INTRASYLLABIC ARTICULATORY CONTROL CONSTRAINTS IN VERBAL WORKING MEMORY

Marc Sato, Jean-Luc Schwartz, Marie-Agnès Cathiard, Christian Abry & Hélène Loevenbruck

Institut de la Communication Parlée (ICP)
INPG / Université Stendhal / CNRS UMR 5009
sato, schwartz, cathiardi, abry, loeven@icp.inpg.fr

ABSTRACT

Verbal transformation effect – an auditory imagery task equivalent to Necker’s cube in visual imagery – recruits a specific working memory, the so-called articulatory or phonological loop. Is this mechanism sensitive to articulatory control constraints, i.e. phase relationships between vowel and consonant gestures? In our experiment, 56 French students repeatedly pronounced aloud non-sense syllables – all combinations of [s] with [p] and [s] – and were asked to stop as soon as they heard a possible syllable transformation. In agreement with our in-phase predictions, the winner is syllable [ps], where all gestures can be launched in synchrony. This experiment demonstrates that verbal working memory – a primary candidate as input memory for word learning – is sensitive to articulatory control of syllable phasing.

1. VERBAL TRANSFORMATIONS AND THE PHONOLOGICAL LOOP

In a set of experiments exploring auditory imagery, Reisberg et al. [Rei89] examined the imagery analogue of the verbal transformation effect [War58]-[War61]. This word game [Tre83] relies on the fact that certain words, if repeated over and over, yield a sound stream compatible with more than one segmentation (e.g. rapid repetitions of the word “life” produce a sound stream fully compatible with the perception that either “life” or “fly” is being repeated). This effect corresponds, in speech, to the depth perceptual rivalry in the Necker’s cube [Cha85].

In their experiment, the authors asked subjects to repeat different stimuli and to detect a possible transformation – the conditions varying by degrees between overt and covert repetition (from loud speech to mental repetition). The authors show that the transformation probability gradually decreases as vocalization declines from one condition to the other, with less and less possibility of “enaction”. They relate enaction to the so-called “phonological loop”, the verbal component of Baddeley’s working memory model [Bad86].

As known, the phonological loop is supposed to involve two distinct components: a phonological store – the “inner-ear” – and a subvocal rehearsal process – the “inner-voice”. In this model, the phonological store is assumed to represent verbal material in a phonological form while the process of subvocal rehearsal aims at “re-enacting” this material, re-presenting it to the store and thus refreshing and preserving its content [Bad90]-[Bad92a]. Recent research suggests that retention of familiar verbal material should be an accidental by-product of the primary function of this system: that is, to serve language along the acquisition of novel phonological forms [Gat93]-[Gup97]-[Bad98]-[Gat99].

In the verbal transformation tasks, the subvocal rehearsal process would not be simply in charge of automatic repetition, but also of the mental search of the transformation (for a discussion about links between verbal imagery and phonological loop, see also [Bad92b]).

2. THE ROLE OF INTERARTICULATORY PHASING

An important issue, not considered by Reisberg et al., concerns the existence of possible verbal transformation asymmetries. For instance, the inverse transformation “fly=life” rather than “life=fly” seems less likely. We assume that this relies on the existence of preferential articulatory phasing. Thus, in “fly”, the speaker can prepare the [i] lingual constriction and the lingual position for the start of the [ai] diphthong within the [f] lip-teeth constriction. Therefore, the speaker can realize the three gestures in synchrony, and particularly the two consonantal gestures inside the [if] syllable onset are in phase. In the “life” case, the synchronization of [i] (syllable onset) and [l] (syllable coda) is of course impossible. Hence the preference for the “life=fly” transformation, from less to more “in-phase” gestures.

This preferential articulatory phasing hypothesis may be related to the properties of dynamical systems studied by Haken, Kelso & Bunz [Hak85]. Their princeps experiment consisted in asking subjects to move vertically the left and right forefingers in an asymmetrical way (producing out-of-phase movements), and to ask them to progressively accelerate. The movement abruptly changed towards a symmetrical mode (in-phase forefinger movements) implying the simultaneous activation of flexion/extension homologous muscles.

This sudden and completely unintentional “phase transition”, modeled by the authors in the framework of coupled non-linear dynamical systems, is interpreted in terms of inter-articulator coupling (though see a recent perceptual rather than motor interpretation by Mechsner et al. [Mec01]).

Tuller & Kelso [Tul90] attempted to apply this principle in speech production. Considering CV [pi] and VC [ni] syllables as test items, the authors pointed out a switch from [ni] to [pi] sequences through acceleration of the repetition rhythm. The VOT index, providing the phasing between the glottal aperture peak and the release of the lips constriction, revealed a specific syllabic structure alternation from VC to CV, with the apparition of the aspiration phase [p]<sup>i</sup>=[p]<sup>i</sup>. However, this result is problematic, since there is in fact a phasing of the lip closing and glottal closing gestures in [ip], while there is no phasing...
between lips and voice in English [pʰ], precisely because of aspiration! Thus, Fuller & Kelso’s data correspond to a language specific VOT control in the consonant oro-laryngeal coordination, and not to a phasing between the consonant and the vowel. Notice that Reisberg et al. then produced a number of transformations also implying the voiced – unvoiced contrast, such as [t] vs. [d] in “stress” vs. “dress” and [k] vs. [g] for “kiss the sky” vs. “kiss this guy”.

Since further literature did not produce firm proposals about cohesion or phasing index in this field [Bro00], we decided to test the contribution of the verbal transformation paradigm to this question.

3. EXPERIMENT

The aim of this experiment was to test the articulatory loop sensitivity to intrasyllabic phase control constraints by showing evidence of preferential transformations during a verbal transformation task.

We consider that external repetition (i.e. aloud repetition) implies internal repetition (i.e. inner voice). Indeed, it is not conceivable to utter, in the same time, an overt and a different covert speech sequence. In fact, mental execution has been shown to imply essentially the same cortical structures as overt execution, with an additional externalization inhibition process [Grê01]. In consequence, we chose the most simple and effective condition for eliciting transformations, that is aloud repetition.

3.1 Stimuli

To avoid possible lexical effects and to focus on intrasyllabic stability – without considering the segmentation issue between syllables ([Con01]; [Dum98]) – we chose to use monosyllabic logatoms. We selected syllables combining two consonants and one vowel, in order to contrast phase relationships between consonants on one hand and also between vowels and consonantal groups on the other hand. The vowel was the neutral vowel [ə] which is likely to be stable during the repetition, whatever the consonantal environment. The consonants were the plosive [p] and the fricative [s], the second having an articulation point posterior to the first one. This allows the possibility to prepare the [s] constriction during the [p] closure, while the inverse is not true. All combinations between these three sounds are possible in French.

The verbal transformation process should induce possible shifts inside two natural groups, that is:
- Group 1 : [ps], [sp], [ps] and [ps] sequences;
- Group 2 : [sp], [ps] and [sp] sequences.

In other words, we predict that subjects, from an orally-presented sequence, will not make all possible combinations of [p], [s] and [ə]. They will rather find transformations within each group, according to a shifting parsing procedure applied to the repeated sequence, that is: …pşşpşşpşş… for Group 1 sequences, and …şşşşşşpşş… for Group 2 ones.

We made a series of predictions on possible transformation preferences. They are reported in Table 1, displaying sequence classification for each group according to an hypothetical global phasing score. This score is based on all the possible synchronizations between the vowel and the consonants in the sequence. In accordance with our initial hypothesis, we take global phasing as a crucial factor determining the preferred transformation during the repetition process. Thus, the better the phasing of a sequence, the less likely its transformation into another one. Conversely, transformation from a poorly-phased sequence to a better-phased one should be the rule.

### Table 1: Sequence classification according to the global phasing degree, and expected preferred transformation.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Phasing</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>psó</td>
<td>Complete: onset [ps] and vowel [ə] phased, phasing of the two consonants in the onset</td>
<td>psó</td>
</tr>
<tr>
<td>sòsp</td>
<td>Average: onset [s] and vowel [ə] phased, coda [p] not phased</td>
<td>sòsp</td>
</tr>
<tr>
<td>ñps</td>
<td>Weak: vowel and consonantal group not phased, phasing of the consonants in the coda</td>
<td>sòsp</td>
</tr>
<tr>
<td>sósp</td>
<td>Average: onset and vowel phased, coda [s] not phased</td>
<td>sòsp</td>
</tr>
<tr>
<td>ñsósp</td>
<td>Very weak: vowel and consonantal coda not phased, [s] and [p] not phased in the coda</td>
<td>sòsp</td>
</tr>
</tbody>
</table>

3.2 Method

Fifty-six students from the Grenoble University participated. All subjects were native French speakers without hearing deficits. They passed the experiment individually without any previous information about it.

After read instructions, the experimenter gave the “life”-“fly” verbal transformation example, to clarify the procedure and to indicate the expected repetition speed (about two repetitions/second, as in other verbal transformation studies). After each orally-presented sequence (the order being randomized among subjects), the subject was asked to repeat it aloud and look for a transformation. If she/he found a transformation, she/he had to stop and indicate it to the experimenter. If no transformation was detected, the experimenter stopped the subject after about thirty seconds, which is once again classical in this paradigm. Of course, no indication about expected transformations was provided. All the experiment was recorded for follow-up analyses.

3.3 Results

Table 2 displays the transformation frequencies summed over the fifty-six subjects, for the six sequences.

### Table 2: Percentage of transformations frequencies.

Sequences in lines, transformations in columns; diagonal indicates no transformation;
transformation inside groups (1 and 2) marked by bold lines.

Overall, most responses occur within groups 1 or 2, which reflects a correct parsing process (80% of the responses). When this is not the case, it is seldom due to transformations from one group to the other: in average, less than 1% from group 1 to group 2 and 5% from group 2 to group 11. It is more often due to diverse (“div.”) transformations such as semantic inputs (e.g. “ça se peut” [sp] – “it may be”), and more frequently transformations involving sequences of a larger length than expected (as [pssp]). Chi-2 tests across observed transformation frequencies lead to the following conclusions:

- There is no significant effect of stimulus order (p=0.4).
- There is a significant stimulus effect (p<0.001).
- There is a significant group effect (p<0.05).
- There is a significant stimulus effect inside group 1 (p=0.05) but not inside group 2 (p=0.4). The negative result for group 2 is due to the balanced pattern of responses between [p] and [sp].
- There is a significant effect of preferential transformation for each stimulus (p<0.05).

4. DISCUSSION

4.1 A pattern of observed transformations compatible with a “phasing” effect

Through each group, our analysis of transformation results is the following:

- Group 1: The [ps] sequence, predicted as very stable, is indeed not transformed for 75% of the subjects, 18% switching towards [sp] and no one towards [ps]. On the other hand, [sp] and [sp] are transformed into [ps] for respectively 50% and 29% of the subjects. We notice that [sp]>[ps] and [ps]>[sp] transformations are nearly non-existent (2% and 0%). For this group, the unexpected result is the stability of [ps] for 64% of the subjects.

- Group 2: [pss] and [sp] sequences, predicted as moderately stable, are transformed into each other for respectively 46% and 43% of the subjects. No transformation of these sequences respectively corresponds to 39% and 41% of the subjects. The [sp] sequence is transformed into [ps] for 20% of subjects and into [sp] to 5%, while there is no opposite transformations from [ps] or [sp] to [ps]. As [sp] in group 1, we note a stability of [sp] sequences for 55% of the subjects.

Altogether, syllables with no consonantal onset [ps] and [sp], though predicted as very unstable, appear rather stable in our results. This can be explained by the presence of a glottal onset produced quite often by the subjects (a post-hoc phonetic analysis confirmed this effect) which prevents them to chain repetition items and hence blocks articulatory phasing [Dep01]. The glottal onset can be considered as a consonant transformation VCC syllables [ps] and [sp] into CVCC syllables [pss] and [sp]. However, when subjects do not add this onset, we note a preferential transformation to the best-phased sequence in group 1 (i.e. [ps]) or to the sequence which appeared more attractive in group 2 (i.e. [pss]).

Overall, this transformation/non-transformation result pattern may be summarized as follows:

- In group 1, we obtain a hierarchy [ps]>[sp]>[sp] in terms of attractiveness. Concerning stability, the winner is [ps]. However, the glottal onset transforming [ps] into the CVCC [pss] makes it more stable than expected.
- In group 2, [sp] is rather stable, as [sp] in group 1 and for the same reasons. There is a balanced pattern between [ps] and [sp] transformations, as expected, though the better global attractiveness of [ps] was not part of our predictions.

4.2 Discarding other possible interpretations

Our experiment displays in an original paradigm a strong bias towards better-phased sequences in the verbal transformation task. Can this result be explained by linguistic constraints, and how can it be related to language inventories?

Firstly, our results cannot be directly applied to typological trends patterning syllabic structure (e.g. CV preferred to CVC, CVC preferred to CV, etc) since our sequences were restricted to CVC, CV and VCC syllabic structures. However, we can argue that VCC and CCV syllables are very infrequent in phonological inventories and that, after CV, it is the CVC syllable which is the most frequent (see the ULSID Database [Val00]: CV=54.48% – CVC=36.15% – CCV=0.5% – VCC=0.05%). Our results argue for the avoidance of syllables with no consonantal onset (the CV universal syllable would not switch towards VC). However, the important number of transformations from CVC to CV in group 1 (if the language, as French in this study, does not impede these structures) is not in favor of a general property of CVC stability.

Another concurrent interpretation of our data could be based on the frequency of each possible sequence of groups 1 and 2 in the subjects' mother-tongue lexicon, which could favor transformations towards more frequent lexical entries. Therefore, we studied [ps] and [sp] frequencies in initial or final word position in three French phonetic dictionaries [Aub88]. We obtain results clearly not in favor of [ps], much less frequent than [sp] though it is much preferred in our data. How then can we explain the small lexical frequency of the well-phased [ps]? Our interpretation is that this sequence is disadvantaged in view of the small perceptual prominence of [p] when it is uttered in a [p] group.

Altogether, neither the syllabic structure nor the lexical frequency can account for the preference for [ps] in our data.

5. CONCLUSION

The preference for better-phased sequences is clearly demonstrated in this experiment, and impossible to explain by linguistic formal criteria. Since the phonological loop seems engaged in verbal transformations (Reisberg et al., 1989), this result suggests that production constraints (here, interarticulatory phasing) are likely to intervene in this component of the verbal memory. In other terms, the “phonological loop” is indeed articulatory (Baddeley, 1990)!

As far as we know, it is the first time that the role of interarticulatory phasing in speech production/perception is clearly demonstrated – considering the caveats we raised about the ip-pi VOT switch displayed by Tuller and Kelso.
consider that this auditory imagery task with stimulus decision, in a “Necker cube” way, depends on the large stimulus ambiguity due to the replay process. Thus, this task is no longer a detection of segmentation cues between iterate syllables but a change of perceptual internal state produced by active auditory perception [Gom95]-[Sch97]. The asymmetry of transformations (also present in the Necker’s cube) is due to the influence of interarticulatory phasing control, with its preferred phased state.

Finally, the fact that neither extra-articulatory nor language-specific constraints can account for our results – in particular, the success of the “all-phased” [ps] – argues in favor of their relevance for the role of production constraints inside verbal working memory.

**Note**

(1) In fact, these differences between Group 1 and Group 2 are probably important to notice. When cautiously analyzing [ps] and [ps] repetitions by the 56 subjects, we found that in almost 50% of the [ps] repetitions, the subjects made pronunciation mistakes leading them to introduce some utterances of [ps] instead of [ps]. The inverse pattern ([ps] instead of [ps]) almost never appeared. This confirms that [ps] sequences are much easier to pronounce than [sp] ones … because they are better phased?

**Acknowledgments**

This study has been realized with the support of the French MESR-ACI-Neuroscience project “Building a Baby”. We thank Pascal Perrier, Anne Vilain and the students of the Grenoble University for their contribution to the realization of this study.

**References**


