], a significant effect of speech rate condition [$F(1,32) = 14.97$, $MS_e = 0.00$, $p < .001$], and a reliable interaction between the two conditions [$F(1,32) = 1{,}328.84$, $MS_e = 0.37$, $p < .001$].

Discussion

In summary, the data for /psə/ and /səp/ sequences fit well with the claim that more temporally clustered stimuli should be more stable and play the role of attractors during a repetitive speech task. Analyses of timing relations between articulatory and acoustic events during a repetitive speech production task showed that, at a low rate, the onset events appear to be more synchronous for /psə/ than for /səp/. In fact, for the /psə/ sequence, consonants in the syllable onset are temporally clustered and hence may be thought of as a tightly synchronized unit (for further evidence on the clustering of CV structures, see de Jong, 2001; MacKay, 1982). At a rate above 2 cps, /psə/ should play the role of an attractor for cycling /səp/, the release of the final [p] becoming synchronous with the tongue-driven onset of the initial [s] in the next utterance. No matter what the repetition rate, [s] and [p] onset events never cohere in cycles for /spə/ and /pəs/, since they are separated by lip closure. However, /pəs/ and /spə/ sequences appear to be barely distinguishable in terms of articulatory-acoustic events at a rhythm above 2 cps. Finally, variations in timing relationships between articulatory-acoustic events also occurred without any rate acceleration. This suggests that a sufficient speaking rate is more crucial than a speeding-up paradigm in the study of articulatory phasing per se.

EXPERIMENT 2

The purpose of Experiment 2 was to test the existence of preferential transformations by contrasting more or less "temporally clustered" syllable stimuli during an overt production variant of the verbal transformation paradigm. As Reisberg et al.'s (1989) study showed, the efficiency of verbal transformations depends on the degree of enactment. Therefore, we first adopted the simplest and most effective condition for eliciting transformations: overt repetition.

We assumed that the repetitive production of the speech stimuli might result in shifted sequences distributed in two groups (i.e., the /psə/, /səp/, and /əps/ sequences in Group 1, and the /spə/, /pəs/ and /əsp/ sequences in Group 2). In other words, we predicted that, when presented with a given sequence, participants would not produce all possible permutations but would naturally be brought to "mentally read" the result of their repetition according to a *shifting* parsing within each group. For example, repetition of the /psə/ sequence would lead to a shifting segmentation according to which a perceptual boundary may be placed after [ə], [p], or [s].

Our claim is that more temporally clustered stimuli should be more stable and play the role of attractors in verbal transformations. The speech production data in Experiment 1 indicate that the /səp/ sequence in Group 1 should be transformed into /psə/, whereas /spə/ and /pəs/ in Group 2 should be equally stable. Furthermore, the assumption concerning VCC sequences in both groups is that they should be least stable, since the consonantal gesture(s) in the coda intervenes long after vowel initiation. This results in the pattern of predictions shown in Table 1, in which it is assumed that the lower the articulatory-acoustic coherence

Table 1
Sequence Classification According to Degree of Articulatory Cohesion Between the Consonantal and Vocalic Gestures, and Expected Transformations During the Repetition Process

Sequence	Degree of Articulatory Cohesion	Prediction
	Group 1	
/psə/	Strong; onset cluster and vowel synchronized, consonants in the onset synchronized	/psə/
/səp/	Average; onset consonant and vowel synchronized, coda desynchronized	/psə/
/əps/	Weak; vowel and consonant cluster desynchronized, consonants in the coda synchronized	/psə/
	Group 2	
/pəs/	Average; onset consonant and vowel synchronized, coda desynchronized	/pəs/ or /spə/
/spə/	Average; onset cluster and vowel synchronized, consonants in the onset desynchronized	/pəs/ or /spə/
/əsp/	Very weak; vowel and consonantal cluster desynchronized, consonants in the coda desynchronized	/pəs/ or /spə/

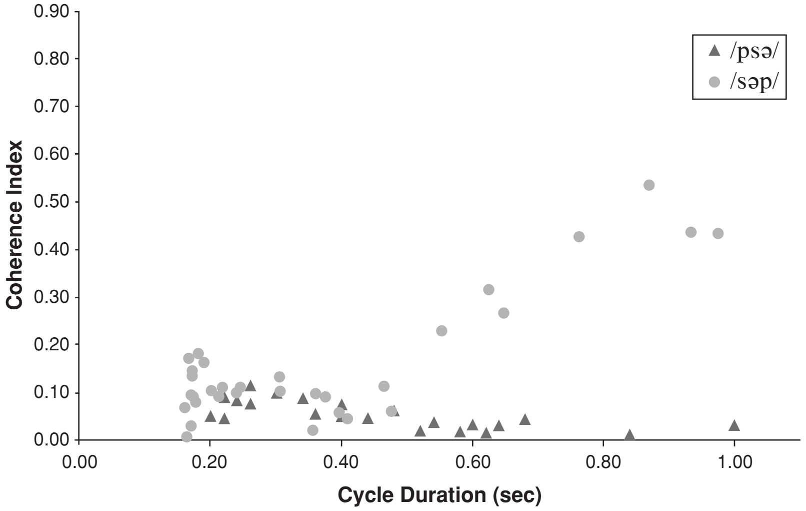

Figure 2A. Variations of the coherence index of articulatory-acoustic events (defined as the time separating the onsets of [p] and [s] gestures divided by the total cycle duration) as a function of cycle duration (in seconds) for the /psə/ and /səp/ speech sequences in the increased rate condition.

(i.e., the less temporally clustered the sequence), the less stable the sequence will be, and the more likely its transformation into a more coherent one.

Method

Participants. Fifty-six students at Grenoble University participated in this experiment. All were native French speakers without hearing or speaking disorders, and all were naive as to the purpose of the experiment.

Apparatus. For follow-up analyses, the experiment was entirely recorded onto a portable audio recorder. The recording was then digitized as individual sound files to the hard disk of a PC computer at a sampling rate of 22.05 kHz with 16-bit quantization.

Procedure. The participants were tested individually. The experiment began with a lengthy briefing, during which the participants were introduced to the verbal transformation task. The participants listened to the experimenter repeat the word *life* at a rate of 2 repetitions/sec and were asked to listen carefully for any changes in the repeated utterance. The experimenter then asked the participants if they had perceived another sequence and, if they had not, explained the possibility of hearing the word *fly*. The purpose of this "bootstrap" example, presented in English rather than in French, was to display the verbal transformation effect on a material that all the participants understood, yet letting them experience the phenomenon later on with their own production and in their own language. The participants were then told that they would repeat a given sequence aloud into a microphone placed in front of them, at a rate of about 2 cps, with no gap between repetitions. If they heard a transformation, they were to stop and report it. If they did not hear any transformation, they were to say nothing; the experimenter would stop them after 30 sec. Finally, it was indicated that changes could be subtle or very noticeable and could correspond to a word or to a nonsense utterance. Furthermore, the participants were assured that there were no correct or incorrect responses.

In the test session, the six sequences—/psə/, /səp/, /əps/, /pəs/, /spə/, and /əsp/—were orally presented by the experimenter in one of six counterbalanced orders (based on the sequence alternation from one group to the other, excluding the successive presentation of two sequences with similar onsets). If the participant made pronunciation mistakes, paused (thereby breaking rhythm), slowed down, or stopped before the 30-sec time period ended without reporting any change, the experimenter asked him or her to resume the ongoing activity. Lengthy breaks were offered between trials.

Results

The transformation frequencies observed for the six sequences and averaged over participants are shown in Table 2A.

The observed transformations from one group to the other were extremely rare (on average 3% for the two groups). The last column of Table 2A (Misc) represents the percentages of unpredicted transformations. For 4% of all responses, they involved lexical transformations (e.g., /sø/ ["these"], /pø/ ["few"], or /saspø/ ["it may be"]), and for 6% of the responses they corresponded to a nonsense word with a larger-than-expected phonological structure (e.g., /psəp/ for /səp/). The total proportion of such transformations, although not trivial, remains small (on average 11%) considering that the participants were not informed about expected transformations. For each stimulus, most responses occurred within its related group, which confirms the shifting parsing hypothesis.

Stabilities and preferential transformations. Global statistical analyses yielded no significant effect of the six counterbalanced stimulus orders [$\chi^2(5) = 4.39$, $p > .05$ unless otherwise stated] and a significant global stimulus effect [$\chi^2(5) = 32.10$, $p < .0005$].

According to the observed shifting parsing process, further statistical comparisons were carried out on each group separately. We tested discrepancies between sequences on two distinct measures. The stability index was calculated by summing the number of times a given sequence was not transformed. The attractivity index, used to evaluate the sequence's capacity to attract, or "capture," the other sequences during the repetition process, was calculated by summing the number of times a given sequence was selected

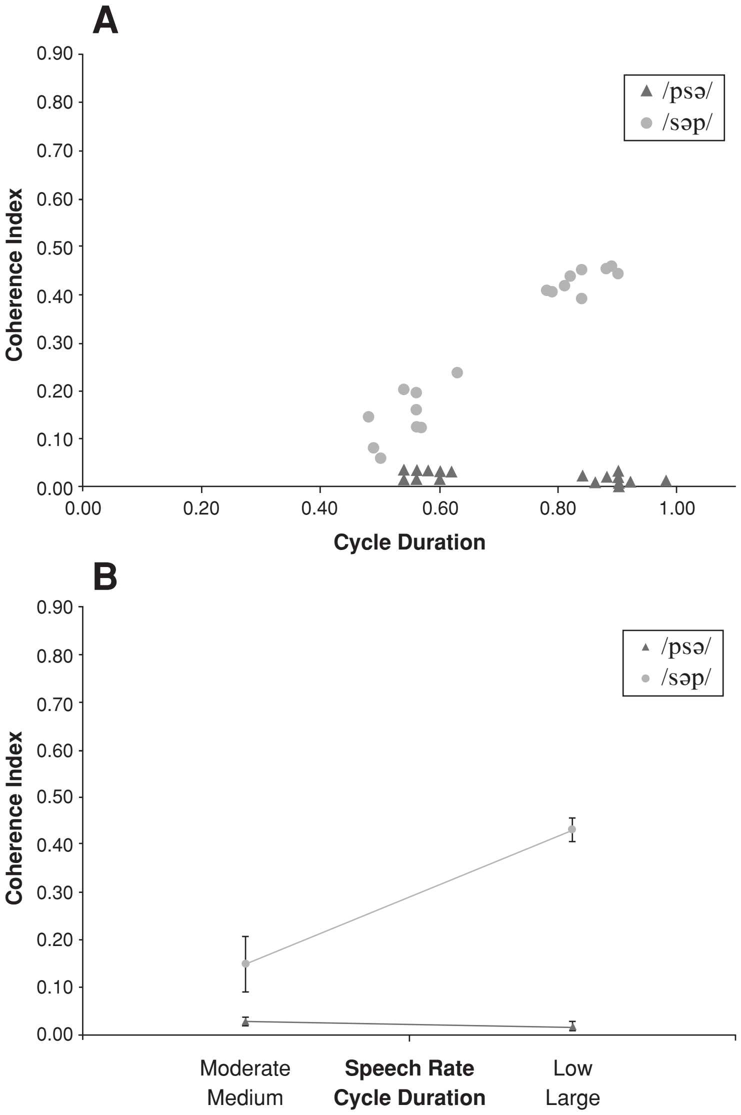

Figure 2B. (A) Variations of the coherence index of articulatory-acoustic events (defined as the time separating the onsets of [p] and [s] gestures divided by the total cycle duration) as a function of cycle duration (in seconds) for the /psə/ and /səp/ speech sequences in the fixed rate condition. (B) Variations of the mean coherence index (with *SD*s) according to moderate (around 2 cps) and low (around 1 cps) speaking rates.

as a transformation within a group, weighted by the number of times it could have been selected as a transformation.

Within Group 1 (see Table 2B), the global comparison of the observed stability per sequence yielded a significant effect [$\chi^2(2) = 24.76$, $p < .0001$]. Analyses of stability across sequences, with a Bonferroni correction (applied in all the following individual comparisons), showed reliable discrepancies between /psə/ and /səp/ [$\chi^2(1) = 22.39$] and between /səp/ and /əps/ [$\chi^2(1) = 12.93$]. The global comparison within Group 1 of the observed attractivity per sequence was reliable [$\chi^2(2) = 64.12$, $p < .0001$], with significant discrepancies between /psə/ and /səp/ [$\chi^2(1) = 18.07$], between /psə/ and /əps/ [$\chi^2(1) = 61.38$], and between /səp/ and /əps/ [$\chi^2(1) = 14.21$]. In summary, within Group 1 /psə/ and /əps/ showed stronger stability than /səp/, whereas /psə/ was the most attractive sequence and /səp/ showed stronger attractivity than /əps/. Within Group 2 (see Table 2B), the global comparison of the observed stability per sequence yielded no significant effect [$\chi^2(2) = 3.51$]. However, the global comparison of the observed attractivity per sequence was reliable [$\chi^2(2) = 59.57$, $p < .0001$], with significant differences between /pəs/ and /əsp/ [$\chi^2(1) = 59.49$] and between /spə/ and /əsp/ [$\chi^2(1) = 39.78$]. In summary, within Group 2 we

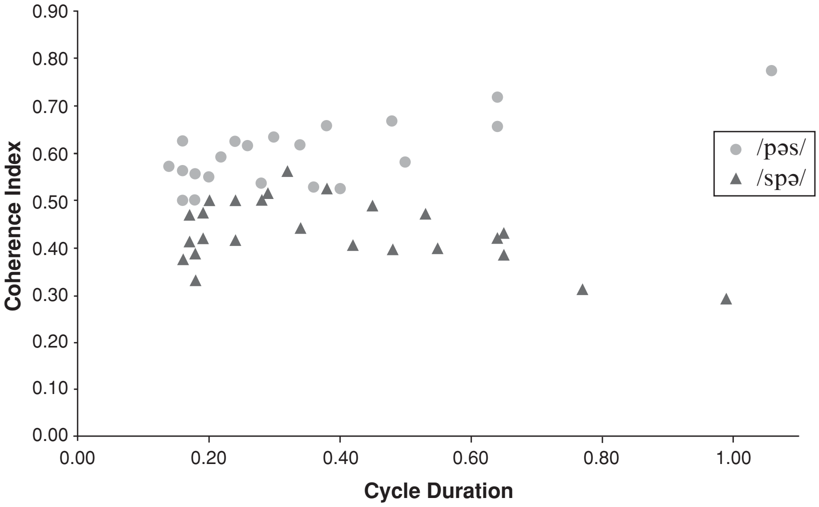

Figure 3A. Variations of the coherence index of articulatory-acoustic events (defined as the time separating the onsets of [p] and [s] gestures divided by the total cycle duration) as a function of cycle duration (in seconds) for the /pəs/ and /spə/ speech sequences in the increased rate condition.

observed no reliable discrepancies between sequences in terms of stability and stronger attractivity for /pəs/ and /spə/ than for /əsp/.

Test of a glottal onset effect. Although they were expected to be very unstable, the /əps/ and /əsp/ syllables with empty onsets appeared rather stable (although not at all attractive) in this experiment. This could be explained by the presence of a glottal stop often produced by the participants at syllable onset. This glottal stop can be considered an additional consonant in the syllabic structure, transforming the VCC syllables /əps/ and /əsp/ into the CVCC syllables /ʔəps/ and /ʔəsp/, respectively. This glottal onset might prevent the fast and connected repetition of items and hence block articulatory synchronization (de Jong, 2001). Consistent with this hypothesis was the longer mean duration rate observed for these two sequences in comparison with the others (we also observed a longer mean duration rate for /əsp/ than for /əps/). Two trained phoneticians conducted a post hoc phonetic analysis consisting in determining the presence or absence of a glottal stop in the final portion of each of the /əps/ and /əsp/ recordings, with no indication of the observed stability/instability of the sequence. This analysis confirmed that 71% and 63% of the participants produced a glottal stop at the end of the repetition process for the /əps/ and /əps/ sequences, respectively. Further comparisons between transformed and untransformed sequences showed that the glottal onset was produced for 86% of the untransformed /əps/ sequences and 45% of the transformed /əps/ sequences, and for 82% of the untransformed /əsp/ sequences and 40% of the transformed /əsp/ sequences. Reanalyses of results excluding participants who produced a glottal onset showed that the /əps/ sequence remained untransformed for 36% of the participants, whereas it was transformed toward /psə/ for 50% of the participants and underwent an unexpected transformation for the remaining 14%. Likewise, the /əsp/ sequence remained untransformed for 27% of the participants, whereas it was transformed toward /pəs/, /spə/, and /psə/ for 41%, 9%, 9%, of the participants, respectively, and underwent an unexpected transformation for the remaining 14%. Taken together, these results confirm that the observed stabilities

Table 2A
Proportions of Transformations Observed in Experiment 2

	Transformation to						
Sequence	/psə/	/səp/	/əps/	/pəs/	/spə/	/əsp/	Misc
/psə/	**.75**	.18			.02		.05
/səp/	.50	**.30**	.02				.18
/əps/	.29		**.64**				.07
/pəs/				**.41**	.43		.16
/spə/	.04			.46	**.39**		.11
/əsp/	.07		.04	.20	.05	**.55**	.09

Note—Entries in boldface represent proportions of stable utterances; all other entries correspond to transformed sequences (*N* = 56). Misc, miscellaneous transformations.

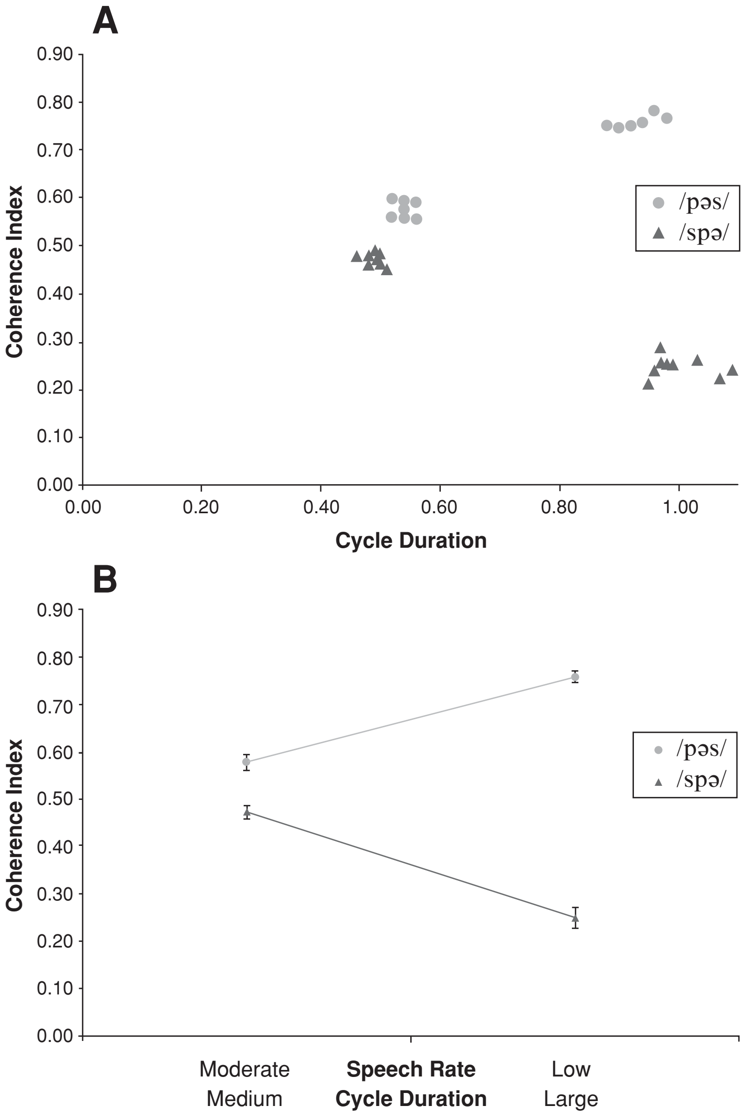

Figure 3B. (A) Variations of the coherence index of articulatory-acoustic events (defined as the time separating the onsets of [p] and [s] gestures divided by the total cycle duration) as a function of cycle duration (in seconds) for the /pəs/ and /spə/ speech sequences in the fixed rate condition. (B) Variations of the mean coherence index (with *SD*s) according to moderate (around 2 cps) and low (around 1 cps) speaking rates.

Table 2B
Degrees of Stability and Weighted Attractivity per Sequence Within Group 1 and Group 2 in Experiment 2

Group 1			Group 2		
Sequence	Stability	Attractivity	Sequence	Stability	Attractivity
/psə/	.75	.75	/**pəs**/	.41	.63
/səp/	.30	.29	**/spə/**	.39	.47
/əps/	.64	.02	/əsp/	.55	.00

Note—The sequences predicted as the most stable and attractive are represented in boldface.

of /əps/ and /əsp/ were due largely to a glottal onset effect and explain why their respective degrees of stability and attractivity differ so much (i.e., high stability vs. low attractivity).

Discussion

Altogether, the results fit reasonably well with the expectations summarized in Table 1. First, shifting parsing seems to be the rule. Second, the hierarchy of attractivities within each group mirrors the proposed hierarchy of articulatory cohesion (i.e., /psə/ > /səp/ > /əps/ in Group 1, and /pəs/ = /spə/ > /əsp/ in Group 2). Stability patterns are less in agreement with our predictions, but the glottal onset effect is the factor mainly responsible for this, and, once it is taken into account, both the stability patterns and the attractivity patterns correspond to the predictions.

Considering the unexpected transformations, a number of previously emphasized contents of transformations were observed in our experiments, including substitution of a phoneme by a phonetically close one (e.g., /səb/ for /səp/ and /əbs/ for /əps/; Warren, 1961; Warren & Meyers, 1987), auditory streaming (e.g., /əp/ for /əps/; Pitt & Shoaf, 2001, 2002), and lexical transformations (e.g., /sø/ ["these"], /pø/ ["few"], or /saspø/ ["it may be"] for /spə/, and /pus/ ["thumb"] for /pəs/; Kaminska et al., 2000; Shoaf & Pitt, 2002; Warren, 1961). However, in the present experiment the majority of observed transformations cannot be related to such lexical or phonological transformation processes. By using a production variant of the classical verbal transformation paradigm, Reisberg et al. (1989) first demonstrated that speech production constraints, specifically speech enactment, could also intervene in the transformation process. The observed asymmetries in the reported transformations reinforced this position by showing that the perceptual stability and attractivity of an uttered sequence might also depend on specific articulatory constraints—that is, on the temporal clustering between intrasyllabic articulatory gestures.

EXPERIMENT 3

According to Reisberg et al.'s (1989) results, the decrease in enactment from overt to covert speech should result in a decrease in the number of transformations, the interpretation being that elaboration of verbal auditory images depends on the degree of subvocalized enactment. The question, however, is whether or not the specific articulatory constraints related to variations in temporal clustering of /p/ and /s/ are preserved in covert speech and produce verbal transformation asymmetries as they do in the overt mode. The purpose of the following experiment was to further examine this hypothesis by testing the persistence of verbal transformation asymmetries in a covert repetition mode.

Method

Phonetic materials. The stimuli used in this experiment were the same as those in Experiment 2. The assumptions about the shifting parsing and the articulatory cohesion hierarchy were therefore the same.

Participants and Task. Twenty-nine new participants were recruited from Grenoble University. All were native speakers of French, had no hearing or speech disorders, and were naive as to the purpose of the experiment.

Procedure. The participants were tested individually. As in Experiment 2, they were first introduced to the verbal transformation task. Then, they were told that they would mentally repeat a given sequence while keeping their mouths closed, at a rate of about 2 cps, with no gaps between repetitions. They were asked to "mentally listen" for any changes in the repeated utterance. If they found a transformation, they were to stop and report it. If they did not hear any transformations, they were to say nothing. It was indicated that changes could be subtle or very noticeable and could correspond to a word or to a nonsense utterance. As previously, the participants were assured that there were no correct or incorrect responses.

In the test session, the six sequences—/psə/, /səp/, /əps/, /spə/, /pəs/, and /əsp/—were orally presented by the experimenter in one of six counterbalanced orders. During the covert repetition, some of the participants happened to move their lips without phonation; in this case, the experimenter asked them to keep their mouths closed and to start again without moving the lips. Lengthy breaks were offered between trials.

Results

The results show that the percentage of the observed transformations from one group to the other were on average 11%, whereas the percentage of unpredicted transformations was on average 13% (see Table 3A). Most responses (76% on average) occurred within each group, according to a shifting parsing process. This percentage, however, was 10% lower than the corresponding percentage in Experiment 2.

Stabilities and preferential transformations. A global statistical analysis of the results displayed no significant effect of stimulus order [$\chi^2(5) = 1.74$] and a significant global stimulus effect [$\chi^2(5) = 12.98$, $p < .05$] (see Table 3B). The statistical comparisons of stability and attractivity patterns showed the following results. Within Group 1, the global comparison of the observed stability per sequence was not significant [$\chi^2(2) = 5.90$]. The global comparison of the observed attractivity per sequence was reliable [$\chi^2(2) = 43.39$, $p < .0001$], with significant differences between /psə/ and /səp/ [$\chi^2(1) = 17.38$] and between /psə/ and /əps/ [$\chi^2(1) = 25.83$]. Within Group 2, the global comparison of the observed stability per sequence was significant [$\chi^2(2) = 6.99$, $p < .05$], with /pəs/ reliably differing from /əsp/ [$\chi^2(1) = 6.90$]. The global comparison of the observed attractivity per sequence was reliable [$\chi^2(2) = 10.81$, $p < .005$], with significant differences between /pəs/ and /əsp/ [$\chi^2(1) = 6.55$] and between /spə/ and /əsp/ [$\chi^2(1) = 8.10$]. In summary, within Group 1 we observed no reliable discrepancies between sequences in terms of stability and stronger attractivity for /psə/ than for /səp/ and /əps/. Within Group 2, we observed stronger stability for /pəs/ than for /əps/ and stronger attractivity for /pəs/ and /spə/ than for /əsp/.

Discussion

In terms of the degree of attractivity per sequence, the covert repetition mode explored in Experiment 3 produced patterns of verbal transformation asymmetries similar to those of the overt mode in Experiment 2 [$\chi^2(5) = 5.75$,

Table 3A
Proportions of Transformations Observed in Experiment 3

	Transformation to						
Sequence	/psə/	/səp/	/əps/	/pəs/	/spə/	/əsp/	Misc
/psə/	**.63**	.10		.07	.10		.10
/səp/	.31	**.62**					.07
/əps/	.42	.07	**.34**				.17
/pəs/	.07			**.69**	.21		.03
/spə/	.10	.03		.11	**.48**		.28
/əsp/	.21	.03	.07	.17	.07	**.35**	.10

Note—Entries in boldface represent proportions of stable utterances; all other entries correspond to transformed sequences ($N = 29$). Misc, miscellaneous transformations.

n.s.]. Indeed, there is a convergence between the results of the two experiments, showing a nearly identical hierarchy (/psə/ > /səp/ ≥ /əps/) within Group 1 and the same hierarchy (/pəs/ = /spə/ > /əsp/) within Group 2. Furthermore, these attractivity patterns largely correspond to our predictions.

However, the patterns of stabilities differ significantly between the two experiments [$\chi^2(5) = 13.20$, $p < .05$]. These differences come first from the higher stability of the /əps/ and /əsp/ sequences in Experiment 2 (in which they displayed an unexpectedly high stability) in comparison with that of the sequences in Experiment 3 (in which they presented a decreased stability more in line with our predictions). As was described previously, we explain the unpredicted stability of the /əps/ and /əsp/ sequences in Experiment 2 by the presence of a glottal stop often produced at syllable onset, hence preventing fast and connected repetitions of items and then blocking articulatory synchronization (de Jong, 2001). Although the stability discrepancies of the sequences observed between the two experiments could be due to differences between participants, we cannot rule out the possibility that the glottal effect might have decreased in Experiment 3, showing dependence on the degree of external articulation. Under covert conditions, some articulatory control constraints (e.g., the temporal clustering between intrasyllabic articulatory gestures) would remain active in the building up of auditory images, whereas the glottal onset effect would disappear. Considering the latter hypothesis (i.e., that of dependence on the degree of external articulation), this might constitute an interesting phenomenon to be further examined in the study of the functional equivalence between overt and covert speech. This would require additional tests, which are not within the scope of the present work.

Another source of stability discrepancies between the two experiments is the smaller number of transformations in the covert repetition condition in Experiment 3. Indeed, if the /əps/ and /əsp/ sequences are discarded because of possible discrepancies among participants or glottal onset effect size, a higher stability of the sequences is observed in the covert than in the overt repetition condition. This is quite similar to what was observed by Reisberg et al. (1989): When contrasting all the displayed transformations, they found an average 38% decrease (according to the transformation results of the monosyllabic word *stress* to *dress*; see p. 635), whereas we found an average 26% decrease. Interestingly, the number of shifting parsing violations is quite a bit larger in the covert repetition condition. Particularly, /psə/ plays the role of an attractor for an important proportion of sequences in Group 2. This suggests that the segmental order of articulatory events of the repeated utterances is more difficult to maintain during a covert repetition, possibly owing to fewer auditory and proprioceptive inputs in the control of the uttered sequence (Murray, 1965).

In summary, when contrasting the results of the two experiments we observed varying levels of stability for the /əps/ and /əsp/ sequences between experiments and stronger stability of the other sequences in the covert mode. It is therefore remarkable that, in spite of these differences, the attractivity pattern, which largely corresponds with our articulatory cohesion predictions, is maintained from the overt to the covert repetition condition. This suggests that the temporal clustering of articulatory-acoustic events can also take place internally, with neither an articulatory nor an auditory external stimulus. This result appears to be consistent with previous behavioral studies that showed some degree of functional equivalence between overt

Table 3B
Degrees of Stability and Weighted Attractivity per Sequence Within Group 1 and Group 2 in Experiment 3

Group 1			Group 2		
Sequence	Stability	Attractivity	Sequence	Stability	Attractivity
/psə/	.63	.70	**/pəs/**	.69	.24
/səp/	.62	.17	**/spə/**	.48	.29
/əps/	.34	.00	/əsp/	.35	.00

Note—The sequences predicted as the most stable and attractive are represented in boldface.

and covert speech (e.g., Landauer, 1962; MacKay, 1982; Postma & Noordanus, 1996) and, more generally, with the burgeoning domain of "motor cognition" (providing strong empirical evidence for a functional coupling between a simulated action and an executed one; for a review, see Jeannerod, 1994). In line with these studies, the persistence of the asymmetric bias from an overt to a covert repetition procedure suggests that constraints from the speech production system seem able to penetrate verbal imagery and participate in the mental analysis and interpretation of phonological forms during the emergence of verbal transformations.

EXPERIMENT 4

Speech is a matter of gestures and sounds resulting in a set of more or less clustered events, which are, to a certain extent, both audibly and articulatorily interpretable. Considering that verbal transformations may involve both perceptual mechanisms (e.g., auditory streaming) and motor constraints, the question is whether the participants' behavior in the verbal transformation tasks of Experiments 2 and 3 was actually driven mainly by articulatory coordination or by auditory templates as well. The following experiment was designed to test a possible perceptual alternative to the articulatory cohesion hypothesis by examining the persistence of verbal transformation asymmetries using a purely perceptual paradigm.

Method

Participants. Twenty-four students at Grenoble University participated in this experiment. All were native French speakers, had no hearing or speech disorders, and were naive as to the purpose of the experiment.

Phonetic material. The /psə/, /səp/, /əps/, /spə/, /pəs/, and /əsp/ sequences were individually recorded into a digital audiotape by a trained phonetician (J.-L.S.) at a fixed speech rate of 2 cps. The items were digitized to the hard disk of a PC computer at a sampling rate of 48 kHz with 16-bit quantization. Each sequence was then reduplicated 100 times in an individual sound file with a 500-msec stimulus onset asynchrony.

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 tested individually in a quiet room. The experiment began with a lengthy briefing, during which the participants were introduced to the verbal transformation task. Then, they were told that they would hear an utterance being played repeatedly and were asked first to report what they heard and then to listen carefully for any changes in the repeated utterance. If the stimulus changed to another form, they were asked to report the transformation. It was indicated that the change could be subtle or very noticeable and could correspond to a word or to a pseudoword. Finally, the participants were assured that there were no correct or incorrect responses and were told that if they did not hear a transformation, they were to say nothing. In the test session, the six stimuli were presented in one of 12 counterbalanced orders. Lengthy breaks were offered between trials.

Results

The data were analyzed by labeling the participants' reports in the response sound files. Overall, 80.6% of the first reported forms matched the veridical repeated sequence. Furthermore, when the participant did not first report the correct sequence (e.g., /psəp/ instead of /psə/), the following transformation corresponded to the repeated utterance in 61% of the cases.

In the analyses presented below, only the transformations following a correct initial identification of the repeated utterances were taken into account. If the participant did not report a transformation during the 30-sec period following the first reported form, the sequence was considered to be stable.

On average, only 8% of the sequences remained stable and 3% were transformed according to a shifting parsing procedure (see Table 4A).[1] Observed transformations from one group to the other were nonexistent. As for the unpredicted transformations (on average 89%; see Table 4B), 22% of the overall responses involved lexical transformations (e.g., /sypɛr/ ["super"] for /spə/), 29% corresponded to a phonetic deviation (e.g., /sop/ for /səp/, /tse/ for /psə/), and 17% involved auditory streaming mechanisms (e.g., /əp/ for /əps/, /əs/ for /əsp/).

Stabilities and preferential transformations. Global statistical analyses yielded no significant effect of stimulus order [$\chi^2(5) = 1.05$] and a significant global stimulus effect [$\chi^2(5) = 15.02$, $p < .05$]. The statistical comparisons of stability and attractivity patterns showed the following results. Within Group 1, the global comparison of the observed stability per sequence was not significant [$\chi^2(2) = 4.14$]. Furthermore, none of the sequences was transformed according to a shifting parsing procedure. Within Group 2, the global comparison of the observed stability per sequence was significant [$\chi^2(2) = 6.38$, $p < .05$], a result largely due to the greater stability of the /pəs/ syllable, although none of the individual comparisons was significant. However, the global comparison of the observed attractivity per sequence was not significant [$\chi^2(2) = 2.70$, $p < .005$].

Miscellaneous transformations. Because of the great number of transformations outside the two groups, we performed three distinct analyses across sequences related to their respective numbers of lexical transformations, phonetic deviations, and transformations attributable to auditory streaming. The results showed no discrepancies across sequences for lexical transformations and phonetic deviations [$\chi^2(5) = 9.32$ and $\chi^2(5) = 8.00$, respectively] and a significant effect across sequences for the transformations involving an auditory streaming mechanism [$\chi^2(5) = 41.80$, $p < .001$], with the /əps/ and /əsp/ sequences showing a large proportion of transformations (62% and 32%, respectively).

Discussion

Taken together, the stability and attractivity patterns of the sequences suggest that distinct constraints act on the elaboration of verbal representations during a perception procedure and a self-repetition procedure.

All the sequences of Experiment 4 showed a lower degree of stability (on average 8%) in comparison with the sequences of Experiments 2 and 3 (on average 51% and

Table 4A
Proportions of Transformations Observed in Experiment 4

	Transformation to						
Sequence	/psə/	/səp/	/əps/	/pəs/	/spə/	/əsp/	Misc
/psə/							1.00
/səp/		**.10**					.90
/əps/							1.00
/pəs/				**.29**	.06		.65
/spə/				.05	**.05**		.90
/əsp/				.11		**.05**	.84

Note—Entries in boldface represent proportions of stable utterances; all other entries correspond to transformed sequences (*N* = 24). Misc, miscellaneous transformations.

52%, respectively), a result in line with previous studies showing that perceptual stability constraints acting on verbal transformations are not fully equivalent for a perception procedure and a production variant. Lackner (1974) first reported that self-produced repetition of monosyllabic nonsense words resulted in far fewer speaker-perceived transformations than when the speaker's productions were played back to them. MacKay et al. (1993) further replicated this finding, showing that participants experienced more transformations when they listened to a repeating word than when they overtly repeated the word. Altogether, these results suggest an increase of perceptual stability during an overt self-repetition mode. To explain this effect, Lackner proposed that perceptual mechanisms during self-repetition are alerted by a corollary discharge, or efference copy, that accompanies the motor execution of a speech sequence (for an explanation of the concept of efference copy to the online monitoring of one's own voice, see Frith, 1992). In the perception condition, in the absence of such a generated signal informing on the forthcoming speech sound, the ability to maintain a stable perceptual representation should decrease.

Another important result of Experiment 4 is the lower degree of attractivity of the target sequences (on average 4% vs. 38% and 36% in Experiments 2 and 3, respectively). Indeed, contrary to Experiments 2 and 3, in which transformations occurred principally within the same group of sequences, 89% of the present transformations corresponded to a lexical transformation, to a phonetic deviation, or to auditory streaming processes—three well-established transformation mechanisms occurring during a perceptual procedure of the verbal transformation paradigm (see, e.g., Kaminska et al., 2000; Pitt & Shoaf, 2001, 2002; Shoaf & Pitt, 2002; Warren, 1961; Warren & Meyers, 1987). Hence, the weak number of transformations relying on a shifting parsing process and the completely different pattern of transformations in comparison with those of Experiments 2 and 3 clearly rule out a purely auditory interpretation of the verbal transformation asymmetries observed in the production conditions. This reinforces the articulatory cohesion hypothesis as the most likely and coherent explanation of the asymmetries displayed in Experiments 2 and 3.

GENERAL DISCUSSION

Investigating the causes of auditory illusions, which are viewed as windows into the linguistic processes that operate during veridical perception, provides a useful framework that can enhance our understanding of the language system by revealing otherwise hidden mechanisms. From this point of view, the verbal transformation effect appears to be well suited as a tool for examining how speech production and perception constraints may intervene in the emergence and analysis of verbal representations.

The goal of the present study was to explore whether or not specific speech production constraints—specifically, temporal coherence between articulatory gestures—may intervene in the verbal transformation effect. Having selected appropriate phonetic material and displayed variations in the temporal clustering of articulatory-acoustic events within this material (Experiment 1), we showed that these variations could lead to perceptual asymmetries in verbal transformations during an overt repetition procedure, thus suggesting that the perceptual stability and attractivity of an uttered sequence might also depend on articulatory cohesion constraints (Experiment 2). The fact that the same transformation trends were found during a

Table 4B
Proportions of the Miscellaneous Transformations in Experiment 4

Sequence	Lexical Transformation	Auditory Streaming	Phonetic Deviation	Other Transformations
/psə/	.26		.53	.21
/səp/	.35		.45	.10
/əps/		.62	.19	.19
/pəs/	.12	.06	.24	.23
/spə/	.35		.20	.35
/əsp/	.21	.32	.16	.15

covert repetition mode confirms some functional coupling between overt and covert speech and suggests that specific articulatory control constraints originating from the motor system may participate in the emergence of verbal representations in the human brain, even without any articulatory or auditory external signal (Experiment 3). Finally, the absence of such asymmetric bias in the transformations observed during the perceptual condition rules out a purely auditory interpretation of the verbal transformation asymmetries observed in the production conditions and shows that distinct constraints may act on the elaboration of verbal representations during a perception procedure and a self-repetition procedure (Experiment 4).

Discarding Purely Lexical and Purely Phonological Interpretations

One important source of influence in the verbal transformation paradigm comes from a set of general or language-specific linguistic constraints. With regard to our articulatory cohesion hypothesis, it is therefore important to check for a possible alternative explanation of the asymmetric bias observed in Experiments 2 and 3.

First, verbal transformations should vary as a function of distinct lexical factors related to the repeated stimulus: lexical status, neighborhood density, and lexical type frequency (see Table 5). Given that none of our speech stimuli occurs in the French lexicon, we consider that lexical status cannot account for the stability and attractivity discrepancies between the speech sequences. Apart from the lexical status of the stimulus, it has been argued that a large number of neighbors (i.e., the number of lexical entries that are phonologically similar to the repeated stimulus—e.g., /pus/ [“thumb”] for /pəs/) should increase the number of possible primed lexical candidates, resulting in a greater number and a wider range of transformations and thus entailing lower stability of the stimulus (MacKay et al., 1993; Yin & MacKay, 1992). However, this neighborhood density effect does not appear to be in accordance with our results. Indeed, the weak stability for the /əps/ and /əsp/ sequences, observed once the glottal stop effect was taken into account in Experiment 2, cannot be explained by the respective neighborhood density values; these sequences show a smaller (or at least a similar) number of neighbors in comparison with the other sequences. Hence, neighborhood density is not likely to provide an explanation for the present asymmetries. A concurrent interpretation could come from the lexical frequency of the speech sequences in the participant’s lexicon. Indeed, this lexical frequency could bias transformations toward a sequence with a greater number of lexical entries (MacKay et al., 1993). Considering the results of Experiments 2 and 3, it appears that in Group 1 this lexical frequency effect is not in accordance with the stronger attractivity of /psə/, /səp/ being the most favored sequence in terms of lexical entries. However, in Group 2 there seems to be a slight advantage of /pəs/ over /spə/ in terms of attractivity, although it was always below the significance threshold. This trend, not included in our predictions, could be due to the fact that /pəs/ displays a greater number of lexical entries than /spə/, thus underscoring a potential effect of lexical frequency during the task. This position is reinforced by the results of Experiment 4, which show a greater (though not significantly so) attractivity of /pəs/ in comparison with the other sequences.

Another hypothesis based on phonetic or phonotactic regularities could be that preferential transformations derive from syllabic structure constraints. The study of typological trends in syllable structure (see Table 6 for an overview of syllabic structure frequencies extracted from a sample of geographically and genetically dispersed languages of the UCLA Lexical and Syllabic Inventory Database [ULSID]; Maddieson, 1984; Vallée, Boë, Maddieson, & Rousset, 2000) shows that VCC and CCV syllables are very infrequent in phonological inventories and that CVC is the most frequent syllable after CV. The results of Experiments 2 and 3 indeed seem to satisfy the largely shared constraint of avoidance of syllables with no consonantal onset: CVC and CCV syllables do not switch toward VCC, just as CV syllables do not switch toward VC syllables (de Jong, 2001; de Jong et al., 2002; Stetson,

Table 5
Lexical Type Frequency (LTF) and Neighborhood Density for Each of the Measures, and the Sum (STF) and Range of Associated Token Frequencies (per Million Occurrences)

Sequence	LTF	STF	Range	ND	STF	Range
/psV/	114	81	0–13	31	11,357	0–5,031
/sVp/	371	812	0–92	59	20,310	0–5,031
/Vps/	131	170	0–118	34	3,206	0–2,096
/spV/	674	1,248	0–119	19	10,160	0–5,031
/pVs/	1,379	10,676	0–6,372	118	54,538	0–16,011
/Vsp/	598	1,539	0–229	14	9,320	0–8,743

Note—LTF is defined as the number of lexical entries incorporating a monosyllabic structure identical to that of the stimulus at any position in a word. Neighborhood density (ND) is defined as the number of phonologically similar words that differ from the stimulus by a single substitution, insertion, or deletion at any position in the target word (Luce, Pisoni, & Goldinger, 1990). All lexical analyses were extracted from VoCoLex, a lexical database for the French language (~105,000 words; Dufour, Peereman, Pallier, & Radeau, 2002).

Table 6
Proportion of Syllabic Structures in ULSID

Type	Proportion	Type	Proportion
CV	.545	CCVC	.013
CVC	.362	CVCC	.006
V	.044	CCV	.005
VC	.025	VCC	.000

Note—From Vallée et al. (2000).

1951; Tuller & Kelso, 1990, 1991). However, the large number of transformations from CVC /səp/ to CCV /psə/ in Group 1, which does violate the constraint of CVC stability, allows us to abandon this alternative interpretation, which is based on syllabic structure regularities.

Hence, none of these linguistic factors can fully explain the observed patterns of both stability and attractivity in our experiments. A final concurrent interpretation must be considered quite seriously, however. Indeed, our results appear to be compatible with several phonological theories of syllabification (for a review, see Goslin & Frauenfelder, 2001) and are relevant to the syllabic segmentation issue in psycholinguistics (see, e.g., Content, Kearns, & Frauenfelder, 2001; Dumay, Frauenfelder, & Content, 2002; Treiman & Danis, 1988). Whatever the group, the observed pattern of attractivity followed the sonority sequencing principle (Clements, 1990), according to which a preferred syllable shows a sonority profile (or sonority scale) that maximally rises toward the nucleus peak and minimally falls toward the end of the syllable. This sonority principle, together with the preference for onsets over codas according to the obligatory onset principle (Hooper, 1972) and the maximal onset principle (Pulgram, 1970), would predict hierarchies such as "/psə/ > /səp/ > /əps/" in Group 1 and "/pəs/ > /spə/ > /əsp/" in Group 2, which are more or less compatible with the observed stability and attractivity patterns. In this respect, the present results might be considered relevant to the syllabic segmentation issue. Particularly, the fact that the patterns of preferential transformations are maintained in covert speech is an important indication of the ability of syllabification mechanisms to intervene in the speaker's brain.

However, the sole explanation based on syllabification theories does not fully account for the observed results. Indeed, syllabification rules should lead to within-groups hierarchies that are not in agreement with our data concerning the nonsignificant difference between /pəs/ and /spə/. Nor can they explain the pervasive trend in the present study toward a shift from Group 2 to Group 1, especially toward the sequence /psə/ (see the many intrusions of /psə/ transformations for most stimuli in Tables 2A and 3A), which happens to be the best sequence in terms of articulatory coherence in our predictions (see Table 1). Moreover, a combination of syllabification mechanisms, syllabic structure constraints, and lexical factors should lead to a weak preference for /psə/ over /səp/ (/səp/ appearing more frequently in the participants' lexicon than /psə/ and corresponding to a "good" CVC syllable in terms of sonority and syllabic structure) and to a large preference for /pəs/ over /spə/ (/pəs/ being preferred in terms of syllabic structure, sonority sequencing, and lexical frequency). This is obviously at odds with the obtained results.

Therefore, the predicted articulatory cohesion scale—which is of course related in some sense to syllabification principles—seems to provide the most likely and coherent explanation of the results of Experiments 2 and 3. Thus, the fact that neither universal nor language-specific constraints can fully account for the data—in particular for the success of the /psə/ sequence—argues in favor of the existence of specific articulatory control constraints acting on verbal transformations during a self-repetition mode.

Relations Between Speech Perception and Production in the Verbal Transformation Effect

A major result of the present set of experiments concerns the pattern of verbal transformations obtained in the production procedure in Experiment 2 and its covert variant in Experiment 3, which are completely different from that obtained in the perception procedure in Experiment 4. There are two major differences. First, the number of transformations is much larger in Experiment 4, in agreement with previous experiments (Lackner, 1974; MacKay et al., 1993). It has been proposed that this is due to a corollary discharge mechanism (Lackner, 1974) or a top-down priming process (MacKay et al., 1993) that stabilizes the speech representation in the production procedure. Second, the *pattern* of transformations is also completely different, and this is clearly a new finding. Actually, some of the transformations are shared by both procedures, as is evidenced by the unexpected transformations in Experiments 2 and 3, which provide a number of previously emphasized contents, including substitution of a phoneme by a phonetically close one (Warren, 1961; Warren & Meyers, 1987), auditory streaming (Pitt & Shoaf, 2001, 2002), and lexical transformations (Kaminska et al., 2000; Shoaf & Pitt, 2002; Warren, 1961). However, although these transformation mechanisms were in the majority in the purely auditory procedure, this was not the case for the self-repetition conditions, in which most of the transformations depended on a shifting parsing process driven by the degree of synchronization between articulatory gestures. In the latter experiments, the more temporally clustered sequences played the role of attractors in verbal transformations (in particular, the "all-phased" monostable /psə/). Taken together, these results suggest that transformation mechanisms are not fully equivalent for a perception procedure and a production variant procedure, with articulatory control constraints acting as a major factor inducing transformations during a self-repetition procedure.

Although our results seem to point to articulatory control constraints as the major factor in transformations during the production experiments, it could be proposed that multisensory representations, combining auditory and proprioceptive components, drove the search for temporal clustering in Experiments 2 and 3. (In the latter, covert, case, multisensory imagery produced by the "inner voice" would play the same role.) Indeed, the speakers in Experiment 2 both produced and perceived gestures through various sensory channels, and it is quite likely that both motor and perceptual requirements shaped their behavior. The same is true of the participants in Experiment 3, in which the perceptuomotor loop between the "inner voice" and the "inner ear" is involved in the brain in a covert mode. This becomes even clearer when one considers a recent functional brain imaging study carried out in our laboratory (Sato et al., 2004). In this fMRI experiment, two conditions were contrasted: a baseline condition involving the simple mental repetition of the speech sequences used in the present study and a verbal transformation condition involving the mental repetition of the same items with an active search for verbal transformation. The contrast between the verbal transformation task and the baseline revealed a left-lateralized network of activations, notably within the inferior frontal gyrus, the supramarginal gyrus, and the superior temporal gyrus—areas considered to be involved in the analysis of articulatory-based representations, in the interfacing between sound-based and articulatory-based representations of speech, and in the self-monitoring of verbal material, respectively. These results thus strongly suggest that the verbal transformation effect has common components of speech perception and speech production and relies on both sound-based and articulatory-based representations. On the other hand, the present results underscore the fact that transformation mechanisms do not act to the same extent for a perception procedure and a production-variant procedure, with articulatory control constraints acting as a major factor inducing transformations during a self-repetition procedure.

In conclusion, the set of experiments described in the present study demonstrate in a coherent way that the perceptual stability and attractivity of an uttered sequence depend on articulatory control constraints, which are hence likely to be involved, together with auditory, phonological, and lexical constraints, in the emergence and analysis of verbal representations in the human brain.

REFERENCES

Abry, C., Cathiard, M.-A., Vilain, A., Laboissière, R., & Schwartz, J.-L. (2004). Some insights in bimodal perception given for free by the natural time course of speech production. In G. Bailly, P. Perrier, & E. Vatikiotis-Bateson (Eds.), *Festschrift Christian Benoît*. Cambridge, MA: MIT Press.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, **6**, 201-251.

Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 317-328.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Between the grammar and the physics of speech* (Papers in Laboratory Phonology, Vol. 1, pp. 283-333). Cambridge: Cambridge University Press.

Content, A., Kearns, R. K., & Frauenfelder, U. H. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory & Language*, **45**, 177-199.

de Jong, K. J. (2001). Rate-induced resyllabification revisited. *Language & Speech*, **44**, 197-216.

de Jong, K. [J.], Nagao, K., & Lim, B.-J. (2002). The interaction of syllabification and voicing perception in American English. *ZAS Papers in Linguistics*, **28**, 27-38.

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.

Dumay, N., Frauenfelder, U. H., & Content, A. (2002). The role of the syllable in lexical segmentation in French: Word-spotting data. *Brain & Language*, **81**, 144-161.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hove, U.K.: Erlbaum.

Gleason, P., Tuller, B., & Kelso, J. A. S. (1996). Syllable affiliation of final consonant clusters undergoes a phase transition over speaking rates. In H. T. Bunnell & W. Idsardi (Eds.), *Proceedings of the 4th International Conference on Spoken Language Processing* (pp. 276-278). Wilmington, DE: Applied Science & Engineering Laboratories.

Goslin, J., & Frauenfelder, U. H. (2001). A comparison of theoretical and human syllabification. *Language & Speech*, **44**, 409-436.

Hooper, J. B. (1972). The syllable in phonological theory. *Language*, **48**, 525-540.

Jakobson, R. (1966). Implications of language universals for linguistics. In J. H. Greenberg (Ed.), *Universals of language* (pp. 263-278). Cambridge, MA: MIT Press.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral & Brain Sciences*, **17**, 187-245.

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

Lackner, J. R. (1974). Speech production: Evidence for corollary-discharge stabilization of perceptual mechanisms. *Perceptual & Motor Skills*, **39**, 899-902.

Lallouache, M. T. (1990). Un poste "visage–parole": Acquisition et traitement de contours labiaux. In *Actes des XVIIIèmes journées d'études sur la parole* (pp. 282-286). Montreal.

Landauer, T. K. (1962). Rate of implicit speech. *Perceptual & Motor Skills*, **15**, 646.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122-147). Cambridge, MA: MIT Press.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review*, **89**, 483-506.

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 & Language*, **32**, 624-646.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral & Brain Sciences*, **21**, 499-546.

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, **288**, 527-531.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Murray, D. J. (1965). Vocalization-at-presentation and immediate recall, with varying presentation-rates. *Quarterly Journal of Experimental Psychology*, **17**, 47-56.

Natsoulas, T. (1965). A study of the verbal-transformation effect. *American Journal of Psychology*, **78**, 257-263.

Pitt, M. A., & Shoaf, L. [C.] (2001). The source of a lexical bias in the verbal transformation effect. *Language & Cognitive Processes*, **16**, 715-721.

Pitt, M. A., & Shoaf, L. [C.] (2002). Linking verbal transformations to their causes. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 150-162.

Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language & Speech*, **39**, 375-392.

Pulgram, E. (1970). *Syllable, word, nexus, cursus*. The Hague: Mouton.

Reisberg, D., Smith, J. D., Baxter, D. A., & Sonenshine, M. (1989). "Enacted" auditory images are ambiguous; "pure" auditory images are not. *Quarterly Journal of Experimental Psychology*, **41A**, 619-641.

Sato, M., Baciu, M., Lœvenbruck, H., Schwartz, J.-L., Cathiard, M.-A., Segebarth, C., & Abry, C. (2004). Multistable representation of speech forms: A functional MRI study of verbal transformations. *NeuroImage*, **23**, 1143-1151.

Shoaf, L. C., & Pitt, M. A. (2002). Does node stability underlie the verbal transformation effect? A test of node structure theory. *Perception & Psychophysics*, **64**, 795-803.

Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, **11**, 1433-1454.

Stetson, R. H. (1951). *Motor phonetics: A study of speech movements in action*. Amsterdam: North-Holland.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition*, **15**, 49-74.

Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory & Language*, **27**, 87-104.

Tuller, B., & Kelso, J. A. S. (1990). Phase transitions in speech production and their perceptual consequences. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 429-452). Hillsdale, NJ: Erlbaum.

Tuller, B., & Kelso, J. A. S. (1991). The production and perception of syllable structure. *Journal of Speech & Hearing Research*, **34**, 501-508.

Vallée, N., Boë, L. J., Maddieson, I., & Rousset, I. (2000). Des lexiques aux syllabes des langues du monde: Typologies et structures. In *Actes des XXIIIèmes journées d'études sur la parole* (pp. 93-96). Aussois, France.

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

Warren, R. M. (1982). *Auditory perception*. New York: Pergamon.

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

Warren, R. M., & Meyers, M. D. (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, March). *Auditory illusions and aging: Transmission of priming in the verbal transformation paradigm*. Paper presented to the 4th Biennial Cognitive Aging Conference, Atlanta.

NOTE

[1] Since all the sequences showed a lower degree of stability and attractivity than in Experiments 2 and 3, the related values were not included in specific subtables as they were for the previous experiments.

(Manuscript received April 14, 2003;
revision accepted for publication May 31, 2005.)