Effects of temporal chunking on speech recall

Annie C. Gilbert[1,2], *Victor J. Boucher*[1], *Boutheina Jemel*[2,3]

[1] Laboratoire de Sciences Phonétiques, Université de Montréal, Canada
[2] Laboratoire de recherche en Neurosciences et Électrophysiologie Cognitive, Centre de Recherche Fernand-Séguin, Canada
[3] École d'orthophonie et d'audiologie, Université de Montréal, Canada
annie.gilbert@umontreal.ca, www.phonetique.info

Abstract

It is established that temporal grouping or "chunking" arises in serial recall as it does in speech. For instance, chunking appears in common tasks like remembering series such as phone numbers. In the present study, we examine how detected chunks in meaningless strings of syllables and meaningful utterances influence memory. We use a Sternberg task where listeners identify whether a heard item was part of a presented context. Such tasks serve to explore if working memory operates in terms of chunks and is influenced by meaning. Observations using evoked potentials ensured that chunks in the heard stimuli were detected by the 20 listeners. The results showed that, for meaningless series, chunk size and position significantly affected listeners' recall and their response times. However, there were no such effects for meaningful utterances. This suggests that memory of novel series operates by chunks. But in dealing with sequences of items that are already in long-term store, chunks may not have a dominant influence on working memory.

Index Terms: speech segmentation, temporal grouping, chunking, working memory, short-term memory.

1. Introduction

It is well known that, in recalling lists such as series of digits or even meaningless syllables, temporal groupings arise spontaneously (e.g. [1-5]) This phenomenon, which Miller [6] called "chunking" is not limited to verbal lists. In fact, it also appears in recalling non-verbal sequences [7]. It is also widely acknowledged that chunks do not generally exceed 3 or 4 items, which reflects the capacity limits of short-term serial recall (for a review of chunk limits, see [8]). What is less recognized in the literature is that temporal grouping also operates in the perception and production of meaningful speech (e.g. [9-11]). In particular, our previous studies using the technique of evoked potentials have shown that listeners chunk speech by reference to temporal groups (TG). But why would listeners chunk speech this way? One reason is that, from the standpoint of the listener, interpreting speech requires that fleeting series of sounds be held in a short-term store, while processing incoming signals. Thus, processing rapidly changing speech sounds implies a serial working memory and, given that serial memory is limited, processing needs to operate by some chunk of signal.

However, temporal chunking obviously marshals different memory processes depending on whether one hears meaningful or meaningless speech. The latter context can be likened to the situation where one is beginning to learn a new language. In such cases, research has shown that prosodic groups determine the acquisition of sequences of sounds constituting novel verbal forms. (See [12] and [13] on the effects of prosody on "statistical learning"). For instance, it has been established that transitional probabilities (TPs) between sounds assist in the learning of novel forms. Thus, syllables that often follow each other are perceived as part of the same form or "word" (e.g. [14]). However, Shukla et al. [13] showed that listeners do not detect TPs marking (artificial) words when they straddle prosodic groups. Similarly, Gilbert, Boucher, and Jemel [15] found that perceived TGs can hinder the learning of forms that straddle groups. In short, listeners learn novel verbal forms by detecting frequently associated sounds *within* TGs. On the other hand, in listening to speech containing recognized forms, temporal grouping serves another function. Specifically, temporal chunking appears essential to accessing forms in long-term memory. For example, Christophe et al. [16] showed that lengthening *grin-* in *le chat grincheux* prevents an access to the form *grincheux* and leads to access *chat* and *grin* as part of the same form *chagrin.* This suggests that a detection of temporal grouping creates associations between recognized items within a group that can map onto meaningful forms in long-term store.

Thus, we know that processing speech requires serial working memory and that listeners temporally chunk speech in conformity to capacity limits on serial memory. However, in dealing with meaningful and meaningless speech sounds one might view chunking as involving different processes (or else a common process). In listening to novel series such as nonsense syllables, chunking can create groups independent of any syntactic structure. On the other hand, in listening to meaningful speech, it is unclear that chunking can operatte independently of semantic-syntactic units. The present study basically asks how temporal grouping bears on memory processes of meaningful and meaningless speech using a modified version of a **Sternberg task[17]**.

This classic task is used to study the scanning of items that are active in working memory. Typically, participants are presented series of items followed by a target and asked to determine if the target was present in the list or not. Usually, the length of a list to remember has an impact on the reaction times: the longer the list, the longer the reaction time. Therefore, if speech is stored in working memory by temporal groups, then the length of these groups (in syllables) should have an impact on reaction times and accuracy of responses in a Sternberg paradigm. With this in mind, we used a modified Sternberg task controlling for the length of TGs and their position in the utterance. The originality of the present study bears on the use of stimuli where the neural responses of listeners showed a detection of temporal chunks. By reference to these contexts, the prediction was that the size and position of chunks in heard meaningless series would affect both the accuracy and speed of recall. However, we expected that none of these effects would extend to meaningful sequences, which would suggest a different storage principle.

2. Methods

2.1. Participants

Twenty native speakers of French, aged from 19 to 42 years (mean = 25.6) were recruited on the campus of the Université de Montréal. All presented normal hearing levels in a standard audiometric evaluation, and normal memory performances on a digit-span test (WAIS; [18]). All were dominant right-handers, with no history of substance abuse (other than tobacco smoking), and no neurological, psychiatric or speech disorder. Participation in the tests was subject to written consent and the research was approved by the ethics committee of the Hôpital Rivière-des-Prairies.

2.2. Stimuli

2.2.1. Stimuli design

Two distinct sets of nine-syllable stimuli were created. One set included 100 meaningless series of syllables and the other included 100 meaningful utterances in French. Both sets were constructed so as to obtain different rhythmic conditions represented in Figure 1 (adapted from [11]). Note that in one condition the initial TG comprises 3 syllables and the internal TG contains 4 syllables. The order is inverted in the other condition where the initial TG contains 4 and the internal TG contains 3 syllables. Thus, the two conditions allow comparisons of the effects of TG length (3 vs 4 syllables) and position (initial vs internal). The final TG has a constant length of two syllables and intonation was also kept constant throughout the stimuli. This way, all stimuli presented two intonation contours over three TGs and the first intonation contour spanned the first two TGs. The second contour spanned only the final TG.

Every stimulus was followed by a target syllable or monosyllabic lexeme taken from either the initial (in 50% of the cases) or the internal temporal group (50%). Filler stimuli were also created to vary the rhythmic, syntactic and intonation patterns, and to balance the presentation of target syllables or lexemes (to avoid the impression that targets were drawn mostly from one or the other TG). Both sets of stimuli were controlled with respect to the following linguistic attributes.

Meaningless series of syllables were created using consonant-vowel (CV) syllables of French. Each series was balanced with no repeated C or V within a series and avoiding the creation of recognizable multisyllabic lexemes. Syllable order was controlled so that no consecutive syllables shared a common point of articulation (to prevent confounding effects on recognition recall). See examples bellow where / indicates a TG boundary: Example (1) represents a stimulus with an initial TG of 3 syllables followed by a TG of 4 syllables; Example (2) presents a stimulus with TGs of 4 and 3 syllables.

(1) [na ʒy wɑ̃ / fœ tø ʒɛ̃ ɲi / zɔ̃ mœ̃]
(2) [ze ʒa bœ̃ ɲɔ̃ / kɛ̃ wø gy / fœ nɑ̃]

As for the meaningful utterances, these stimuli were created using monosyllabic lexemes and functors with a high index of familiarity in French [19]. These were arranged in a given syntactic structure so that the initial TG always contained the subject, the internal TG contained a complement to the subject and the final TG contained the verb phrase. All interpretations were kept literal. The following Examples of utterances (3) and (4) illustrate the same TG patterns (3-4 vs 4-3) as in the above sequences of syllables [Ex. (1) and (2)].

(3) Les eaux sales / de cette marre calme / gèlent tôt.
(4) La grande poêle creuse / à fond plat / reste chaude.

2.2.2. Stimuli recording

The stimuli were recorded using a pacing technique where a speaker is asked to produce contexts while listening to series of pure tones providing a metronome-like signal. Using headphones, a native speaker of French listened to a continuous playback of the metronome while repeating each stimulus. This technique enables the speaker to produce specific rhythm and intonation patterns. In the present case, it served to obtain productions of TGs of constant durations (4-syll TG = 1,150 ms, 3-syll. TG = 900 ms, 2-syll. TG = 650 ms.) marked by a lengthening of the last syllable corresponding to French natural prosody (1.6 times longer than non-final syllables [20]. Recordings were performed in a sound-treated booth using an external sound card (M-Audio Fast-Track Pro, 44,1kHz, 16 bits, mono). Every stimulus series was saved in an individual sound file and amplitudes were normalized. Filler stimuli (meaningful and meaningless utterances) were also recorded following different prosodic patterns to vary the presentations.

2.2.3. Stimuli validation

To ensure that the recordings matched the desired prosodic patterns, pitch and energy contours were measured for each of the stimulus (see Figure 1, adapted from [11]). Top panels represent F0 contours of every recorded stimulus with regard to rhythm patterns (initial TG of 3 or 4 syllables). One can see that all stimuli present similar intonation contours with only one intonation reset (at about 2,250 ms). As for the energy contours, these show the location of the relative lengthenings marking the end of TGs. A substantial difference can be seen between rhythm conditions 3-4 vs 4-3 for both meaningful and meaningless utterances. Overall, there is little variability between stimuli from the same prosodic condition, and a clear demarcation between conditions.

Furthermore, electroencephalographic recordings were acquired while participants listened to the stimuli. Analyses of these recordings confirmed that the TGs evoked *Closure Positive Shifts* (CPS) (see [11]). The occurrence of this neural component demonstrates that presented TGs were actually detected on-line by the listeners (see Figure 1, bottom panel) -- therefore ensuring that measured effects on recall are attributable to the presented TGs.

2.3. Procedures

Presentation of both sets of stimuli was counterbalanced with half the participants hearing the meaningless stimuli first. Stimuli were presented using insert earphones (*Eartone* 3A, EAR Auditory Systems). Participants were instructed to listen to the stimuli and indicate, via a key press and as fast as possible, whether the prompt was part of the preceding stimulus. Sound files were played back via *E-Prime 2.0* (Psychology Software Tools) in random blocks divided by rest pauses. The sounds were delivered at a constant intensity (peak levels of 68 dBA).

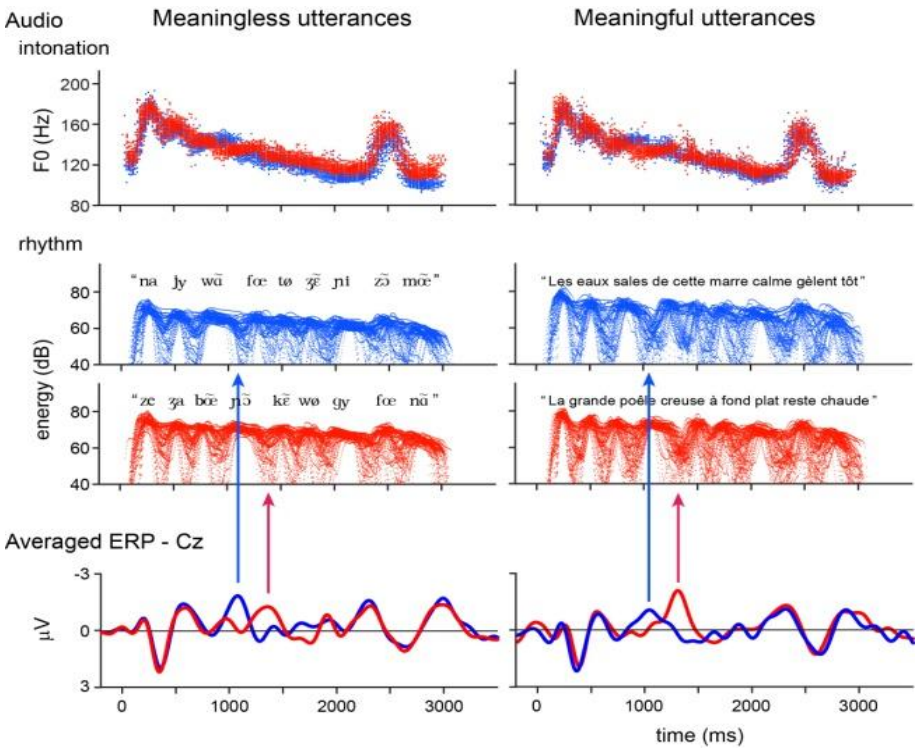

Figure 1: *Measured acoustic attributes of the stimuli and averaged ERPs at Cz. Blue traces correspond to stimuli with initial TGs of 3 syllables, red traces represent stimuli with initial TGs of 4 syllables.*

3. Results

3.1. Accuracy of recognition

Average rates of accurate recognition to both meaningful and meaningless utterances are presented in Figure 2 as a function of TG length and position. As one can see from the standard errors, there is substantial variation in the accuracy of item recall in heard meaningless contexts, and a near ceiling effect for meaningful utterances. We used two separate ANOVAs to compare TG length and position effects on recall accuracy. For meaningless series, the results showed significant main effects of TG length [$F(1,19)$ = 13,064, $p < .005$, $\eta^2 = .407$] and position [$F(1,19) = 24.818$, $p < .001$, $\eta^2 = .566$], as well as significant interaction [$F(1,19) = 6.332$, $p < .03$, $\eta^2 = .25$]. These main effects, however, did not appear with meaningful contexts. In this case, there were no significant effects of TG length or position ([$F(1,19) = 3.963$, $p > .05$, $\eta^2 = .163$], [$F(1,19) = 3.003$, $p > .05$, $\eta^2 = .136$]), but only a significant interaction of factors [$F(1,19) = 14.648$, $p < .002$, $\eta^2 = .435$].

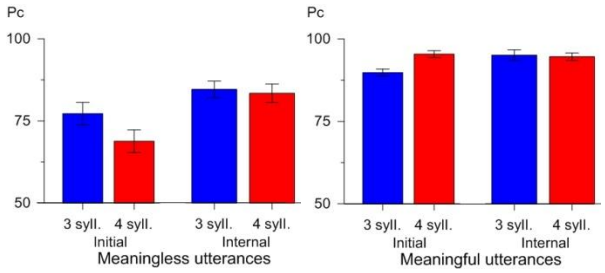

Figure 2: *Average accuracy of recognition rates (Percent correct* Pc*) as a function of TG length and position for both sets of stimuli.*

3.2. Reaction times

Visual inspection of graphs in Figure 3 reveals a similar variability between participant's performances to both meaningless and meaningful stimuli but overall shorter reaction times for meaningful compared to meaningless utterances. Reaction times were analyzed using the same 2 X 2 Anovas presented earlier. Similar to accuracy rates, significant effects of TG length [$F(1,19) = 13.402$, $p < .005$, $\eta^2 = .414$] and position [$F(1,19) = 4.85$, $p < .05$, $\eta^2 = .203$] were found for meaningless utterances, but no significant interaction were revealed [$F(1,19) = .701$, $p > .1$, $\eta^2 = .036$]. As for meaningful utterances, no TG length effect was found [$F(1,19) = 2.949$, $p > .1$, $\eta^2 = .134$], but a significant effect of position [$F(1,19) = 10.262$, $p < .01$, $\eta^2 = .351$] and a significant interaction were revealed [$F(1,19) = 4.393$, $p = .05$, $\eta^2 = .188$].

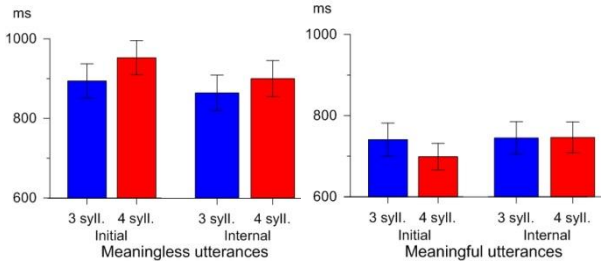

Figure 3: *Average reaction time (ms) as a function of TG length and position for both sets of stimuli.*

To further elucidate the effect of meaning on the recognition task, we pooled the results from meaningless and

meaningful stimuli to perform a three factor ANOVA (2 X 2 X 2) with respect to TG length, TG position, and meaningfulness of presented contexts. The *F* statistics of Table 1 show that the presentation of meaning has a significant effect in itself, and yields significant interactions.

Table 1: *F statistics from an ANOVA comparing effects of length (3 vs 4), position (initial vs internal), and meaning (utterances vs meaningless series)*

	Reaction times		
	F(1,1,19)	*p*	η^2
TG length	2.065	.167	.098
TG position	.553	.466	.028
Meaning	54.419	***.000***	.73
TG length TG position*	.314	.582	.016
TG length Meaning*	17.285	***.001***	.476
TG position Meaning*	11.513	***.003***	.377
TG length TG position* Meaning*	4.59	***.045***	.195

4. Discussion / Conclusion

It will be recalled that, in our Sternberg task, listeners were asked to determine whether a prompt was part of a heard utterance or meaningless series of syllables. In such a task, listeners must scan their working memory to determine if the presented target was part of the stored items or not. Typically, reaction times will vary according to the number of items kept active in working memory. The more items to scan, the longer it takes to respond.

What is particular about the present tests is that sets of presented contexts contained temporal chunks *that were detected by the listeners* (see Figure 1). The results show that, for meaningless contexts, reaction times of the Sternberg task (and accuracy of recall) varied with the length of the chunk (number of syllables) and its position. Hence, these observations suggest that speech is processed by chunks in listeners working memory. On the other hand, recall times and accuracy did not vary much when subjects listened to meaningful utterances. In this case, the responses suggest that listeners quickly recognize forms that are in long-term memory and that chunking had little influence. Hence, chunking appears to have a major impact in learning novel series – as when learning new verbal expressions in an unknown language -- but a minor influence in recognizing or remembering series of forms in meaningful utterances. Of course, temporal grouping may have a fundamentally different role in that in the latter case chunks not only constitute prosodic units but also serve to segment speech in meaning units.

5. References

[1] Broadbent, D. E. and Broadbent, M. H. P., "Grouping strategies in short-term memory for alpha-numeric lists," Bull. of the British Psychological Society, 26:135, 1973.

[2] Frick, R. W., "Explanation of grouping in immediate ordered recall," Memory and Cognition, 17:551-562, 1989.

[3] Ryan, J., "Temporal grouping, rehearsal and short-term memory," Quarterly J. of Experimental Psychology, 21:148-155, 1969.

[4] Wickelgren, W. A., "Rehearsal grouping and hierarchical organization of serial position cues in short-term memory," Quarterly J. of Experimental Psychology, 19:97-102, 1967.

[5] Winzenz, D. and Bower, G. H., "Subject-imposed coding and memory for digit series," J. of Experimental Psychology, 83:52-56, 1970.

[6] Miller, G. A., "The magical number seven, plus or minus two: some limits on our capacity for processing information," Psychological Review, 63:81-97, 1956.

[7] Terrace, H. S., Jaswal, V., Brannon, E., and Chen, S., "What is a chunk? Ask a monkey.," Abstracts of Psychonomic Society, 1:35, 1996.

[8] Cowan, N., "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," Behavioral and Brain Sciences, 24:87-185, 2000.

[9] Boucher, V. J., "On the function of stress rhythms in speech: Evidence of a link with grouping effects on serial memory," Language and Speech, 49:495-519, 2006.

[10] Gilbert, A. C., Boucher, V. J., and Jemel, B., "Exploring the rhythmic segmentation of heard speech using evoked potentials," Proc. of the 5th Conference on Speech Prosody. 1-3 vol. 100 334, Chicago, USA, 2010.

[11] Gilbert, A. C., Boucher, V. J., and Jemel, B., "How listeners chunk speech: brain responses reveal a rhythm-based segmentation," submitted.

[12] Christophe, A., Gout, A., Peperkamp, S., and Morgan, J., "Discovering words in the continuous speech stream: The role of prosody," J. of Phonetics, 31:585-598, 2003.

[13] Shukla, M., Nespor, M., and Mehler, J., "An interaction between prosody and statistics in the segmentation of fluent speech," Cognitive Psychology, 54:1-32, 2007.

[14] Saffran, J. R., Newport, E. L., and Aslin, R. N., "Word segmentation: The role of distributional cues," J. of Memory and Language, 35:606-621, 1996.

[15] Gilbert, A. C. and Boucher, V. J., "The role of rhythmic chunking in speech: Synthesis of findings and evidence from statistical learning," in L. W.-S. and E. Zee, Eds. The 17th Intl. Congress of Phonetic Sciences. 747-750, Hong Kong, China: City University of Hong Kong, 2011.

[16] Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J., "Phonological phrase boundaries constrain lexical access I. Adult data," J. of Memory and Language, 51:523-547, 2004.

[17] Sternberg, S., "High-speed scanning in human memory," Science, 153:652-654, 1966.

[18] Wechsler, D., Wechsler Adult Intelligence Scale - Fourth Edition. San Antonio, TX: Pearson, 2008.

[19] Desrochers, A., "OMNILEX : Une base de données sur le lexique du français contemporain," Cahiers Linguistiques d'Ottawa, 34:25-34, 2006.

[20] Delattre, P., "A comparison of syllable length conditioning among languages," International Rev. of Applied Linguistics, 4:183-198, 1966.