

Université de Montréal

**Le chunking perceptif de la parole :
Sur la nature du groupement temporel et son effet
sur la mémoire immédiate**

par

Annie C. Gilbert

Département de Linguistique et de Traduction
Faculté des Arts et des Sciences

Thèse présentée à la Faculté des Arts et des Sciences
en vue de l'obtention du grade de docteur
en linguistique

mars 2012

© Annie C. Gilbert, 2012

Université de Montréal
Faculté des études supérieures et postdoctorales

Cette thèse intitulée :

Le chunking perceptif de la parole :
Sur la nature du groupement temporel et son effet
sur la mémoire immédiate

Présentée par :
Annie Gilbert

a été évaluée par un jury composé des personnes suivantes :

Richard Patry, président-rapporteur
Victor J. Boucher, directeur de recherche
Boutheina Jemel, co-directrice
Pheadra Royle, membre du jury
Sonja Kotz, examinatrice externe

.....
Représentant du doyen de la FES

Résumé

Dans de nombreux comportements qui reposent sur le rappel et la production de séquences, des groupements temporels émergent spontanément, créés par des délais ou des allongements. Ce « chunking » a été observé tant chez les humains que chez certains animaux et plusieurs auteurs l'attribuent à un processus général de *chunking* perceptif qui est conforme à la capacité de la mémoire à court terme. Cependant, aucune étude n'a établi comment ce *chunking* perceptif s'applique à la parole. Nous présentons une recension de la littérature qui fait ressortir certains problèmes critiques qui ont nui à la recherche sur cette question. C'est en revoyant ces problèmes qu'on propose une démonstration spécifique du *chunking* perceptif de la parole et de l'effet de ce processus sur la mémoire immédiate (ou mémoire de travail). Ces deux thèmes de notre thèse sont présentés séparément dans deux articles.

Article 1 : *The perceptual chunking of speech: a demonstration using ERPs*

Afin d'observer le *chunking* de la parole en temps réel, nous avons utilisé un paradigme de potentiels évoqués (PÉ) propice à susciter la *Closure Positive Shift* (CPS), une composante associée, entre autres, au traitement de marques de groupes prosodiques. Nos stimuli consistaient en des énoncés et des séries de syllabes sans sens comprenant des groupes intonatifs et des marques de groupements temporels qui pouvaient concorder, ou non, avec les marques de groupes intonatifs. Les analyses démontrent que la CPS est suscitée spécifiquement par les allongements marquant la fin des groupes temporels, indépendamment des autres variables. Notons que ces marques d'allongement, qui apparaissent universellement dans la langue parlée, créent le même type de *chunking* que celui qui émerge lors de l'apprentissage de séquences par des humains et des animaux. Nos résultats appuient donc l'idée que l'auditeur chunk la parole en groupes temporels et que ce *chunking* perceptif opère de façon similaire avec des comportements verbaux et non verbaux. Par ailleurs, les observations de l'Article 1 remettent en question des études où on

associe la CPS au traitement de syntagmes intonatifs sans considérer les effets de marques temporels.

Article 2 : Perceptual chunking and its effect on memory in speech processing: ERP and behavioral evidence

Nous avons aussi observé comment le *chunking* perceptif d'énoncés en groupes temporels de différentes tailles influence la mémoire immédiate d'éléments entendus. Afin d'observer ces effets, nous avons utilisé des mesures comportementales et des PÉ, dont la composante N400 qui permettait d'évaluer la *qualité* de la trace mnésique d'éléments cibles étendus dans des groupes temporels. La modulation de l'amplitude relative de la N400 montre que les cibles présentées dans des groupes de 3 syllabes ont bénéficié d'une meilleure mise en mémoire immédiate que celles présentées dans des groupes plus longs. D'autres mesures comportementales et une analyse de la composante P300 ont aussi permis d'isoler l'effet de la position du groupe temporel (dans l'énoncé) sur les processus de mise en mémoire.

Les études ci-dessus sont les premières à démontrer le *chunking* perceptif de la parole en temps réel et ses effets sur la mémoire immédiate d'éléments entendus. Dans l'ensemble, nos résultats suggèrent qu'un processus général de *chunking* perceptif favorise la mise en mémoire d'information séquentielle et une interprétation de la parole « chunk par chunk ».

Mots-clés : Segmentation de la parole, chunking, prosodie, allongements, potentiels évoqués, CPS, N400, P300, mémoire à court terme, mémoire immédiate.

Abstract

The perceptual chunking of speech: On the nature of temporal grouping and its effect on immediate memory

In numerous behaviors involving the learning and production of sequences, temporal groups emerge spontaneously, created by delays or a lengthening of elements. This chunking has been observed across behaviors of both humans and animals and is taken to reflect a general process of perceptual chunking that conforms to capacity limits of short-term memory. Yet, no research has determined how perceptual chunking applies to speech. We provide a literature review that bears out critical problems, which have hampered research on this question. Consideration of these problems motivates a principled demonstration that aims to show how perceptual chunking applies to speech and the effect of this process on immediate memory (or “working memory”). These two themes are presented in separate papers in the format of journal articles.

Paper 1: The perceptual chunking of speech: a demonstration using ERPs

To observe perceptual chunking on line, we use event-related potentials (ERPs) and refer to the neural component of *Closure Positive Shift* (CPS), which is known to capture listeners’ responses to marks of prosodic groups. The speech stimuli were utterances and sequences of nonsense syllables, which contained intonation phrases marked by pitch, and both phrase-internal and phrase-final temporal groups marked by lengthening. Analyses of CPSs show that, across conditions, listeners specifically perceive speech in terms of chunks marked by lengthening. These lengthening marks, which appear universally in languages, create the same type of chunking as that which emerges in sequence learning by humans and animals. This finding supports the view that listeners chunk speech in temporal groups and that this perceptual chunking operates similarly for speech and non-verbal behaviors. Moreover, the results question reports that relate CPS to intonation phrasing without considering the effects of temporal marks.

Paper 2: Perceptual chunking and its effect on memory in speech processing: ERP and behavioral evidence

We examined how the perceptual chunking of utterances in terms of temporal groups of differing size influences immediate memory of heard speech. To weigh these effects, we used behavioural measures and ERPs, especially the N400 component, which served to evaluate the *quality* of the memory trace for target lexemes heard in the temporal groups. Variations in the amplitude of the N400 showed a better memory trace for lexemes presented in groups of 3 syllables compared to those in groups of 4 syllables. Response times along with P300 components revealed effects of position of the chunk in the utterance.

This is the first study to demonstrate the perceptual chunking of speech on-line and its effects on immediate memory of heard elements. Taken together the results suggest that a general perceptual chunking enhances a buffering of sequential information and a processing of speech on a chunk-by-chunk basis.

Keywords: Speech segmentation, chunking, prosody, pre-boundary lengthening, event-related potentials, CPS, N400, P300, short-term memory

Table des matières

Liste des tableaux	vii
Liste des figures	viii
Liste des sigles et abréviations	x
Remerciements	xii
1. Introduction	1
1.1. Problématique et objectifs	1
1.1.1. Définition du chunk	1
1.1.2. Une définition opérationnelle du <i>chunking</i> : deux types de <i>chunking</i>	4
1.1.3. Le <i>chunking</i> de la parole : mise en perspective des approches linguistiques et stochastiques.	8
1.1.3.1. L'Approche linguistique et les marques prosodiques du <i>chunking</i>	12
1.1.3.2. L'approche stochastique et le <i>chunking</i>	16
1.1.4. La présente étude : le <i>chunking</i> perceptif de la parole et ses effets sur la mémoire 18	
1.1.4.1. Un indice du <i>chunking</i> perceptif de la parole : la <i>Closure Positive Shift</i>	20
1.1.4.2. Indices des effets du <i>chunking</i> sur la mémoire : les composantes N400, P300 et les mesures comportementales.	22
1.2. Sommaire des objectifs et hypothèses spécifiques.....	25
1.2.1. Objectif 1 : Démontrer que, lors de l'écoute, l'auditeur segmente (<i>chunk</i>) la parole selon des groupes temporels	25
1.2.2. Objectif 2 : Démontrer que le <i>chunking</i> de la parole reflète bien un <i>chunking</i> « d'entrée » et non « de sortie ».	25
1.2.3. Objectif 3 : Isoler les effets du <i>chunking</i> perceptif sur la MCT	26
2. Méthodologie générale.....	28
2.1. Participants	28
2.2. Stimuli	29
2.2.1. Conception et caractéristiques générales des stimuli	29

2.2.1.1.	Contrôles spécifiques appliqués aux contextes de l'ensemble 1 : Énoncés sensés	32
2.2.1.2.	Contrôles spécifiques appliqués aux contextes de l'ensemble 2 : Séries de syllabes	34
2.2.2.	Enregistrement et élaboration des stimuli	35
2.2.3.	Validation des stimuli	37
2.3.	Procédure et tâche expérimentale	39
2.4.	Enregistrement EEG et analyse des PÉ	40
2.5.	Analyses spécifiques à l'article 1	42
2.6.	Analyses spécifiques à l'article 2	43
3.	Résultats	45
3.1.	Article 1	46
3.2.	Article 2	78
4.	Discussion	98
5.	Conclusion	108
	Annexe 1 : Formulaire de consentement	xiii
	Annexe 2 : Profil général des participants	xx

Liste des tableaux

Corps de la thèse

Tableau 1 : <i>Expressions coréennes et leur équivalent français</i>	12
Tableau 2 : <i>Exemples d'énoncés</i>	33
Tableau 3 : <i>Structure syntaxique des groupes temporels utilisés</i>	33
Tableau 4 : <i>Exemples d'énoncés distracteurs</i>	34
Tableau 5 : <i>Exemples de séries de syllabes</i>	34

Article 1

Table 1: <i>Behavioural results as a function of TG length and position.</i>	89
--	----

Liste des figures

Corps de la thèse

Figure 1 : Analyse acoustique de la récitation d'un numéro de téléphone en français, anglais et allemand.	7
Figure 2 : Représentation schématique du déroulement d'un essai.	29
Figure 3 : Représentation schématique des schèmes prosodiques des stimuli.	30
Figure 4 : Représentation schématique des schèmes prosodiques des stimuli distracteurs.	31
Figure 5 : Représentation schématique de la distribution des cibles selon les conditions de longueur du GT (3 vs 4 syll.) et de position du GT (1 ^{er} vs 2 ^e).	32
Figure 6 : Exemple de calcul du centre perceptif (P-center).	37
Figure 7 : Comparaison des analyses acoustiques des guides prosodiques et des stimuli enregistrés (pointillés = stimuli enregistrés, losanges = guide prosodique)	38
Figure 8 : Déroulement d'un essai.	40
Figure 9 : Montage d'électrodes utilisé. Système 10-20 augmenté avec localisation de la mise à terre (G) et des références mastoïdes (M1, M2). Adapté de la documentation fournie par le fabricant (Easy Cap, www.easycap.de).	42

Article 1 (titres abrégés)

Figure 1. Temporal grouping in German, English, and French recitations of a ten-digit series. Note that grouping entails a relative lengthening of syllables at the ends of groups.	51
Figure 2. Schematic representation summarizing the structural prosody of stimuli in the four experimental conditions.	59
Figure 3. Measured acoustic attributes of the utterances	61
Figure 4. Acoustic stimuli with changing TGs and effects on ERPs.	65
Figure 5. Acoustic stimuli with changing IGs and effects on ERPs (cdns 1 vs. 3).	66

Figure 6. Acoustic stimuli with changing IGs and effects on ERPs (cdns 2 vs. 4).....	67
Figure 7. ERPs across trials and onset latencies of the initial TGs.....	68
Figure 8. Measured acoustic attributes of meaningless sequences of syllables.....	70
Figure 9. Acoustic stimuli (nonsense syllables) with changing TGs and effects on ERPs..	71
Figure 10. ERPs across trials and onset latencies of the initial TGs (nonsense syllables)..	72

Article 2 (titres abrégés)

Figure 1. Schematic representation of intended prosodic structure of the stimuli and the origin of the targets.	86
Figure 2. Time-course of a trial.....	87
Figure 3. Topographies and averaged ERPs for targets from TGs in 1 st and 2 nd positions.	91
Figure 4. Average amplitude (n=20) and latency of the P300.	91
Figure 5. Topography and averaged ERPs for targets from TGs of different lengths (4- vs 3-syll.).....	92
Figure 6. Average amplitude (n=20) of N400.....	93

Liste des sigles et abréviations

APF	Allongement précédant la finale
dB	Décibel
F ₀	Fréquence fondamentale
GT	Groupe temporel
GI	Groupe intonatif
MCT	Mémoire à court terme
MLT	Mémoire à long terme
ms	Millisecondes
PT	Probabilité de transition
RdI	Région d'intérêt
s	Seconde
syll.	Syllabe
TR	Temps de réaction
μV	Microvolts

À Monsieur B. et Madame J.

Remerciements

Merci aux membres du Laboratoire de sciences phonétiques de l'Université de Montréal et du Laboratoire de recherche en neuroscience et électrophysiologie cognitive de l'Hôpital Rivière-des-Prairies pour leur support.

Les présents travaux ont été financés par des bourses de recherche doctorales du Conseil de Recherches en Sciences Humaines du Canada (CRSH) et du Fonds Québécois de Recherche sur la Société et la Culture (FQRSC).

1. Introduction

1.1. Problématique et objectifs

Lorsqu'on étudie le langage oral, on doit toujours tenir compte du fait que le traitement d'un stimulus acoustique (la parole entendue) est soumis à des contraintes qui structurent les processus de traitement. Une contrainte évidente découle du fait que le signal sonore est transitoire ou éphémère : afin d'interpréter le signal, l'auditeur doit garder en mémoire à court terme, ou « mémoire de travail » (MdeT), les séquences de sons entendues. Toutefois, nos capacités mnésiques et attentionnelles sont limitées en ce qu'on ne peut maintenir efficacement en MdeT plus de trois ou quatre « items » à la fois (voir Chen & Cowan, 2005; Cowan, 2000, 2005). Cette limite impose donc une segmentation, ou *chunking*, de l'information séquentielle de la parole en groupes ou en unités qui pourront être stockés en MdeT et maintenus actifs dans le « foyer attentionnel » (Cowan, 2000, 2005). Dans les pages qui suivent, nous appellerons ces unités des chunks. Les chercheurs s'entendent sur le besoin de segmenter le signal entrant (Kurby & Zacks, 2008) selon les contraintes de la mémoire en général¹ (Buschke, 1976). Cependant, il n'y a pas de consensus sur la manière dont l'auditeur découpe les séries de sons lorsqu'il écoute de la parole. En effet, contrairement à l'écrit où les espaces et la ponctuation servent à séparer les « mots » et les « syntagmes », aucun indice acoustique connu ne permet de segmenter l'oral aussi facilement. C'est ce problème qui est au centre du présent travail. L'objectif général de la présente recherche est donc de définir ce qui guide le processus de *chunking* dans la perception de la parole.

1.1.1. Définition du chunk

L'utilité générale du *chunking* dans l'apprentissage et la production de comportements moteurs sériels fait déjà l'objet de consensus dans la littérature. Le *chunking* se manifeste par un regroupement d'items perçus (ou produits) et on peut facilement l'observer à l'aide de tâches de rappel sériel. Par exemple, lorsqu'un locuteur

¹ « It appears that spontaneous clustering of items into small, well-defined chunks, which also are organized, is a basic process in natural free recall » (Buschke, 1976, p.316)

doit rappeler des séries de lettres ou de chiffres, il ne les produit pas de façon isochrone, mais crée plutôt des groupements de quelques items (voir, p. ex., Boucher, 2006; Frick, 1989; Ryan, 1969a, 1969b; Thrope & Rowland, 1965; Wickelgren, 1967; Wilkes, 1975; Wilkes, Lloyd, & Simpson, 1972; Winzenz & Bower, 1970). Les auteurs qui tentent d'expliquer ce phénomène font le plus souvent référence à l'article classique de Miller sur le *chunking* (Miller, 1956). L'idée de Miller est que la mémoire à court terme (MCT, ou mémoire de travail, MdeT, Baddeley, 2010) est limitée. Néanmoins, un individu peut réussir à mémoriser un grand nombre d'items en les recodant en blocs avant de les stocker en mémoire à long terme (MLT). Par exemple, la séquence de chiffres « 1, 3, 2, 0, 1, 6, 5, 9 » peut être mémorisée telle quelle (soit huit items), ou par blocs recodés (13, 20, 16, 59; donnant quatre items à mémoriser). Un autre exemple connu de ce recodage est la séquence de lettres « *I, B, M, F, B, I, C, I, A, I, R, S,* » qui peut être mémorisée en créant des blocs recodés selon des acronymes connus comme *I.B.M., F.B.I., C.I.A.* et *I.R.S.* Cette conception du bloc en tant qu' « unité recodée » a eu un impact considérable sur la recherche, tout particulièrement sur les travaux portant sur l'organisation des comportements sériels. Dans cette optique, le *chunking* est souvent considéré comme un moyen de regrouper des séquences motrices en unités pour répondre aux contraintes de la MCT. Ces processus sont considérés comme essentiels à la compréhension des comportements impliquant un contrôle sériel, allant de la production de mouvements simples comme la préhension d'objets jusqu'à la production de la parole (Lashley, 1951).

Cependant, certains critiques font remarquer qu'un concept de chunk basé sur un principe de recodage verbal, tel que décrit par Miller, est inapplicable au *chunking* observé lors de l'acquisition de séquences non verbales, tant chez l'humain que chez l'animal (Terrace, 2001). De telles observations indiquent que les processus de *chunking* relèvent plutôt d'un principe plus fondamental pouvant s'appliquer à tout un ensemble de comportements, y compris le traitement de la parole.

Sur l'existence d'un principe général de *chunking*, il est intéressant de noter qu'on porte peu d'attention à un autre article de Miller où l'auteur a proposé une autre conception

du *chunking* (Miller, 1962). Dans cet article, Miller soutient que les auditeurs n'interprètent pas la parole « son par son », au fur et à mesure qu'elle leur parvient. Ils attendent plutôt d'avoir perçu une certaine étendue de signal constituant une « unité perceptive » qui, selon Miller (1962), serait de la taille d'un syntagme (*phrase*, Miller, 1962, p.82). Toutefois, cette forme de *chunking* n'implique pas de recodage des items entendus, mais plutôt un regroupement d'items séquentiels dans les premières étapes de traitement du signal entendu. Notons que cette conception du *chunking* s'applique aussi bien aux séquences verbales que non verbales et se rapproche d'une conception plus générale, présentée dans les paragraphes subséquents, et qui réfère à un *chunking* perceptif ou « sensori-moteur » (Graybiel, 1998). Cependant, sans une définition formelle du chunk, il demeure difficile de rattacher un principe de *chunking* à des unités observables dans la parole.

Dans les pages qui suivent, nous présenterons une recension critique de la littérature qui porte sur le *chunking* et son application à la perception de la parole. L'objectif général de cette recension est de démontrer les principes motivant notre hypothèse d'un *chunking* perceptif de la parole correspondant à des marques de groupement temporel qui sont généralisables à un ensemble de comportements verbaux et non verbaux. La démonstration se fait en trois parties. Une première section sert à clarifier la notion de *chunking* de façon à rendre ce terme opérationnel par rapport à des marques observables au plan des comportements (section 1.1.2.). Par la suite, nous clarifions les problèmes conceptuels et techniques qui se présentent lorsqu'on tente de démontrer comment le *chunking* perceptif s'applique à la parole. En particulier, il est fondamental de mettre au clair la difficulté de relier des indices universels de *chunking* perceptif aux unités linguistiques traditionnelles qui guident certaines conceptions du *chunking* dans des approches axées sur l'apprentissage stochastique (section 1.1.3.). Ces considérations nous amènent à élaborer une démonstration spécifique du *chunking* perceptif et de ses effets sur la MdeT au moyen de la technique de potentiels évoqués (PÉ; section 1.1.4.). L'utilisation de la technique des PÉ est largement motivée par le fait que le *chunking* perceptif est un processus qui opère lors de l'écoute. En somme, les PÉ, par l'intermédiaire de certaines composantes, nous permettent

d'établir comment un entendant perçoit la parole « en blocs » et comment ce processus facilite la mise en mémoire immédiate d'un signal changeant.

1.1.2. Une définition opérationnelle du *chunking* : deux types de *chunking*

Dans le but de clarifier la terminologie utilisée dans le présent travail, il est important de noter que le concept de *chunking* que l'on retrouve dans des études sur les comportements non verbaux (chez l'humain ou l'animal) diffère considérablement de celui introduit par Miller (1956). Rappelons que l'approche de Miller était plutôt intuitive et imprécise quant à ce qui constitue un chunk d'information (comme l'auteur le rappelle lui-même : « [...] we are not very definite about what constitutes a chunk of information. » Miller 1956, p. 93). Certains critiques déplorent d'ailleurs que les définitions courantes du chunk soient tout aussi imprécises que celle de Miller (voir surtout les analyses de Mathy & Feldman, 2012, et Terrace, 2001). Par exemple, des auteurs caractérisent les chunks comme des « groupes d'éléments » (Anderson & Matessa, 1997; Cowan, 2000; Cowan, Chen, & Rouder, 2004; Farrell, 2008; Ng & Maybery, 2002) ou réfèrent à un principe de recodage sémantique (Hitch, Burgess, Towse, & Culpin, 1996; Mathy & Feldman, 2012; Shiffrin & Nosofsky, 1994; Simon, 1974; Stark & Calfee, 1970). Il va de soi que ce principe ne peut s'appliquer au *chunking* que l'on observe dans des comportements non verbaux. D'autres définitions sont circulaires comme celle voulant qu'un chunk soit un ensemble ou groupe d'éléments qui sont fortement associés entre eux, mais faiblement associés avec les éléments d'autres chunks² (Cowan, 2000; Gobet, et coll., 2001; Perruchet & Pacton, 2006; Servan-Schreiber & Anderson, 1990; Simon, 1974.)

En reconnaissant ces imprécisions et les limites du principe de recodage sémantique, Terrace (2001) insiste sur la nécessité de faire une distinction entre les chunks « d'entrée » (*input chunks*), imposés par les limites de la MCT, et les chunks dits « de sortie » (*output chunks*), relevant de formes en MLT. Les chunks de sortie reflètent des mécanismes où des formes stockées en MLT peuvent être activées pour produire une action

² « [...] a collection of elements having strong associations with one another, but weak associations with elements within other chunks », (Gobet et al., 2001, p. 236).

ou pour reconnaître, dans un stimulus, des schèmes ou des groupements unitaires d'items ayant déjà fait l'objet d'un *chunking* à une étape antérieure. En revanche, le chunk d'entrée en lien avec des processus plus fondamentaux de *chunking perceptif* concorde avec les contraintes de la MCT et du foyer attentionnel (voir Cowan, 2000 pour une discussion du rôle du foyer attentionnel). Ce processus fait que l'on peut percevoir des éléments séquentiels non pas un à la fois, mais « par groupe ». Il s'agit d'un *chunking* que Graybiel (1998) a qualifié de « sensori-moteur » présent chez certains animaux. Ce processus s'appliquerait généralement lors du traitement de stimuli séquentiels (verbaux ou non) et contribuerait à la formation d'unités (de chunks) primaires pouvant être stockées en MLT, ou qui faciliteraient l'accès à des unités stockées. Cette division du *chunking* en deux mécanismes distincts est aussi adoptée par Gobet et coll. (2001) pour pallier au caractère circulaire de leur définition du chunk portant sur les associations entre les éléments composant un chunk. Notons qu'une définition axée sur les associations entre les éléments présuppose un *chunking* perceptif. Autrement dit, pour que des liens entre des éléments séquentiels se créent et forment un chunk en MLT, il faut d'abord que les séquences soient perçues en unités. En reconnaissant ce principe, Gobet et coll. réfèrent à deux formes de *chunking*, dont l'une est consciente et ciblée (angl. *goal-oriented chunking*), alors que l'autre repose sur des processus dits plus automatiques de *chunking* perceptif (angl. *perceptual chunking*, correspondant au *chunking* d'entrée de Terrace, 2001). Pour illustrer les processus du *chunking* perceptif, Gobet et coll. donnent l'exemple de la lecture, où il y a un regroupement des lettres comprises entre deux espaces (*chunking* perceptif) pour former et reconnaître un mot (*chunking* ciblé). Il demeure toutefois difficile d'appliquer cette conception du *chunking* au traitement des stimuli acoustiques tel la parole qui ne contient pas de marques artificielles telles que des espaces entre les mots. La parole contient plutôt des marques naturelles pouvant refléter un *chunking* perceptif ou sensori-moteur. En fait, l'identification de ces marques devient centrale à une définition fonctionnelle du chunk perceptif.

Sur cette question, Terrace (2001) souligne que l'application superficielle du concept de *chunking* dans plusieurs domaines d'études a fait que bon nombre d'auteurs ne

reconnaissent pas les marques universelles du chunk. Selon lui, le chunk d'entrée possède des indices observables en termes des latences entre les éléments, ou ce que Terrace (2001) appelle « l'intervalle interréponses » (angl. *inter-response times*). L'allongement de cet intervalle par des délais entre les éléments séquentiels ou des prolongements d'éléments crée des groupements et de tels chunks ont été signalés dans nombre de comportements, chez l'humain et chez l'animal (pour des exemples, voir Graybiel, 1998; Terrace, 2001; Terrace & Chen, 1991; Terrace, Jaswal, Brannon, & Chen, 1996).

Dans le but d'illustrer comment de tels marqueurs temporels peuvent s'appliquer à la parole, la Figure 1 ci-dessous présente une analyse acoustique d'énoncés représentant la récitation d'un numéro de téléphone en trois langues différentes. La production de telles séquences s'accompagne habituellement d'un *chunking* spontané où un locuteur produira des groupements de deux à quatre items. Notons que de tels groupements apparaissent aussi avec des séquences de syllabes sans sens ou des gestes, et donc, ne reflètent pas un codage linguistique ou culturel (voir, p. ex., Boucher, 2006). Si on examine les attributs acoustiques des signaux de la figure 1, on peut observer les paramètres du *chunking* auxquels Terrace réfère.

Dans cette figure, l'intonation est représentée par la courbe de fréquence fondamentale (F_0) alors que la durée relative des chiffres (monosyllabes) est représentée par les traits horizontaux. En examinant ces deux paramètres, on voit qu'ils peuvent varier de façon indépendante. Les schèmes de la F_0 fluctuent d'une langue à l'autre. Par contre on voit que, de façon systématique, il y a des allongements relatifs des monosyllabes en finale de groupes (voir p.ex. les chiffres 1, 2, 3, 4 et 8 qui sont plus longs à la fin qu'à l'intérieur des groupes). Ces allongements sont subtils et peuvent être accompagnés de pauses. Cependant, il faut noter que ces changements en final de groupe modifient l'intervalle entre les syllabes successives, et donc, correspondent à un changement dans l'intervalle interréponse. Les marques de groupes temporels en parole se comportent donc comme les marques de *chunking* dans la production des séquences non verbales identifiées par Terrace (2001). Une telle correspondance suggère fortement que le *chunking* de la parole obéit aux

mêmes contraintes que le *chunking* non verbal, pouvant refléter un principe général de *chunking* perceptif.

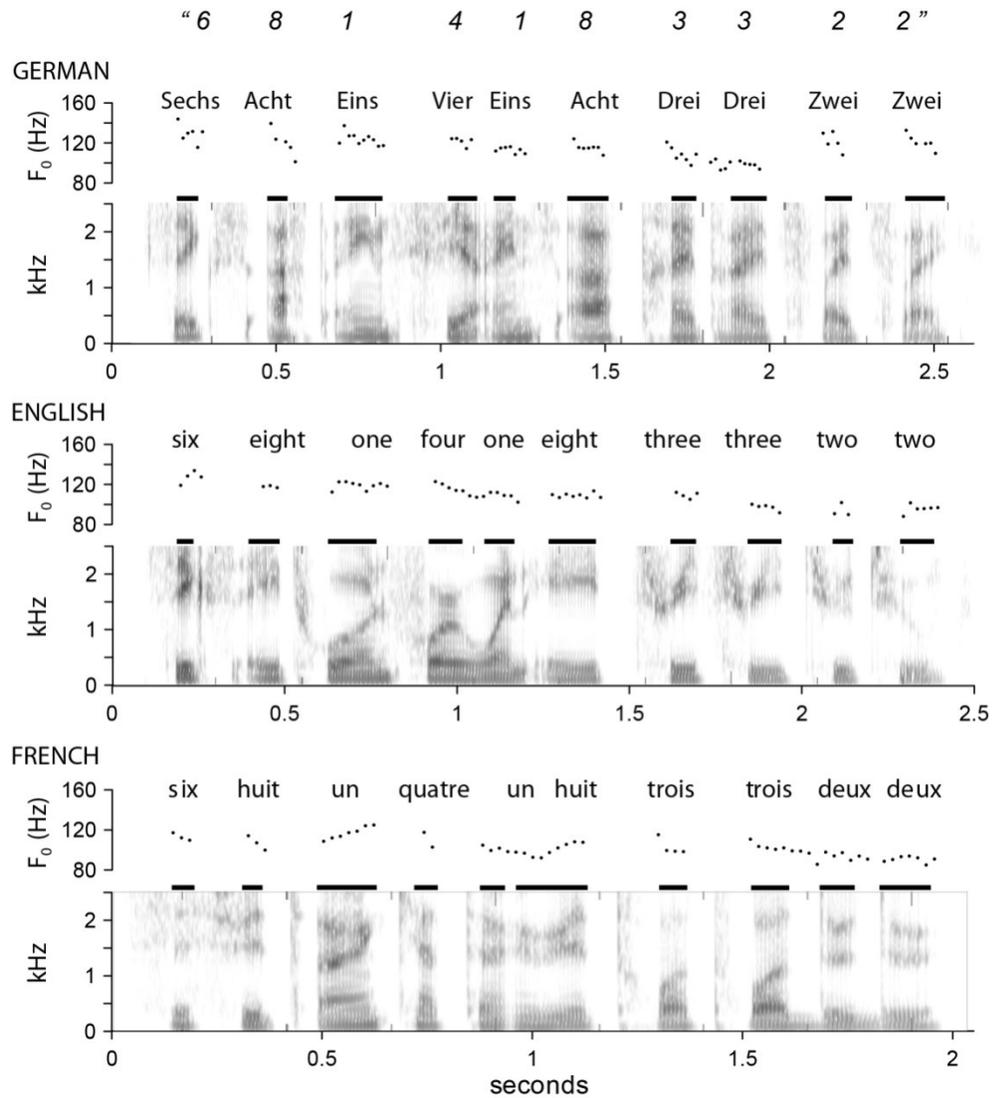


Figure 1 : Analyse acoustique de la récitation d’un numéro de téléphone en français, anglais et allemand.

En regard de ces observations, la présente thèse a pour but de déterminer si les auditeurs effectuent un *chunking* perceptif de la parole selon des groupements temporels comme ceux illustrés à la figure 1. Si le *chunking* de la parole relève d'un phénomène général de *chunking* perceptif, alors il devrait s'appliquer de la même manière à des énoncés avec ou sans sens. Toutefois, certains pourraient argumenter que la parole sensée n'est pas traitée comme une séquence aléatoire de chiffres ou de syllabes, mais implique la reconnaissance d'éléments stockés en MLT et des unités syntaxiques familières, comme des « mots » et des « syntagmes » (ou *phrase* comme l'entend Miller, 1962). Si tel est le cas, alors l'écoute d'énoncés sensés et de séquences de syllabes sans sens devrait impliquer différents types de *chunking*, soit un *chunking* « de sortie » dans le cas d'énoncés, et un *chunking* perceptif ou « d'entrée » dans le cas de séquences sans sens. Bon nombre de travaux portant sur la segmentation de la parole ont tenu pour acquis la première option et présument que la parole est perçue « en mots » ou « en syntagmes ». Ce présupposé, qui caractérise les approches linguistiques et stochastiques de la segmentation, ignore le rôle central du *chunking* perceptif, comme nous l'expliquons dans les sections suivantes.

1.1.3. Le *chunking* de la parole : mise en perspective des approches linguistiques et stochastiques.

Les arguments ci-dessus suggèrent que les allongements marquant les groupes temporels en parole peuvent être associés à un processus universel de *chunking* perceptif. Si tel est le cas, les marques d'allongement final de groupe devraient être identifiables dans l'ensemble des langues naturelles. Plusieurs études suggèrent en effet le statut universel de ces allongements, qui sont le plus souvent identifiés comme des marques prosodiques d'« allongements précédant la frontière » (APF, angl. : *pre-boundary lengthening*). Les APF ont été observés dans plusieurs langues, dont l'anglais (Lehiste, Olive, & Streeter, 1976), le français (Fletcher, 1991), le suédois (Lindblom, 1968), le néerlandais (Gussenhoven & Rietveld, 1992), l'espagnol (Delattre, 1966a), l'allemand (Delattre, 1966a), le russe (Makarova, 2001), le mandarin (Chow, 2008), le cantonais (Chow, 2008), le japonais (Fon, 2002), le guoyu (Fon, 2002), le putonghua (Fon, 2002), de même que dans des langues

finno-ougriennes comme le hongrois, où la durée est distinctive (Hockey & Fagyal, 1998; voir aussi Vaissière, 1983). Du point de vue de l'ontogenèse, il n'y a pas de consensus quant à la production d'APF chez l'enfant (voir entre autres Bion, Benavides-Varela, & Nespor, 2011). En fait, l'étude de la parole enfantine soulève des difficultés, car on néglige souvent de distinguer les APF des allongements en fin d'énoncés. Une exception récente est l'étude de DePaolis et coll. portant sur la parole enfantine en quatre langues (anglais, finnois, français et gallois; DePaolis, Vihman, & Kunnari, 2008). Cette étude a révélé que l'APF marque généralement des formes bisyllabiques à une étape où les enfants commencent à acquérir un vocabulaire (« as infants begin to use words », DePaolis, et coll., 2008, p.409). Autrement dit, alors que les enfants commencent à apprendre des séquences de syllabes constituant de nouvelles formes verbales, des marques de groupement apparaissent sous la forme d'APF. En revanche, l'étude a révélé qu'il n'y avait aucune variation systématique de la F_0 ou de l'amplitude marquant les formes bisyllabiques. Par ailleurs, il n'y avait aucun patron systématique de la F_0 du type « iambique » ou « trochaïque » (« faible-fort » ou « fort-faible ») selon les langues. Dans l'ensemble, donc, l'APF semble constituer un marqueur prosodique généralement présent dans la parole et qui apparaît avec l'apprentissage de séquences constituant des unités verbales. Toutefois, aucune étude n'a démontré l'existence d'un lien entre ce marqueur et un processus de *chunking* perceptif, et ce, même si ce processus constitue un principe essentiel de segmentation de la parole (voir Beckner & Bybee, 2009). Ce problème de segmentation a, depuis peu, gagné en importance, surtout dans le cadre de travaux sur l'acquisition de la parole (Endress & Hauser, 2010). Cependant, de nouvelles données issues d'études de corpus appuient un changement fondamental de perspective, remettant en question certains présupposés des approches linguistiques et stochastiques.

Pour saisir l'impact de ces nouvelles données, rappelons que la recherche sur la segmentation de la parole visait depuis longtemps à mettre à jour les indices acoustiques qui permettent de diviser des unités linguistiques comme le « mot » (Cutler, Mehler, Norris, & Segui, 1986, 1992; Cutler & Norris, 1988; Endress & Hauser, 2010; Houston, Santelmann, & Jusczyk, 2004; Johnson & Jusczyk, 2001; Johnson, Jusczyk, Cutler, &

Norris, 2003; Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Houston, & Newsome, 1999; Mattys, Jusczyk, Luce, & Morgan, 1999; Mattys & Samuel, 1997, 2000; McQueen, Otake, & Cutler, 2001; Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006; Norris, McQueen, Cutler, & Butterfield, 1997; Norris, McQueen, Cutler, Butterfield, & Kearns, 2001; Otake, Hatano, Cutler, & Mehler, 1993). Face à l'absence d'indices acoustiques généraux permettant d'isoler les mots, des auteurs ont affirmé que ces unités ne se retrouvent tout simplement pas dans le signal de parole (p. ex., Jackendoff, 2007, p. 378). De même, plusieurs études ont révélé que les enfants ne manipulent pas plusieurs des unités qui sont généralement assumées dans une analyse linguistique traditionnelle.

Par exemple, il a été démontré que, même à trois ans, les enfants ne manipulent pas des unités séparées correspondant à des verbes, des pronoms et des déterminants, mais produisent des expressions lexicalisées ou formules en *chunks* (Dabrowska, 2004; Lieven, Behrens, Speares, & Tomasello, 2003; Lieven, Pine, & Baldwin, 1997; Pine & Lieven, 1997; Pine & Martindale, 1996; Tomasello, 1992, 2000a, 2000b). Ces données ont incité certains auteurs à émettre des réserves quant à l'utilisation d'unités linguistiques *a priori* associées à l'écriture alphabétique (Parsse, 2005; Tomasello, 2003, p. 18). Les tenants des approches basées sur l'usage (*usage-based grammar*) ont d'ailleurs souligné le fait que plusieurs formules ou expressions figées n'impliquent pas des unités séparées, mais fonctionnent plutôt comme des « chunks lexicaux » (Bates & Goodman, 2001; Bybee, 2006; Dabrowska, 2004; Wray, 2002, 2008). Des formules ou expressions lexicalisées comme *vient t'en, j pense que X, tsé (tu sais), ça va, s'il vous plaît, ça s peut pas, mettre en branle, prends garde, marché conclu*, etc. forment la majeure partie de l'expression verbale des adultes (Dabrowska, 2004). Altenberg avance même qu'elles représentent 70 % de l'expression verbale d'un locuteur natif adulte (Altenberg, 1990). Pour décrire ces blocs, une variété de termes est apparue récemment dans la littérature spécialisée; par exemple, « collocations », « composites », « formules figées », « expressions préfabriquées », etc. (voir p. ex. Wray & Perkins, 2000, qui présentent une liste d'une quarantaine de termes).

Précisons toutefois que certains analystes voient dans les expressions idiomatiques (comme *tomber dans les pommes*) ou mots composés (comme *porte-monnaie*) des élaborations syntaxiques ou des formes analysables. Nous ne faisons pas référence ici à ce type d'expressions qui peuvent se diviser, mais plutôt à des blocs qui sont stockés et activés tels quels lors de la production et non pas générés grammaticalement³ (Altenberg, 1990; Beckner & Bybee, 2009; Bybee, 2010; Dabrowska, 2004; Wray, 2008; Wray & Perkins, 2000). L'acquisition de telles unités suppose un *chunking* perceptif à l'entrée, lors de l'apprentissage des séquences verbales. Ainsi, on ne peut supposer que l'enfant ou l'adulte qui apprend une langue seconde soit aussitôt en mesure de reconnaître les catégories grammaticales qui pourraient être contenues dans ces expressions et reconnues par un locuteur expérimenté. En guise d'illustration, imaginons un apprenant exposé aux expressions coréennes du tableau 1. Il serait difficile de présumer qu'un apprenant puisse de prime abord déterminer la frontière des « mots » dans ces expressions ou le rôle syntaxique des constituants. Il ne peut au plus que reconnaître des chunks et formules (*An-nyeong / h- X-yo*) et leur assigner une interprétation générale en contexte. En fait, ce n'est qu'après avoir stocké un très grand nombre de chunks qu'il pourra en analyser les formes (Beckner & Bybee, 2009; Dabrowska, 2004; Wray, 2008; Wray & Perkins, 2000).

Dans cette perspective, plusieurs études de corpus suggèrent l'existence dans la langue usuelle d'innombrables chunks ou formules figées qui ne correspondent pas à une division traditionnelle de la parole en « mots ». Cela amène à définir un processus général pour expliquer l'émergence de ces chunks (comme le signalent Beckner & Bybee, 2009). Dans les sections suivantes, nous situons brièvement comment cette nouvelle problématique remet en cause les perspectives linguistiques et stochastiques de la segmentation de la parole.

³ Soit « a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. » (Wray & Perkins, 2000, p.1).

Tableau 1 : Expressions coréennes et leur équivalent français. (tiré de « cours de coréen », consulté le 20 Juillet, 2012 <http://www.cours-coreen.fr>)

Expressions coréennes (translittérations françaises)	Équivalent français
<i>An-nyeong ha-se-yo!</i>	Bonjour !
<i>An-nyeong-hi ju-mu-se-yo!</i>	Bonne nuit ! (dormez bien!)
<i>An-nyeong-hi ga-se-yo!</i>	Au revoir ! (à celui qui part : "partez en paix")
<i>An-nyeong-hi gye-se-yo!</i>	Au revoir ! (à celui qui reste : "restez en paix")

1.1.3.1. L'Approche linguistique et les marques prosodiques du *chunking*

Par rapport aux recherches ci-dessus, la multiplication des termes pour désigner les chunks ou formules figées souligne la difficulté de conceptualiser les unités du langage oral par rapport à des concepts linguistiques traditionnels comme le mot. En effet, il est difficile de relier les chunks perceptifs de la parole formés par les APF aux unités de la linguistique formelle. Cette difficulté se manifeste à plusieurs niveaux. D'abord, il est possible que les chunks universaux ne correspondent pas aux mêmes unités morphologiques d'une langue à l'autre. De plus, certaines langues présentent des marques prosodiques codées qui peuvent compliquer l'observation des APF. Prenons l'exemple de langues à accent lexical comme l'anglais. Ce type d'accent est marqué par une montée d'amplitude et de F₀, de même que par un allongement de la syllabe accentuée (Fry, 1955). Les énoncés de cette langue présentent donc deux types d'allongements utilisés en parallèle (lexical et AFP) qui peuvent s'influencer (voir, entre autres, Cutler, Dahan, & van Donselaar, 1997; Delattre, 1966a; Fox, 2000). Néanmoins, des études ont démontré que l'APF s'observe en contrôlant la position des accents lexicaux (Klatt, 1975; Lindblom & Rapp, 1973; Oller, 1973). Il a aussi été démontré que les allongements (en général) sont plus efficaces comme indices de frontière d'unités que les variations de F₀ et d'amplitude (Fon, 2002; Nooteboom, Brokx, &

Rooij, 1978; Rooij, 1978; Wagner & Watson, 2010). En somme, l'identification des marques d'APF exige le contrôle de plusieurs facteurs confondants, dont l'accent lexical.

En particulier, il est important de noter que, dans une approche linguistique, les marques prosodiques (APF ou autres) sont interprétées en fonction d'unités syntaxiques qui varient d'une langue à l'autre. Cette approche rend très difficile une association entre des marques universelles d'APF et un principe général de *chunking*. En fait, les descriptions linguistiques présentent le plus souvent de multiples interprétations des marques prosodiques universelles. Par exemple, l'APF est interprété comme un marqueur de fin de « mot » (Umeda & Quinn, 1981), de « syntagme » (Ladd, 1986), de « phrase » (Klatt, 1975), ou de « paragraphe » (Vaissière & Michaud, 2006). Cette perspective d'un lien entre la prosodie et les unités syntaxiques est reprise dans des modèles phonologiques qui postulent une structure prosodique hiérarchique similaire à la structure syntaxique formelle⁴. Ainsi, certains phonologues conçoivent l'APF comme un marqueur de syntagme intonatif (angl. *intonation phrase*; p. ex. Ladd, 1986, 1996; Nespor & Vogel, 1986; Selkirk, 1982). Cependant, l'APF apparaît aussi à l'intérieur des syntagmes (Beckman & Edwards, 1990) de sorte qu'il ne marque pas spécifiquement des fins de syntagmes (ou d'énoncés), mais peut en fait refléter des groupes temporels plus courts. Par contre, il n'y a pas de consensus sur les unités prosodiques marquées par l'APF. Plusieurs termes sont utilisés et varient d'un analyste à l'autre et selon la langue décrite (par exemple : le « mot phonologique » en anglais, le « groupe rythmique » en français, le « syntagma » en russe ou le « buntetsu » en japonais, Vaissière, 1983). Notons aussi que les unités de la hiérarchie prosodique mentionnée ci-haut ne s'appliquent pas à toutes les langues. Par exemple, le mot phonologique a été défini comme étant une unité pouvant porter un accent lexical ou « accent de mot » et correspondant environ à un mot⁵ (Selkirk, 1984). Or plusieurs langues,

⁴ Selon ces modèles, un « énoncé » est constitué d'au moins un « syntagme intonatif » (*intonation phrase*), qui englobent au moins un « mot phonologique » (*phonological word*) parmi d'autres unités (voir Nespor & Vogel, 1986, p. 16; Selkirk, 1986, p. 384; ainsi que Beckman & Edwards, 1990; Jun & Fougeron, 1997; Ladd & Campbell, 1991).

⁵ Prosodic word : « a roughly word-sized unit » (Selkirk, 1984, p. 30); Phonological word : « any level of prosodic constituent structure that may include one or more major category words » (idem, p.29)

dont le français, n'ont pas d'accent lexical. En somme, les descriptions linguistiques des marques d'APF ne permettent pas d'isoler ou de conceptualiser le lien entre les marques de groupes temporels et un processus universel de *chunking* perceptif.

Par ailleurs, il est utile de mentionner que certaines analyses linguistiques portant sur la prosodie intonative tendent à omettre toute description d'APF en tant que marques indépendantes (Beckman & Pierrehumbert, 1986; Ladd, 1983, 1986, 1996). C'est le cas des analyses phonologiques de langues comme l'anglais, où le rythme est souvent défini par rapport à l'accent lexical (pour des détails historiques, voir Fox, 2000). Suivant cette tradition, les phonologues ont modélisé le rythme de la parole (ou sa métrique) en se basant sur des langues à accent lexical où la tendance est de créer des alternances de syllabes « fortes » et « faibles » (accentuées et non accentuées) regroupées à l'intérieur d'un pied phonologique (Liberman, 1975; Liberman & Prince, 1977). Ainsi, bien qu'un accent lexical soit marqué par plusieurs paramètres acoustiques (Fry, 1955; Lehiste, 1970), les tenants d'approches intonatives et métriques tendent à concevoir l'accent lexical comme étant essentiellement associé à des variations de F_0 (p. ex. : Zheng & Pierrehumbert, 2010). Ce faisant, ils perpétuent l'idée du « *stress-as-accent* » de Bolinger (1958, pp. 111 et 149). Cette préférence pour la F_0 se base sur la présomption que les syntagmes contiennent des marques de proéminence, ou « accents intonatifs » (*pitch accents*), qui peuvent coïncider ou non avec les syllabes portant l'accent lexical (Beckman & Pierrehumbert, 1986; Pierrehumbert, 2000). Les analyses prosodiques faites dans ce cadre théorique ignorent les APF comme marques séparées. Ainsi, le « rythme » y est décrit par référence à des alternances de tons, hauts ou bas, portés par les syllabes accentuées, et aucune mention n'est faite des APF qui ne coïncideraient pas avec une frontière de syntagme intonatif (ou *intonation phrase*). Ce biais se manifeste d'ailleurs dans le système de notation *Tone and Break Indices* (ToBI, Beckman & Ayers Elam, 1997) utilisé pour transcrire la prosodie. Ce système n'offre aucun moyen de noter les APF et repose sur les impressions de

proéminences perçues⁶. Ce type d'analyses prosodiques n'est donc pas approprié pour observer les APF comme marqueur de chunk dans la parole.

Par rapport aux approches ci-dessus, notons qu'il est démontré que l'APF et la F_0 sont contrôlés et perçus de façon indépendante (Cooper & Sorensen, 1981; Loeb & Allen, 1993; Seddoh, 1999; Snow, 1994). Un exemple évident provient des langues à tons où les tons et la durée sont contrôlés et appris séparément (Gandour, Petty, & Dardarananda, 1988; Li & Thompson, 1978; So & Dodd, 1995). Dans un même ordre d'idées, un locuteur est en mesure de créer des modulations d'intonation à la fin d'un énoncé à des fins expressives sans qu'elles ne soient influencées par les marques d'allongement (Cooper & Sorensen, 1981). Enfin, plusieurs études cliniques démontrent que l'intonation et la durée sont contrôlées séparément. Par exemple, les locuteurs présentant une aphasie non fluente manifestent généralement des déficits sévères dans le contrôle des durées en parole, mais conservent une intonation à peu près normale (voir p. ex. Gandour, 1998a, 1998b; Seddoh, 2004; Seddoh & Robin, 2001). Aussi, plusieurs auteurs appuient la thèse selon laquelle le contrôle des facteurs temporels se fait dans l'hémisphère gauche, alors que le contrôle de l'intonation se fait dans l'hémisphère droit (Alcock, Wade, Anslow, & Passingham, 2000; Zatorre & Belin, 2001). L'ensemble de ces études tend à démontrer que la durée et l'intonation sont des variables indépendantes, ce qui remet en question l'exactitude des systèmes de description linguistique comme ToBI.

C'est en regard des considérations ci-dessus que nous n'avons pas référé aux approches linguistiques et phonologiques dans le présent travail. Notre objectif étant d'observer le *chunking* perceptif en fonction des marques d'APF, il aurait été inutile de nous attarder à des analyses ne permettant pas d'isoler ces marques ni de concevoir leurs liens avec un *chunking* perceptif qui, rappelons-le, n'est pas un principe « linguistique »,

⁶ Pour une bonne mesure de la subjectivité dans la description des proéminences, voir le guide d'application du système ToBI qui justifie la notation d'une proéminence dans certains exemples sonores (disponibles sur le web) en disant qu'ils « donnent une forte impression d'accent » (« sounds very prominent, giving a strong subjective impression of accent. » p.28) ou qu'ils « sont ressentis comme plus proéminents » (« [...] in utterance [...] for example, the word *marmalade* feels more prominent than Marianna. » p.11). (Beckman & Ayers Elam, 1997)

mais un processus général qui s'applique à des comportements verbaux et non verbaux. Il va sans dire que notre perspective n'exclut pas la possibilité que d'autres paramètres puissent *aussi* être utilisés comme indices de segmentation du signal. De précédentes études ont d'ailleurs démontré que les auditeurs peuvent segmenter la parole selon une variété d'attributs comme le contour intonatif (Frankish, 1989, 1995), l'emphase (Reeves, Schauder, & Morris, 2000), les accents lexicaux (Reeves, et coll., 2000) et même le changement de voix (Parmentier & Maybery, 2008). Toutefois, contrairement à ces paramètres, qui sont liés à des comportements vocaux, les APF correspondent à des marqueurs temporels de chunks applicables à de multiples comportements (Terrace, 2001). Les APF offrent ainsi la possibilité de relier les marques de groupes temporels de la parole à un processus général de *chunking*.

1.1.3.2. L'approche stochastique et le *chunking*

Au même titre que les approches linguistiques, les approches basées sur l'apprentissage stochastique tiennent pour acquis que les auditeurs segmentent la parole en unités comme les « mots ». Dans cette perspective, plusieurs études utilisant des minilangues artificielles ont démontré que des auditeurs exposés à des séquences répétées de sons se servent des probabilités de transition (PT) entre ceux-ci pour localiser les frontières de « mots artificiels » (Aslin, 1993; Aslin, Saffran, & Newport, 1998; Batchelder, 2002; Brent & Cartwright, 1996; Brent & Siskind, 2001; Fernandes, Ventura, & Kolinsky, 2007; Mirman, Magnuson, Graf Estes, & Dixon, 2008; Monaghan, Chater, & Christiansen, 2005; Pena, Bonatti, Nespors, & Mehler, 2002; Perruchet & Vinter, 1998; Saffran, 2001; Saffran, Newport, & Aslin, 1996; Thiessen & Saffran, 2003). Formellement, la PT entre deux sons ou deux syllabes réfère au nombre d'occurrences d'une suite xy par rapport au nombre total d'occurrences de l'item initial x à l'intérieur d'un corpus, soit :

$$PT = \frac{\text{Nombre d'occurrences de la suite } xy}{\text{Nombre d'occurrences de } x}$$

Ainsi, si un item x était toujours suivi du même item y , la PT entre ces deux items serait de 1, ce qui indiquerait qu'ils font partie d'un seul et même « mot ». Plus la PT est

élevée, donc, plus les chances sont grandes que les items appartiennent à une même unité. (Bien sûr, les PT n'atteignent pas 1 en parole normale.) Dans cette optique, la segmentation de la parole de même que le *chunking* sont souvent définis en fonction de fréquences de cooccurrences (Knowlton & Squire, 1994, 1996; Perruchet & Pacteau, 1990). Toutefois, il s'agit d'une définition circulaire (le *chunking* reposerait sur des cooccurrences, mais pour que l'on produise ou perçoive des cooccurrences, il faut qu'il y ait eu un *chunking* perceptif des séquences de prime abord). En fait, le *chunking* d'entrée implique un principe qui n'est pas de nature statistique.

Pour clarifier ce principe, notons que les effets des PT sur la segmentation de mots artificiels ont le plus souvent été étudiés par le biais d'expériences impliquant de la parole synthétique présentée en boucle et sans marques prosodiques. Cependant, il a été démontré que les effets des PT deviennent nuls lorsqu'on ajoute des marques prosodiques (Endress, Nespors, & Mehler, 2009; Fernandes, et coll., 2007; Mattys, et coll., 1999; Morgan & Saffran, 1995; Shukla, Nespors, & Mehler, 2007). En particulier, Shukla et coll. ont démontré que, dans ce type de paradigme, les mots artificiels qui chevauchent des groupes prosodiques ne sont pas détectés. Autrement dit, le *chunking* ou la consolidation d'unités en mémoire n'est pas uniquement fonction des fréquences d'occurrence, mais repose sur la perception de marques prosodiques associées à des groupements. Notons que les marques prosodiques sont beaucoup plus redondantes en parole que ne peuvent l'être les différentes combinaisons de syllabes et leurs PT. Ces marqueurs sont aussi reconnus dès la naissance, et donc précèdent l'acquisition de séquences constituant des formes verbales (Bion, et coll., 2011; Christophe, Mehler, & Sebastián-Gallés, 2001; Nazzi, Bertoncini, & Mehler, 1998; Nazzi, et coll., 2006; Ramus, Hauser, Miller, Morris, & Mehler, 2000).

À la lumière de ces données, certains auteurs ont conclu que l'apprentissage de formes verbales implique d'emblée une perception de marques de groupes prosodiques permettant une détection de régularités statistiques à l'intérieur de chunks définis (Christophe, Millotte, Bernal, & Lidz, 2008). Ainsi, selon ce point de vue, la consolidation d'éléments séquentiels en unités verbales *impliquerait d'abord* un *chunking* perceptif

suivant des marques de groupes, avant qu'il y ait apprentissage de formes par le biais de la détection de cooccurrences. Notons que cette perspective s'accorde avec l'effet bénéfique du groupement sur la mémoire sérielle, un effet bien établi. Plusieurs études, maintenant classiques, ont démontré que des groupements de 3 à 5 éléments favorisent un rappel sériel optimal (Broadbent, 1975; Crowder, 1976; Wickelgren, 1964, 1967). Suite à une vaste recension, Cowan (2000) conclut que la MCT sérielle, tout comme le foyer attentionnel, peut contenir et traiter un maximum de quatre items à la fois. Fait intéressant, cette limite de capacité correspond aussi à la longueur maximale moyenne (en syllabes) des unités lexicales libres trouvée par Rousset pour quelque 32 langues (Rousset, 2004). La même limite de quatre syllabes s'applique aussi à la longueur de groupes prosodiques marqués par l'APF, appelés aussi « groupes d'accent » ou « groupes rythmiques » (Boucher, 2006; Dauer, 1983; Fletcher, 1991; Lea, 1974; Martin, 1999; Nootboom, 1997; Pasdeloup, 1992; Wenk & Wioland, 1982; Wioland, 1984). Bien sûr, les formes lexicalisées relevées par Rousset peuvent toutes contenir des sons cooccurents. Mais une segmentation basée sur les PT ou les cooccurrences de sons n'explique en rien une limite de longueur sur les formes apprises. Par contre, le *chunking* perceptif s'associe à des groupes d'éléments dont le nombre n'excède pas la capacité de la MCT (soit environ quatre éléments), ce qui pourrait expliquer une contrainte de longueur sur les formes lexicales libres.

1.1.4. La présente étude : le *chunking* perceptif de la parole et ses effets sur la mémoire

Les points essentiels des sections précédentes peuvent se résumer comme suit. En dépit de l'absence d'une définition du *chunking* qui fasse l'unanimité, il existe des indices opérationnels qui permettent d'identifier le chunk perceptif et qui s'appliquent à tous les comportements impliquant la production et l'apprentissage de séquences (voir surtout Terrace, 2001). Ces indices portent sur des délais ou des allongements d'éléments créant des groupes temporels. Aussi, le fait que de tels groupes émergent spontanément lors de l'apprentissage de nouvelles séquences suggère qu'il s'agit d'un processus perceptif ou un « *chunking* d'entrée » en lien avec la MCT. En ce qui a trait à la parole, comportement

séquentiel s'il en est un, on y retrouve des marques d'allongements ou d'APF qui s'associent au *chunking* et se retrouvent de façon universelle dans des comportements verbaux et non verbaux. Cependant, depuis les suggestions de Miller (1962), aucune étude n'a démontré un *chunking* perceptif de la parole en fonction des marques d'APF. Outre le fait qu'on ne faisait pas de distinction entre les différents types de *chunking* et qu'on tentait d'interpréter les APF par rapport à une variété d'unités culturelles et linguistiques, une difficulté méthodologique permet d'expliquer en grande partie cette lacune dans la recherche.

En particulier, les paradigmes utilisés pour étudier le *chunking* de la parole ont le plus souvent impliqué des tâches de rappel. Ces paradigmes ne peuvent servir à examiner le *chunking* perceptif dans le cas de la parole. Le problème qui se présente est que le rappel d'énoncés fait intervenir, en plus d'un *chunking* perceptif, un *chunking* « de sortie » impliquant la reconnaissance de schèmes prosodiques et de formes syntaxiques en MLT. L'influence de la syntaxe sur la segmentation de la parole a d'ailleurs été parmi les premiers thèmes de recherche en psycholinguistique (Geers, 1978; Lehiste, 1973; Lehiste, et coll., 1976; Leonard, 1974; Miller & Isard, 1963). Parmi les paradigmes utilisés alors, notons la localisation de clics (Geers, 1978), le masquage (Miller & Isard, 1963) et la désambiguïsation syntaxique (Lehiste, 1973), qui reposent tous sur une forme de rappel de stimuli. Or, lorsqu'un participant attend la fin d'un énoncé avant de répondre, la réponse fournie ne dépend pas uniquement des processus de *chunking* perceptif mais est aussi influencé par le traitement subséquent d'information lexicale et syntaxique (Cutler & Butterfield, 1992) p. 161). Ainsi, l'étude du *chunking* perceptif de la parole exige une méthode permettant de capter les réponses de l'auditeur *durant* l'écoute de la parole. De telles mesures « en ligne » sont possibles, grâce à l'électroencéphalographie (EEG), qui permet d'observer les potentiels évoqués⁷ (PÉ) par un stimulus durant l'écoute. Les sections suivantes présentent les principales composantes de PÉ qui feront l'objet d'hypothèses spécifiques en tant qu'indices de *chunking* perceptif et d'effort mnésique. À la

⁷ Pour une introduction à la méthodologie électroencéphalographique, consulter (Handy, 2005; Luck, 2005)

section qui suit (1.1.4.1.), on décrit une composante EEG particulière qui s'associe à la détection de marques prosodiques. Cette composante sera utilisée comme un indicateur de *chunking* perceptif. Par la suite (section 1.1.4.2.), nous présentons d'autres composantes ainsi que des variables comportementales qui permettent d'observer les effets du *chunking* sur la mémoire immédiate.

1.1.4.1. Un indice du *chunking* perceptif de la parole : la *Closure Positive Shift*

Des études utilisant la technique des PÉ ont déjà démontré l'existence de composantes associées à un *chunking* perceptif de la parole impliquant la détection de marques de groupes prosodiques (Steinhauer, Alter, & Friederici, 1999; Steinhauer & Friederici, 2001). Ces travaux ont montré que les contours de F_0 marquant des syntagmes intonatifs suscitent une augmentation graduelle de la négativité (le long de la ligne centromédiale) suivie d'une large déflexion positive coïncidant avec la frontière de syntagmes intonatifs (Steinhauer, et coll., 1999). On nomme cette composante *Closure Positive Shift* (CPS). Bien que la CPS montre que l'auditeur applique un *chunking* d'entrée par rapport à une marque de groupe, elle peut aussi être associée à un processus de *chunking* de sortie associé à la reconnaissance de formes (sémantiques ou syntaxiques), de patrons prévisibles de sons ou d'éléments verbaux (Brown, Salverda, Dilley, & Tanenhaus, 2011; Dilley, Mattys, & Vinke, 2010). Par exemple, des études ont montré que la CPS peut être influencée par la présence de frontières lexicales même en l'absence de marqueurs prosodiques (Bögels, Schriefers, Vonk, Chwilla, & Kerkhofs, 2010; Itzhak, Pauker, Drury, Baum, & Steinhauer, 2010; Kerkhofs, Vonk, Schriefers, & Chwilla, 2009; Pauker, Itzhak, Baum, & Steinhauer, 2011). À l'opposé, d'autres études ont démontré que la CPS peut être évoquée par les contours intonatifs dans des séquences sans aucune forme lexicale (comme des séries de syllabes sans sens : Pannekamp, Toepel, Alter, Hahne, & Friederici, 2005; ou par les contours de tons en musique : Knösche et coll., 2005; Nan, Knösche, & Friederici, 2006; Neuhaus, Knösche, & Friederici, 2006). Holzgrefe et coll. (2011) ont d'ailleurs démontré que la CPS ne reflète pas la simple détection d'un indice prosodique (allongement), mais plutôt une forme de cumul suivi d'une décharge de potentiels

(Holzgreffe, Schröder, Höhle, & Wartenburger, 2011). Le seul point commun entre les différents facteurs qui suscitent l'apparition de la CPS est qu'ils représentent tous une forme de *chunking* (d'entrée ou de sortie).

Dans la présente étude, la CPS sera utilisée afin d'observer le *chunking* perceptif évoqué par les APF marqueurs de groupes temporels. Ces marques prosodiques n'ont pas encore fait l'objet d'observations EEG pour la simple raison que la plupart des études sur la segmentation de la parole se basent sur des analyses linguistiques et des transcriptions en ToBI. Rappelons que ces analyses ne tiennent pas compte du fait que l'APF constitue une marque contrôlée séparément de l'intonation (dans cette optique, voir entre autres Bögels, Schriefers, Vonk, & Chwilla, 2011a, 2011b; Hruska & Alter, 2004; Kerkhofs, Vonk, Schriefers, & Chwilla, 2008). De ce fait, plusieurs auteurs considèrent que les frontières de syntagmes intonatifs sont marquées à la fois par une variation de la F_0 et par un APF (incluant même une pause optionnelle). Une telle conception des structures prosodiques ne reconnaît pas le fait que des marques d'APF apparaissent à l'intérieur des syntagmes intonatifs, dénotant la présence d'autres groupes prosodiques (Boucher, 2006). Pour cette raison, aucune des études citées ci-haut n'a examiné le rôle des APF qui ne concordent pas avec le syntagme intonatif. Cette omission dans les analyses pourrait s'expliquer par le fait que les premières études sur la CPS portaient sur des langues comme l'anglais et l'allemand, qui possèdent un accent lexical. Dans ces langues, les marques d'APF à l'intérieur des syntagmes intonatifs sont moins marquées que l'accent lexical. Ce dernier peut interagir avec les APF et ainsi nuire à l'observation de ces marques. Afin de contourner ce problème, la présente étude utilise le français qui a l'avantage de ne pas avoir d'accent lexical. Cela permet d'observer et de contrôler plus facilement les APF marquant la fin de groupes temporels, et ainsi de clarifier leur rôle dans le *chunking* perceptif de la parole.

L'hypothèse générale de la présente thèse porte sur le *chunking* perceptif de la parole par rapport à l'APF, qui constitue une marque universelle de groupement temporel, et sur l'effet bénéfique de ce groupement ou *chunking* sur la mémoire immédiate. En ce qui

concerne le *chunking* perceptif, nous utilisons la CPS comme principal indicateur ou variable dépendante. Dans cette optique, notre hypothèse est qu'une CPS sera évoquée par les marques d'APF indépendamment des groupes intonatifs (GI) dans lesquels ces marques sont présentées ou du type de séquences entendues (des énoncés ou des séquences de syllabes sans sens). On étayera de façon opérationnelle cette hypothèse à la section 1.2. Par rapport aux effets du *chunking* sur la mémoire immédiate, nous utilisons les composantes P300 et N400, ainsi que des mesures comportementales. Nous présenterons brièvement ces dernières variables avant de décrire nos hypothèses spécifiques.

1.1.4.2. Indices des effets du *chunking* sur la mémoire : les composantes N400, P300 et les mesures comportementales.

Il est bien établi que les groupements dans des séquences de stimuli favorisent leur rappel exact dans la mesure où les groupes ne dépassent pas environ quatre éléments (Broadbent, 1975; Cowan, 2000; Crowder, 1976; Wickelgren, 1964, 1967). Ces observations ont été obtenues à l'aide de paradigmes comportementaux impliquant des mesures de rappel sériel et de temps de réaction (TR). Bien que ces mesures permettent d'observer la présence ou l'absence de représentation en mémoire, elles ne peuvent servir à évaluer la *qualité* de la trace mnésique au moment de la perception des séquences. Il s'agit d'un aspect central puisque la trace de l'information séquentielle permet de voir l'effet du *chunking* lors de la perception, c'est-à-dire avant l'interprétation des séquences en fonction de formes en MLT. Afin d'évaluer la qualité de la trace des éléments séquentiels, nous jumelons des mesures comportementales à des observations de PÉ. Parmi les nombreux PÉ qui ont un lien avec la mémoire, nous ciblerons deux composantes en particulier, soit la N400 et la P300. Les paragraphes suivants expliquent sommairement la pertinence de ces composantes dans l'évaluation de l'effet du *chunking* sur la mémoire immédiate de la parole.

La N400 constitue une des composantes les plus étudiées en électrophysiologie du langage. Cette composante se présente comme une déflexion négative des potentiels qui atteint son maximum environ 400 ms après un stimulus. La N400 a d'abord été associée au

traitement sémantique des lexèmes d'un énoncé (Kutas & Hillyard, 1980, 1983). Elle a aussi été associée à une variété de phénomènes allant de la reconnaissance de stimuli visuels comme des séquences de formes (Abla & Okanoya, 2009) et de visages (Chaby, Jemel, George, Renault, & Fiori, 2001) aux effets d'amorçage (p. ex. : Barrett & Rugg, 1989; Bentin & McCarthy, 1994; voir Kutas & Iragui, 1998, pour une recension de travaux en rapport avec la linguistique). Toutefois, l'interprétation de la N400 a beaucoup changé au fil des ans et cette composante est désormais associée à *l'effort* nécessaire pour activer un item en mémoire⁸ (voir, entre autres, Astésano, Besson, & Alter, 2004; Finnigan, Humphreys, Dennis, & Geffen, 2002; Holcomb, 1993; Kutas, 1993; Kutas & Federmeier, 2000; Picton, 1988). Ainsi, une N400 de *moindre* amplitude signale qu'un *moins grand effort* est nécessaire pour activer un item. Dans une tâche de rappel immédiat, un item ayant bénéficié d'une bonne mise en mémoire devrait être plus facile à réactiver, donc susciter une N400 de moindre amplitude qu'un item moins bien stocké.

Ainsi, la composante N400 permet d'observer de manière indirecte, mais objective la *qualité* de la mise en mémoire des éléments en fonction d'un *chunking* perceptif. Nous jumelons ces observations de la N400 à des mesures comportementales où on demande aux participants de déterminer si un élément entendu (une cible) faisait partie d'un énoncé ou d'une séquence entendue. En général, nous prédisons que le *chunking* par groupe de trois items sera optimal par rapport à un *chunking* par groupe de quatre items, ce qui concordera avec les données générales sur les effets de groupement dans le rappel immédiat (Broadbent, 1975; Cowan, 2000; Crowder, 1976; Wickelgren, 1964, 1967). Dans cette optique, nous utilisons une version de la tâche de Sternberg pour démontrer l'effet du *chunking* (Sternberg, 1966). Dans la version classique du paradigme de Sternberg, un participant écoute d'abord une séquence de stimuli suivie d'une cible. La tâche consiste à déterminer le plus rapidement possible si la cible était présente ou non dans la séquence présentée. Par rapport à ce type de tâche, nous prévoyons une plus grande exactitude et un TR inférieur dans la reconnaissance de cibles présentées dans des groupes de trois éléments

⁸ "The N400 wave varies with the amount of memory search necessary to obtain the meaning of a stimulus." (Picton, (1988, p.261).

comparativement aux groupes de quatre. Nous nous attendons aussi à ce que les cibles provenant de groupes de quatre items bénéficient d'une moins bonne trace mnésique, et donc suscitent des N400 de plus grande amplitude.

Notons, par ailleurs, que la tâche de Sternberg permet de vérifier d'autres processus, qui sont secondaires par rapport au thème principal de la présente recherche, mais qui permettent de préciser les effets du *chunking* sur la mémoire immédiate de la parole. En particulier, Sternberg a démontré que le temps de reconnaissance d'une cible est proportionnel au nombre d'items contenus dans la séquence à mémoriser. Ainsi, plus une séquence est longue, plus l'auditeur prendra de temps avant de déterminer si la cible était présente dans la séquence. Cette constante indique que l'auditeur doit balayer (*scanner*) l'ensemble du contenu de sa mémoire de travail (MdeT) avant de fournir sa réponse, et ce, peu importe la position de l'item cible dans la séquence. Ce balayage a aussi été associé à un PÉ, soit la P300, apparaissant sous la forme d'un pic positif à environ 300 ms du début du stimulus. La P300 est d'ailleurs associée au traitement de l'information contenue dans un stimulus lors de tâches d'évaluation ou de classification de stimuli (Kutas, McCarthy, & Donchin, 1977; Magliero, Bashore, Coles, & Donchin, 1984). Selon différentes études, la longueur des séquences à balayer peut influencer soit l'amplitude de la P300 ou sa latence. Ainsi, plus le nombre d'items à balayer est grand, moins la P300 est ample (Houlihan, Stelmack, & Campbell, 1998; Karrer, McDonough, Warren, & Cone, 1980; Marsh, 1975; Pelosi, Hayward, & Blumhardt, 1995; Pratt, Michalewski, Barrett, & Starr, 1989; Starr & Barrett, 1987; Wolach & Pratt, 2001) ou plus elle est tardive (Adam & Collins, 1978; Ford, Pfefferbaum, Tinklenberg, & Kopell, 1982; Ford, Roth, Mobs, Hopkins, & Kopell, 1979; Kramer, Schneider, Fisk, & Donchin, 1986; Marsh, 1975; Pfefferbaum, Ford, Roth, & Kopell, 1980; Pratt, et coll., 1989; Starr & Barrett, 1987). En résumé, notre version de la tâche de Sternberg permettra d'évaluer les effets du *chunking* sur la mémoire immédiate selon des mesures de PÉ associés à la qualité de la trace mnésique (N400) et des mesures comportementales d'efficacité du rappel. Par ailleurs, la tâche servira à évaluer, de façon secondaire, comment le *chunking* peut affecter le balayage de la mémoire illustré par la composante P300.

1.2. Sommaire des objectifs et hypothèses spécifiques.

Rappelons que notre thèse vise à observer le *chunking* perceptif de la parole par rapport à des marques d'APF de groupement et d'examiner les effets de ce *chunking* sur la mémoire immédiate. Par rapport aux mesures décrites ci-dessus nous soumettons des hypothèses spécifiques répondant à trois objectifs.

1.2.1. Objectif 1 : Démontrer que, lors de l'écoute, l'auditeur segmente (*chunk*) la parole selon des groupes temporels

On vise à démontrer que l'auditeur écoute la parole « par blocs » en détectant les marques d'APF de groupes temporels. Nous avons donc utilisé des batteries d'énoncés où on a contrôlé les marques de groupes intonatifs et de groupes temporels, de façon à déterminer l'effet de ces marques sur la composante CPS. L'obtention d'une CPS dont la latence correspondrait aux marques d'APF de groupes temporels démontrerait pour la première fois un *chunking* perceptif de la parole en fonction de marques générales que l'on peut aussi observer dans des comportements non verbaux.

Hypothèse 1 : lors de l'écoute de la parole, la CPS est évoquée par les marques d'APF de groupes temporels, et ce, indépendamment des changements de F_0 marquant les groupes intonatifs.

1.2.2. Objectif 2 : Démontrer que le *chunking* de la parole reflète bien un *chunking* « d'entrée » et non « de sortie ».

Advenant que la CPS soit évoquée spécifiquement par les marques de groupes temporels, il faut démontrer que ces réponses reflètent un *chunking* perceptif et non un *chunking* de sortie (voir Terrace, 2001). On peut différencier ces deux processus en comparant les CPS obtenues lors de l'écoute d'énoncés et de séries de syllabes sans sens. Puisque les séquences de syllabes ne réfèrent à aucune forme verbale en MLT, une CPS évoquée dans ce contexte reflètera un *chunking* d'entrée en groupes temporels. En

supposant que ce même type de *chunking* s'applique lors de l'écoute d'énoncés sensés, alors il n'y aura aucun changement significatif des CPS évoquées.

Hypothèse 2 : les CPS évoquées par les marques d'APF dans les énoncés et dans les séquences de syllabes ne présenteront aucune différence significative.

1.2.3. Objectif 3 : Isoler les effets du *chunking* perceptif sur la MCT

Cet objectif vient compléter les deux premiers en ce qu'il porte sur les effets du *chunking* perceptif sur la mémoire immédiate d'éléments entendus. Du fait que la parole représente un signal éphémère, l'auditeur doit mettre en mémoire des séquences entendues. Aussi, il est bien connu que le *chunking* facilite la mise en mémoire dans la mesure où le groupe ou chunk utilisé n'excède pas la capacité de la MCT, qui est d'environ quatre items. Notre objectif sera de démontrer que le chunk marqué par l'APF facilite la mise en mémoire et que cet effet est fonction de la taille (ou longueur) des groupes temporels détectés. Ainsi, l'auditeur entend, dans l'ordre, des groupes de 3 et de 4 syllabes, ou des groupes de 4 et 3 syllabes. Avec ce type de présentation, nous explorons non seulement l'effet de la taille des groupes, mais aussi celui de l'ordre de présentation de ces derniers. Ces effets sont évalués en utilisant une version modifiée de la *tâche de Sternberg*, telle que décrite ci-dessus. Notre version de cette tâche permet non seulement d'évaluer les effets des groupes temporels sur la mise en mémoire, mais aussi de mesurer la *qualité* de la trace mnésique.

Hypothèse 3 : Les groupes temporels courts (trois syllabes) permettent une meilleure mise en mémoire (selon la N400) et un meilleur rappel que les groupes plus longs (quatre syllabes). La *qualité* de la mise en mémoire sera évaluée suivant trois mesures :

1) *Le nombre de bonnes réponses*

L'hypothèse spécifique est que la perception de groupes courts de trois items (mots monosyllabiques) permettra une meilleure mise en mémoire que la perception de groupes de quatre items, qui sont à la limite de la capacité de la MCT (Broadbent & Broadbent, 1973; Chen & Cowan, 2005; Cowan, 2000; Frankish, 1989, 1995; Hitch, et coll., 1996;

Mayberry, Lock, & Kazmi, 2002; Reeves, et coll., 2000; Wickelgren, 1964). Ceci devrait se refléter dans un plus grand nombre de réponses correctes pour les cibles provenant des groupes courts. On pourrait aussi s'attendre à des fluctuations dans le taux de bonnes réponses selon l'emplacement du groupe temporel dans l'énoncé; le premier groupe temporel pourrait être mieux rappelé que le deuxième, car il bénéficie d'une mémoire libre (Frankish, 1989).

2) Le temps de réaction

En utilisant une version modifiée de la tâche de Sternberg (1966), nous pourrions aussi observer la mise en mémoire des chunks par le biais des TR. Si la parole est stockée sans égard aux chunks, alors les différentes longueurs de groupes ou leur position dans l'énoncé ne devraient pas influencer les TR. Seule la longueur totale de l'énoncé devrait influencer les TR, comme dans la version classique de la tâche de Sternberg. Dans ce cas, le TR sera proportionnel au nombre d'items dans la séquence ce qui indiquera un balayage complet de la MCT. Toutefois, si l'on observe un quelconque effet de position des groupes sur les TR, nous saurons que les stimuli ne sont pas traités comme de simples séquences de stimuli lors du balayage de la MCT, mais « par groupes ».

3) Les potentiels évoqués

Nous utilisons aussi les PÉ, en particulier la composante N400, pour déterminer de façon plus précise la qualité de la trace mnésique en fonction des groupes temporels perçus. Rappelons que, dans notre tâche, un participant entend un énoncé ou une séquence de syllabes suivi d'un item-cible. Selon notre prédiction, les cibles qui ont été entendues dans des groupes courts évoqueront une N400 moins ample que cibles entendues dans des groupes longs. Un effet de position du groupe dans l'énoncé pourrait également être observé sur l'amplitude ou la latence de la P300 associée au balayage de la MdeT. Un tel effet confirmerait un balayage de la MdeT groupe par groupe.

2. Méthodologie générale

Les données de notre thèse sont présentées séparément dans deux articles, mais réfèrent à une même collecte de données. Le nombre de participants varie d'un article à l'autre du fait que certains ont dû être retirés de l'analyse en raison de problèmes techniques (bruit excessif dans les enregistrements EEG, mauvaise compréhension des consignes, etc.). Par ailleurs, les objectifs et hypothèses spécifiques ont été formulés différemment dans les articles en fonction des revues scientifiques visées.

2.1. Participants

Vingt individus (13 femmes, 7 hommes) ont été recrutés sur le campus de l'Université de Montréal. Tous étaient âgés de 19 à 41 ans (moyenne 25,55 ans, écart type 6,48 ans), avaient le français comme *seule* langue maternelle et étaient droitiers. (Le test de *Edinburgh*, Oldfield, 1971) a été administré pour confirmer la dominance manuelle.) Les participants ont aussi été soumis à un dépistage audiométrique qui a établi que tous avaient des seuils d'audition normale (à 15dBHL ou moins) pour les fréquences entre 250 et 4 000 Hz (fréquences significatives pour la perception de la parole). Les participants présentant une perte auditive plus grande que 15dBHL à 4 000 Hz ont été conservés parce que la perte n'affectait qu'une oreille et ne dépassant pas 30dBHL. Un test de mémoire de travail a aussi été administré. Le sous-test d'empan numérique du Wechsler (Wechsler, 1997) a établi que tous les participants présentaient un empan normal avec des scores égaux ou supérieurs à la moyenne normalisée⁹ (entre 8 et 16, moyenne de 10,16, écart type de 2,43 en score normalisé pour l'âge). Enfin, un consentement écrit a été obtenu de chacun des participants suite à la lecture d'un formulaire approuvé par le comité d'éthique de la recherche de l'Hôpital Rivière-des-Prairies (Montréal, Québec) et présenté à l'annexe 1. Une compensation de 40 \$ a été remise à chacun pour sa participation. L'annexe 2 résume

⁹ Les scores sont pondérés pour l'âge et rapportés sur 19. La valeur moyenne en scores pondérés est de 10 et les valeurs dites « normales » se situent entre 6 et 14. (6 à 8 = moyenne faible, 8 à 12 = moyenne, 12 à 14 = moyenne élevée.)

les caractéristiques des participants. Par rapport à cette annexe, mentionnons que le participant 1 a été enregistré séparément des autres, ce qui explique l'absence de données sur sa MdeT et son audition. Quant au participant 2, les données sur son audition n'ont pu être recueillies pour des raisons techniques. Les résultats de ces deux individus ont néanmoins été conservés, car ils concordaient avec ceux des autres participants.

2.2. Stimuli

2.2.1. Conception et caractéristiques générales des stimuli

Chaque essai était composé d'une séquence (de mots monosyllabiques ou de syllabes sans sens) suivie d'une cible à reconnaître tel qu'illustré schématiquement à la figure 2. Dans les sections suivantes, nous utilisons le terme « stimulus » pour référer à une séquence d'items, peu importe son contenu, et le terme « cible » pour désigner l'item à reconnaître, peu importe sa nature. Comme ces deux parties (stimulus et cible) ont fait l'objet de contrôles distincts, elles seront présentées séparément.

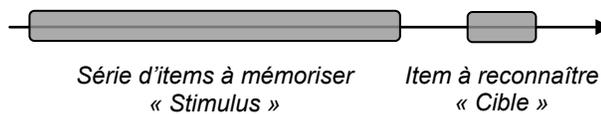


Figure 2 : Représentation schématique du déroulement d'un essai.

Deux ensembles de stimuli ont été conçus : le premier ensemble comporte 100 énoncés sensés (ensemble 1) et le deuxième, 100 séries de syllabes sans sens (ensemble 2). Tous les stimuli contenaient 9 syllabes et ont été construits avec deux contours intonatifs et trois groupes temporels (GT) tel qu'illustré à la figure 3. Le premier contour intonatif comprenait deux des trois GT, alors que le deuxième contour en contenait un seul. Pour chacun des ensembles, 50 stimuli suivaient un schème rythmique de type 3-4-2 (bleu) et 50

stimuli suivaient un schème 4-3-2 (rouge). Notons que la longueur des GT utilisés ne dépasse pas le nombre maximal d'items pouvant être maintenus actifs dans le foyer attentionnel (Cowan, 2000) et respecte la longueur maximale moyenne des GT observés en parole spontanée (Boucher, 2006; Dauer, 1983; Fletcher, 1991; Lea, 1974; Martin, 1999; Nootboom, 1997; Padeloup, 1992; Wenk & Wioland, 1982; Wioland, 1984). Ceci permet de supposer que tous les groupes présentés peuvent être traités par des participants normaux et bénéficier d'une représentation en MdeT. On s'assure ainsi que les résultats obtenus ne sont pas dus à la présence ou l'absence de trace mnésique, mais plutôt à la *qualité* de celle-ci.

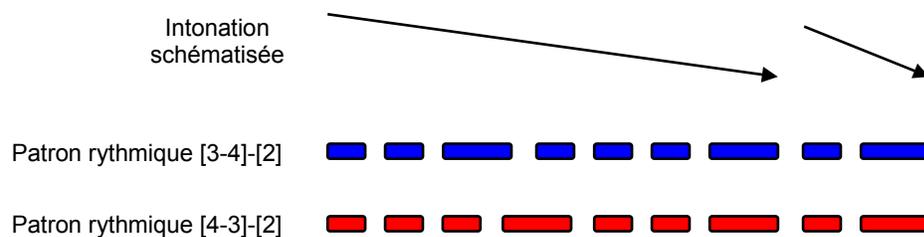


Figure 3 : Représentation schématique des schèmes prosodiques des stimuli.

Des stimuli « distracteurs » ont aussi été construits pour les deux ensembles de contextes en suivant des patrons prosodiques distincts illustrés à la figure 4. Les stimuli distracteurs ont été conçus de façon à maximiser les variations prosodiques (et grammaticales). Tel qu'illustré à la figure 4, les schèmes prosodiques varient soit sur le plan de la portée des contours intonatifs (sur un seul ou deux GT), soit sur le plan de l'ordre des GT, ou encore en fonction d'une combinaison de ces facteurs. Bien qu'il n'était pas prévu d'analyser les réponses à ces stimuli, certains ont tout de même été utilisés dans l'Article 1 afin de répondre à des commentaires émis par un réviseur.

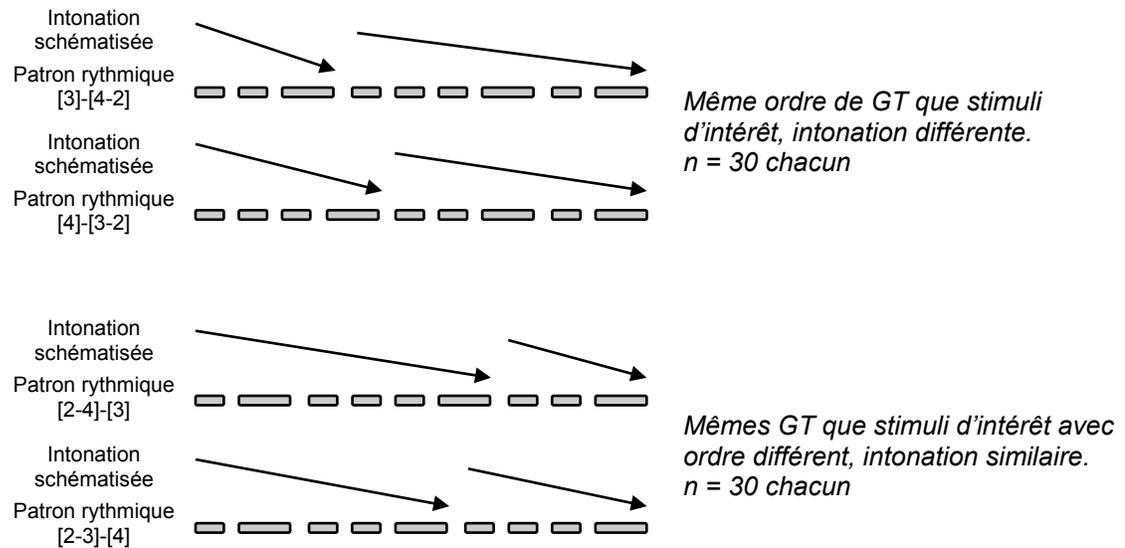


Figure 4 : Représentation schématique des schèmes prosodiques des stimuli distracteurs.

Au cours de l'expérience, chaque stimulus (énoncé ou séquence de syllabes) était suivi d'une cible à laquelle le participant devait répondre. Les cibles ont toutes été choisies parmi les éléments (lexèmes monosyllabiques ou syllabes sans sens) contenus dans le stimulus précédent. Ces cibles reflétaient des items contenus dans les deux premiers GT des énoncés : 50 % des cibles provenaient du 1^{er} GT et 50 % du 2^e GT. Les cibles ont aussi été réparties en fonction de la longueur des GT de façon à permettre une ANOVA avec comme facteurs la *position* des GT (1^{er} vs 2^e) et la *longueur* des GT (3 vs 4 syll.). Nous avons adopté cette stratégie afin de maximiser le nombre d'occurrences de rappel potentiel. Les stimuli distracteurs ont pour leur part été associés à des cibles présentes dans le 3^e GT ou absente du stimulus. Ce faisant, nous avons équilibré la provenance des cibles de sorte que 50 % étaient présentes dans le stimulus précédent et 50 % étaient absentes du stimulus.

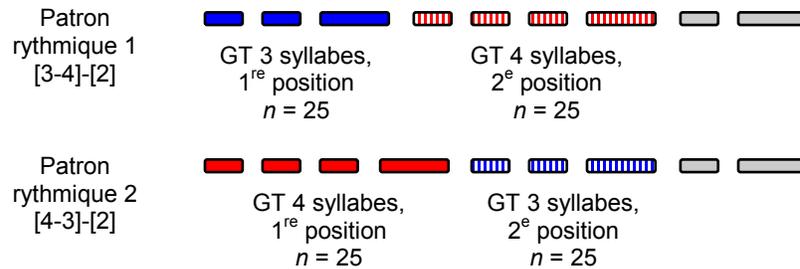


Figure 5 : Représentation schématique de la distribution des cibles selon les conditions de longueur du GT (3 vs 4 syll.) et de position du GT (1^{er} vs 2^e).

2.2.1.1. Contrôles spécifiques appliqués aux contextes de l'ensemble 1 : Énoncés sensés

L'ensemble 1 comprenait des énoncés sensés composés de lexèmes et foncteurs monosyllabiques du français. Nous nous sommes assurés que les lexèmes utilisés dans l'élaboration des différentes conditions avaient une fréquence d'occurrence comparable en langue parlée (Desrochers, 2006). Aussi, la structure syntaxique et intonative des énoncés était maintenue constante. Le premier contour intonatif comprenait toujours le groupe sujet et en deux GT, et le second contour intonatif comprenait toujours le groupe verbal en un GT (voir exemples du tableau 2).

En ce qui a trait à la structure grammaticale, les énoncés contenaient des GT comportant des éléments syntaxiques fixes, tel que spécifié au tableau 3. Les énoncés étaient toujours déclaratifs, neutres et interprétables au sens propre (aucun élément figuré ou métaphorique). Seuls des noms communs ont été choisis comme cibles afin d'éviter que la catégorie lexicale n'ait d'influence sur le rappel (Halgren et coll., 2002).

Tableau 2 : Exemples d'énoncés (où / représente la frontière de GT)

Condition prosodique	Exemples
Condition 1 [3 - 4] - [2]	Les eaux sales / de cette mare calme / gèlent tôt. Vos frais fixes / pour ces cours libres / coûtent cher. Ton thé froid / à la menthe fraîche / goûte fort.
Condition 2 [4 - 3] - [2]	La grande poêle creuse / à fond plat / reste chaude. Nos six chiots noirs / à poils drus / jappent trop. Toutes tes planches neuves / en pin mou / brûlent bien.

Tableau 3 : Structure syntaxique des groupes temporels utilisés

Ordre et rôle grammatical des GT	N ^{bre} . syllabes	Ordre des syllabes			
		1 ^{re} ,	2 ^e ,	3 ^e ,	4 ^e ,
GT ₁ Sujet	3	Dét. + Nom + Adj.; ex : « Les eaux sales » ...			
	4	Dét. + Dét. + Nom + Adj.; ex : « La grande poêle creuse » ...			
GT ₂ Compl. du nom sujet	3	Prép. + Nom + Adj; ex : ... « à fond plat » ...			
	4	Prép. + Dét. + Nom + Adj; ex : ... « de cette mare calme » ...			
GT ₃ Groupe verbal	2	Verbe + Adv.; ex : ... « gèlent tôt ». ... « reste chaude ».			

Dét. = déterminant, Adj. = adjectif, Prép. = préposition, Adv. = adverbe

La structure des énoncés distracteurs était moins rigide que celle indiquée au tableau 3. Certains énoncés distracteurs comprenaient aussi des lexèmes bisyllabiques et des cibles provenant d'autres catégories lexicales (voir exemples au tableau 4).

Tableau 4 : *Exemples d'énoncés distracteurs (où / représente la frontière de GT)*

Condition prosodique	Exemples
[3] - [4 - 2]	Même GM / brasse des affaires / en juin.
[4] - [3 - 2]	Cette pluie battante / frappe la tôle / du toit.
[2] - [4 - 3]	La course / vers le cap Horne / dure deux s(e)maines.

2.2.1.2. Contrôles spécifiques appliqués aux contextes de l'ensemble 2 : Séries de syllabes

L'ensemble 2 était composé de séries de syllabes du français de type CV du français choisies parmi un ensemble de 165 syllabes. Plusieurs précautions ont été prises dans l'élaboration de ces séries afin que la distribution des syllabes ne biaise pas le rappel des éléments. Premièrement, chaque consonne et chaque voyelle n'a été utilisée qu'une fois par séquence, et aucune suite de syllabes ne partageait le même point d'articulation. On évitait ainsi des effets d'assonance et d'allitération qui affecte la mémoire des séries (Lindstromberg & Boers, 2008). Aussi, aucune des séquences utilisées ne constituait un lexème polysyllabique ni une structure syntaxique reconnaissable en français. (voir exemples du tableau 5).

Tableau 5 : *Exemples de séries de syllabes (où / représente la frontière de GT)*

Condition prosodique	Exemples
Condition 1 [3 - 4] - [2]	[na jy wã / fœ tø zẽ ji / zõ mœ] [bu ti lø / do gã sœ ja / vœ mõ] [pa kã fi / zo me zu lẽ / gœ nø]
Condition 2 [4 - 3] - [2]	[ze za bæjõ / kẽ wø gy / fœ nã] [pẽ zã no mø / di vy ju / gõ te] [ba ku de ji / vẽ sø pã / gõ lë]

Pour l'ensemble 2, les cibles ont été choisies parmi les syllabes du stimulus précédant. Aucune syllabe n'a été utilisée plus de deux fois comme cible (92 syllabes ont été utilisées une seule fois contre quatre syllabes utilisées deux fois).

2.2.2. Enregistrement et élaboration des stimuli

L'élaboration des stimuli a impliqué l'enregistrement d'un locuteur natif du français québécois âgé de 23 ans. Ce dernier ne présentait aucun trouble de la parole ni d'idiosyncrasie articulatoire. Ce locuteur a une formation en pose de voix et en chant, de même que plusieurs années d'expérience en performance sur scène. Les enregistrements ont eu lieu dans une cabine à l'épreuve du bruit et le locuteur pouvait prendre des pauses à sa guise.

Notons que nous avons préféré utiliser des enregistrements de parole naturelle au lieu de parole de synthèse, car cette dernière présente généralement une qualité peu fiable. La qualité des stimuli acoustique est d'autant plus importante que nous faisons appel à la technique des PÉ et que toute anomalie dans les stimuli risquerait de susciter des réponses neuronales confondantes.

Le contrôle de la structure prosodique des stimuli s'est fait suivant une technique d'entraînement développée au *Laboratoire de sciences phonétiques de l'Université de Montréal*. Cette technique exploite les effets d'entraînement du rythme et de l'intonation en présentant des séquences de tons artificiels représentatives de la structure rythmique et intonative désirée pour les stimuli. Suivant cette méthode, le locuteur écoute d'abord ces séquences en boucle durant quelques secondes pour se familiariser avec la structure visée. Ensuite, l'amplitude du signal est diminuée, et le locuteur produit les contextes tout en continuant d'écouter le guide prosodique via un seul écouteur sur l'oreille gauche. Chacun des contextes est répété au moins trois fois, ce qui permet à l'expérimentateur de choisir l'occurrence qui correspond le mieux à la structure désirée. Nous avons construit un guide prosodique différent pour chaque condition rythmique (incluant les distracteurs) en tenant compte de la prosodie naturelle du français (Delattre, 1966b; Fant, Kruckensberg, & Nord,

1991; Selkirk, 1984). Chaque guide prosodique était construit de façon à susciter la production de syllabes d'une durée de 450 ms en position finale de GT et de 250 ms partout ailleurs. On guidait ainsi la production de GT de durée fixe d'une condition à l'autre (GT de 3 syllabes = 950 ms, 4 syllabes = 1 200 ms et 2 syllabes = 700 ms). L'intonation donnée par la fréquence fondamentale (F_0) des tons était contrôlée de manière à guider la production de deux GI séparés par remontée de F_0 d'au moins 6 demi-tons. Enfin, un guide prosodique distinct a été utilisé pour l'enregistrement des cibles. Celui-ci ne contenait pas de variation de F_0 et guidait la production de syllabes d'une durée de 450 ms. Nous avons enregistré les cibles séparément au lieu de les découper à l'intérieur des autres enregistrements pour éviter les possibles effets de coarticulation et obtenir une prosodie plus représentative de la production de lexèmes isolés en français.

L'ensemble des enregistrements a été fait au moyen d'une carte de son externe de 16-bit (*Fast-track Ultra*, M-Audio) à un taux d'échantillonnage de 44,1 kHz et stocké dans des fichiers *.wav* distincts. Une analyse acoustique de l'intonation et de l'énergie relative a été effectuée pour chaque stimulus à l'aide de *MultiSpeech 3700* (KayPENTAX) afin de nous assurer que les enregistrements ainsi obtenus représentaient réellement les structures prosodiques désirées. Les stimuli qui divergeaient de la structure du guide étaient réenregistrés jusqu'à ce qu'ils soient jugés satisfaisants.

Des précautions ont aussi été prises pour tenir compte du marquage des stimuli dans l'enregistrement EEG. Ce marquage est essentiel pour situer les PÉ par les stimuli. Notons, que les routines de collecte de données EEG enregistrent automatiquement des marqueurs temporels dans l'EEG, mais que ces marqueurs correspondent au début des fichiers sonores et non au début perceptif des stimuli. Pour tenir compte de cet écart, il est essentiel de contrôler la durée de l'intervalle entre le début du fichier *.wav* et le moment où la première syllabe du fichier est perçue. Pour ce faire, nous avons localisé le centre perceptif (*P-center*, Marcus, 1981) de la première syllabe de chaque stimulus et de chaque cible, et nous avons fixé l'intervalle entre le *P-center* et le début du fichier (stimuli : 100 ms; cibles : 200 ms). Le *P-center* représente le point à partir duquel on perçoit un stimulus acoustique

comme étant une *syllabe* (par opposition à la perception d'un simple phone) et peut être localisé à l'aide d'un calcul simple prenant en compte la durée de son attaque et de sa rime (voir Marcus, 1981). La figure 6 illustre le calcul du *P-center* d'une syllabe simple. Le montage acoustique nécessaire au contrôle de l'intervalle entre le début du fichier et le *P-center* de même que la normalisation des amplitudes ont été faits à l'aide de *GoldWave* (GoldWave inc., v5.58). La version finale des fichiers sonores contenant les stimuli avait une durée moyenne de 3 224 ms (minimum 3 167 ms, maximum 3 280 ms) alors que les fichiers contenant les cibles avaient une durée de 500 ms (cibles d'une durée moyenne de 447 ms, min. = 425 ms, max. = 477 ms).

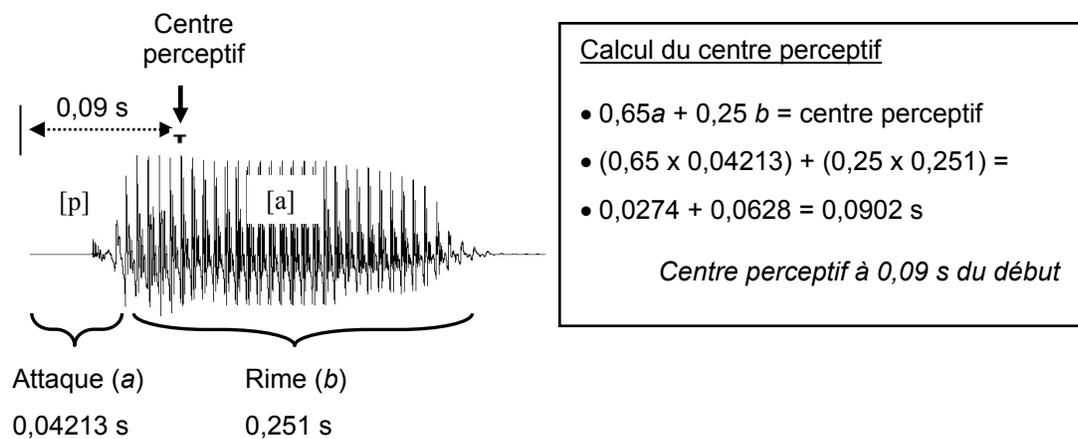


Figure 6 : Exemple de calcul du centre perceptif (*P-center*).

2.2.3. Validation des stimuli

Une analyse acoustique a été faite sur les stimuli finaux afin de vérifier leur correspondance aux structures prosodiques recherchées. Afin d'évaluer l'efficacité des contrôles appliqués, nous avons comparé les analyses faites sur les stimuli à une analyse des guides prosodiques utilisés lors de leur enregistrement. La figure 7 présente les résultats

des analyses de fréquence fondamentale (F_0) et d'amplitude des différents contextes (énoncés sensés et séries de syllabes) et des guides prosodiques correspondants.

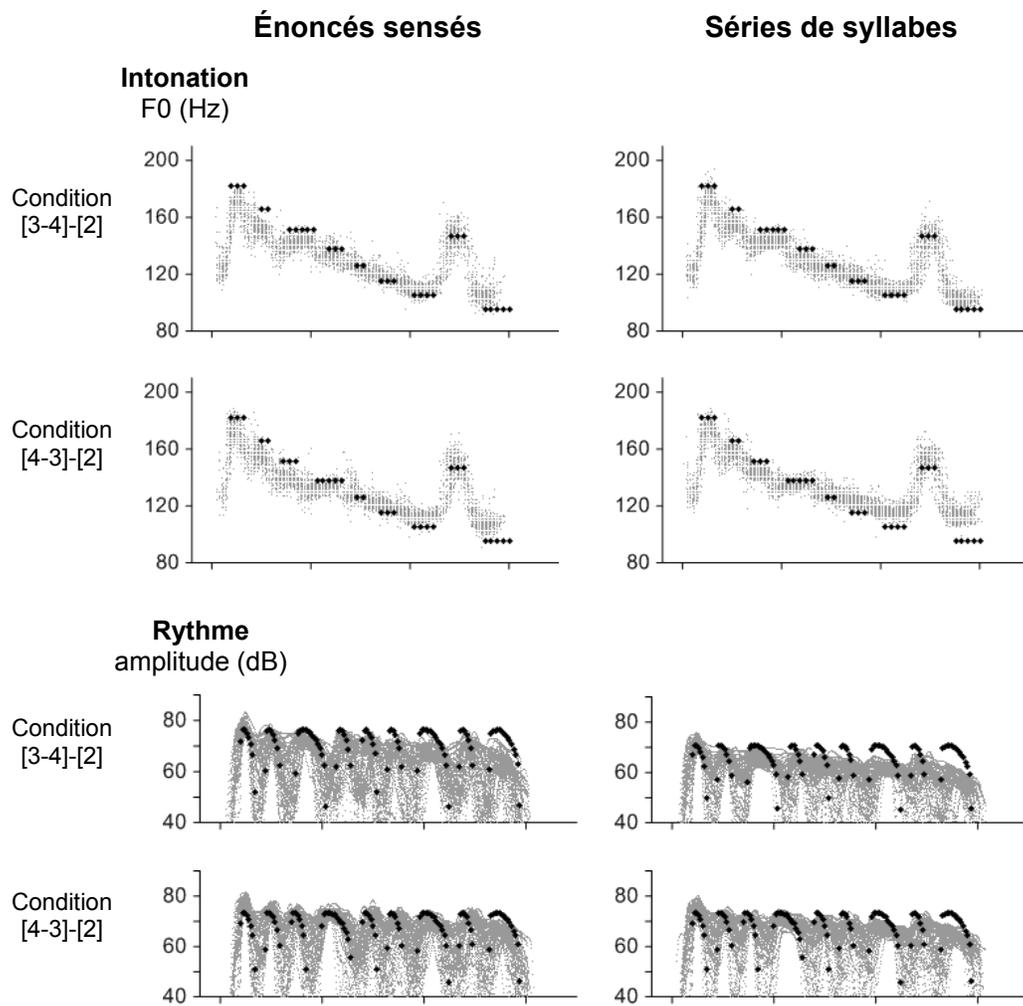


Figure 7 : Comparaison des analyses acoustiques des guides prosodiques et des stimuli enregistrés (pointillés = stimuli enregistrés, losanges = guide prosodique)

Dans cette figure, chaque ligne pointillée représente un stimulus alors que les losanges noirs représentent le guide prosodique utilisé lors de la production des contextes. Notons d'abord la correspondance étroite entre la courbe d'intonation du guide prosodique et les courbes représentant les enregistrements du locuteur. Les courbes d'amplitude démontrent aussi l'efficacité de la technique d'entraînement, car à chaque pic d'amplitude du guide prosodique correspond un pic dans le stimulus enregistré. On peut aussi voir la présence de syllabes allongées qui marquent la fin des GT et l'absence de pauses. On remarque toutefois que l'effet d'entraînement syllabique semble plus stable lors de la production de syllabes sans sens que lors de la production d'énoncés sensés. Cette différence est en fait due à la plus grande variété de structures syllabiques dans les énoncés (syllabes fermées, attaques complexes...) que dans les séries de syllabes (CV). En général, cependant, les deux types de contextes manifestent une synchronicité au point où les différences sont imperceptibles.

2.3. Procédure et tâche expérimentale

Les participants avaient pour directive d'écouter attentivement les stimuli (les énoncés et les séquences de syllabes sans sens) suivis des cibles et d'indiquer, le plus rapidement possible, si la cible était présente ou non dans le stimulus précédent. Dans nos directives, nous insistions sur l'importance de répondre le plus rapidement possible tout en faisant le moins d'erreurs possible. Les participants donnaient leurs réponses par le biais d'une boîte de réponse standard en appuyant sur les boutons de droite (avec l'index de la main droite) ou de gauche (avec l'index de la main gauche). Cet ordre de réponses a été équilibré entre les participants (50 % ont répondu que la cible était présente en utilisant le bouton de droite, 50 % avec le bouton de gauche). On demandait aussi aux participants de fixer leur regard sur un point devant eux et d'éviter de cligner des yeux durant un stimulus afin de limiter les artefacts dus à l'activité musculaire.

La présentation des deux ensembles de stimuli (syllabes sans sens et énoncés) s'est faite séparément et de façon équilibrée en ce que la moitié des participants ont entendu

l'ensemble 1 suivi de l'ensemble 2, alors que l'autre moitié a entendu l'ensemble 2 en premier. Tous les participants ont entendu les stimuli et les cibles par le biais d'écouteurs à insert (*Eartone 3A*, EAR Auditory Systems). L'amplitude du son à la sortie de l'insert a été calibrée à l'aide d'un sonomètre de façon à obtenir un niveau maximal de 68 dBA (amplitude approximative d'une conversation normale à 1 m). Les stimuli ont été présentés par blocs entrecoupés de pauses au moyen du logiciel *E-prime 1.0* (Psychology Software Tools). L'ordre des stimuli à l'intérieur des blocs était aléatoire avec la restriction que deux stimuli consécutifs ne pouvaient pas présenter le même schème prosodique. Chaque bloc comprenait en moyenne 78 stimuli (17 d'intérêt et 61 distracteurs faisant partie d'une autre expérience).

Chacun des essais était présenté selon la séquence temporelle illustrée à la figure 8, ci-dessous. L'intervalle entre le stimulus et sa cible variait (entre 750 et 1 250 ms) afin que le participant ne puisse pas anticiper le moment de donner sa réponse et pour nous assurer que les PÉ par les stimuli contaminent le moins possible les PÉ associés à la cible. L'enregistrement d'une réponse déclenchait la présentation d'un autre essai. En l'absence de réponse, le prochain essai était automatiquement lancé après 1,5 seconde.

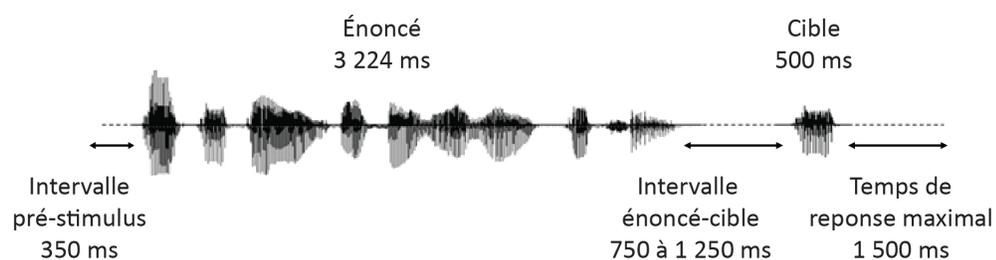


Figure 8 : Déroulement d'un essai.

2.4. Enregistrement EEG et analyse des PÉ

L'ensemble de la collecte de données a eu lieu au *Laboratoire de recherche en neurosciences et électrophysiologie cognitive* de l'Hôpital Rivière-des-Prairies. Les

enregistrements EEG ont été faits à l'aide de 58 électrodes Ag/CL intégrées dans un casque élastique (*Easy Cap*) et distribuées selon le système 10-20 augmenté (Sharbrough et coll., 1991). La figure 9 illustre la localisation relative des électrodes utilisées. Les enregistrements, le prétraitement des signaux et les analyses de bases ont été effectués à l'aide de *Cognitrace* et *EEprobe* (A.N.T.). Des électrodes bipolaires ont été installées au-dessus et en dessous de l'œil dominant (EOG vertical ou EOGV) et au canthus extérieur de chaque œil (EOG horizontal ou EOGH) pour enregistrer les mouvements oculaires et les clignements. Une référence au mastoïde gauche a été utilisée durant l'enregistrement et l'électrode AFz a servi de mise à terre. L'EEG incluant les EOG a été filtré à l'aide d'un filtre passe-bande de DC à 100 Hz avant d'être enregistré avec un taux d'échantillonnage de 512 Hz et stocké avec des marqueurs codés indiquant le début des stimuli et les réponses des participants. Les signaux EEG ont par la suite été filtrés et reréférencés hors-ligne selon les besoins des analyses prévues et tel que présenté dans les sections suivantes (2.5 Analyses spécifiques à l'article 1 et 2.6 Analyses spécifiques à l'article 2). Les clignements d'yeux ont été corrigés en soustrayant la composante EOG obtenue par une *Principal Component Analysis* (PCA) pour chaque électrode pondérée selon un facteur de propagation de l'EOGV obtenu par régression linéaire. Les segments contenant d'autres types d'artéfacts ont été rejetés. Les segments d'EEG sans artéfacts (ou corrigés) ont par la suite été moyennés différemment selon que l'objectif était d'observer les PÉ lors de l'écoute du stimulus ou de la cible. Ces paramètres, de même que les analyses qui ont été appliquées, sont décrits séparément dans les sections suivantes.

Les potentiels évoqués par l'écoute du stimulus (énoncé sensé ou série de syllabes sans sens) seront présentés dans l'article 1. On y comparera les différentes conditions rythmiques afin d'observer si les APF suscitent l'apparition de CPS indépendamment des groupements intonatifs et du contenu de l'énoncé. Les réponses associées à la reconnaissance des cibles (PÉ et comportemental) seront par la suite présentées dans l'article 2. Une attention particulière sera portée à la longueur et la position du GT d'où proviennent les cibles.

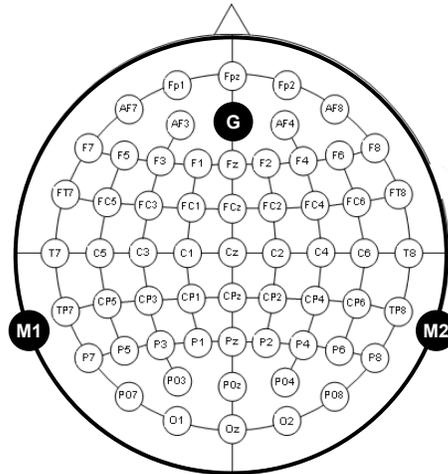


Figure 9 : Montage d'électrodes utilisé. Système 10-20 augmenté avec localisation de la mise à terre (G) et des références mastoïdes (M1, M2). Adapté de la documentation fournie par le fabricant (Easy Cap, www.easycap.de).

2.5. Analyses spécifiques à l'article 1

Pour analyser les potentiels évoqués (PÉ) par l'écoute des stimuli (et non par les cibles), les signaux sans artéfacts (ou corrigés) ont d'abord été re-référencés à la moyenne des deux mastoïdes et filtrés à l'aide de filtres passe-bas (15 Hz). Les signaux ainsi filtrés ont par la suite été regroupés par condition et moyennés en fonction du début du fichier sonore (200 ms avant et 4 000 ms après le début du fichier). La correction de la ligne de base (angl. *baseline*) a été effectuée en fonction de l'amplitude moyenne de l'intervalle de 200 ms préstimulus. Nous avons choisi d'appliquer ici une forme de moyennage inspirée des premiers travaux sur la CPS (entre autres, ceux de Steinhauer, et coll., 1999 et de Pannekamp, et coll., 2005) malgré le fait que d'autres méthodes sont aussi couramment appliquées (Kerkhofs, et coll., 2008; Männel & Friederici, 2009; Steinhauer, 2003). Ces méthodes, bien qu'elles entraînent des problèmes en ce qui a trait à la correction de la ligne de base, sont rendues nécessaires par l'utilisation de stimuli contenant des pauses. Comme

nos propres stimuli ne contiennent pas de pauses, il demeure plus avantageux d'utiliser les méthodes de moyennage originales (voir l'article 1 pour plus de détails sur le choix du point de moyennage).

Nous avons par la suite quantifié la CPS en appliquant des mesures d'amplitude et de latence inspirées de celles appliquées dans le premier article sur la CPS (Steinhauer, et coll., 1999) sur les moyennes des électrodes d'une région d'intérêt (RdI) fronto-centrales (FC3, FC1, FCz, C3, C1, Cz.). Le choix de la RdI a été déterminé à partir de la topographie des activités de PÉ associées à la CPS. Les mesures ont été prises en deux étapes. Tout d'abord, une valeur d'amplitude moyenne représentant une fenêtre de 250 ms (de 1 250 à 1 500 ms du début du stimulus) a été obtenue pour chaque participant. Cette fenêtre temporelle correspond au moment de l'écoute de la quatrième syllabe. Par la suite, des mesures d'amplitude pic-à-pic (angl. *peak to peak*) ont été appliquées sur les PÉ moyennés et normalisés (en score-*z*) pour chaque participant. Les analyses statistiques sur les données ainsi obtenues ont été effectuées avec *SPSS 17.0* (SPSS inc.). Notons enfin que nous avons effectué une analyse supplémentaire sur un sous-ensemble des conditions afin d'observer l'effet des microvariations de la latence des syllabes. Pour ce faire, nous avons représenté l'ensemble des données brutes (avant moyennage) à l'aide de la fonction *ERP Image* de *EEGLAB* (Delorme & Makeig, 2004). Avant de pouvoir appliquer ces fonctions, il nous fallait normaliser l'amplitude de l'EEG associée à chacun des essais (en score-*z*) et regrouper dans un même fichier la totalité des essais de tous les participants. Ensuite, les essais ont été reclassés en fonction de la latence exacte du *P-center* de la syllabe accentuée marquant la fin du premier GT.

2.6. Analyses spécifiques à l'article 2

Pour analyser les PÉ lors de l'écoute des cibles, les signaux sans artéfacts ont été re-référencés à la moyenne de l'ensemble des électrodes et filtrés à l'aide de filtres passe-bas de 10 Hz. Seuls les essais où les participants ont correctement reconnu les cibles ont été conservés pour les analyses subséquentes. Les signaux filtrés ont ensuite été regroupés par

condition et moyennés en fonction du début du fichier sonore de la cible (100 ms avant et 700 ms après). Une correction de la ligne de base a été appliquée en fonction de l'amplitude moyenne d'un intervalle de 100 ms avant le stimulus. En plus d'un moyennage par condition (4 combinaisons de *position* des GT (1^{er} vs 2^e) et *longueur* des GT (3 vs 4 syll.)), les PÉ ont aussi été moyennés en tenant compte d'une seule variable à la fois. Ce moyennage a permis le calcul de courbes de différence permettant d'observer les effets de la modulation d'une variable indépendamment de l'autre (effet de la position des GT sans égards à leur longueur et vice-versa). Ces courbes de différences ont été utilisées pour produire des séries chronologiques de topographies qui ont servi à localiser les RdI où se manifestent les effets et leur latence. Suite à la localisation de RdI, les enregistrements de chaque participant ont été remoyennés en fonction des RdI et quantifiés à l'aide de mesures d'amplitude (moyenne et maximum) et de latence. Toutes les analyses statistiques ont été faites à l'aide des routines d'ANOVA de SPSS 17.0.

3. Résultats

Les résultats obtenus sont présentés dans les deux articles suivants.

3.1. Article 1

L'article 1 présente les données d'une expérience conçue et réalisée par Annie C. Gilbert, sous la supervision de Victor J. Boucher (Université de Montréal) et de Boutheina Jemel (Université de Montréal). Monsieur Boucher a contribué au choix des conditions à utiliser et a supervisé la construction des stimuli. Madame Jemel a pour sa part supervisé la collecte de données et les analyses EEG. Les trois auteurs ont contribué également à la rédaction de l'article. Les figures ont été insérées dans le corps du texte pour faciliter la lecture et les références citées se retrouvent dans la bibliographie générale. La présente version de l'article sera soumise à *Brain and Language*.

Running head: PERCEPTUAL CHUNKING OF SPEECH

The perceptual chunking of speech: a demonstration using ERPs

Annie C. Gilbert^{a,b,*}, Victor J. Boucher^a, and Boutheina Jemel^{b,c}.

*Corresponding author: Annie C Gilbert

^a Laboratoire de Sciences Phonétiques, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, QC, Canada, H3C 3J7

^b Laboratoire de Recherche en Neurosciences et Électrophysiologie Cognitive, Hôpital Rivière-des-Prairies, 7070, boul. Perras, Montréal, QC, Canada, H1E 1A4

^c Centre de recherche Fernand-Seguin, Département de psychiatrie, Université de Montréal.

Abstract

In learning novel sequences, whether verbal or non-verbal, temporal groups emerge spontaneously marked by delays or a lengthening of elements. Such grouping, observed across human and animal behaviours, has been linked to a process of *perceptual* or *input chunking*. Yet, no study has shown how this chunking applies to speech. We provide a review of critical problems that have hampered research on input chunks. This leads to a principled demonstration of perceptual chunking using ERPs and the component called Closure Positive Shift (CPS), which captures listeners' responses to various groups in speech. The stimuli were utterances and sequences of nonsense syllables. These contexts contained intonation phrases marked by pitch changes, and temporal groups marked by final lengthening. Analyses of CPSs show that, across conditions, listeners specifically respond to lengthening marks, indicating a chunking of speech input in temporal groups. The results accord with a domain-general perceptual chunking of sequential information.

Keywords

Chunking, speech segmentation, prosody, ERP, EEG, Closure Positive Shift (CPS)

1. Introduction

It is widely recognized that chunks generally arise in learning and producing sequences. This is commonly observed in tasks of serial recall. For instance, when learning novel sequences of syllables or lists of digits as in a phone number, speakers spontaneously create groups (Boucher, 2006; Ryan, 1969a, 1969b; Thrope & Rowland, 1965; Wilkes, 1975; among others). In explaining such behavior, perhaps the most often cited reference is Miller's (1956) paper on "chunking". Miller's idea was that short-term memory (STM) has a limited capacity but that large numbers of items can be learned if they are recoded into chunks before they enter long-term memory (LTM). A classic example is the series of letters *I, B, M, F, B, I, C, I, A, I, R, S*, which can be learned by semantically recoding items according to the acronyms *I.B.M., F.B.I., C.I.A., I.R.S.* This idea of chunking has been highly influential in psychology and neurophysiological studies of sequence learning (e.g., Graybiel, 1998; Koch & Hoffmann, 2000; Tremblay et al., 2009). Many researchers see chunking as a means by which subjects organize sequences in units overcome limitations in STM. In this light, some have claimed that chunking is crucial to understanding a variety of sequential behaviours, from simple motions like reaching to more complex activities like speaking (Lashley, 1951). Yet, critics note that a view of chunking based on semantic recoding hardly applies when learning novel lists of sounds or non-verbal sequences, as in some studies involving animals (Terrace, 2001). Such observations indeed imply a more general chunking principle.

On this latter idea, it is interesting that relatively little attention is paid to another paper by Miller (1962), which suggests a different concept of chunking. In that paper, Miller argued that, in processing speech, listeners are not making immediate decisions on a sound-by-sound basis. Instead, they are delaying decisions over some "perceptual unit" spanning a number of elements (Miller, 1962). This presents another concept of chunks that does not involve a recoding but a basic process of *perceptual* or *input* chunking (Graybiel, 1998). The following sections offer a three-part literature review of this concept of chunking and how it applies to speech perception. The first part focuses on the definition of

input chunks and their marks in speech (section 1.1). We then briefly clarify several fundamental issues in attempting to relate input chunks to linguistic units (section 1.2) and stochastic views of speech segmentation (section 1.3). Finally, we discuss the particular neural components that are used in the present study to demonstrate input chunking as it applies to speech perception.

1.1. An operational definition of input chunking

Following Miller (1956), who admitted that he was “...not definite about what constitutes a chunk of information” (Miller, 1956; p.82), the concept of chunk has remained rather intuitive (e.g. Mathy & Feldman, 2012; Terrace, 2001). The few definitions that have been proposed are unspecific (e.g., chunks as “meaning groups” or “groups of elements” in Anderson & Matessa, 1997; Cowan, 2000; Cowan, et al., 2004; Farrell, 2008; Ng & Maybery, 2002) or else assume that chunking involves verbal recoding (e.g., Hitch, et al., 1996; Mathy & Feldman, 2012; Shiffrin & Nosofsky, 1994; Simon, 1974; Stark & Calfee, 1970). Obviously, the latter concept is of little use when trying to account for chunks that emerge in learning non-verbal sequences. Considering this problem, Terrace (2001) emphasized the need to distinguish between *output* and *input* chunks. Generally, output chunks can be taken to reflect top-down processes where learned forms or patterns in LTM underlie the recognition of units in sequences. By contrast, input chunks relate to a perceptual unitization of sequential stimuli that conforms to the capacity limit of STM and the focus of attention (which is in the order of 4 ± 1 items, according to Cowan, 2000). This dual concept is also suggested by Gobet et al. (2001) who distinguish between a *goal-oriented chunking* and an “automatic” *perceptual chunking*. For instance, the authors liken perceptual chunking to how letters are grouped into space-divided words. The difficulty with such analogies, however, is that markers of input chunks in speech (as in other behaviours) are not artificial like spaces in text but would reflect a sensori-motor process (Graybiel, 1998). This brings the question of just what these markers are.

On this pivotal issue, Terrace (2001) notes that the uncritical application of the concept of chunking in various disciplines has led to neglect the fact that there are objective

marks of chunks across behaviours. These relate to latencies in sequences, or what Terrace calls “inter-response times” (IRTs), which refer to delays or a lengthening of elements that create temporal groups. Such chunking is reported in numerous animal and human behaviours (for examples, see Graybiel, 1998; Terrace, 2001; Terrace, et al., 1996). To illustrate how these marks extend to the learning verbal sequences, we refer to Figure 1.

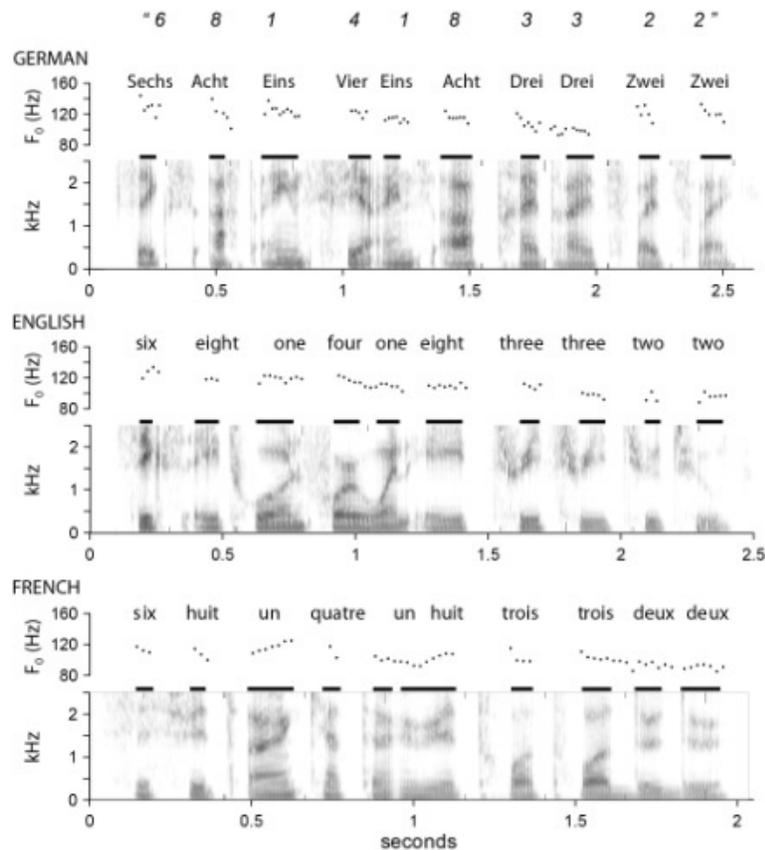


Figure 1. Temporal grouping in German, English, and French recitations of a ten-digit series. Note that grouping entails a relative lengthening of syllables at the ends of groups.

The figure shows the acoustic patterns of a recited sequence of digits in three languages. In producing such lists, chunks usually emerge in terms of temporal groups of 2

to 4 items (note that such chunks appear just as well in recalling unstructured series of nonsense syllables; see Boucher, 2006). A look at the acoustic patterns clarifies the grouping marks. In the figure, intonation is shown by fundamental frequency (F0), while the relative duration of elements (monosyllables) is shown by the lines over the spectrograms. One can see that, while F0 contours vary extensively, a relative lengthening of digits systematically occurs at the ends of groups (i.e., the digits 1, 2, 3, 8 are longer at the ends than within groups). These timing changes can be quite subtle and brief pauses may also be added. The significant point, however, is that such grouping arises even with presented sequences that have no internal structure or meaning, and the timing marks are similar to IRTs that create chunks in learning motor sequences. Hence, these similar marks that arise across verbal and non-verbal behaviours suggest some domain-general, sensori-motor chunking of sequential information (Graybiel, 1998). In this light, the present report addresses the question of how this general chunking involving particular timing marks applies to speech perception. Of course, one might rebut that speech is not simply a sequence of motor events but involves output chunking in relation to learned forms like words and phrases. Such assumed units have guided linguistic and stochastic views of speech segmentation. However, the following sections clarify that these views may not capture input chunking.

1.2. The perceptual chunking of speech and linguistic assumptions

The above discussion suggests that temporal groups which arise in sequence learning can reflect a general process of perceptual chunking. Thus, one might expect that timing changes marking groups would appear across languages and, in fact, reports suggest the universality of “preboundary lengthening” (PBL). For instance, PBL has been observed in Cantonese and Mandarin (Chow, 2008), Dutch (Gussenhoven & Rietveld, 1992), English (Lehiste, et al., 1976), French (Delattre, 1966a), German (Delattre, 1966a), Guoyu (Fon, 2002), Hungarian (Hockey & Fagyal, 1998), Japanese (Fon, 2002), Spanish (Delattre, 1966a), Swedish (Lindblom, 1968), Russian (Makarova, 2001), among other languages (Vaissière, 1983). There has been some controversy on whether young children use PBL

(see, e.g., Bion, et al., 2011). However, a study by DePaolis et al. (2008) addressed this issue by comparing children's speech in four languages (English, Finnish, French, and Welsh) as they begin to learn verbal forms. The study showed consistent PBL production in disyllabic forms across languages. Overall, then, PBL appears as a general mark that arises with the learning of unitized sequences, which suggest some general process of input chunking. However, the way in which sequences are chunked may not always conform to notional units like words.

To explain this briefly, one should note that much research has been guided by the idea that speech contains cues serving to segment words (Cutler, et al., 1986; Cutler & Norris, 1988; Endress & Hauser, 2010; Houston, et al., 2004; Johnson & Jusczyk, 2001; Johnson, et al., 2003; Jusczyk, et al., 1993; Mattys, et al., 1999; Mattys & Samuel, 1997; McQueen, et al., 2001; Nazzi, et al., 2006; Norris, et al., 1997; Otake, et al., 1993). Yet, numerous studies of language usage report that countless expressions like *gi'me, see ya, hold on, look out, thank you* (etc.), which make up a major part of speakers' verbal output, may not involve separate words, as their written forms suggest, but can function as lexical chunks or formulas (Bates & Goodman, 2001; Bybee, 2006; Dabrowska, 2004; Wray, 2002, 2008; Zeschel, 2008). These findings bring to light the difficulty of attempting to identify chunks using traditional units of linguistic analysis, and the need to explain the chunking of verbal expression by some domain-general process that is not based on syntactic assumptions (Beckner & Bybee, 2009; Bybee, 2010).

More confusion may arise from attempts to analyze marks of chunks using linguistic systems such as *Tone and Break Indices* (ToBI; Beckman & Ayers Elam, 1997), which overlook PBL (Beckman & Pierrehumbert, 1986; Pierrehumbert, 2000; see, e.g. Frazier, Carlson, & Clifton Jr, 2006). These systems draw from models of English phonology where rhythm is often conceived in relation to "word" stress and the tendency to alternate stressed and unstressed syllables in metric-like feet (Lieberman, 1975; Lieberman & Prince, 1977). Stress, in this view, is linked to the "pitch accents" of phrases, described in terms of varying high and low tones. Hence, using a system like ToBI, one may come to

conceive speech rhythm as melodic alternations of tones, and to disregard PBL as separately controlled marks. For instance, using ToBI would not serve to capture the temporal chunks seen in Figure 1. Yet, it is important to note that these chunks, which can create rhythms, are not linked to word stress or pitch accents (nor do they “map syntactic structures”). There is also much evidence that timing marks, like PBL, are processed separately from tonal patterns (Cooper & Sorensen, 1981; Loeb & Allen, 1993; Seddoh, 1999; Snow, 1994). This is clear in tone languages, where tones and speech timing are independently learned and controlled (Gandour, et al., 1988; Li & Thompson, 1978; So & Dodd, 1995). PBL also occurs *within* intonation phrases, independently of F_0 changes marking phrase boundaries (Beckman & Ayers Elam, 1997; Boucher, 2006). Finally, clinical studies further support a separate processing of PBL and intonation. For instance, speakers presenting a non-fluent aphasia often manifest severe abnormalities of speech timing but relatively normal intonation (Baum & Pell, 1999; Gandour, 1998a, 1998b; Seddoh, 2004; Seddoh & Robin, 2001). In short, traditional linguistic analyses and systems such as ToBI may not serve to identify input chunks.

1.3. Perceptual chunking and statistical learning

As in a linguistic approach, research on statistical learning of verbal forms often assumes a priori that speech is segmented into words. Thus, numerous reports using small artificial languages have shown that “artificial words” can be learned by a repeated exposure to co-occurring speech sounds (Aslin, 1993; Brent & Cartwright, 1996; Brent & Siskind, 2001; Endress & Hauser, 2010; Monaghan, et al., 2005; Saffran, et al., 1996; Swingley, 2005). In this view, “chunking” is often defined as a function of the repetition (Bybee, 2010) or frequency of co-occurring elements (Knowlton & Squire, 1994, 1996; Perruchet & Pacteau, 1990). Yet recent studies of the effects of marks of prosodic groups question these ideas. In fact, statistical learning of artificial forms has most often been studied via the presentation of repeated streams of speech sounds -- but without any marks of prosodic groups. Recent experiments that include such marks (Endress & Hauser, 2010; Shukla, et al., 2007) show that the perception of groups can override co-occurrence effects.

For instance, when repeatedly presented artificial words containing co-occurring sounds straddle the boundary marks of prosodic groups, they are not learned (Shukla, et al., 2007). In reviewing such findings, some authors have concluded that the parsing of prosodic group markers must operate *before* listeners detect the co-occurrences of sounds within groups that serve to divide forms (Christophe, et al., 2008). By this account, then, the acquisition of verbal forms not only arises from frequently co-occurring sounds but rests on some perceptual chunking of speech in terms of prosodic group markers.

In support of this latter view, it is well established that grouping sequential items in sets of 3 or 4, as in Figure 1, enhances learning (Broadbent, 1975; Crowder, 1976; Wickelgren, 1964, 1967). Interestingly, these size limits on grouping effects concord with observations by Rousset (2004) that stand-alone lexical forms in some 32 languages do not tend to exceed 4 syllables on average. Similar limitations have been reported for temporal groups marked by PBL in speech (Boucher, 2006; Dauer, 1983; Lea, 1974; Martin, 1999; Nootboom, 1997; Padeloup, 1992; Wenk & Wioland, 1982; Wioland, 1984). Taken together, these size limits on learned forms may reflect input chunking in relation to capacity limits of STM (Cowan, 2000). However, it is important to acknowledge that this perceptual chunking in learning and producing novel sequences may not rely on exposure to frequent co-occurrences. For instance, in learning sequences of digits or syllables (as in Figure 1), input chunks arise even after a single presentation and, hence, reflect a process that is not statistical in nature.

2. The present demonstration: using ERPs to capture speech chunking on-line

The main points of the above discussion can be summarized as follows. There are observable markers of chunks that appear across behaviours. These marks refer to delays or length changes which create temporal groups in sequence learning and production. Moreover, since such groups emerge in learning novel series, it is surmised that they reflect a process of input or perceptual chunking. Yet, one does not find a demonstration that perception operates in relation to temporal groups, not even for speech. A central methodological problem may explain this lack of evidence.

In particular, chunking has largely been investigated via experiments involving recall, which may not capture *input* chunking (Terrace, 2001). This is especially evident in the case of speech. As some of the earliest work in psycholinguistics has shown, in tasks where subjects listen to utterances before responding, input chunking can be overridden by latter-stage processing of syntactic and lexical information (Cutler, et al., 1997; Geers, 1978; Lehiste, 1973; Miller & Isard, 1963). Put differently, to investigate input chunking, one has to capture listeners' responses to speech *on-line*. This is made possible by using ERPs in regard of certain neural components. Importantly, studies using ERPs have suggested a neural correlate of chunking associated with prosodic marks (Steinhauer, et al., 1999; Steinhauer & Friederici, 2001). This work has shown, for instance, that marks of intonation phrases in English, German, and Dutch, can evoke an incremental negativity over midline central sites that terminates with a *Closure Positive Shift* (CPS). This rising negativity across sequences of elements suggests a buffering of information that ends with a positive shift corresponding to a group-final mark (see, e.g., Holzgrefe, et al., 2011). Recently, it has been revealed that CPSs can also reflect processes of output chunking relating to syntactic-semantic forms in presented utterances, or even expected sound patterns (e.g., Brown, et al., 2011; Dilley, et al., 2010). For example, reports show that the CPS is influenced by the boundaries of verbal forms with no prosodic marks (Bögels, et al., 2010; Itzhak, et al., 2010; Kerkhofs, et al., 2009; Pauker, et al., 2011). Conversely, studies show that the CPS can be evoked by intonation contours in series of nonsense syllables (Pannekamp, et al., 2005) or by tonal phrasing in music (Knösche, et al., 2005; Nan, et al., 2006; Neuhaus, et al., 2006). However, the present study uses the CPS to examine chunking in terms of groups marked by PBL only. Regarding the PBL, it is important to mention that several studies involving the CPS refer to analyses like ToBI which, as noted earlier, overlook PBL as a separate prosodic marker (e.g., Bögels, et al., 2011a, 2011b; Carlson, Frazier, & Clifton Jr., 2009; Frazier, et al., 2006; Hruska & Alter, 2004; Kerkhofs, et al., 2008). These analyses assume, for instance, that intonation-phrase boundaries involve *co-occurring* changes in F_0 and PBL, and fail to recognize that PBL can occur within phrases independently of any change in intonation. Thus, none of the aforementioned

studies have examined the specific effects of temporal groups marked by PBL, perhaps because such marks can be influenced by word stress in languages like English and German. Stress patterns can complicate observations of the effects of phrase-internal PBL on CPSs. For this reason, the following experiments will use French which does not have word stress.

In using CPS, our hypothesis is that this response can capture perceptual chunking in terms of temporal groups independently of intonation contours in presented utterances and sequences of nonsense syllables. A validation of this prediction offers a principled demonstration of perceptual chunking on two critical points. One point bears on the need to clarify that the CPS can capture a chunking process that relates specifically to marks of temporal groups and not to intonation changes marking phrase boundaries. Another point is that any demonstration of perceptual chunking requires that one distinguish between input and output processes. As noted, output chunking relates to how familiar forms or patterns in LTM lead to recognize groups in sequential elements. Input chunking, however, relates to how sequential elements are perceived in groups regardless of the presence of familiar verbal forms. In this light, if the CPS response captures input chunking rather than top down effects, then responses should be similar for presented utterances and sequences of nonsense syllables.

To test these predictions, we designed two experiments using utterances (Experiment 1) and meaningless series of syllables (Experiment 2) presented while recording brain potentials. Each set of stimuli was constructed with intonation phrases that contained PBLs. In a subset of contexts, the phrase-internal PBL was shifted by one syllable to determine the effect of PBLs on the evoked CPSs. Hence, the two experiments served to evaluate whether CPSs capture a perceptual chunking in temporal groups that applies independently of intonation or the presence of verbal forms in the stimuli.

3. Experiment 1

3.1. Materials and Methods

3.1.1. Participants

The participants were twenty native speakers of French, aged 19 to 42 years (mean = 25.6). These individuals presented normal hearing levels in a standard audiometric screening and normal memory performances on a digit-span test (WAIS; Wechsler, 1997). All were dominant right-handers, with no history of substance abuse (other than tobacco smoking), neurological, or psychiatric disorder. Participation was subject to written consent and followed the ethics guidelines of the *Hôpital Rivière-des-Prairies* (Montréal, QC).

3.1.2. Stimuli

The stimuli were natural-sounding utterances, including 160 target contexts and a number of fillers, which were produced by a native speaker of French following a pacer technique that allows a precise control of prosodic variables. Using this technique, the speaker listens to a series of pure tones presenting F_0 and rhythm patterns and then produces given utterances with the desired prosody. Two sets of utterances were produced containing three temporal groups (TGs) marked by a lengthening of syllables: one set of 80 utterances with a 3-4-2 rhythm pattern and another set of 80 utterances with a 4-3-2 rhythm. Both these sets contained two intonation groups (IGs). Figure 2 summarizes the design of the stimuli in terms of their prosodic attributes (filler contexts containing different prosodic patterns are not described here). Looking at conditions 1 and 2, the first TGs in the utterances changes from a 3-4 to a 4-3 pattern but this does not coincide with a change in IG boundary location. Hence, the main comparison of conditions 1 and 2 served to determine if the CPS responses suggest a chunking of phrase-internal TGs independent of intonation changes marking IGs. As a secondary verification, conditions 3 and 4 aimed to determine if coinciding boundaries of TGs and IGs created differing effects on CPSs. To control for the duration of TGs in the stimuli, the pacers were built to elicit productions of 250 ms syllables and lengthened syllables of 450 ms (1.8 times longer, which reflects the ratio found in French speech; Fant, et al., 1991). Thus, a 2-syllable TG lasted 700 ms, a 3-

syllable TG 950 ms, and 4-syllable TG 1,200 ms. As for the control of IGs, the F_0 contour of the pacer was adjusted to fit the speaker's pitch and was modulated to elicit a phrase-initial F_0 reset of 6 semi-tones (initial resets have been reported to occur in French as in other languages; Delattre, 1966b; Selkirk, 1984). The resulting productions were recorded using a 16-bit external soundcard (*Fast-track Ultra*, M-Audio) at a sampling rate of 44.1 kHz. We analysed the F_0 and energy contours for every target utterance (using *MultiSpeech*, v. 3.7, KayLab) to insure that the productions matched the desired prosodic patterns of the pacers. In cases where the recorded utterance differed from the required pattern, the utterance was re-recorded.

In terms of the grammatical content of the stimuli, all target utterances were composed of monosyllabic lexemes and functors with comparable indexes of familiarity (Desrochers, 2006). As summarized in Figure 2, the first IG always contained a subject noun phrase (NP) whereas the second IG always contained a verb phrase (VP).

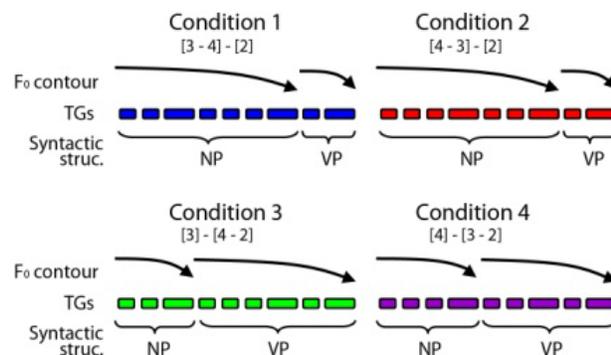


Figure 2. Schematic representation summarizing the structural prosody of stimuli in the four experimental conditions. *Arrows represent schematized F_0 contours marking IGs; dashes indicate the relative durations of syllables with group-final lengthening marking TGs.*

As an added precaution, we aligned the audio files according to the perceptual-center (*P-center*) of the first syllable of each utterance. The *P-center* constitutes the point in

time where a syllable is *perceived* to begin, which can be located in the waveform by means of a simple calculation (see Marcus, 1981). In the present case, we ensured that the *P-center* of the first syllable of each utterance occurred 100 ms from the onset of the audio file. Finally, all audio files were amplitude normalized.

To fathom the degree to which the pacing technique served to elaborate controlled speech contexts, Figure 3 presents the results of an analysis of the F_0 and energy contours of every stimulus. Each dotted line in the figure represents one stimulus so that the superimposed lines show the acoustic properties of the complete set of stimuli. One sees that the intonation patterns were relatively constant for conditions 1 and 2 (blue and red lines) and presented two IGs divided by an F_0 reset of about six semitones. Within the utterances, only the TGs changed, and this is seen in the energy contours (blue and red lines). Each energy peak in these contours represents the vocalic part of a syllable, and one sees that a lengthening marking the end of first TG creates a change from a 3-4 pattern (blue) to a 4-3 pattern (red). (It should also be noted that the energy contours show that no pauses were present between TGs.) Hence, in conditions 1 and 2, TGs change *within* a phrase and independently of F_0 marks of intonation phrases. To verify the possible effects of combining both marks of F_0 and lengthening, Figure 3 shows that conditions 3 and 4 (green and purple lines) contain a F_0 reset occurring on the third (green) or fourth (purple) syllable, creating two IGs that correspond to the lengthening marks of two TGs. Overall, one can see in the figure that the intended prosodic patterns were elicited by the pacing technique with very little jitter considering that these represent natural and not synthesized utterances.

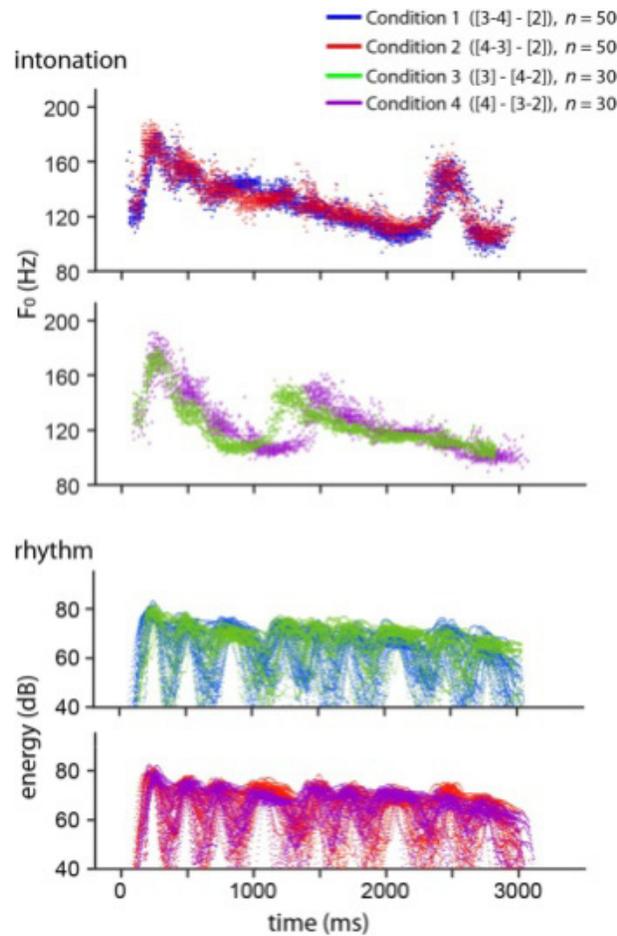


Figure 3. Measured acoustic attributes of the utterances

3.1.3. Procedure, EEG recording, and ERP analysis

Participant screening and data collection were performed at the Laboratoire de recherche en neuroscience et électrophysiologie cognitive of Hôpital Rivière-des-Prairies (Montréal, Canada). All utterances were presented using insert earphones (*Eartone 3A*, EAR Auditory Systems) and each context was followed by a separate one-syllable “target” (a lexeme). Participants were instructed to listen to the stimuli while maintaining their gaze at a fixation point on a blank screen. Their task was to indicate via a key press, whether a heard target was part of the preceding utterance. The behavioural and EEG data relative to

the target are not discussed here as they are not relevant to the present report. Sound files were played back via *E-prime 2.0* (Psychology Software Tools) in random blocks divided by rest pauses. The sounds files were delivered at a fixed intensity with peak levels of 68 dBA.

The EEG data were recorded continuously (with a band-pass from DC to 100 Hz and at a 512 Hz sampling rate) through 58 shielded electrodes embedded in an elastic cap (*Easy Cap*) according to the enhanced 10-20 system (Sharbrough, et al., 1991). Two bipolar electrodes placed above and below the dominant eye (vertical EOG), and at the outer canthus of each eye (horizontal EOG) served to record eye movements and blinks. A left mastoid electrode was used as online reference for all scalp electrodes and AFz served as the ground. Electrode impedances were kept below 5 K Ω .

The EEG signal was offline re-referenced to the averaged mastoids (left and right) and filtered using a digital band-pass filter (0.03-15Hz). Due to a lesser number of trials, conditions 3 and 4 were filtered with a lower cutoff (10 Hz instead of 15 Hz). Eye-blinks were corrected by subtracting from the EEG the PCA-transformed EOG components for each electrode, weighed according to VEOG propagation factors (computed using linear regression). Segments with other artifacts were rejected. The resulting EEG segments were time-locked to the onset of the sound files and averaged from 200 ms before to 4000 ms after the onset of a stimulus file for each condition. We chose to average the EEG segments according to a stimulus sound-file (utterance) as published in earlier works on CPS (Pannekamp, et al., 2005; Steinhauer, et al., 1999) instead of averaging according to the onset of a following constituent, as in recent studies (Kerkhofs, et al., 2008; Männel & Friederici, 2009; Steinhauer, 2003). Note that, compared to these studies, our utterance contexts do not contain pauses, which interrupt the acoustic stimulation and may lead to the elicitation of early acoustic components when the stimulation resumes. These components can affect the positivity of the CPS, but are only elicited if pauses are used (see, e.g., Steinhauer, 2003 on the impact of the N1-P2 complex associated with the onset of acoustic stimulation on the CPS). Furthermore, averaging in terms of a pre-utterance point where

there is no stimulus-induced activity provides a better baseline correction than baselining *within* the stimulus, as has been done in some recent work (Kerkhofs, et al., 2008; Männel & Friederici, 2009; Steinhauer, 2003). Following the aforementioned type of averaging and baseline correction, the CPSs were quantified using both amplitude and latency measures (Steinhauer, et al., 1999). The amplitude measures were obtained by averaging, for each participant, the ERP amplitudes over a fronto-central region of interest (ROI; including FC3, FC1, FCz, C3, C1, and Cz) in a 250 ms time window encompassing the CPSs elicited by first TGs of each utterance condition (from 1,250 to 1,500 ms post-onset of a sound file). Latency data were obtained via peak-to-peak measures performed on the *z*-score normalized amplitude of averaged ERPs for individual participants. Finally, we performed raw data analyses on a subset of conditions to examine the possible effects of slight jitter of syllable beats in the stimuli. This was done using *ERP Image* plots in *EEGLAB* (Delorme & Makeig, 2004). Before plotting, individual single EEG data epochs were RMS-*z* transformed, then pooled together and sorted according to the exact latency of acoustic *P-centers* of the lengthened syllable marking the end of the first TG in an utterance (shortest latency down the *y*-axis).

3.2. Results

The bottom panels of Figures 4, 5, and 6 show topographical representations of the ERPs over the entire scalp (250 ms time-windows) and averaged ERPs on a fronto-central ROI during the presentation of entire utterances. The top panels of the figures summarize the results of the acoustic analysis of all the stimuli as shown in Figure 3. Note that IGs are marked by an F_0 reset whereas TGs are marked by a lengthening of group-final syllables, as seen in the dB contours where each rise represents a syllable beat. In comparing the three figures, one can see that the general patterns of ERPs confirm a processing of speech input in terms of TGs. Specifically, across all conditions, three CPSs were elicited, which is consistent with the number of TGs in the stimuli but not the number of IGs. It can also be seen that the CPSs are *time-locked* to the ends of TGs across conditions. In other words, as detailed in the Method section, comparisons across conditions 1 and 2 (reflecting a change

in the length of the first TG from 3 to 4 syllables) aimed to isolate the specific effects of TGs. Figure 4 shows that changing the length of the TG by one syllable (from a 3-4 to a 4-3 pattern) modulates the latency of the CPSs, creating a delay commensurate with the duration of a syllable. This suggests an on-line chunking in terms of TGs.

As for the effects of intonation patterns, Figures 5 and 6 show ERPs in conditions where the presented utterances share the same TGs but different IGs (conditions 1 vs 3, and 2 vs 4). In these conditions, there is no change in the evoked CPSs: the responses remain time-locked to the TGs regardless of whether or not the TG matches an F_0 reset marking the boundary of an IG. In short, CPSs link specifically to TGs and not to IGs.

As a verification of the effects of TGs and IGs, statistical analyses were carried out on the mean ERP amplitudes within a designated 250 ms time-window of the fronto-central ROI (see Methods). A 2x2 ANOVA was applied comparing the effects of the length of the first TG (3 vs 4 syll.) and the span of the IG (F_0 contour ending on the first TG vs the second TG). This revealed significant main effects of TG length [$F(1,19) = 41.811$; $p < .001$; $\eta^2 = .688$] but not of IG span [$F(1,19) = .817$; $p = .377$; $\eta^2 = .041$], and no significant interactions [$F(1,19) = .095$; $p = .762$; $\eta^2 = .005$]. In other words, the analyses confirmed that the CPS responses were specifically related to heard TGs, with intonation phrasing having no marked effect.

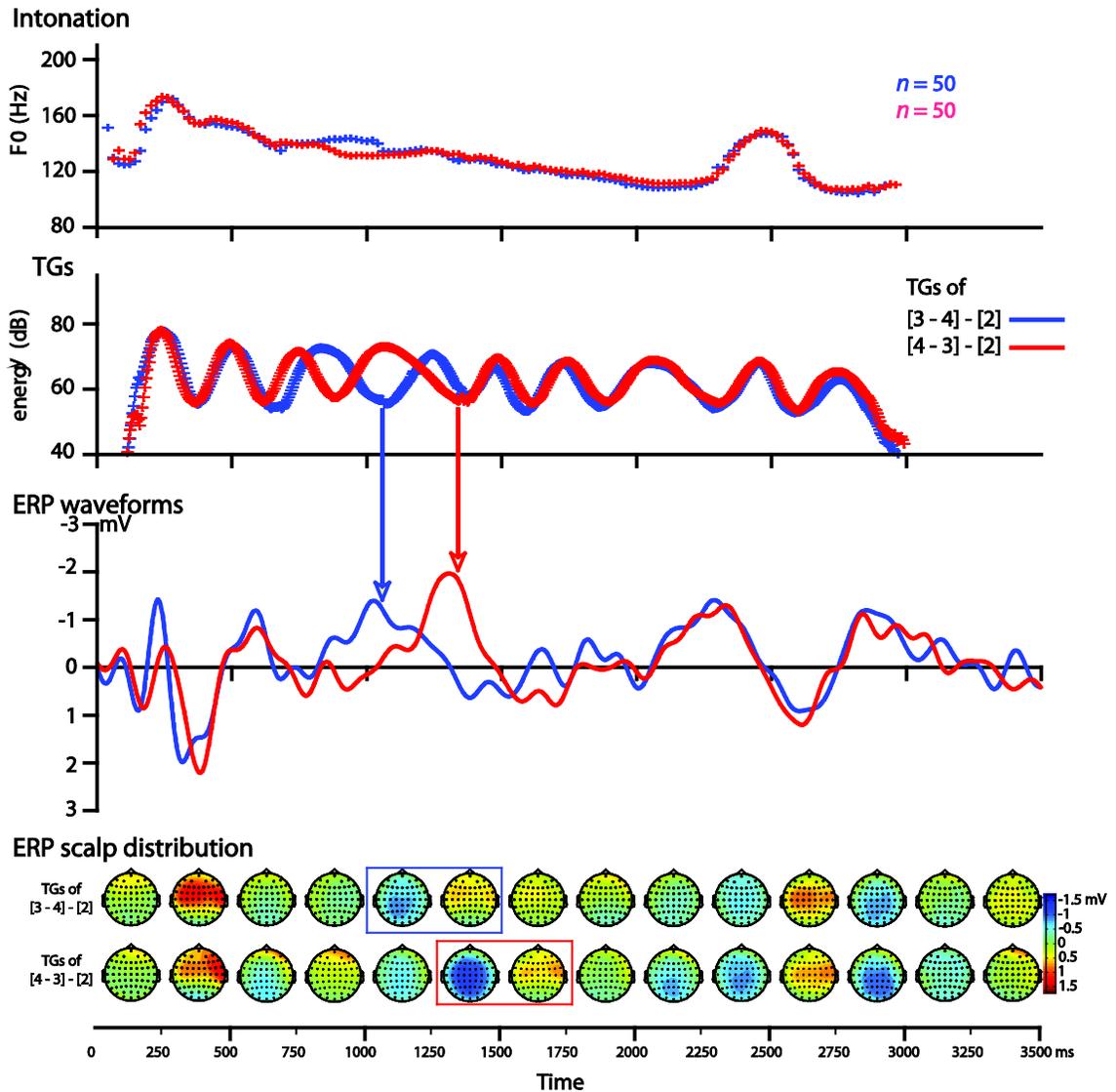


Figure 4. Acoustic stimuli with changing TGs and effects on ERPs. Mean F_0 and dB of utterance stimuli with changing TGs of 3-4 vs. 4-3 syllables (cdns 1 vs. 2); averaged ERPs over fronto-central ROI and topographical representation of the ERP over the entire scalp. Each topography represents a 250 ms time-window.

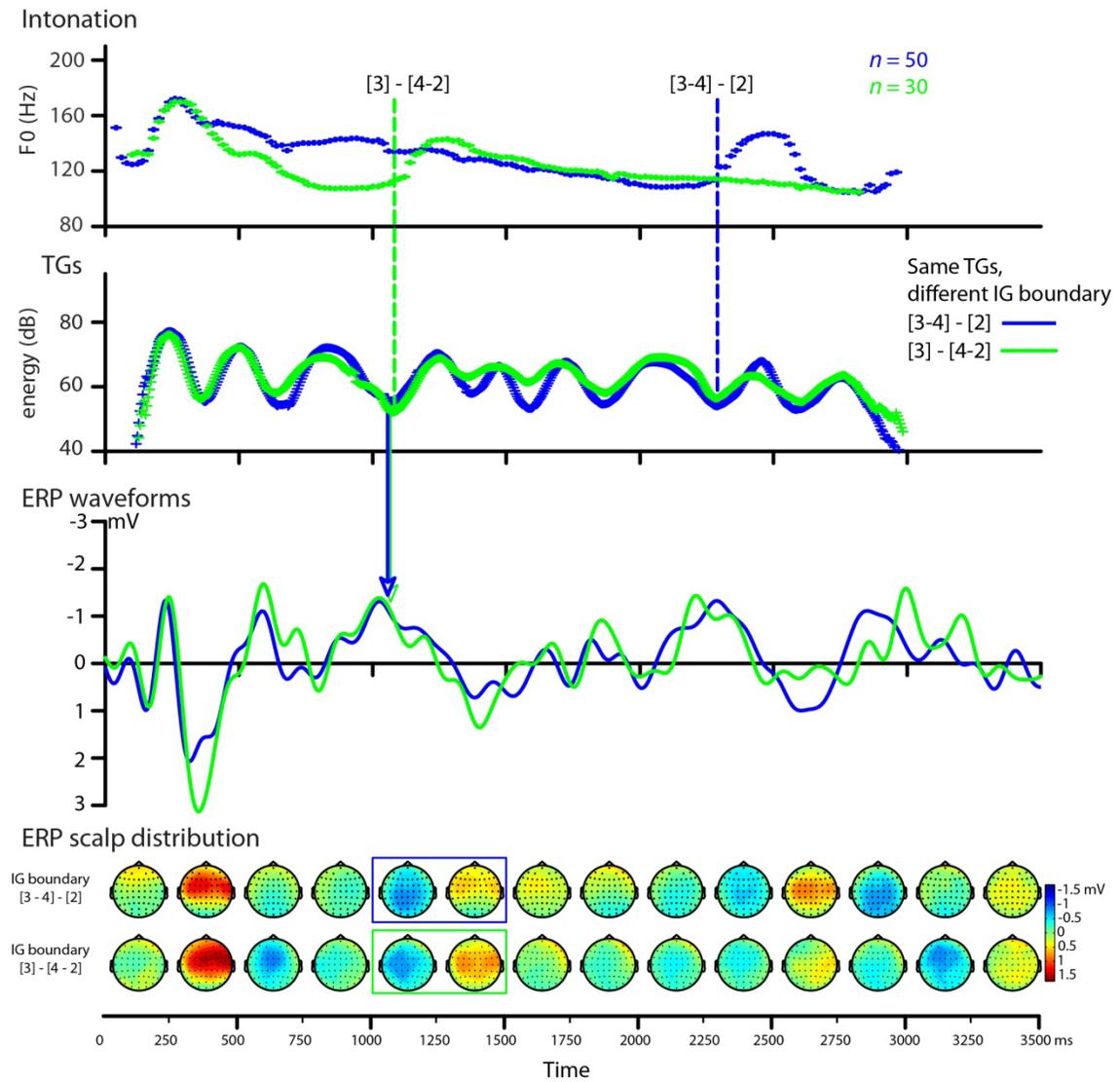


Figure 5. Acoustic stimuli with changing IGs and effects on ERPs (cdns 1 vs. 3). Mean F₀ and dB of the utterance stimuli with changing IGs; averaged ERPs over fronto-central ROI and topographical representation of the ERP over the scalp. Each topography represents a 250 ms time-window.

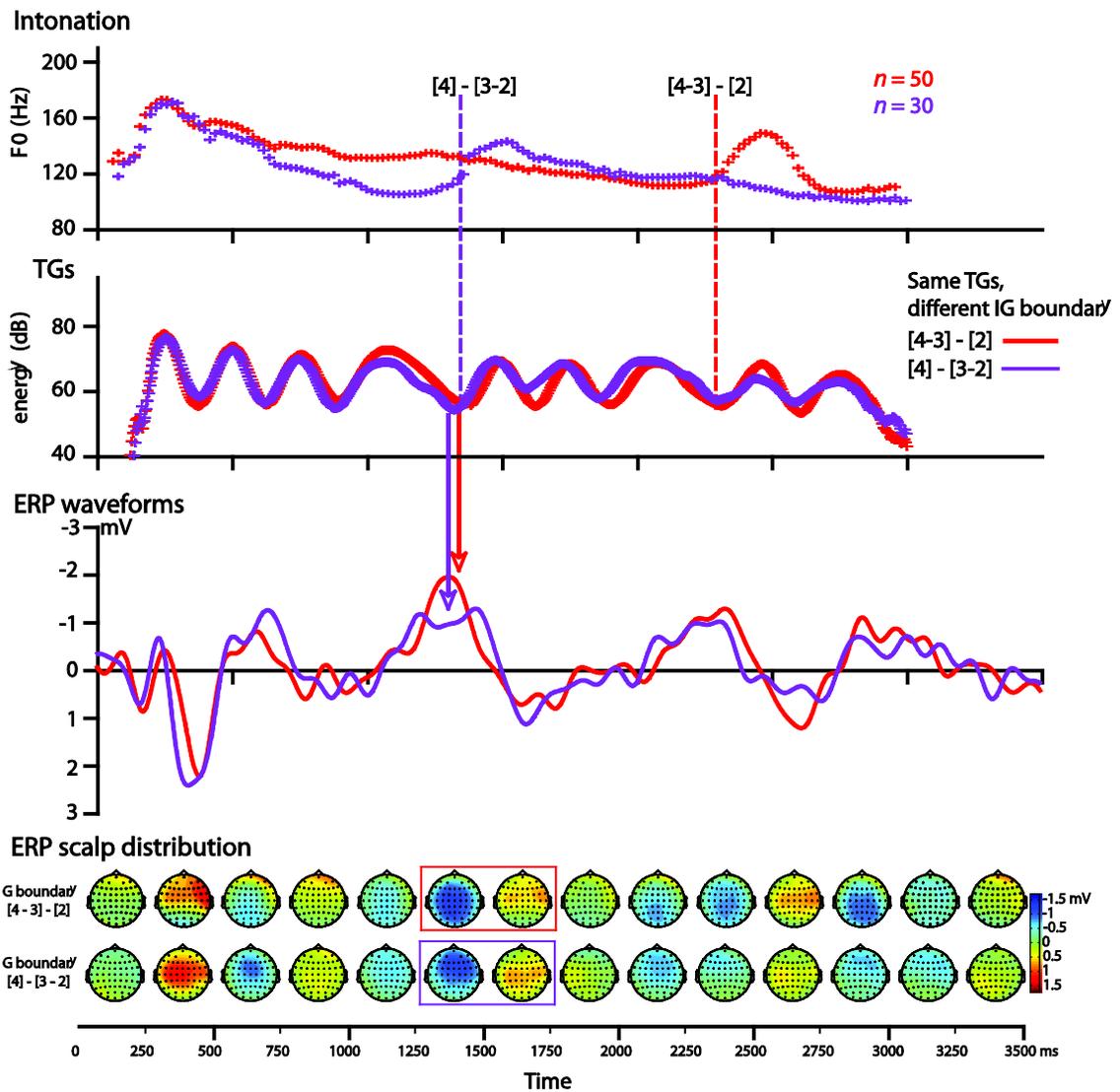


Figure 6. Acoustic stimuli with changing IGs and effects on ERPs (cdns 2 vs. 4). Mean F_0 and dB of the utterance stimuli; averaged ERPs over fronto-central ROI and topographical representation of the ERP over the scalp. Each topography represents a 250 ms time-window.

A number of complementary analyses were carried out on the results representing the main conditions (1 and 2, see Figure 2). First, statistical comparisons (t -tests) of the latency of the initial CPS confirmed a significant effect of TG length with 3-syllable TGs

eliciting earlier CPSs than 4-syllable TGs [$t(19) = -26.247$; $p < .001$; $\eta^2 = .973$]. Second, to evaluate the possible effect of jitter in the spoken stimuli, we sorted single artifact-free trials of EEG signals at Cz according to the latency of the acoustic *P-center* (Marcus, 1981) of the long syllable marking the end of the initial TG. The results of this analysis are displayed in Figure 7 (note that the shortest latency in this plot is down the y -axis). The overlaid dotted lines show that across trials, as the *P-center* latency increases, so does the latency of the negative maxima preceding the CPS. Hence, one sees that jitter in the stimuli had little impact on the CPS response.

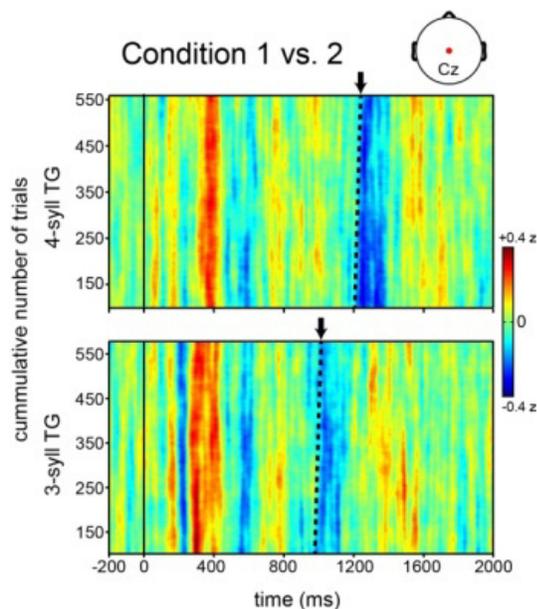


Figure 7. ERPs across trials and onset latencies of the initial TGs. Cumulative EEG trials of the 20 participants sorted according to the onset of the P-center of long syllables marking the first TG in utterances (shortest latency down the y axis). The dotted lines show that the changing latency of the CPS corresponds to the changing length of the TG.

Even though the above results confirm the effects of TGs on CPSs, one cannot rule out that these results may relate to the semantic-syntactic attributes of the stimuli that could

incidentally match TGs. Studies have demonstrated that the CPS arises independently of meaning or syntax of heard speech (Pannekamp, et al., 2005), but the CPS in this work was not linked to TGs. Experiment 2 aimed to disentangle possible semantic-syntactic influences on CPSs from TG effects by examining neural responses to nonsense sequences.

4. Experiment 2

4.1. Materials and Methods

4.1.1. Participants.

The participants described in the above Experiment 1 also took part in Experiment 2.

4.1.2. Stimuli and procedures

The stimuli consisted of 100 meaningless sequences of nine consonant-vowel (CV) combinations. These contexts were produced by a native speaker of French using the pacing technique described in Experiment 1. Each sequence was balanced across conditions with no repeated C or V element within a sequence. Moreover, none of the sequences contained combinations creating recognizable lexemes (for example: [na jy wã fœ tø zẽ ni zõ moẽ]). The stimuli bore the same prosodic patterns used in conditions 1 and 2 of Experiment 1 (see Figure 2). The F_0 and energy contours of the sequences are presented in Figure 8 where each overlaid dotted line represents one stimulus. The energy contours show that the sequences contained groups marked by a final long syllable without pauses, and these groups changed from a 3-4 pattern (blue lines) to a 4-3 pattern (red). These changes representing, respectively, conditions 1 and 2 appear in contexts where F_0 remained similar across contexts. Specifically, one can see that all the sequences bore two IGs marked by a reset of about 6 semitones. Finally, as with the utterance stimuli, the acoustic analyses of Figure 8 show that the pacing technique used to elicit the speech stimuli served to create contexts with specific prosodic patterns and relatively little jitter. Experiment 2 followed the same procedures as Experiment 1 and the same parameters were used for EEG recording and ERP analysis.

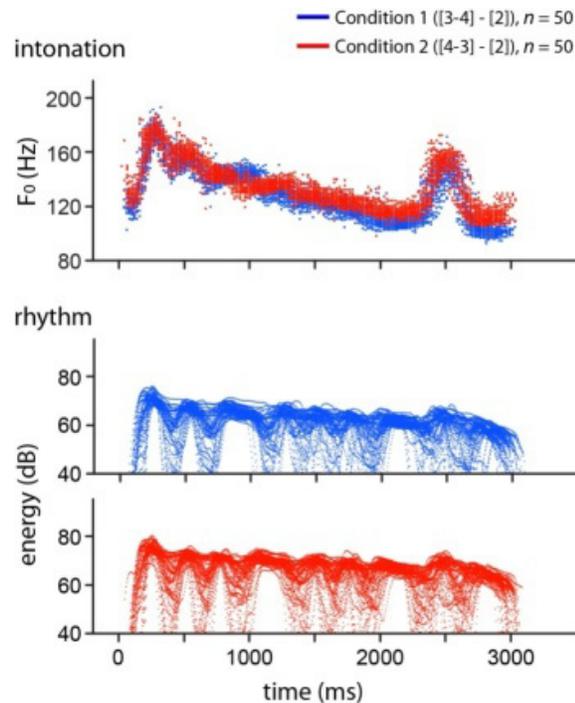


Figure 8. Measured acoustic attributes of meaningless sequences of syllables.

4.2. Results

The results overall resemble those of Experiment 1. First, as in the above test, the stimuli elicited three distinct CPSs matching the number and latencies of TGs but not IGs (compare Figures 4 and 9). As with meaningful utterances, we examined the effects of changing TG lengths (conditions 1 vs 2) by comparing mean ERP amplitudes within a designated 250 ms time-window of a fronto-central ROI (see Methods). A statistical analysis (t -test) confirmed that ERP amplitudes differed significantly across conditions [$t(19) = 4.934$; $p < .001$; $\eta^2 = .562$] suggesting that the length of the initial TGs influenced the responses. Second, a comparison of the onset latencies of the first CPS of each listener confirmed a significant difference between responses in conditions 1 and 2 [$t(19) > -16.511$; $p < .001$; $\eta^2 = .935$]. Finally, we examined possible effects of jitter in the stimuli and found that, as in Experiment 1, changes in the latency of the P -center of the final syllable of a TG leads to consistent changes in the latency of the CPS (see Figure 10). As it appears, jitter in the spoken stimuli did not have a major effect on the timing of CPSs.

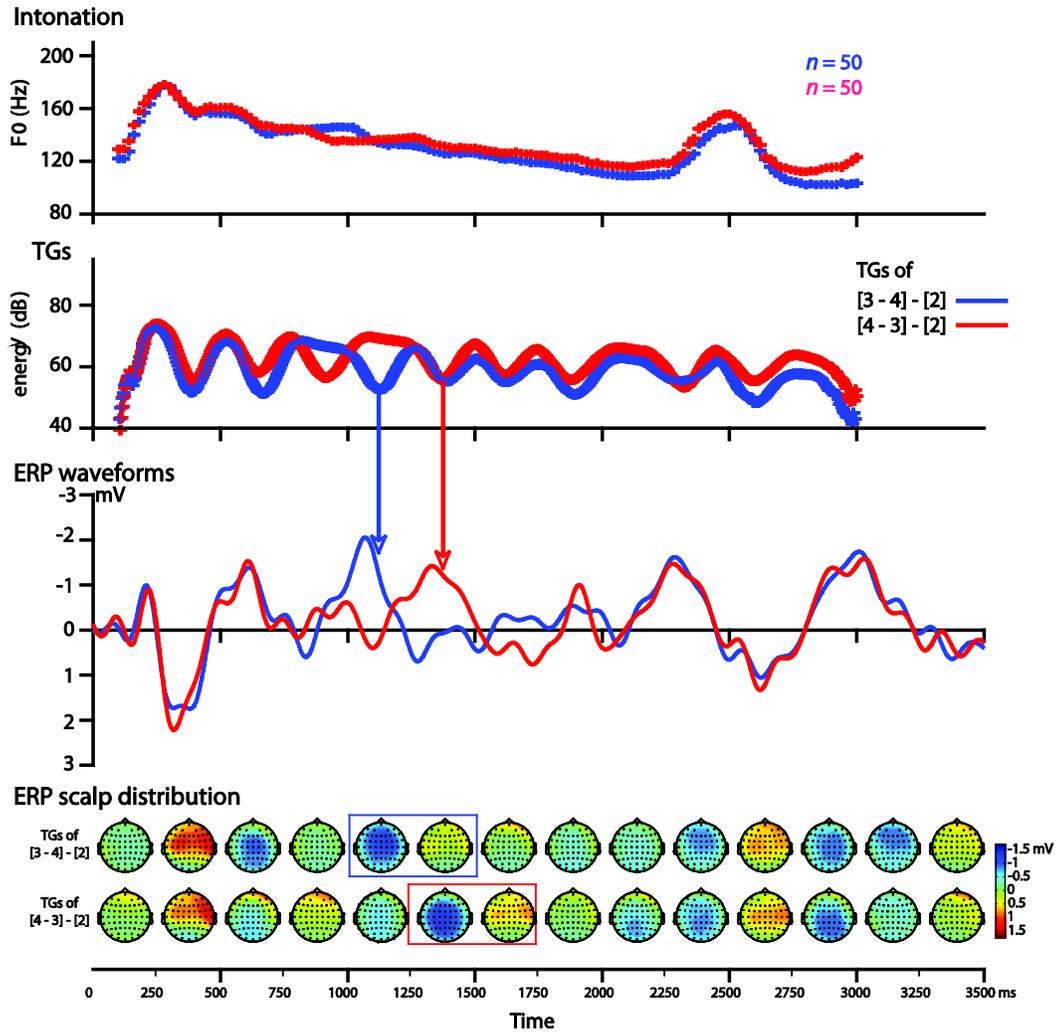


Figure 9. Acoustic stimuli (nonsense syllables) with changing TGs and effects on ERPs. Mean F_0 and dB of presented sequences of nonsense syllables (cdns 1 vs. 2); averaged ERPs over fronto-central ROI and topographical representation of the ERP over the scalp. Each topography represents a 250 ms time-window.

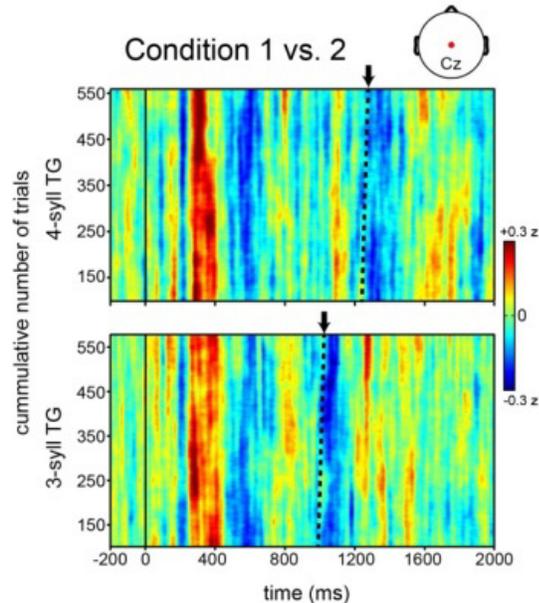


Figure 10. ERPs across trials and onset latencies of the initial TGs (nonsense syllables). Cumulative EEG trials of the 20 participants sorted according to the onset of the P-center of long syllables marking the first TG in presented nonsense sequences (shortest latency down the y axis). The dotted lines show that the changing latency of the CPS corresponds to the changing length of the TG.

4.2.1. The effect of semantic-syntactic information

Finally, to address the central question of the overall effect of semantic-syntactic information, a comparison across experiments was performed. A 2x2 ANOVA comparing the *type* of stimuli (utterances of Exp. 1 vs. syllable sequences of Exp. 2) by the *length* of the initial TGs (3 vs. 4 syllables) was applied to the mean ERP amplitudes within the same 250 ms time-window used in previous analyses and over the same ROI (see Method). Cross-comparisons of ERP amplitudes yielded a main effect of TG *length* [$F(1,19) = 56.795$; $p < .001$; $\eta^2 = .749$] but no significant effect of stimulus *type* [$F(1,19) = .21$; $p =$

.652; $\eta^2 = .011$], and there was no significant interaction [$F(1,19) = 1.761$; $p = .2$; $\eta^2 = .085$]. In short, the results agree with previous work on the point that CPSs may not link principally to the semantic-syntactic content of speech since such responses arise similarly in hearing utterances or meaningless sequences (Pannekamp, et al., 2005). Moreover, with reference to previous studies using prosody, the present results clarify that CPS may specifically reflect TG chunking and not F_0 contours of intonation phrases.

5. General discussion and conclusions

It is recognized that in numerous behaviors involving sequence learning and production, chunks emerge spontaneously, appearing as temporal groups marked by a lengthening of elements. Such grouping is thought to reflect a domain-general process of *input or perceptual chunking* (e.g. Gobet, et al., 2001; Graybiel, 1998; Terrace, 2001). The purpose of the present study was to examine how such chunking applies to speech. We used ERPs, specifically the CPS, to capture this process and hypothesized that listeners perceptually chunk speech in terms of temporal groups independently of other variables. The results of Experiment 1 confirm that the CPS component is evoked by temporal groups, regardless of the intonational phrasing of speech (see Figures 4 to 6). Similar findings were obtained in Experiment 2 using sequences of nonsense syllables where, again, the CPS response remained time-locked to temporal groups regardless of changing intonation. Moreover, no significant difference was found between CPSs evoked in syntactically well-formed utterances (Exp. 1) and those evoked in non-sense series of syllables (Exp. 2), which rules out potential effects of *output* chunking involving forms in LTM. Taken together, the results of the experiments show that neither the intonational aspects nor the syntactic-semantic attributes of the stimuli significantly affected the amplitudes and latencies of CPS. Hence, the observations indicate overall that CPSs were elicited by temporal groups, supporting the idea that listeners perceptually chunk incoming speech in terms of these groups. To discuss the implications of these findings, we draw attention to basic problems that have arisen in the literature with respect to received views of how listeners segment speech.

As we noted, research has traditionally assumed that listeners divide speech into units such as words. However, many corpus-based studies show that, in their daily language, speakers use vast numbers of expressions that may not involve separate words as in writing but function as lexical or formulaic chunks (Bates & Goodman, 2001; Bybee, 2006; Dabrowska, 2004; Wray, 2002, 2008; Zeschel, 2008). Such observations lead one to ask how novel sequences constituting verbal expressions come to be learned and produced as units -- which implies some “pre-linguistic” or “domain-general processes of chunking” (Beckner & Bybee, 2009, p. 28; Bybee, 2010). On this question, Miller’s (1962) idea that speech is chunked in perceptual units the size of *phrases* does not reflect a domain-general unit. There are, though, observable input chunks that appear across behaviors in the form of temporal groups marked by a lengthening of elements (see above introductory section 1.1.). Yet no previous research has demonstrated a relationship between these input chunks and listeners’ perception of speech. As suggested in our literature review, a central methodological problem may explain the lack of work on this question.

In particular, much of what is known about chunking stems from studies involving recall. However, investigating *input* chunking requires observations of on-line responses. This is partly because responses based on recall can entail an access to forms in LTM leading to an *output* chunking. In considering this technical problem, the present study used ERPs and the CPS component, which is known to capture listeners’ on-line responses to marks of various groups in speech. This CPS component is particular. It presents a rising negativity across heard sequential elements, suggestive of a buffering of information, followed by a rapid positive shift coinciding with a group-ending mark. In the above tests, the CPSs served to capture listeners’ response to temporal groups, intonation phrases, and semantic-syntactic information. The findings show that, across conditions, CPSs are specifically elicited by temporal groups. This is best illustrated by comparing Figures 4 and 9, which show averaged ERPs for presented utterances and nonsense sequences. Note that all the contexts contain two intonation phrases. Yet, the figures show three CPS peaks throughout, which conforms to the three temporal groups in the stimuli. This supports the idea that listeners are perceptually chunking incoming signals in groups marked by length

changes. Importantly, these length marks create the same type of chunking that emerges in reciting lists (e.g. Figure 1) or in sequence learning by humans and animals (cf. “inter-response times” as a general mark of chunking: Terrace, 2001). This latter finding is central in that it suggests that input chunking can operate similarly across verbal and non-verbal domains.

On the problem of distinguishing between input and output chunking, the above results show that listeners’ responses to temporal groups, as reflected by CPSs, do not differ significantly across utterances or meaningless sequences. Thus, the observations suggest that CPSs reflect perceptual chunking and not effects of output chunking related to recognized lexical forms or syntactic structures in the stimuli. This should not be taken to imply that there is no output chunking in speech processing. On the contrary, there is necessarily some interaction between output and input chunking in speech as in other behaviours (see, e.g., Gobet, et al., 2001). However, in the present results, CPSs likely captured listeners’ input chunking prior to syntactic and semantic interpretation. It should also be noted that the components were not significantly affected either by semantic-syntactic information or intonational phrasing, and this latter observation bears implications with respect to previous studies of the CPS elicited by prosodic marks.

Overall, then, the present results support the view that listeners process speech in perceptual units corresponding to temporal groups. It should be noted that this perceptual chunking does not involve verbal recoding (Miller, 1956) but points to a basic sensori-motor process likely involving subcortical structures (Graybiel, 1998). In this light, one should notice that such notions of chunks as recoded “meaning groups” or “phrases” have little value in understanding *input chunking*, and that research on this process may best be oriented by referring to observable marks of temporal grouping, which apply across behaviours (Terrace, 2001).

Acknowledgments

This research was partly funded by SSHRC grant number 410-2008-1732 awarded to Victor J. Boucher and by SSHRC and FQRSC scholarships awarded to Annie C. Gilbert.

References

[References to the present paper were included in the general bibliography]

Figure Captions

[Figures and figure captions were included in the text.]

3.2. Article 2

L'article 2 présente les données d'une expérience conçue et réalisée par Annie C. Gilbert, sous la supervision de Victor J. Boucher (Université de Montréal) et de Boutheina Jemel (Université de Montréal). Monsieur Boucher a contribué au choix des conditions à utiliser et a supervisé la construction des stimuli. Madame Jemel a pour sa part supervisé la collecte de données et les analyses EEG. Les trois auteurs ont contribué également à la rédaction de l'article. Les figures ont été insérées dans le corps du texte pour faciliter la lecture et les références citées se retrouvent dans la bibliographie générale. La présente version de l'article sera soumise à *Brain and Language*.

Running head: SPEECH CHUNKING AND MEMORY

**Perceptual chunking and its effect on memory in speech processing:
ERP and behavioral evidence**

Annie C. Gilbert^{a,b,*}, Victor J. Boucher^a, and Boutheina Jemel^{b,c}.

*Corresponding author: Annie C Gilbert

^a Laboratoire de Sciences Phonétiques, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, QC, Canada, H3C 3J7

^b Laboratoire de Recherche en Neurosciences et Électrophysiologie Cognitive, Hôpital Rivière-des-Prairies, 7070, boul. Perras, Montréal, QC, Canada, H1E 1A4

^c Centre de recherche Fernand-Seguin, Département de psychiatrie, Université de Montréal.

Abstract

We examined how perceptual chunks of varying size in utterances can influence immediate memory of heard items (monosyllabic words). Using behavioral measures and ERPs (N400) we evaluated the *quality* of the memory trace for targets taken from perceived temporal groups of 3 and 4 items. Variations in the amplitude of the N400 showed a better memory trace for items presented in temporal groups of 3 compared to those in groups of 4. Analyses of behavioral responses along with P300 components revealed effects of chunk position in the utterance. This is the first study to measure the on-line effects of perceptual chunks on the memory trace of spoken items. Taken together, the results demonstrate that perceptual chunking of speech facilitates information buffering and allow interpretation on a chunk-by-chunk basis.

Keywords

Chunking, working memory, prosody, ERP, N400, P300.

1. Introduction

The ability to interpret speech sounds inherently requires that rapidly changing sequences be kept in short-term memory (or “working memory”, Baddeley, 2010). However, since short-term memory is limited, one must assume that speech processing operates over some chunk of signal that fits this limited store (Kurby & Zacks, 2008). In fact, a number of authors argue that what Miller (1956) called “chunking” can imply a basic perceptual process that groups stimuli (see Gobet, et al., 2001; Graybiel, 1998; Terrace, 2001). Such grouping is commonly observed in sequence learning tasks. For instance, in learning and producing novel sequences of digits or nonsense syllables, temporal groups arise spontaneously (Terrace, 2001). These groups are associated with a domain-general *perceptual chunking* and are marked by characteristic delays or a lengthening of inter-response times at the ends of groups (Gobet, et al., 2001; Terrace, 2001). It is known that perceptual chunks or groups do not tend to exceed four items, which reflects the capacity limits of short-term serial memory (for an extensive discussion of chunk limits in sequence recall, see Cowan, 2000). Moreover, numerous studies have shown that grouping items in three or four benefits sequence memory (Boucher, 2006; Broadbent & Broadbent, 1973; Chen & Cowan, 2005; Cowan, 2000; Frankish, 1989, 1995; Hitch, et al., 1996; Mayberry, et al., 2002; Reeves, et al., 2000; Terrace, 2001; Wickelgren, 1964). What is less recognized, though, is that temporal grouping marked by length changes also operates in speech production and perception. For instance, a number of studies attest to the production of variably termed “accentual” groups in natural speech that do not tend to exceed four syllables on average (e.g., Boucher, 2006; Dauer, 1983; Martin, 1999), while other studies suggest that listeners detect such groups in speech (Boucher, 2006; Christophe, Gout, Peperkamp, & Morgan, 2003; Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Gilbert, Boucher, & Jemel, 2010, submitted; Nooteboom, 1997).

Aside from behavioural observations, a recent study by Gilbert et al. (Gilbert, et al., 2010, submitted) using ERPs has provided more direct evidence of a perceptual chunking of heard speech in terms of temporal groups. In particular, the study showed that a neural

component called *Closure Positive Shift* (CPS) is evoked by group final lengthening – independently of the intonation patterns or the syntactic-semantic content of utterances. In other words, the evoked CPS response revealed that listeners chunk speech in terms of temporal groups similar to those that appear spontaneously in sequence learning across behaviors (Terrace, 2001). As such, the CPS component is characterized by an incremental negativity over midline central sites followed by rapid positive deflection. This rising negativity across elements in a temporal group is suggestive of a buffering of information that ends with a positive shift corresponding to a group-final mark. The CPS has previously been associated with intonation and semantic-syntactic units (cf. Bögels, et al., 2010; Brown, et al., 2011; Dilley, et al., 2010; Itzhak, et al., 2010; Kerkhofs, et al., 2009; Pauker, et al., 2011; Steinhauer, et al., 1999; Steinhauer & Friederici, 2001). Although evidence from Gilbert et al. clarifies that the CPS can specifically reflect listeners’ perceptual chunking of speech in temporal groups, it remains unclear that such chunks link to processes of working memory. If indeed this is the case, then one might expect that temporal groups of differing size would influence listeners’ immediate memory of speech.

In assessing this hypothesis, we focus on how perceptual groups in utterances variably affect the memory trace of heard items. In our test, we use an adapted Sternberg task where listeners hear an utterance followed by a target word and are asked to indicate, as quickly as possible, whether or not the word was part of the utterance. The effects of perceptual groups on immediate memory are evaluated by way of listeners’ recall and characteristics of the N400 and P300 evoked by the target. We briefly explain in the following paragraphs the relevance of these components.

It should be reminded that a Sternberg task is a memory scanning paradigm. Results on this task generally show that increasing the number of verbal items in a presented set leads to an increase in response times to given targets by about 30 to 40 ms per item (Sternberg, 1966). This linear relation between the number of items in a set and response times is taken to show that participants are scanning the entire content of their working memory before responding. Such scanning does not imply an activation of every stored

item but does imply some process of comparison between a target and a memory set. Only when a match is established is there *activation* and retrieval of an element leading to a response. In terms of the two components used in the present report, the literature generally shows that P300 can be related to the scanning and comparison process, whereas N400 may mostly relate to the activation of items. To further clarify, we consider each component in turn.

In terms of the P300, some of the earliest reports of this component referred to memory scanning in a Sternberg task. Research has shown that the P300 is associated with stimulus comparison and classification during the scanning process (Kutas, et al., 1977; Magliero, et al., 1984). A number of studies have reported that the amplitude of P300 evoked by a target decreases as the size of the memory set increases (see, e.g., Gomer, Spicuzza, & O'Donnell, 1976; Houlihan, et al., 1998; Karrer, et al., 1980; Marsh, 1975; Pelosi, et al., 1995; Pratt, et al., 1989; Starr & Barrett, 1987; Wolach & Pratt, 2001). P300 latencies have also been related to stimulus classification times (Duncan-Johnson & Donchin, 1982; Kutas, et al., 1977; Magliero, et al., 1984; McCarthy & Donchin, 1983), though there is some disagreement on this point. Several studies show significant increases in P300 latency for a target when sets increase in size (Adam & Collins, 1978; Ford, et al., 1982; Ford, et al., 1979; Gomer, et al., 1976; Kramer, et al., 1986; Marsh, 1975; Pfefferbaum, et al., 1980; Pratt, et al., 1989; Starr & Barrett, 1987) but some show no significant change beyond sets of two or three items (Pelosi, et al., 1995; Pelosi et al., 1992; Pfefferbaum, et al., 1980). In the present report, the P300 component is specifically used to evaluate effects of perceptual chunking on working memory scanning.

As for the N400, though this component was first thought to reflect the degree of semantic integration of an item to a preceding sentence (Kutas & Hillyard, 1980), it is now associated with effort relating to the activation of an item in a context (for a review, see Kutas & Federmeier, 2000). More specifically, the amplitude of the N400 varies inversely with the degree of pre-activation of an item: the more restricted a context is, the more an item gets pre-activated and the smaller the N400 effect (Kutas & Federmeier, 2000). This

means that the relative amplitude of the N400 can reflect, among other things, the quality of the memory trace of an item. An item with a strong memory trace would be easier to recall and generate a smaller-amplitude N400 than an item with a poorer memory trace. Hence, the N400 can serve to evaluate effects of perceptual chunks of varying size on the quality of the memory trace of heard items.

In view of the general nature of chunking and the above evidence from speech production and perception, we hypothesize that listeners perceptually chunk utterances in temporal groups and that this process links to immediate memory of heard speech. To demonstrate this link, we varied the size of temporal groups of three and four items, which are known to affect memory differently (with groups of three having optimal benefits for sequence recall; Boucher, 2006; Cowan, 2000; Mathy & Feldman, 2012). In using a Sternberg task to study these effects, we predict that the accuracy of target recognition, response times, and N400 amplitudes will vary with the size of temporal groups. Furthermore, as we noted earlier, response times in a typical Sternberg task vary with the number of items in a set while item position is generally expected to have little effect. This is generally interpreted as suggesting that subjects are scanning the entire content of their working memory before responding. However, in our task, the presented stimuli are utterances containing temporal groups. In these contexts, then, items may be perceived in groups so that a scanning may operate by consecutive chunks. In other words, a match between a target and an item of the first chunk in working memory may lead to an earlier activation compared to a target in a following chunk. If this is the case, then response times to targets should vary in terms of whether they occur in the first or second perceptual group in an utterance, and this may be reflected in the P300 characteristics. Taken together, a validation of the above hypotheses provides a way of determining how perceptual chunking can influence working memory in speech processing.

2. Materials and methods

2.1. Participants

The participants were 16 native speakers of French, aged from 19 to 42 years (mean age = 25.6), who presented normal hearing levels following a standard audiometric screening. All were dominant right-handers (Oldfield, 1971), with no history of substance abuse (other than tobacco smoking), and no history of neurological or psychiatric disorder. All showed normal memory performances on the digit-span test of the WAIS (Wechsler, 1997; overall, average normalized score: 10.16, std dev.: 2.4). Written consent was obtained from every participant, and the research protocol was approved by the ethics committee of the *Hôpital Rivière-des-Prairies* (Montréal, Canada).

2.2. Stimuli

The experimental stimuli for the present study consisted of 100 pairs of French utterances and target lexemes. The target lexemes were all monosyllabic nouns drawn from the utterance context. Only nouns were used as targets to avoid any confounding effect of lexical class on the retrieval processes (Halgren, et al., 2002). The utterances, on the other hand, were made up of monosyllabic lexemes and functors in a similar sentence structure. This structure allowed a speaker to produce the contexts with specific prosodic patterns and temporal groups (TGs), as illustrated in Figure 1. A pacing technique (described below) served to control our experimental variables. These variables involved changing *TG length* (groups of 3 and 4 syllables, see blue and red lines in Figure 1) and *TG position* (1st and 2nd group in the utterance, see dark and light lines). The target lexemes were distributed in equal numbers across conditions so that 25 targets were placed in each combination of TG length and TG position (giving four different sets of 25 stimuli). As for the potential effects of syntax, the grammatical role of the TGs was maintained across utterances: the first TG was always a subject NP, and the second a complement to the subject NP (the third TG, which was VP, is not relevant in the present study). In all TGs, the target noun was a pre-final element, which served to avoid potential recency effects. Moreover, across conditions, the target nouns used had comparable frequency indexes in French (Desrochers, 2006).

In order to create natural sounding contexts, the recording of the target nouns and utterances was done by a male native speaker of French using a pacing technique. By this technique, the speaker follows heard series of rhythmic tones (a metronome-like pacer) in producing utterances with the desired prosody. (For further details of the pacing technique,

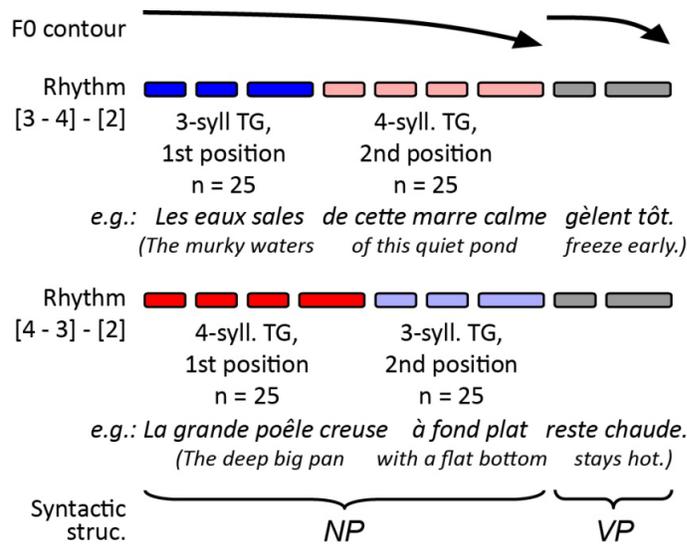


Figure 1. Schematic representation of intended prosodic structure of the stimuli and the origin of the targets.

we refer to Gilbert, et al., submitted.) The speaker's productions were recorded using a 16-bit external soundcard (*Fast-track Ultra*, M-Audio) at a sampling rate of 44.1 kHz and stored in separate .wav files. The target lexemes were recorded separately (and not spliced from the original utterance to avoid coarticulation effects), and this provided natural sounding prosody that was appropriate for isolated lexemes in French (Fant, et al., 1991). The target lexemes lasted 447.32 ms on average (minimum 425 ms, maximum 477 ms) and bore a neutral tone.

Finally, to tag the onset of the stimuli correctly in the EEG, we controlled the delay between the beginning of a sound file and the onset of a stimulus (target word). Such measures are required to minimize the latency jitter when averaging the ERPs and to allow

for precise calculation of intervals between the target and the participant's response. In particular, we made sure that constant intervals were maintained between the Perceptual-Center (*P-center*) of the target lexeme and both the onset and offset of the sound file. The *P-center* constitutes the point in time where a syllable is perceived, and it can be localized on the waveform by means of a simple calculation (Marcus, 1981). In the present case, the *P-center* of the monosyllabic target was always 200 ms from the beginning and 300 ms from the end of the file. Audio file editing and amplitude normalization was performed using *GoldWave* (GoldWave inc., v5.58).

2.3. Procedures

Figure 2 represents the time-course of a typical trial. All contexts were presented using insert earphones (*Eartone 3A*, EAR Auditory Systems). The sounds were delivered at a constant intensity, which was calibrated with a sonometer to obtain peak levels of 68 dBA (i.e., conversation-level sound volume) at the inserts. Participants were instructed to listen to the stimuli while maintaining their gaze at a fixation point on a blank screen. They were also instructed to keep index fingers of both hands on two buttons of a response box. Their task was to indicate, as fast as possible by a key press, whether or not the prompt was part of the preceding utterance or not. Half the participants answered affirmatively by pressing the right button, half by pressing the left button. Sound files were played back via *E-prime 1.0* (Psychology Software Tools) in random blocks divided by rest pauses. Each block contained on average 78 stimuli (17 targets and 61 distracters) in random order with the restriction that no consecutive stimuli presented the same prosodic pattern. The interval between the end of an utterance and the beginning of the prompt file varied from 750 ms to 1,200 ms in steps of 50 ms (11 different intervals). Presentation of successive stimuli was initiated by the participants' response or automatically after 1,500 ms if no response was received.

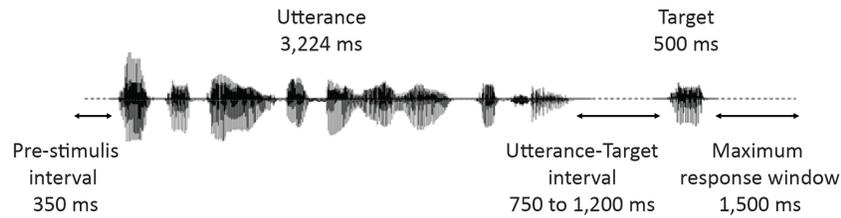


Figure 2. Time-course of a trial.

2.4. EEG recording and ERP analysis

The EEG data were recorded through shielded electrodes embedded in an elastic cap (*Easy Cap*) according to the enhanced 10-20 system (Sharbrough, et al., 1991). Two bipolar electrodes placed above and below the dominant eye (vertical EOG) and at the outer canthus of each eye (horizontal EOG) served to record eye movements and blinks. A left mastoid electrode was used as an online reference for all scalp electrodes, and AFz served as the ground. The EEG including EOG signals were recorded continuously with a band-pass from DC to 100 Hz at a sampling rate of 512 Hz, and stored with trigger codes. The EEG signal was average-referenced off-line and low-pass filtered at 10 Hz using a digital filter. Only trials corresponding to correct responses were kept and submitted to the further procedures. Eye-blinks were corrected by subtracting from the EEG the PCA-transformed EOG components for each electrode, weighed according to the vertical EOG propagation factors (computed using linear regression). Segments with other artefacts were rejected. The obtained artefact-free EEG segments, which were time-locked to the onset of a target sound-file, were then averaged for each group length and group position, from 100 ms before to 700 ms after the onset of a target sound-file. The ERPs were also pooled together according to the TG length irrespective of TG position and vice-versa. This allowed for the calculation of difference waves isolating one factor from the other. Time-sequence topographies of the differential waveforms were used to identify the regions of interest (ROI) and time windows showing differences between experimental conditions. ROIs were delimited according to the distribution of the maximum differences on the scalp. Each participant's ERP was re-averaged according to the ROIs and these were used to obtain

quantitative measures of peak latency and amplitude (maximal and averaged within a time-window). All statistical analyses were computed using 2X2 ANOVAS with TG length (3 vs 4) and position in the utterance (1st vs 2nd) as implemented in SPSS (version 17.0).

3. Results

3.1. Behavioral data

The recognition task was easy and correct recognition was high, with scores varying from 90% to 96% correct (see Table 1). A 2X2 ANOVA showed no main effects of TG length or position [$F(1,15) < 3.88$; $p > .068$; $\eta^2 < .205$] but there were significant interactions between the two factors [$F(1,15) = 8.934$; $p = .009$; $\eta^2 = .373$]. As for the reaction times (RTs) for correct responses, a significant main effect was found for the position of the TGs [$F(1,15) = 9.124$; $p = .009$; $\eta^2 = .378$], the first group triggering shorter RT than the second. However, there were no significant effects of TG length [$F(1,15) = 1.348$; $p = .264$; $\eta^2 = .082$], and no significant interaction between TG length and position [$F(1,15) = 2.452$; $p = .138$; $\eta^2 = .140$]. Overall, the analyses showed no marked influence of the size of chunks on subjects' behavioural responses, only an effect of TG position in the utterance.

Table 1: Behavioural results (Accuracy of recognition and RT) as a function of TG length and position.

TG position	TG length			
	3 syll.		4 syll.	
	Accuracy (%) <i>SD</i>	RT (ms) <i>SD</i>	Accuracy (%) <i>SD</i>	RT (ms) <i>SD</i>
1 st	89.8 <i>0.053</i>	711.21 <i>182.27</i>	95.8 <i>0.052</i>	677.92 <i>147.06</i>
2 nd	94.9 <i>0.076</i>	716.29 <i>166.78</i>	95.3 <i>0.049</i>	722.94 <i>170.92</i>

3.2. ERP data

3.2.1. P300

Since we expected that the P300 would be affected by the position of TGs, topographical maps of the differences between targets from the two TG positions (1st and 2nd in the utterance) were used to identify ROI and relevant time-windows. A visual inspection showed that the largest difference was in a central-parietal region which includes Cz, C1, C2, CPz, CP1 and CP2. Figure 3 presents the difference topography for TG positions in the time-window of the P300 and the N400 as well as ERPs re-averaged according to the identified ROI (black line = TG in 1st position, gray line = TG in 2nd position). One can see that the difference between the conditions extends outside the range of the P300. In fact, TGs in 2nd position present, overall, a sustained greater negativity between 200 ms and 600 ms than TGs in the 1st position (black line). A 2X2 ANOVA comparing TG length (3- vs 4-syll.) and position (1st vs 2nd) revealed no significant main effects or interactions ($[F(1,15) < 2.8; p > .1; \eta^2 < .157]$) even though the difference between these conditions looks salient in the averaged ERPs.

To evaluate the effect of the experimental conditions on these differences, we focused on the P300 component and measured the average amplitude (per participant) over 5 consecutive 25 ms time-windows starting at 200 ms post onset. For all of these measures, a 2X2 ANOVA comparing TG length (3- vs 4-syll.) and position (1st vs 2nd) showed no significant main effects ($[F(1,15) < 2.879; p > .1; \eta^2 < .161]$) or interactions ($[F(1,15) < 3.989; p > .06; \eta^2 < .21]$). Peak amplitude and latency measures were also taken on the P300 for each participant. The P300 was identified on the average waveform as the positive peak immediately preceding the negative deflexion of the N400 as indicated by the arrows in Figure 3. The results of these latter measures are displayed in Figure 4.

One can see in the left panel of this figure that although targets appear to elicit greater P300 amplitudes when they are heard in the 2nd position, a 2X2 ANOVA showed no significant main effects or interaction ($[F(1,15) < 3.374; p > .08; \eta^2 < .206]$). Moreover, what appears as an effect of TG length in the 1st position was found to be non significant in terms of a repeated measure *t*-test ($[T(1,15) = 1.249; p = .231.; \eta^2 = .094]$). On the other

hand, P300 peak latencies (seen in the right panel of Figure 4) present marked differences with respect to position. A 2X2 ANOVA confirmed that only TG position had a significant impact on P300 latency (TG position = [$F(1,15) = 10.865$; $p = .006$; $\eta^2 = .455$], whereas effects of TG length and interactions were not significant [$F(1,15) < .098$; $p > .759$; $\eta^2 < .008$]).

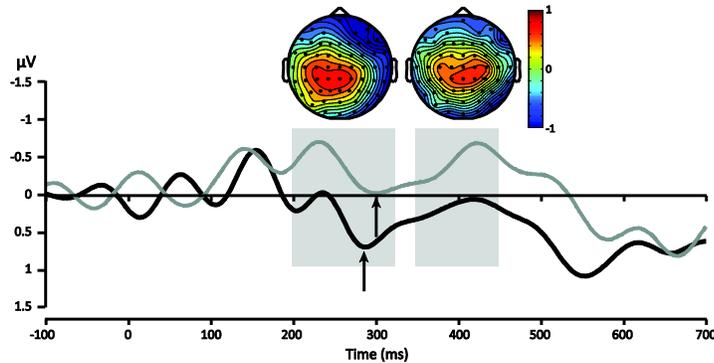


Figure 3. Top: topographies of the average differences between ERPs for targets from TGs in 1st and 2nd positions in the utterance, for 2 time-windows (described in the text). Below: averaged responses for centro-parietal ROI: black line = 1st TG, gray line = 2nd TG; arrows indicate the peak of the P300.

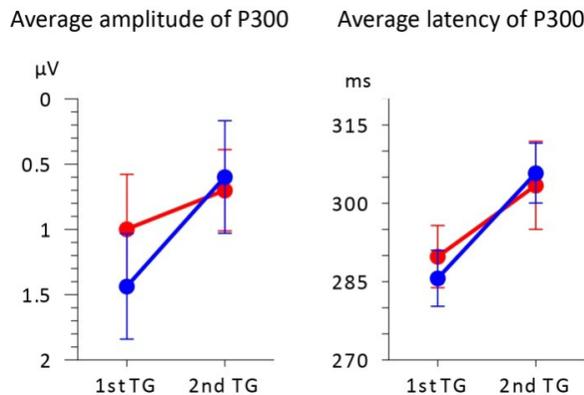


Figure 4. Average amplitude ($n=20$) and latency of the P300 peak plotted according to TG position (1st vs 2nd) and TG length (blue = 3-syll. vs red = 4-syll.). Vertical bars represent standard errors.

3.2.2. N400

Visual inspections of the topographical difference between TGs of 3 and 4 syllables revealed that the greatest difference was found in a region comprising electrodes FCz, Cz and C2 (ROI) and at latencies within the usual time-window of the N400 effect (from 300 to 500 ms, (Kutas & Federmeier, 2000; Kutas, et al., 1977)). Figure 5 shows a topography of the difference between TG lengths (3 vs 4) for a 100 ms time-window from 350 to 450 ms and the ERPs re-averaged according to the ROI. Line colour represents TG length (3-syll. in blue, 4-syll. in red). Note that the 4-syllable TGs elicit a larger N400 than the 3-syllable TGs.

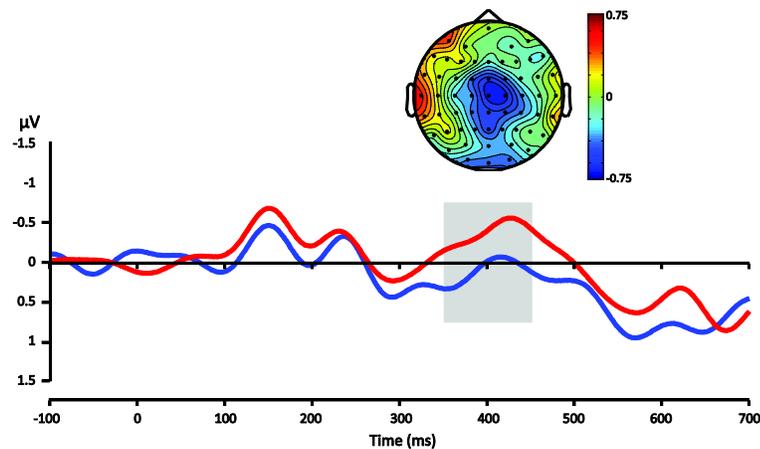


Figure 5. Top: topographies of the average differences between ERPs for targets from TGs of different lengths (4- vs 3-syll.) in the 350-450 ms time-window. Below: averaged responses for the fronto-central ROI; blue lines = 3-syll TG, red lines = 4-syll TG.

This difference in N400 amplitude is further illustrated in Figure 6 representing the average amplitudes in the 100 ms time-window according to both TG length and position. This figure illustrates two main findings. First, TGs of 3 syllables display a smaller

negativity then TGs of 4 syllables when controlling for position. Second, that the averaged amplitude is greater overall for TGs in the 2nd position than from 1st position. Therefore, it seems that two factors influence the amplitude of the N400. However, a 2X2 ANOVA comparing TG length (3- vs 4-syll) and position (1st vs 2nd in utterance) showed a significant effect only for TG length [$F(1,15) = 8.506$; $p = .011$; $\eta^2 = .362$]. Main effects of TG position and interactions were not significant: [$F(1,15) < 1.152$; $p > .3$; $\eta^2 < .071$].

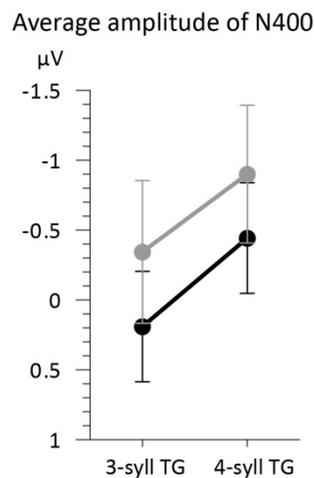


Figure 6. Average amplitude ($n=20$) in the 350-450 ms time-window according to both TG length (3- vs 4-syll.) and position (black = 1st TG, gray = 2nd TG). Vertical bars represent standard errors.

4. Discussion and conclusion

In the above task, listeners heard utterances followed by a word item and were asked to rapidly indicate whether or not the item was part of the heard utterances. Targets within the contexts were placed in perceptual chunks of different size (3 or 4 syllables) and these chunks appeared in different positions (1st or 2nd) in an utterance. The results show that TG size had a significant impact on the quality of the memory trace as reflected in the

amplitude of the N400. However, this effect did not carry over in behavioral responses of recognition accuracy. This can be explained by the fact that the task of remembering items in a heard utterance is quite easy. Moreover, groups of 3 and 4 items used in utterances are found in natural speech and conform to working memory span (Cowan, 2000). Hence, the task was hardly demanding, and this led to high recognition scores with only slight fluctuations across conditions. Nonetheless, the size the TGs had, as expected, a significant effect on the amplitude of the N400 component validating the view that perceptual chunks in utterances link to immediate memory of speech. Additional effects were studied in terms of the memory scanning task that confirmed that speech in working memory is processed on a chunk-by-chunk basis.

To clarify this last point, it should be reminded that the Sternberg task provides a means of assessing how items are processed in working memory. The above results show significant effects of the position of TGs on reaction times. Thus, even though the utterances were of the same length (9 monosyllabic words), the recognition times were shorter when target items were heard in the first group than when they were heard in the second group. To account for such differences, one has to assume that the content of working memory was scanned in terms of ordered chunks and stopped at a point when there was a match between the target and an item in a chunk. Hence, it appears that utterance-related information is stored in serially-ordered chunks and that the scanning of this information in working memory proceeds on a chunk-by-chunk basis. Observations of the P300 also appear to support this interpretation.

The P300 is generally associated with memory scanning processes, and its latency is believed by some to be linked to scanning duration (e.g. Pelosi, et al., 1995). In the above observations, P300 latencies show effects of TG position that conform to duration differences in response times: essentially, TGs heard first in the utterances lead to shorter response times and earlier P300s compared to later-occurring TGs. It may be useful to mention that these results agree with reports of a relationship between P300 latencies and “set size” if one understands that, in the present experiment, sets are given by TGs (see.

Adams, 1915; Ford, et al., 1982; Ford, et al., 1979; Gomer, et al., 1976; Kramer, et al., 1986; Marsh, 1975; Pelosi, et al., 1995; Pfefferbaum, et al., 1980; Pratt, et al., 1989; Starr & Barrett, 1987). That is, the difference with respect to previous work is that the Sternberg task in the present study includes grouped stimuli. Moreover, the P300 responses strongly suggest a scanning of working memory based on ordered groups or chunks.

Overall, the above results, combined with the earlier findings of Gilbert et al. (2010, submitted) offer a perspective on how heard speech is chunked and stored on-line. Basically, the findings suggest that listeners perceive speech in TGs, and these perceptual chunks are stored in working memory in a serial order. That this perceptual chunking of speech in TGs inherently links to memory processes is demonstrated by the effects of group size on the N400 amplitude. In weighing these findings, it is useful to remark that, despite the large body of research on chunking (for a recent review, see Cowan, 2000), there is a paucity of work on the role of perceptual chunking in speech processing. This can be explained by the fact that research on chunking is largely based on experiments involving recall whereas perceptual chunking requires a method that captures *on-line* responses to incoming stimuli (Terrace, 2001). In this context, the above research using ERPs provides a novel demonstration of how perceptual chunking influences immediate memory of heard speech. It should also be emphasized that the evidence indicating that listeners perceive speech in groups should not be equated with the idea that listeners are detecting prosodic phrases that map syntactic forms (e.g. Frazier, et al., 2006). In fact, groups created by delays or a lengthening of elements generally emerge in learning and producing novel sequences in numerous behaviors both verbal and non-verbal (for examples, see Graybiel, 1998; Terrace, 2001). Hence, these universal temporal groups (TGs) do not “map” onto syntactic or semantic structure. Instead, they seem to reflect a domain-general sensori-motor chunking that links to working memory (Graybiel, 1998).

Acknowledgments

This research was partly funded by SSHRC grant number 410-2008-1732 awarded to Victor J. Boucher and by SSHRC and FQRSC scholarships awarded to Annie C. Gilbert.

References

[References to the present paper were included in the general bibliography]

Figure Captions

[Figures and figure captions were included in the text.]

4. Discussion

Les présents travaux avaient trois objectifs complémentaires portant sur le *chunking* perceptif de la parole et ses effets sur la MdeT. Bien que chacun des articles ci-dessus contienne une discussion des résultats, nous regroupons ici les principaux éléments qui répondent aux objectifs spécifiques énumérés à la section 1.2, en vue de mettre en évidence certaines implications dans le domaine de la linguistique.

Objectif 1 : Démontrer que, lors de l'écoute, l'auditeur segmente (chunk) la parole selon des groupes temporels

Les données empiriques

Dans les expériences présentées précédemment, nous avons observé les PÉ de participants lors de l'écoute de la parole afin de déterminer s'il y a un *chunking* perceptif des signaux en fonction de groupes temporels (GT) comme ceux qui apparaissent spontanément dans le rappel de séquences (voir figure 1). Notre hypothèse était que les GT évoquent des CPS, soit une composante qui a été associée à différents types de *chunking* perceptif et ciblé (tels que définis à la section 1.1.2.). Les résultats présentés dans l'article 1 (Expérience 1) confirment cette prédiction sans équivoque. En particulier, on a observé dans les PÉ moyennés que le nombre de CPS correspondait au nombre de GT dans les énoncés (trois) et non au nombre de groupes intonatifs (deux). Qui plus est, la *latence* des CPS correspond étroitement aux marques d'allongement des GT. Ainsi, rappelons que dans l'Expérience 1 les auditeurs ont entendu des énoncés où le premier GT comprend trois syllabes (condition 1) ou quatre syllabes (condition 2). Or, cette différence d'une syllabe dans la longueur de GT s'est reflétée dans les latences de la CPS où il y avait un décalage d'environ 250 ms. Il s'agit là d'une correspondance assez étroite attestant un *chunking* d'entrée. Dans l'essentiel, donc, ces données confirment l'idée que l'auditeur perçoit la parole « par blocs » en découpant le signal selon des GT.

Il est aussi important de noter que les données montrent que le *chunking* perceptif opère spécifiquement en termes de GT, indépendamment des variations majeures au plan de l'intonation et du contenu sémantique-syntaxique des stimuli. Nos contextes expérimentaux ont permis d'évaluer ces effets. Considérons d'abord les effets de groupes

intonatifs. Les conditions 1 et 2 reflètent des contextes où les deux premiers GT sont contenus à l'intérieur d'un seul groupe intonatif et d'un seul syntagme nominal. Par contre, dans les conditions 3 et 4, les GT sont divisés à la fois par une frontière de groupe intonatif (marquée par une remontée de la F_0 de 6 demi-tons) et une frontière syntaxique majeure (entre un syntagme nominal sujet et un syntagme verbal). En observant les CPS pour l'ensemble des quatre conditions (voir figures 4 à 6 de l'article 1), on retrouve trois CPS dont les latences correspondent aux marques des GT. Par ailleurs, des analyses statistiques confirment que seule la longueur des GT a un impact significatif sur la CPS. Autrement dit, la présence de structures intonatives ou sémantiques-syntaxiques différentes n'a eu aucun effet significatif sur les CPS évoquées par les GT.

Implications

Comme on le soulignait dans l'introduction, en linguistique, les marques d'allongement de GT sont souvent interprétées selon différentes unités syntaxiques *a priori* (mots, syntagmes, phrases) ou différents schèmes prosodiques que l'on relie à des unités syntaxiques (mot phonologique, syntagme intonatif, etc.; voir, p. ex. Selkirk, 1982). Toutefois, ce type d'analyse s'applique difficilement aux marques des GT, et ce, pour deux raisons.

Notons d'abord que les marques de GT, bien qu'universelles, peuvent correspondre à une variété d'unités selon la langue, et parfois à aucune unité selon certaines analyses linguistiques. Ainsi, nous avons vu qu'avec un système d'analyse comme ToBI (Beckman & Ayers Elam, 1997) seules les marques d'allongement correspondant à des groupes intonatifs sont considérées. On ne tient pas compte des marques de GT qui apparaissent à l'intérieur des syntagmes intonatifs ni du fait que ces marques sont perçues et contrôlées indépendamment de l'intonation (p. ex. Ladd, 1996; Nespor & Vogel, 1986; Selkirk, 1995). Par exemple, avec un tel système de description, on ne peut analyser les GT qui apparaissent lors de l'apprentissage de séquences (comme les groupes que l'on produit en rappelant un numéro de téléphone).

Cette absence de correspondance entre le GT et une unité linguistique particulière amène à voir l'inapplicabilité des analyses de type linguistique dans l'identification de chunks perceptifs, comme l'entendait Miller (1962). On se souviendra que Miller (1962) avait vu qu'un auditeur ne traite pas l'information de la parole un son à la fois, mais procède par « unité perceptive » qui, selon lui, était de la taille d'un syntagme (angl. *phrase*). Cependant, l'équivalence entre un chunk perceptif et une unité linguistique quelconque ne tient pas pour la simple raison que le *chunking* perceptif n'est pas de nature linguistique. Par exemple, le groupement ou *chunking* apparaît avec l'apprentissage de séquences de syllabes sans sens ou de simples gestes, c'est-à-dire des séquences non verbales ou préverbales qui ne contiennent *aucune unité sémantique ou syntaxique*. On ne peut, en somme, analyser (et encore moins expliquer l'origine) des groupements non linguistiques issus d'un processus de *chunking* avec des méthodes de description qui tiennent pour acquises ces unités linguistiques.

Par rapport à ces problèmes et aux présupposés des analyses prosodiques, nos résultats ont deux implications fondamentales quant au rôle du *chunking* perceptif de la parole. La première implication porte sur l'effet des GT par rapport à d'autres marques prosodiques. Le fait que les GT soient détectés lors de l'écoute, sans égard aux contours intonatifs de syntagmes, démontre qu'ils sont impliqués dans les étapes initiales du *chunking* de la parole. Ces résultats ne nient pas le traitement de l'intonation, mais indiquent seulement que l'intonation n'influence pas le *chunking* perceptif du signal. Par contre, la détection de marques de GT *à l'intérieur de groupes d'intonation*, tel que démontré dans l'article 1, remet en question la justesse d'analyses linguistiques basées sur des systèmes comme ToBI. Ces systèmes ne reconnaissent pas que l'allongement constitue une marque prosodique contrôlée de façon indépendante et pouvant apparaître à l'intérieur de groupes intonatifs. Notons que de tels *a priori* au plan de l'analyse de la prosodie mènent à une interprétation ambiguë, sinon fautive, des marques qui évoquent une CPS. Aussi, la présente recherche suggère que, sans un contrôle étroit des marques d'allongement et de F_0 , on ne peut présumer que ce soit les courbes d'intonation qui évoquent une CPS. Notons par ailleurs que nos résultats démontrant la primauté des effets

des GT concordent avec un ensemble d'études étayant l'universalité des marques d'allongement et le rôle du groupement dans l'apprentissage de formes verbales (Dabrowska, 2004; DePaolis, et coll., 2008; Lieven, et coll., 2003; Lieven, et coll., 1997; Pine & Lieven, 1997; Pine & Martindale, 1996; Tomasello, 1992, 2000a, 2000b).

Une deuxième implication de nos résultats porte sur la nature des processus de *chunking* révélés par la CPS. Rappelons que la CPS ne reflète pas une simple détection de marques d'allongement, mais présente plutôt un cumul d'information suivi d'une décharge de potentiels (Holzgrefe, et coll., 2011). Nous sommes donc en présence d'une composante dont le comportement suggère une accumulation d'information en mémoire tampon sur l'étendue d'un groupe marqué par un allongement (soit un changement de l'intervalle interréponse). On peut aussi voir que ce processus de *chunking* rejoint des études sur les effets d'allongement sur l'accès lexical (Christophe, et coll., 2004) et sur la résolution d'ambiguïtés syntaxiques (Pauker, et coll., 2011). Bien que ces deux types d'études utilisent une autre terminologie pour désigner les chunks perceptifs, la principale variable considérée était l'allongement relatif des items. Aussi, dans ces deux cas la présence d'un allongement contraignait le traitement (lexical ou syntaxique) du signal entendu, ce qui peut s'interpréter comme reflétant un *chunking* perceptif. Ainsi, en déplaçant ou en ajoutant un allongement, ces études modifiaient le *chunking*, ce qui affectait la façon dont l'information était stockée en mémoire immédiate et son interprétation.

Objectif 2 : Démontrer que le chunking de la parole reflète bien un chunking « d'entrée » et non « de sortie ».

Les données empiriques

Pour répondre à ce deuxième objectif, nous avons conçu à l'Expérience 2 (article 1) des séquences de syllabes sans sens présentant les mêmes conditions prosodiques que les énoncés de l'Expérience 1 (article 1). Notons que cet objectif répond au problème des effets possibles d'un *chunking* « de sortie ». Il faut considérer que les résultats de l'Expérience 1 ne permettaient pas d'exclure la possibilité que les CPS obtenues soient dues à un *chunking* basé sur la reconnaissance de formes lexicales ou syntaxiques contenues dans la mémoire à

long terme. Nos comparaisons des réponses aux deux types de contextes (énoncés sensés et syllabes sans sens) ont démontré que la CPS est associée à un *chunking perceptif* en termes de GT indépendamment de la présence de formes verbales dans les stimuli. Plus spécifiquement, tout comme les réponses évoquées avec les énoncés, les réponses obtenues avec les séries de syllabes contiennent trois CPS distinctes dont les latences correspondent étroitement aux allongements des GT. D'ailleurs, en examinant les figures 4 (énoncés sensés) et 9 (syllabes sans sens, expérience 1, article 1), on constate la ressemblance entre les deux séries de PÉ. Nos analyses statistiques ont d'ailleurs confirmé que seule la longueur des GT a eu un impact significatif sur l'amplitude moyenne des potentiels et sur la latence de la CPS. Dans l'ensemble, ces comparaisons démontrent que les GT dans les deux types de stimuli ont été traités de manière similaire. En résumé, ces résultats appuient notre hypothèse voulant que les réponses CPS reflètent un *chunking* perceptif basé sur les GT et non pas un *chunking* basé sur une reconnaissance de formes sémantiques ou syntaxiques.

Ces résultats concordent avec ceux de (Pannekamp, et coll., 2005) qui avaient obtenu des CPS similaires avec des énoncés sensés et des énoncés dépourvus de contenu linguistique. Toutefois, Pannekamp et coll. avaient utilisé des constituants phonologiques pour créer leurs stimuli (syntagmes intonatifs) et avaient donc manipulé simultanément la F_0 et les allongements, en plus d'insérer une pause entre les groupes. Ainsi, bien que Pannekamp et coll. aient démontré que la CPS s'associe à la prosodie, leur méthode ne permettait pas d'élucider *laquelle* de ces variables acoustiques causait la CPS. Par contre, dans l'élaboration de nos stimuli, nous avons contrôlé différentes variables de sorte que les résultats permettent de clarifier le fait que la CPS s'associe spécifiquement aux GT plutôt qu'aux structures syntaxiques ou intonatives. Les stimuli que nous avons conçus pour la présente thèse possèdent l'avantage de contrôler les différentes variables acoustiques séparément. Ceci nous permet d'affirmer que les APF marquant les groupes temporels sont suffisants pour déclencher le *chunking* perceptif et évoquer une CPS.

Implications

La pertinence d'étudier le *chunking* perceptif de séquences de syllabes sans sens peut se concevoir sur deux plans. Premièrement, il faut voir que la parole est un signal éphémère où des séquences de sons changent rapidement. Toute interprétation de ce signal éphémère implique qu'il y ait *d'abord* une mise en mémoire temporaire d'une certaine étendue de ce signal, soit un « chunk ». On n'interprète pas la parole, après tout, un son à la fois, mais plutôt sur une « unité perceptive » (Miller, 1962). Ainsi, avant qu'on puisse interpréter la parole en fonction d'unités verbales qui ont été stockées en mémoire à long terme, il faut qu'il y ait eu un *chunking* perceptif du signal. Nos comparaisons permettent de voir que ce *chunking* perceptif ou « *chunking* d'entrée » du signal en GT s'applique aussi à la parole sensée et pas seulement à des séquences de syllabes ou de chiffres

L'autre utilité des présentes données porte sur le fait que l'apprentissage d'une langue repose sur l'acquisition de séquences de syllabes constituant de nouvelles expressions. Or il est connu que la mémoire des séquences est limitée et qu'un processus général « prélinguistique » de *chunking* perceptif s'applique nécessairement à l'apprentissage de formes verbales. Dans cette perspective, il est important de signaler que le travail ci-dessus permet de répondre à un besoin d'explication du *chunking*, comme l'entend Bybee (2010; voir aussi Lindblom, MacNeilage, & Studdert-Kennedy, 1984). En effet, de nombreuses recherches ont démontré que les enfants acquièrent le langage par chunks, sans nécessairement être conscients de leur structure interne (Peters, 1977, 1983). Ce n'est qu'après avoir mémorisé un certain nombre de chunks qu'ils peuvent les comparer et en extraire des unités indépendantes (Bybee, 2010; Dabrowska & Lieven, 2005; Lieven, et coll., 1997; Locke, 1993, 1995, 1997; Pine & Lieven, 1993; Wray & Perkins, 2000). Dans l'ensemble, cependant, aucune étude n'avait précisé comment un processus général de *chunking* pouvait s'appliquer à la parole. Les données ci-dessus clarifient le fait que le *chunking* perceptif s'applique autant au traitement d'énoncés qu'au traitement de séquences sans sens. On peut voir, en somme, qu'un processus général de *chunking* sensori-moteur

s'applique à l'apprentissage et la production de séquences à travers de nombreux comportements (Gobet, et coll., 2001; Graybiel, 1998; Terrace, 2001).

Enfin, par rapport aux théories d'apprentissage statistique, il est utile de noter que nos résultats démontrent que le *chunking* perceptif est lié aux GT et ne découle pas de régularités statistiques dans les stimuli. La conception que le *chunking* découle de répétitions ou de régularités stochastiques est souvent utilisée pour expliquer la segmentation des mots par l'enfant. Par exemple, Beckner et Bybee (2009) vont jusqu'à proposer que la répétition représente l'unique facteur de *chunking*¹⁰ (voir aussi Bybee, 2010). Cependant, nos résultats indiquent sans ambiguïté qu'il y a un *chunking* perceptif selon des indices de GT même lorsque les stimuli ne sont présentés qu'une seule fois. Autrement dit, nos données et l'observation de comportements habituels d'apprentissage de séquences indiquent que le *chunking* n'est pas de nature statistique, mais reflèterait plutôt un processus de segmentation sensori-moteur (Graybiel, 1998).

Objectif 3 : Isoler les effets du chunking perceptif sur la MCT

Les données empiriques

Notre troisième hypothèse portait sur les effets du *chunking* perceptif de la parole sur la mémoire immédiate. Le *chunking* est associé depuis longtemps à la mémoire de travail (MdeT) et il est accepté que le groupement ou *chunking* d'éléments facilite le rappel à différents degrés selon la taille des groupes (voir Cowan, 2000). Dans cette perspective, nous avons prédit que si le *chunking* perceptif de la parole en GT est lié à la MdeT, alors la perception de GT courts (trois mots monosyllabiques ou items) permettra une meilleure mise en mémoire des items entendus que la perception de GT longs (quatre items). Pour tester cette hypothèse, nous avons utilisé une tâche de Sternberg et évalué la mise en mémoire d'éléments en termes de la reconnaissance de cibles, du temps réponse, de même que par l'amplitude et la latence de certains PÉ spécifiques (P300 et N400).

¹⁰ « Repetition is the factor that leads to chunking... » (Beckner & Bybee, 2009, p.30).

Les résultats présentés dans l'article 2 appuient notre hypothèse, sans toutefois confirmer toutes nos prédictions. Ainsi, nous avons observé une modulation de l'amplitude de la N400 en fonction de la longueur des GT : les cibles provenant de GT de 4 syllabes ont suscité des N400 plus amples que celles provenant de GT de 3 syllabes, peu importe leur position dans l'énoncé. La N400 est reconnue comme étant associée à l'effort fourni pour activer (ou réactiver) un item en mémoire. En ce sens, des cibles issues de GT de 4 syllabes ont demandé un plus grand effort de réactivation que des cibles provenant de GT de 3 syllabes. Ce résultat confirme l'idée que le *chunking* perceptif en GT est lié à la MdeT et indique que la taille des GT affecte bel et bien la qualité de la trace mnésique. Cet effet observable au plan des PÉ ne s'est toutefois pas traduit par des différences comportementales (ni sur l'exactitude de la reconnaissance des cibles, ni sur les temps de réaction qui n'ont été affectés que par la position du GT dans l'énoncé) ce qui, en toute vraisemblance, s'explique par la facilité de la tâche. En fait, la reconnaissance d'un mot que l'on vient tout juste d'entendre dans un énoncé ne présente généralement aucune difficulté pour un locuteur natif. Toutefois, il est intéressant de noter que ce type de tâche a tout de même révélé des différences significatives causées par la position des GT, autant sur les temps de réaction que les réponses P300.

En fait, les analyses des fluctuations d'amplitude de la N400 ont révélé que la position du GT dans l'énoncé n'avait pas d'effet significatif sur la qualité de la trace mnésique. Autrement dit, un élément entendu dans le premier GT n'était pas significativement plus facile à réactiver qu'un élément entendu dans le deuxième GT. Par contre, la position des GT a eu un impact significatif sur les temps de réaction : les cibles provenant de groupes en 1^{re} position sont reconnues plus rapidement que celles provenant de groupes en 2^e position (ce qui tend à infirmer un effet de récence généralement présent dans les tâches de rappel). Cet effet de position ne s'accorde pas avec le modèle de Sternberg, qui postule que le rappel procède par un balayage complet du contenu de la MdeT avant de produire une réponse (Sternberg, 1966). Notons que les énoncés utilisés dans notre tâche contenaient tous le même nombre de lexèmes et de syllabes. Selon la théorie du balayage complet de Sternberg, les temps de réponse devraient donc être

comparables d'une condition à l'autre, ce qui n'est pas le cas. Nous observons plutôt un effet marqué de la position des GT, ce qui tend à démontrer que le balayage a été interrompu une fois la cible localisée dans un groupe perceptif.

Dans ce même ordre d'idée, on observe aussi un effet de position des GT sur la P300, une composante associée au balayage de la MdeT (voir Pelosi, et coll., 1995). Plus particulièrement, les cibles dans les GT en 1^{re} position ont généré des P300 plus précoces que les cibles dans les GT en 2^e position. Cet effet de position sur la P300 concorde avec des temps de réaction plus courts pour les GT en 1^{re} position. En somme, dans une tâche classique de Sternberg où on présente des séries verbales sans groupement, la latence de la P300 est généralement associée au temps nécessaire pour balayer le contenu de la MdeT en entier. Toutefois, dans une tâche où l'on présente des énoncés avec des groupes perceptifs, il y a un stockage et un balayage ordonné de la MdeT « chunk par chunk », ce qui expliquerait des temps de réponse variables selon l'ordre des chunks.

Implications

De façon générale, les données de l'article 2 confirment une idée reçue voulant que le *chunking* perceptif opère en fonction des limites de la mémoire à court terme ou MdeT qui est de l'ordre de trois ou quatre items (voir p. ex. Broadbent & Broadbent, 1973; Chen & Cowan, 2005; Cowan, 2000; Frankish, 1989, 1995; Hitch, et coll., 1996; Mayberry, et coll., 2002; Reeves, et coll., 2000; Terrace, 2001; Wickelgren, 1964). Toutefois, l'article 1 constitue la première démonstration que ce principe de *chunking* s'applique non seulement à des séries de chiffres ou de syllabes, mais aussi à la perception d'énoncés. Cette démonstration du lien entre les processus généraux de *chunking* et la perception de la parole permet d'expliquer nombre d'observations en apparence indépendantes les unes des autres.

Tout d'abord, étant donné les limites de la MdeT, l'apprentissage de séquences verbales ou non verbales implique un *chunking* perceptif qui soit conforme aux limites de la mémoire. Comme l'a souligné Terrace (2001), la production porte nécessairement les marques du *chunking* perceptif et, en fait, on retrouve dans tout comportement impliquant l'apprentissage de séquences des GT de trois ou quatre éléments marqués par des délais ou

des allongements. Ainsi, sur le plan de la parole, des études rapportent la présence de GT dans de nombreuses langues (aussi appelés groupes « rythmiques » ou « accentuels ») ne dépassant généralement pas quatre syllabes (Boucher, 2006; Dauer, 1983; Fletcher, 1991; Lea, 1974; Martin, 1999; Nootboom, 1997; Padeloup, 1992; Wenk & Wioland, 1982; Wioland, 1984). Rappelons aussi que le *chunking* perceptif et le stockage en mémoire immédiate constituent les premières étapes du stockage de formes verbales dans la mémoire à long terme (MLT). Les contraintes sur la mémoire sérielle ou immédiate structurent donc les formes verbales pouvant être conservées en MLT. Les résultats de Rousset (2004) viennent illustrer cette contrainte en ce qu'ils révèlent que la longueur des lexèmes libres ne dépasse pas 4 syllabes en moyenne (selon une recension portant sur 32 langues). Il va sans dire qu'un locuteur est en mesure d'apprendre des unités linguistiques de plus de 4 syllabes, mais la basse fréquence de telles formes pourrait être expliquée par l'effort mnésique supplémentaire qu'elles exigent. Les résultats de Lalonde (2011) abondent en ce sens et montrent que les lexèmes ou expressions lexicalisées de 3 syllabes et plus ne représentent généralement pas plus de 35 % des lexèmes utilisés en parole spontanée par des adultes locuteurs natifs du français¹¹. Ainsi, 65 % des lexèmes utilisés dans un discours spontané normal d'adulte ne dépassent pas deux syllabes (Lalonde, 2011).

¹¹ Il faut d'ailleurs attendre l'adolescence avant que cette proportion ne dépasse 20 % (Lalonde, 2011, p.33).

5. Conclusion

La recherche ci-dessus visait à démontrer comment un processus général de *chunking* perceptif s'applique à la parole et influence la mémoire. Il faut noter que ce processus général, qui est aussi appelé *chunking* d'entrée (Terrace, 2001) est un principe qui s'applique à une multitude de comportements animaux et humains impliquant l'apprentissage et la production de séquences. En fait, dans tout comportement impliquant l'apprentissage de nouvelles séries, des groupes temporels émergent et sont marqués par des délais et des allongements en fin de groupe. Ces marques de groupes sont universelles et apparaissent aussi en parole (soit l'« allongement précédent la finale » ou APF). La contribution centrale du présent travail est d'avoir démontré pour la première fois, à notre connaissance, que la perception de la parole opère spécifiquement par rapport à des marques universelles de groupes temporels. Il s'agit des mêmes marques de *chunking* qui apparaissent dans l'apprentissage ou la production de séquences non verbales tant chez les humains que chez certains animaux. En somme, notre travail démontre que le sujet perçoit la parole « *par chunk* » suivant un processus perceptif général qui est conforme aux contraintes de la MCT et du foyer attentionnel (Cowan, 2000). Rappelons que la parole ne peut être interprétée un son à la fois, mais exige la mise en mémoire d'un nombre d'éléments séquentiels sur une étendue du signal qui, nécessairement, sera contraint par les limites de la MdeT. Le *chunking* perceptif représente justement cette fonction fondamentale qui permet la saisie d'une séquence.

Notre travail présente aussi une démonstration particulière des effets du *chunking* perceptif. Il est bien connu que le *chunking* et la présentation d'éléments par groupes facilite la mise en mémoire. Plusieurs études, maintenant classiques, ont démontré l'effet bénéfique du groupement sur le rappel sériel. Cependant, dans tous les cas, il s'agissait d'études faisant usage de tâches de *rappel* uniquement, soit des tâches où l'individu devait attendre la fin d'un stimulus avant de répondre. Ce genre de tâche ne permet aucunement l'étude d'un *chunking* perceptif, qui exige une technique permettant d'observer les réponses d'un individu *lors de la réception du stimulus*. En utilisant la technique des PÉ, nous avons démontré l'application d'un *chunking* perceptif basé sur la détection des groupes temporels.

Nous avons par la suite pu observer l'effet de la longueur de ces groupes sur la mémoire immédiate des éléments entendus. Il s'agit là aussi d'une première puisque, à notre connaissance, aucune étude n'a démontré un *chunking* perceptif de la parole par le biais de PÉ.

Pris dans leur ensemble, les résultats de la présente étude contribuent à une meilleure compréhension des processus impliqués dans la segmentation et la mise en mémoire de la parole basée sur l'application d'un principe général de *chunking* sensori-moteur applicable à un ensemble de comportements (verbaux et non verbaux). En présentant une démonstration de l'application du *chunking* perceptif à la parole, nous avons aussi répondu à un besoin *d'explication du chunking de formes verbales* comme l'entend Bybee (2010; voir aussi Lindblom, et coll., 1984). Il y a aussi, dans notre travail, des éléments de réponses à une variété de questions fondamentales que se pose tout étudiant du langage oral. Par exemple, pourquoi l'allongement en fin de groupe est-il universel en parole ? Pourquoi le groupement apparaît-il généralement lorsqu'on apprend des séquences, verbales ou non verbales ? Pourquoi apprend-on de nouvelles expressions *en chunk* qui ne dépassent pas environ quatre syllabes ? De telles questions ne sont pas souvent formulées en linguistique où on tient souvent pour acquises certaines marques et unités d'analyse. Toutefois, il s'agit de problèmes fondamentaux qui sont centraux à toute explication du langage humain.

Bibliographie

- Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience Research*, *64*, 185-190.
- Adam, N., & Collins, G. I. (1978). Late components of visual evoked potential to search in short term memory. *Electroencephalography and clinical Neurophysiology*, *44*, 147-156.
- Adams, H. F. (1915). A note on the effect of rhythm on memory. *Psychological Review*, *22*(4), 289-298.
- Alcock, K. J., Wade, D., Anslow, P., & Passingham, R. E. (2000). Pitch and Timing Abilities in Adult Left-Hemisphere-Dysphasic and Right-Hemisphere-Damaged Subjects. *Brain and Language*, *75*(1), 47-65.
- Altenberg, B. (1990). Speech as linear composition. Dans G. Caie, K. Haastrup, A. L. Jakobsen, J. E. Nielsen, J. Sevaldsen, H. Specht & A. Zettersten (Éds.), *Proceedings from the Fourth Nordic Conference for English Studies* (Vol. 1, pp. 133-143). Copenhagen: Department of English, University of Copenhagen.
- Anderson, J. R., & Matessa, M. P. (1997). A production system theory of serial memory. *Psychological Review*, *104*, 728-748.
- Aslin, R. N. (1993). Segmentation of fluent speech into words: Learning models and the role of maternal input. Dans B. d. Boysson-Bardies, S. d. Schonen, P. Jusczyk, P. McNeilage & J. Morton (Éds.), *Developmental neurocognition: Speech and face processing in the first year of life* (Vol. 69, pp. 305-315). Boston: Kluwer Academic.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*, 321-324.
- Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, *18*, 172-184.
- Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136-R140.
- Barrett, S. E., & Rugg, M. D. (1989). Event-related potentials and the semantic matching of faces. *Neuropsychologia*, *27*, 913-922.

- Batchelder, E. O. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition*, 83, 167-206.
- Bates, E., & Goodman, J. C. (2001). On the inseparability of grammar and the lexicon: Evidence from acquisition. Dans M. Tomasello & E. Bates (Éds.), *Language development: The essential readings* (pp. 134-162). Malden, MA: Blackwell.
- Baum, S. R., & Pell, M. D. (1999). The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology*, 13(8), 581-608.
- Beckman, M. E., & Ayers Elam, G. (1997). Guidelines for ToBI Labelling: The Ohio State University Research Foundation.
- Beckman, M. E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. Dans J. Kingston & M. Beckman (Éds.), *Papers in Laboratory Phonology 1: Between the Grammar and the Physics of Speech* (pp. 152-178). Cambridge: Cambridge University Press.
- Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255-309.
- Beckner, C., & Bybee, J. (2009). A Usage-Based Account of Constituency and Reanalysis. *Language Learning*, 59, 27-46.
- Bentin, S., & McCarthy, G. (1994). The effects of immediate stimulus repetition on reaction time and event-related potentials in tasks of different complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 130-149.
- Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2011). Acoustic Markers of Prominence Influence Infants' and Adults' Segmentation of Speech Sequences. *Language and Speech*, 54, 123-140.
- Bögels, S., Schriefers, H., Vonk, W., & Chwilla, D. J. (2011a). Prosodic Breaks in Sentence Processing Investigated by Event-Related Potentials. *Language and Linguistics Compass*, 5/7, 424-440.
- Bögels, S., Schriefers, H., Vonk, W., & Chwilla, D. J. (2011b). The Role of Prosodic Breaks and Pitch Accents in Grouping Words during On-line Sentence Processing. *Journal of Cognitive Neuroscience*, 23(9), 2447-2467.

- Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2010). The Interplay between Prosody and Syntax in Sentence Processing: The Case of Subject- and Object-control Verbs. *Journal of Cognitive Neuroscience*, *22*(5), 1036-1053.
- Bolinger, D. (1958). A theory of pitch accent in English. *Word*, *14*, 109-149.
- Boucher, V. J. (2006). On the function of stress rhythms in speech: Evidence of a link with grouping effects on serial memory. *Language and Speech*, *49*, 495-519.
- Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93-125.
- Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33-B44.
- Broadbent, D. E. (1975). The magic number seven after fifteen years. Dans A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 3-18). New York: Wiley.
- Broadbent, D. E., & Broadbent, M. H. P. (1973). Grouping strategies in short-term memory for alpha-numeric lists. *Bulletin of the British Psychological Society*, *26*, 135.
- Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin and Review*, *18*, 1189-1196.
- Buschke, H. (1976). Learning is organized by chunking. *Journal of Verbal Learning and Verbal Behavior*, *15*, 313-324.
- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*, 711-733.
- Bybee, J. (2010). *Language, Usage and Cognition: Cambridge University Press Textbooks*.
- Carlson, K., Frazier, L., & Clifton Jr., C. (2009). How prosody constrains comprehension: A limited effect of prosodic packaging. *Lingua*, *119*, 1066-1082.
- Chaby, L., Jemel, B., George, N., Renault, B., & Fiori, N. (2001). An ERP study of famous face incongruity detection in middle age. *Brain and Cognition*, *45*, 357-377.
- Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1235-1249.

- Chow, I. (2008). Quantitative analysis of preboundary lengthening in Cantonese. Dans P. A. Barbosa, S. Madureira & C. Reis (Éds.), *Proceedings of the Speech Prosody 2008 Conference* (pp. 543-546). Campinas, Brazil: Editoria RG/CNPq.
- Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*, 31, 585-598.
- Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of Prosodic Boundary Correlates by Newborn Infants. *Infancy*, 2(3), 385-394.
- Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech*, 51, 61-75.
- Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language*, 51, 523-547.
- Cooper, W. E., & Sorensen, J. M. (1981). *Fundamental frequency in sentence production*. New-York: Springer-Verlag.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87-185.
- Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.
- Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological Science*, 15, 634-640.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum.
- Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218-236.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech*, 40(2), 141-201.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25(4), 385-400.

- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, *24*, 381-410.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113-121.
- Dabrowska, E. (2004). *Language, mind and brain: Some psychological and neurological constraints on theories of grammar*. Edinburgh: Edinburgh University Press.
- Dabrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, *16*, 437-474.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51-62.
- Delattre, P. (1966a). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics*, *4*, 183-198.
- Delattre, P. (1966b). Les dix intonations de base du français. *The French Review*, *40*(1), 1-14.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9-21.
- DePaolis, R. A., Vihman, M. M., & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics*, *36*, 406-422.
- Desrochers, A. (2006). OMNILEX : Une base de données sur le lexique du français contemporain. *Cahiers Linguistiques d'Ottawa*, *34*, 25-34.
- Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63*(3), 274-294.
- Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology*, *14*, 1-52.

- Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology, 61*(2), 177-199.
- Endress, A. D., Nespors, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences, 13*(8), 348-353.
- Fant, G., Kruckensberg, A., & Nord, L. (1991). Durational correlates of stress in Swedish, French and English. *Journal of Phonetics, 19*, 351-365.
- Farrell, S. (2008). Multiple roles for time in short-term memory: Evidence from serial recall of order and timing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 128-145.
- Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception & Psychophysics, 69*(6), 856-864.
- Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP "old/new" effects: Memory strength and decisional factor(s). *Neuropsychologia, 40*, 2288-2304.
- Fletcher, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics, 19*, 193-212.
- Fon, Y.-J. J. (2002). *A cross-linguistic study on syntactic and discourse boundary cues in spontaneous speech*. Ph.D. dissertation, Ohio State University.
- Ford, J. M., Pfefferbaum, A., Tinklenberg, J. R., & Kopell, B. S. (1982). Effects of perceptual and cognitive difficulty on P3 and RT in young and old adults. *Electroencephalography and clinical Neurophysiology, 54*, 311-321.
- Ford, J. M., Roth, W. T., Mobs, R. C., Hopkins, W. F., & Kopell, B. S. (1979). Event-related potentials recorded from young and old adults during a memory retrieval task. *Electroencephalography and clinical Neurophysiology, 47*, 450-459.
- Fox, A. (2000). *Prosodic features and prosodic structure: the phonology of suprasegmentals*. Oxford: University Press.
- Frankish, C. (1989). Perceptual organization and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(3), 469-479.

- Frankish, C. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology*, 9(special issue), S5-S22.
- Frazier, L., Carlson, K., & Clifton Jr, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10(6), 244-249.
- Frick, R. W. (1989). Explanation of grouping in immediate ordered recall. *Memory & Cognition*, 17(5), 551-562.
- Fry, D. B. (1955). Duration and Intensity as Physical Correlates of Linguistic Stress. *The Journal of the Acoustical Society of America*, 27(4), 765-768.
- Gandour, J. (1998a). Aphasia in tone languages. Dans P. Coppens, Y. Lebrun & A. Basso (Éds.), *Aphasia in Atypical Populations* (pp. 117-141). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gandour, J. (1998b). Phonetics and phonology. Dans B. Stemmer & H. A. Whitaker (Éds.), *Handbook of Neurolinguistics* (pp. 207-219). San Diego, CA: Academic Press.
- Gandour, J., Petty, S. H., & Dardarananda, R. (1988). Perception and production of tone in aphasia. *Brain and Language*, 35(2), 201-240.
- Geers, A. E. (1978). Intonation contour and syntactic structure as predictors of apparent segmentation. *Journal of experimental Psychology: Human Perception and Performance*, 4(2), 273-283.
- Gilbert, A. C., Boucher, V. J., & Jemel, B. (2010). Exploring the rhythmic segmentation of heard speech using evoked potentials *Proceedings of the 5th Conference on Speech Prosody* (Vol. 100 334, pp. 1-3). Chicago, USA.
- Gilbert, A. C., Boucher, V. J., & Jemel, B. (soumis). The perceptual chunking of speech: a demonstration using ERPs.
- Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., et coll. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), 236-243.
- Gomer, F. E., Spicuzza, R. J., & O'Donnell, R. D. (1976). Evoked potential correlates of visual item recognition during memory scanning tasks. *Physiological Psychology*, 4, 61-65.

- Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and memory*, 70, 119-136.
- Gussenhoven, C., & Rietveld, A. C. M. (1992). Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics*, 20, 283-303.
- Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., et coll. (2002). N400-like Magnetoencephalography Responses Modulated by Semantic Context, Word Frequency, and Lexical Class in Sentences. *NeuroImage*, 17(3), 1101-1116.
- Handy, T. C. (Éd.). (2005). *Event-related potentials: A methods handbook*. Cambridge: MIT Press.
- Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *The Quarterly Journal of Experimental Psychology*, 49A(1), 116-139.
- Hockey, B. A., & Fagyal, Z. (1998). Pre-Boundary Lengthening: Universal or Language-Specific? The Case of Hungarian. *Working Papers in Linguistics (University of Pennsylvania)*, 5(1), 71-82.
- Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, 30, 47-61.
- Holzgrefe, J., Schröder, C., Höhle, B., & Wartenburger, I. (2011). *Neurophysiological investigations on the processing of prosodic boundary cues*. Paper presented at the Experimental and Theoretical Advances in Prosody 2, Montréal, Canada.
- Houlihan, M., Stelmack, R., & Campbell, K. (1998). Intelligence and the effects of perceptual processing demands, task difficulty and processing speed on P300, reaction time and movement time. *Intelligence*, 26(1), 9-25.
- Houston, D., Santelmann, L., & Jusczyk, P. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes*, 19(1), 97-136.

- Hruska, C., & Alter, K. (2004). How prosody can influence sentence perception. Dans A. Steube (Éd.), *Information Structure: Theoretical and Empirical Aspects* (Vol. 1, pp. 211–226). Berlin: Gruyter.
- Itzhak, I., Pauker, E., Drury, J. E., Baum, S. R., & Steinhauer, K. (2010). Event-related potentials show online influence of lexical biases on prosodic processing. *NeuroReport*, *21*, 8-13.
- Jackendoff, R. (2007). Linguistics in cognitive science: the state of the art *The Linguistic Review*, *24*, 347-401.
- Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language*, *44*(4), 548-567.
- Johnson, E. K., Jusczyk, P. W., Cutler, A., & Norris, D. (2003). Lexical viability constraints on speech segmentation by infants. *Cognitive Psychology*, *46*, 65-97.
- Jun, S.-A., & Fougeron, C. (1997). *A Phonological model of French intonation*. Paper presented at the The ETRW ESCA Workshop on Intonation, Athènes.
- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' Preference for the Predominant Stress Patterns of English Words. *Child Development*, *64*(3), 675-687.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology*, *39*(3–4), 159-207.
- Karrer, R., McDonough, B., Warren, C., & Cone, R. (1980). CNV during memory retrieval by normal and retarded adults. Dans H. H. Kornhuber & L. Deecke (Éds.), *Motivation, motor and sensory processes of the brain: electrical potentials, behavior and clinical use, Progress in brain research* (Vol. 54, pp. 668-672). Amsterdam: Elsevier.
- Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D. J. (2008). Sentence processing in the visual and auditory modality: Do comma and prosodic break have parallel functions? *Brain Research*, *1224*, 102-118.

- Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D. J. (2009). Discourse, Syntax, and Prosody: The Brain Reveals an Immediate Interaction. *Journal of Cognitive Neuroscience, 19*(9), 1421-1434.
- Klatt, D. H. (1975). Vowel Lengthening is Syntactically Determined in a Connected Discourse. *Journal of Phonetics, 3*(3), 129-140.
- Knösche, T. R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O. W., et coll. (2005). Perception of phrase structure in music. *Human Brain Mapping, 24*, 259-273.
- Knowlton, B. J., & Squire, L. R. (1994). Artificial Grammar Learning and Implicit Memory: Reply to Higham and Vokey. *Psychological Science, 5*, 61.
- Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 169-181.
- Koch, I., & Hoffmann, J. (2000). Patterns, chunks, and hierarchies in serial reaction-time tasks. *Psychological Research, 63*, 22-35.
- Kramer, A., Schneider, W., Fisk, A., & Donchin, E. (1986). The effects of practice and task structure on the components of the event-related potential. *Psychophysiology, 23*, 33-47.
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12*(2), 72-79.
- Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes, 8*(4), 533-572.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*, 203-205.
- Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition, 11*(5), 539-550.

- Kutas, M., & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and clinical Neurophysiology*, 108, 456-471.
- Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: P300 as a measure of stimulus evaluation time. *Science*, 197, 792-795.
- Ladd, D. R. (1983). Peak features and overall slope. Dans A. Cutler & D. R. Ladd (Éds.), *Prosody: Models and measurements* (pp. 39-52). New-York: Springer-Verlag.
- Ladd, D. R. (1986). Intonational phrasing: The Case for recursive prosodic structure. *Phonology Yearbook*, 3, 311-340.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Ladd, D. R., & Campbell, N. (1991). Theories of prosodic structure: Evidence from syllable duration. *Proceedings of the 12th International Congress of Phonetic Science*, 2, 290-293.
- Lalonde, B. (2011). *Effets de la croissance des capacités respiratoires sur la longueur des énoncés et sur la diversité lexicale*. Mémoire de Maîtrise, Université de Montréal, Montréal.
- Lashley, K. S. (1951). The problem of serial order in behavior. Dans L. A. Jeffries (Éd.), *Cerebral Mechanisms in Behavior* (pp. 112-136). New York: John Wiley & Sons.
- Lea, W. A. (1974). Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding. *Univac Repport No. PX10791*. St.Paul, MN: Sperry Univac.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge MA: The M.I.T. Press.
- Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7, 107-122.
- Lehiste, I., Olive, J. P., & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *The Journal of the Acoustical Society of America*, 60(5), 1199-1202.
- Leonard, L. B. (1974). The role of intonation in the recall of various linguistic stimuli. *Language and Speech*, 16, 327-335.
- Li, C. N., & Thompson, S. A. (1978). The acquisition of tone. Dans V. A. Fromkin (Éd.), *Tone: A linguistic survey* (pp. 271-284). New York: Academic Press.

- Liberman, M. (1975). *The intonational system of English*. Unpublished PdD dissertation. MIT. Cambridge. <http://dspace.mit.edu/handle/1721.1/27376>.
- Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249-336.
- Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, 30, 333-370.
- Lieven, E., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187-220.
- Lindblom, B. (1968). Temporal organization of syllable production. *Quarterly Progress and Status Report*, 9(2-3), 1-5.
- Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of phonological universals. Dans B. Butterworth, B. Comrie & Ö. Dahl (Éds.), *Explanations for language universals* (pp. 181-203). New York, NY: Mouton.
- Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish *Papers from the Institute of Linguistics (PILUS)* (Vol. 21). Stockholm: University of Stockholm.
- Lindstromberg, S., & Boers, F. (2008). Phonetic repetition and the learning of lexical chunks: The power of assonance. *System*, 36, 423-436.
- Locke, J. L. (1993). *The Child's Path to Spoken Language*. Cambridge, MA: Harvard University Press.
- Locke, J. L. (1995). Development of the capacity for spoken language. Dans P. Fletcher & B. MacWhinney (Éds.), *The Handbook of Child Language* (pp. 278-302). Oxford: Blackwell.
- Locke, J. L. (1997). A theory of neurolinguistic development. *Brain and Language*, 58, 265-326.
- Loeb, D. F., & Allen, G. D. (1993). Preschoolers' Imitation of Intonation Contours. *Journal of Speech and Hearing Research*, 36(1), 4-13.

- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge: MIT Press.
- Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology*, *21*, 171–186.
- Makarova, V. (2001). Perceptual correlates of sentence-type intonation in Russian and Japanese. *Journal of Phonetics*, *29*(2), 137-154.
- Männel, C., & Friederici, A. D. (2009). Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. *Journal of Cognitive Neuroscience*, *21*(10), 1988-2006.
- Marcus, S. M. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics*, *30*, 247-256.
- Marsh, G. R. (1975). Age differences in evoked potentials correlates of a memory scanning process. *Experimental Aging Research*, *1*, 3-16.
- Martin, P. (1999). Intonation of spontaneous speech in French. *Proceedings of the 14th International Congress of Phonetic Sciences*, 17-20.
- Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, *122*(3), 346-362.
- Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*, 465-494.
- Mattys, S. L., & Samuel, A. G. (1997). How Lexical Stress Affects Speech Segmentation and Interactivity: Evidence from the Migration Paradigm. *Journal of Memory and Language*, *36*, 87-116.
- Mattys, S. L., & Samuel, A. G. (2000). Implications of stress-pattern differences in spoken-word recognition. *Journal of Memory and Language*, *42*, 571-596.
- Mayberry, R. I., Lock, E., & Kazmi, H. (2002). Linguistic ability and early language exposure. *Nature*, *417*, 38.

- McCarthy, G., & Donchin, E. (1983). Chronometric analysis of human information processing. Dans A. W. K. Gaillard & W. Ritter (Éds.), *Tutorials in ERP Research: Endogenous Components* (pp. 251-268). Amsterdam: Holland Publications.
- McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic Cues and Possible-Word Constraints in Japanese Speech Segmentation. *Journal of Memory and Language*, 45(1), 103-132.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-97.
- Miller, G. A. (1962). Decision units in the perception of speech. *IRE Transactions on Information Theory*, 8(2), 81-83.
- Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 2, 217-228.
- Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, 108, 271-280.
- Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, 96(2), 143-182.
- Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, 66, 911-936.
- Nan, Y., Knösche, T. R., & Friederici, A. D. (2006). The perception of musical phrase structure: A cross-cultural ERP study. *Brain Research*, 1094(1), 179-191.
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 756-766.
- Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, 54(3), 283-299.
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

- Neuhaus, C., Knösche, T. R., & Friederici, A. D. (2006). Effects of musical expertise and boundary markers on phrase perception in music. *Journal of Cognitive Neuroscience, 18*, 1-22.
- Ng, H. L. H., & Maybery, M. T. (2002). Grouping in short-term verbal memory: Is position coded temporally? *The Quarterly Journal of Experimental Psychology Section A, 55*(2), 391-424.
- Nooteboom, S. G. (1997). The prosody of speech: Melody and rhythm. Dans W. J. Hardcastle & J. Laver (Éds.), *Handbook of Phonetic Sciences* (pp. 640-673). Oxford: Blackwell Publishers.
- Nooteboom, S. G., Brokx, J. P. L., & Rooij, J. J. d. (1978). Contributions of prosody to speech perception. Dans W. J. M. Levelt & G. B. Flores d'Arcais (Éds.), *Studies in the Perception of Language* (pp. 75-107). New York: Wiley.
- Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The Possible-Word Constraint in the Segmentation of Continuous Speech. *Cognitive Psychology, 34*, 191-243.
- Norris, D., McQueen, J. M., Cutler, A., Butterfield, S., & Kearns, R. (2001). Language-universal constraints on speech segmentation. *Language and Cognitive Processes, 16*(5/6), 637-660.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*, 97-113.
- Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America, 54*(5), 1235-1247.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or Syllable? Speech Segmentation in Japanese. *Journal of Memory and Language, 32*(2), 258-278.
- Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience, 17*, 407-421.
- Parisse, C. (2005). New perspectives on language development and the innateness of grammatical knowledge. *Language Sciences, 27*, 383-401.

- Parmentier, F. B. R., & Maybery, M. T. (2008). Equivalent effects of grouping by time, voice, and location on response timing in verbal serial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1349-1355.
- Pasdeloup, V. (1992). Durée intersyllabique dans le groupe accentuel en français. *Actes des 19èmes journées d'étude sur la parole* (pp. 19-22). Bruxelles: S.F.A.
- Pauker, E., Itzhak, I., Baum, S. R., & Steinhauer, K. (2011). Effects of Cooperating and Conflicting Prosody in Spoken English Garden Path Sentences: ERP Evidence for the Boundary Deletion Hypothesis. *Journal of Cognitive Neuroscience*, 23(10), 2731-2751.
- Pelosi, L., Hayward, M., & Blumhardt, L. D. (1995). Is “memory-scanning” time in the Sternberg paradigm reflected in the latency of event-related potentials? *Electroencephalography and Clinical Neurophysiology*, 96, 44-55.
- Pelosi, L., Holly, M., Slade, T., Hayward, M., Barrett, G., & Blumhardt, L. D. (1992). Wave form variations in auditory event-related potentials evoked by a memory-scanning task and their relationship with tests of intellectual function. *Electroencephalography and clinical Neurophysiology*, 84, 344-352.
- Pena, M., Bonatti, L. L., Nespors, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298, 604-607.
- Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, 119(3), 264-275.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, 10(5), 233-238.
- Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, 39, 246-263.
- Peters, A. M. (1977). Language learning strategies: does the whole equal the sum of the parts? *Language*, 53(3), 560-573.
- Peters, A. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.

- Pfefferbaum, A., Ford, J. M., Roth, W. T., & Kopell, B. S. (1980). Age differences in P3-reaction time associations. *Electroencephalography and clinical Neurophysiology*, *49*, 257-265.
- Picton, T. W. (1988). The endogenous evoked potentials. Dans E. Basar (Éd.), *Dynamics of sensory of sensory and cognitive processes in the brain* (pp. 258-265). New York: Springer-Verlag.
- Pierrehumbert, J. B. (2000). Tonal elements and their alignment *Prosody: Theory and experiment: Studies presented to Gosta Bruce* (pp. 11-36). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Pine, J. M., & Lieven, E. (1993). Reanalysing rote-learned phrases: individual differences in the transition to multiword speech. *Journal of Child Language*, *20*, 551-571.
- Pine, J. M., & Lieven, E. V. M. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, *18*, 123-138.
- Pine, J. M., & Martindale, H. (1996). Syntactic categories in the speech of young children: The case of the determiner. *Journal of Child Language*, *23*, 369-395.
- Pratt, H., Michalewski, H. J., Barrett, G., & Starr, A. (1989). Brain potentials in a memory-scanning task. I. Modality and task effects on potentials to the probes. *Electroencephalography and clinical Neurophysiology*, *72*, 407-421.
- Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language Discrimination by Human Newborns and by Cotton-Top Tamarin Monkeys. *Science*, *288*(5464), 349-351.
- Reeves, C., Schauder, A. R., & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1638-1654.
- Rooij, J. J. d. (1978). *Speech Punctuation. An Acoustic and Perceptual Study of some Aspects of Speech Prosody in Dutch*. Unpublished Doctor's thesis. University of Utrecht. Utrecht.

- Rousset, I. (2004). *Structures syllabiques et lexicales des langues du monde: Données, typologies, tendances universelles et contraintes substantielles*. PhD, Université Grenoble III, Grenoble.
- Ryan, J. (1969a). Grouping and short-term memory: Different means and patterns of groups. *Quarterly Journal of Experimental Psychology*, 21, 137-147.
- Ryan, J. (1969b). Temporal grouping, rehearsal and short-term memory. *Quarterly Journal of Experimental Psychology*, 21, 148-155.
- Saffran, J. R. (2001). Words in a sea of sounds: the output of infant statistical learning. *Cognition*, 81, 149-169.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606-621.
- Seddoh, S. A. K. (1999). Intonation and speech timing: Association or dissociation? *The Journal of the Acoustical Society of America*, 106(4), 2246-2246.
- Seddoh, S. A. K. (2004). Prosodic disturbance in aphasia: Speech timing versus intonation production. *Clinical Linguistics & Phonetics*, 18(1), 17-38.
- Seddoh, S. A. K., & Robin, D. A. (2001). Neurogenic disorders of prosody. Dans D. Vogel & M. Cannito (Éds.), *Treating Disordered Speech Motor Control: For Clinicians by Clinicians* (pp. 277-320). Austin, TX: Pro-Ed.
- Selkirk, E. O. (1982). *The Syntax of Words*. Cambridge, MA: MIT Press.
- Selkirk, E. O. (1984). *Phonology and syntax: the relationship between sound and structure*. Cambridge: MIT Press.
- Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology*, 3, 371-405.
- Selkirk, E. O. (1995). Sentence prosody: intonation, stress, and phrasing. Dans J. A. Goldsmith (Éd.), *The handbook of phonological theory* (pp. 550-569). London: Blackwell.
- Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 592-608.

- Sharbrough, F., Chatrian, G.-E., Lesser, R. P., Lüders, H., Nuwer, M., & Picton, T. W. (1991). American electroencephalographic society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *8*, 200-202.
- Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review*, *101*(2), 357-361.
- Shukla, M., Nespors, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*, 1-32.
- Simon, H. A. (1974). How big is a chunk? *Science*, *183*(4124), 482-488.
- Snow, D. (1994). Phrase-Final Syllable Lengthening and Intonation in Early Child Speech. *Journal of Speech and Hearing Research*, *37*(4), 831-840.
- So, L. K. H., & Dodd, B. J. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language*, *22*(3), 473-495.
- Stark, K., & Calfee, R. C. (1970). Recoding strategies in short-term memory. *Journal of Experimental Psychology*, *85*(1), 36-39.
- Starr, A., & Barrett, G. (1987). Disordered auditory short term memory in man and event-related potentials. *Brain*, *110*, 935-959.
- Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, *86*, 142-164.
- Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*(2), 191-196.
- Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERP's as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*(3), 267-295.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*(3736), 652-654.
- Swingle, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86-132.

- Terrace, H. S. (2001). Chunking and serially organized behavior in pigeons, monkeys and humans. Dans R. G. Cook (Éd.), *Avian visual cognition*. Medford, MA: Comparative Cognition Press.
- Terrace, H. S., & Chen, S. (1991). Chunking during serial learning by a pigeon: II. Integrity of a chunk on a new list. *Journal of Experimental Psychology: Animal Behavior Processes*, 17(1), 94-106.
- Terrace, H. S., Jaswal, V., Brannon, E., & Chen, S. (1996). What is a chunk? Ask a monkey. *Abstracts of Psychonomic Society*, 1, 35.
- Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, 39, 706-716.
- Thrope, C., & Rowland, G. E. (1965). The Effect of "Natural" Grouping of Numerals on Short Term Memory. *Human Factors*, 7, 38-44.
- Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.
- Tomasello, M. (2000a). Do young children have adult syntactic competence? *Cognition*, 74, 209-253.
- Tomasello, M. (2000b). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, 4, 156-163.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.
- Tremblay, P. L., Bedard, M. A., Levesque, M., Chebli, M., Parent, M., Courtemanche, R., et coll. (2009). Motor sequence learning in primate: role of the D2 receptor in movement chunking during consolidation. *Behavioral Brain Research*, 198, 231-239.
- Umeda, N., & Quinn, A. M. S. (1981). Word duration as an acoustic measure of boundary perception. *Journal of Phonetics*, 9, 18-28.

- Vaissière, J. (1983). Language-independent similarities. Dans A. Cutler & D. R. Ladd (Éds.), *Prosody: Models and measurements* (pp. 53-66). New-York: Springer-Verlag.
- Vaissière, J., & Michaud, A. (2006). Prosodic constituents in French: a data-driven approach. Dans I. Fónagy, Y. Kawaguchi & T. Moriguchi (Éds.), *Prosody and syntax* (pp. 47-64). Amsterdam: John Benjamins.
- Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25(7-9), 905-945.
- Wechsler, D. (1997). *WAIS-III Administration and scoring manual*. San Antonio, TX: Psychological Corporation.
- Wenk, B. J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, 10, 193-216.
- Wickelgren, W. A. (1964). Size of rehearsal group and short-term memory. *Journal of Experimental Psychology*, 68, 413-419.
- Wickelgren, W. A. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology*, 19, 97-102.
- Wilkes, A. L. (1975). Encoding processes and pausing behaviour. Dans A. Wilkes & A. Kennedy (Éds.), *Studies in long-term memory* (pp. 19-42). Oxford: John Wiley & Sons.
- Wilkes, A. L., Lloyd, P., & Simpson, I. (1972). Pause measures during reading and recall in serial list learning. *Quarterly Journal of Experimental Psychology*, 24, 48-54.
- Winzenz, D., & Bower, G. H. (1970). Subject-imposed coding and memory for digit series. *Journal of Experimental Psychology*, 83(1), 52-56.
- Wioland, F. (1984). Organisation temporelle des structures rythmiques du français parlé: Étude d'un cas. *Bulletin de la section linguistique de la faculté des lettres de Lausanne*, 6, 293-322.

- Wolach, I., & Pratt, H. (2001). The mode of short-term memory encoding as indicated by event-related potentials in a memory scanning task with distractions. *Clinical Neurophysiology*, *112*, 186-197.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, A. (2008). *Formulaic language: Pushing the boundaries*. New York: Oxford University Press.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language & Communication*, *20*, 1-28.
- Zatorre, R. J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. *Cerebral Cortex*, *11*(10), 946-953.
- Zeschel, A. (2008). Lexical chunking effects in syntactic processing. *Cognitive Linguistics*, *19*(3), 427-446.
- Zheng, X., & Pierrehumbert, J. B. (2010). The effects of prosodic prominence and serial position on duration perception. *The Journal of the Acoustical Society of America*, *128*(2), 851-859.

Annexe 1 : Formulaire de consentement



**Hôpital
Rivière-des-Prairies**
Centre hospitalier de soins psychiatriques

Université 
de Montréal

Université 
de Montréal

Laboratoire de sciences phonétiques
Département de linguistique et de traduction
<http://www.ling.umontreal.ca/phonetique/>

FORMULAIRE DE CONSENTEMENT À VOTRE PARTICIPATION À UN PROJET DE RECHERCHE

TITRE DU PROJET

Potentiels évoqués par la mise en mémoire des structures prosodiques de la parole : Étude exploratoire d'un paradigme expérimental.

RESPONSABLE(S) DU PROJET

Boutheina Jemel - Laboratoire de Recherche en Neurosciences et Electrophysiologie Cognitive, Hôpital Rivière-des-Prairies

Victor J. Boucher - Laboratoire de sciences phonétiques - Département de linguistique - Université de Montréal.

CO-CHERCHEURS

Annie Gilbert - Laboratoire de sciences phonétiques - Département de linguistique - Université de Montréal.

ORGANISME SUBVENTIONNAIRE

Conseil de Recherches en Sciences Humaines (CRSH)

PRÉSENTATION DU PROJET ET OBJECTIFS

Nous vous demandons de participer à un projet de recherche. Cependant, avant d'accepter de participer à ce projet de recherche, veuillez prendre le temps de lire, de comprendre et de considérer attentivement les renseignements qui suivent.

Ce formulaire de consentement vous explique le but de cette étude, les procédures, les avantages, les risques et inconvénients, de même que les personnes avec qui communiquer au besoin.

Le présent formulaire de consentement peut contenir des mots que vous ne comprenez pas. Nous vous invitons à poser toutes les questions que vous jugerez utiles.

La présente étude vise à déterminer si la prosodie (rythme et intonation de la parole) influence la mémoire immédiate et la perception de la parole de façon similaire et permettra d'élucider le rôle fondamental de la mémoire immédiate dans la compréhension de la parole.

NATURE ET DURÉE DE VOTRE PARTICIPATION

Pour participer, vous devez avoir le français comme langue maternelle, n'avoir aucun trouble d'ouïe, ne pas être musicien et être droitier (ce qui sera vérifié par un court questionnaire).

Les tâches se divisent en deux sessions d'environ 35 minutes chaque et impliquent le port d'un casque (voir illustration) et la pose d'électrodes faisant partie d'une technique standard utilisée en électro-encéphalographie. Lors des tâches, on entend des stimuli constitués de phrases et de séries de sons. Pour chaque stimulus entendu, on doit répondre le plus rapidement possible si «oui » ou «non» un mot ou un son était présent à l'écoute en tapant sur des touches de clavier.

Ces tâches sont exigeantes en ce qu'on doit maintenir une attention constante. Pour minimiser l'inconfort, on prévoit des pauses où on laisse le participant décider à quel moment reprendre le test. Il y a aussi l'inconfort physique de porter un casque pendant plus d'une heure et l'application de petites couches de graisse conductrice sur le cuir chevelu. Cette graisse n'est pas nécessairement visible et le participant pourra rincer ses cheveux dans une salle de toilette individuelle près du lieu du test.



AVANTAGES POUVANT DÉCOULER DE VOTRE PARTICIPATION

Vous ne retirerez aucun avantage de votre participation à ce projet si ce n'est votre contribution à l'avancement des connaissances scientifiques dans le domaine des sciences de la parole et du langage.

INCONVÉNIENTS POUVANT DÉCOULER DE VOTRE PARTICIPATION

Aucun inconvénient majeur ne peut découler de votre participation. Cependant, outre le temps et le déplacement consacrés à votre participation, il est possible que vous ressentiez un certain état de frustration, de stress ou de fatigue.

RISQUES

Votre participation à ce projet de recherche ne vous fait courir, sur le plan médical, aucun risque que ce soit. Il est également entendu que votre participation au projet n'aura aucun effet sur tout traitement auquel vous serez éventuellement soumis.

COMPENSATION FINANCIÈRE

Une compensation de 40\$ est offerte aux participants pour deux sessions de test.

INFORMATION CONCERNANT LE PROJET

On répondra à votre satisfaction à toute question que vous poserez à propos du projet de recherche auquel vous acceptez de participer.

RETRAIT DE VOTRE PARTICIPATION

Il est entendu que votre participation au projet de recherche décrit ci-dessus est tout à fait volontaire, et que vous êtes à tout moment libre de mettre fin à celle-ci sans avoir ni à motiver votre décision ni à subir de préjudice de quelque nature que ce soit. Le retrait de votre participation n'affectera d'aucune façon les services ou les traitements ultérieurs qui vous seront offerts. Dans le cas de votre retrait ou de l'arrêt du projet, la compensation sera au pro rata des tâches accomplies.

En cas de retrait de votre part, les données qui vous concernent pourront être détruites à votre demande.

ARRÊT DU PROJET PAR LE CHERCHEUR

Le projet de recherche peut être interrompu par le chercheur pour différents motifs ou dans certaines circonstances, par exemple, des contre-indications d'ordre éthique ultérieures ou l'établissement de nouveaux critères de sélection auxquels le sujet ne répondrait plus

CONFIDENTIALITÉ

Les renseignements personnels (nom, adresse ou toute autre indication) vous concernant seront gardés confidentiels dans les limites prévues par la loi. Ils seront codifiés et gardés au Laboratoire de sciences phonétiques de l'Université de Montréal dans un classeur sous clé, auquel seuls les responsables du projet auront accès. De plus, toutes les données vous concernant, seront conservées pendant cinq ans suite à la publication des résultats et détruits à la fin de ce délai. En cas de présentation des résultats de cette recherche ou de publication dans des revues spécialisées, rien ne pourra permettre de vous identifier ou de vous retracer.

Une exception sera faite dans les cas où votre dossier devrait être révisé par un comité de déontologie, le comité d'éthique de la recherche de l'Institut universitaire de gériatrie de Montréal ou par les organismes qui subventionnent ou commanditent cette recherche. Les membres de ces comités sont tenus de respecter les exigences de confidentialité.

SIGNATURES :

Je déclare avoir lu et pris connaissance du projet, de la nature et de l'ampleur de ma participation, ainsi que des risques auxquels je m'expose tels qu'exprimés dans le présent formulaire. Une copie de ce formulaire de consentement me sera remise.

Nom du sujet

Signature du sujet

Fait à _____, le _____.

Je, soussigné(e) _____, certifie :

- a) avoir expliqué au signataire intéressé les termes du présent formulaire;
- b) avoir répondu aux questions qu'il m'a posées à cet égard;
- c) lui avoir clairement indiqué qu'il reste à tout moment libre de mettre un terme à sa participation au projet de recherche décrit ci-dessus.

Nom du chercheur ou de son représentant

Signature du chercheur ou de son représentant

Fait à _____, le _____.

ACCÈS AUX CHERCHEURS

Si vous avez besoin de plus amples renseignements au sujet de ce projet de recherche vous pourrez communiquer avec les chercheurs principaux : Victor J. Boucher, Tél. 514-343-6019; [REDACTED], ou Boutheina Jemel, Tél. 514-323-7260 poste: 2724.

EN CAS DE PLAINTÉ

Si vous avez des plaintes ou des commentaires à formuler, ou si vous avez des questions concernant vos droits, et pour tout problème concernant les conditions dans lesquelles se déroule votre participation à ce projet, vous pouvez, après en avoir discuté avec la personne responsable du projet, faire part de vos préoccupations à l'ombudsman de l'Hôpital Rivière-des-Prairies, Mme Hélène Bousquet, au (514) 323-7260 poste 2154.

INFORMATION SUR LA SURVEILLANCE ÉTHIQUE

Le comité d'éthique de la recherche de l'Hôpital Rivière des prairies a approuvé ce projet de recherche et s'assure du respect des règles éthiques durant tout le déroulement de la recherche. Pour toute information, vous pouvez rejoindre le secrétariat du Comité d'éthique de la recherche au (514) 323-7260 poste 2281.

Comité d'éthique de la recherche
Hôpital Rivière-des-Prairies
Projet accepté le: 22 juillet 08
Signature: [REDACTED]

Annexe 2 : Profil général des participants

# Participant	Sexe	Âge	% Droitier	Score mémoire standardisé	Audition
					● oreille droite + oreille gauche
1	F	39	Non disponible	Non disponible	Non disponible
2	M	28	80	9	Non disponible
3	F	20	80	12	
4	F	27	79	10	
5	M	26	90	9	
6	F	22	89,5	9	
7	F	21	89,5	8	
8	F	19	90	11	

# Participant	Sexe	Âge	% Droitier	Score mémoire standardisé	Audition	
					● oreille droite	+ oreille gauche
9	M	21	100	16		
10	M	28	100	9		
11	F	21	100	11		
12	F	20	100	10		
13	M	29	100	10		
14	F	41	80	12		
15	F	19	100	8		
16	F	24	87,5	8		

# Participant	Sexe	Âge	% Droitier	Score mémoire standardisé	Audition	
					● oreille droite	✚ oreille gauche
17	F	24	100	16		
18	M	24	87,5	8		
19	F	22	68	9		
20	M	36	100	8		
Minimum		19	68	8		
Maximum		41	100	16		
Moyenne		25,55	90,58	10,16		
Écart type		6,48	9,75	2,43		