Electrophysiological Investigations of Age Differences in Phrasal Interpretation: The Time Course of Cross-modal Interactions

Shani Haviva Abada

School of Communication Sciences and Disorders
McGill University, Montréal, Canada

August, 2010

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

© Shani Haviva Abada, 2010

Table of Contents

Figures and Tables

Figures and Tables

DEDICATION

To my parents,
Maralyn and Raphael Abada,
and my grandparents,
for support and encouragement always

To Craig Zern,
for unending patience and so much more

ABSTRACT

The current research examines the neural correlates of younger and older adults' processing of prosodic cues as they relate to phrase groupings and the influence of visual context on prosodic perception. Studies investigating linguistic prosodic perception in older adults show that these individuals remain sensitive to prosody, but allude to subtle processing differences. The use of event-related brain potentials (ERPs) is a particularly useful means of investigating prosodic processing because ERPs permit an analysis of prosodic processing in real time.

Here, ERPs were recorded from younger (ages 18 to 25 years; n = 20) and older (ages 65 to 80 years; n = 11) subjects when presented with phrases such as 'bag and bed and cup', with pauses inserted so as to create a phrasal grouping with an early boundary ('bag # and bed and cup') or a late boundary ('bag and bed # and cup'). Visual displays of the items were presented simultaneous with the onset of the auditory phrases. These pictures corresponded to the phrases (match), differed in the phrase grouping depicted (prosodic mismatch), differed by the center item (semantic mismatch), or differed in both phrase grouping and the second item (double mismatch). Participants were asked to determine whether the auditory and visual stimuli matched.

We found that older and younger participants were able to successfully integrate auditory and visual prosodic and semantic information. Both age groups showed increased difficulty detecting prosodic mismatches, though this was particularly difficult for older adults. Prosodic and semantic mismatches were reflected in N400 and P600 electrophysiological responses, providing important insight into the interpretation of these components. Interestingly, many young adults and all older adults displayed a specific pattern of eye movement which also influenced neural responses. Together, the ERP, eye movement and behavioral findings suggest that older and younger adults display similar sensitivity to prosody in early processing stages but may differ in performance at later stages of integration.

RÉSUMÉ

La présente recherche a pour objet l'étude des corrélats neuronaux, chez des adultes jeunes et plus âgés, du traitement de signaux prosodiques liés à des groupements syntagmatiques ainsi que de l'influence du contexte visuel sur la perception de la prosodie. Les études sur la perception de la prosodie linguistique chez des adultes âgés montrent que ces individus demeurent sensibles à la prosodie, mais font état de différences subtiles dans le traitement prosodique. Le recours aux potentiels évoqués (PÉs) s'avère être un moyen utile d'investigation sur le traitement prosodique dans la mesure où ceux-ci permettent une analyse du traitement prosodique en temps réel.

Notre étude se penche sur les PÉs acquis auprès de sujets jeunes (18 à 25 ans, n = 20) et plus âgés (65 à 80 ans, n = 11) à qui ont été présentés des syntagmes du type «sac et lit et tasse» séparés par des pauses de telle façon à créer un groupement syntagmatique à frontière tôt («sac # et lit et tasse») ou à frontière tardive («sac et lit # et tasse»). Ces syntagmes étaient présentés de façon auditive en même temps que des images représentant les objets mentionnés. Ces images correspondaient aux syntagmes (concordance), différaient des groupements syntagmatiques présentés (discordance prosodique), différaient au niveau de l'objet central (discordance sémantique), ou encore au niveau à la fois du groupement syntagmatique et du deuxième objet (discordance double). La tâche des participants consistait à déterminer si les stimuli auditifs et les stimuli visuels concordaient.

Nos données montrent que les participants jeunes et âgés étaient en mesure d'intégrer les stimuli prosodiques et visuels. Les deux groupes avaient plus de

difficulté à détecter les discordances prosodiques, quoique cette difficulté ait été bien plus importante pour les adultes plus âgés. Les discordances prosodiques et sémantiques se reflétaient par des réponses neurophysiologiques correspondant à la N400 et à la P600, ce qui apporte d'importantes indications sur l'interprétation de ces composantes. Il est intéressant de noter qu'il y avait des patrons visuels spécifiques chez de nombreux jeunes adultes et tous les adultes âgés, lesquels influençaient aussi les réponses neuronales. L'ensemble des PÉs et des données comportementales indique une sensibilité prosodique similaire chez les adultes plus jeunes et plus âgés dans les premiers stades du traitement, mais une différence entre ces deux groupes dans les stades ultérieurs d'intégration.

ACKNOWLEDGEMENTS

I have been fortunate to have tremendous support and guidance throughout the course of my PhD experience. This dissertation would not have been what it is without the guidance of my supervisor, Dr. Shari Baum, and committee members, Dr. Karsten Steinhauer and Dr. John E. Drury. I am grateful to Shari for giving me the freedom to pursue my research interests – and for her flexibility each time plans changed. I thank Shari for making sure projects and marathon meetings stayed on track, always being available whether in person, by e-mail, phone or Skype, and providing feedback on so many drafts of this work with incredible speed and attention to detail. I am also grateful for her support, motivation, sarcasm, and humor throughout every stage of the PhD and life-cycle events I've shared with her over these years.

I thank Karsten for instilling in me an appreciation of SAS that I never thought possible, for teaching me how to carefully examine ERP waveforms, and for his enthusiasm regarding my ideas and findings. I strive to reach a point where, like him, I find it easier to track eye movement using EEG rather than an eye tracker. I thank John for the many, many hours, many, many e-mails, and many, many Excel spreadsheets required to counterbalance these stimuli and to create a strong design that I feel proud of. I am grateful for the life-changing knowledge of Pivot tables and if-then statements John has given me. Thanks are also due for all the encouragement and for one day asking: "You OK?"

I am grateful to all those who contributed to this project. I thank Erin Vensel for recording the auditory stimuli and for all the fun in (and out of) the lab. I thank Eric Zentner for creating the visual stimuli in an efficient way with an

incredible attention to detail. I am grateful to Dr. Erin Bell, Jonathan Eidelman, and Dr. Hyekyung Hwang who assisted in data collection. I thank Gill Abada for regular technical support that has kept my laptop running long after its natural lifespan. I am grateful to Nathaniel Zaharia and Christine Seuss for their Illustrator tutorials. I also thank Chrissy for opening her home and computer to me during crunch time. I am appreciative of all my professors and colleagues throughout my time in Beatty Hall who provided an intellectually stimulating environment, coffee, conversation, and stress-relief; and special thanks to Nicolas Bourguignon for translating the abstract. I owe gratitude to Antoinette Sommer, Lili Saran, and Karen Cavanaugh, who have made every step of the PhD so much easier and who have made Beatty Hall so much brighter.

Thanks are also due to those who sent me on this path and taught me the basics of quality research: Dr. Rosalind Thornton, Dr. Stephen Crain, Dr. Colin Phillips, Dr. David Poeppel, Dr. Andrea Gualmini and Dr. Luisa Meroni. I am thankful for the many friends who supported me throughout my graduate career in various ways, especially: Tara Bognar, Laura Copeland, Scott Gaslow, Dr. Christopher Grindrod, Scott Horowitz, Jon Jon Mitchell, Robin Nobel, Jen Rich, Ally Tash, Judith Wolfthal, and my daily writing partner, Nathan Feiles. I am also appreciative of everyone at Young Israel of White Plains for weekly check-ins and keeping us well fed, with special thanks to Rabbi Shmuel Greenberg, Mrs. Ahuva Greenberg, and Dr. Shalom Wind. And of course, thanks to Dr. Jorge Cham for creating PhD Comics.

I am deeply indebted to Imanuel Muller for returning to Montreal at the perfect time, housing me, providing statistical counseling, Second Cup writing

sessions, Crapette tournaments, life-saving laptop chargers, friendship, sarcasm, and much more.

I thank Linda and Marty Zern, for placing me in a controlled environment and for being so supportive and sharing all the frustration and excitement. I am so appreciative and thankful for my family, especially my parents, Maralyn and Raphael, who supported me through this whole journey in more ways than I can describe; I would not have seen this through without their love and encouragement. A special thanks to my mother for all the "Crisafi Cosher" meals and snacks. Thanks also to Sasson for all the rides and coming "all the way to White Plains".

Finally, to my husband, Craig Zern, for waiting, memorizing all the exits on I-87, showing me what's important, and for not only believing that I could, but that I would, and for making sure that I did, even when it would have been easier if I didn't. And for making sure there was always coffee.

FUNDING

This work was funded by CIHR grants to Dr. Shari Baum (MOP-11290) and Dr. Karsten Steinhauer (MOP-74575) as well as CRC/CFI awards to Dr. Steinhauer (#201876, #209843).

STATEMENT OF ORIGINALITY

The studies conducted for this dissertation employ a novel cross-modal design to examine the electrophysiological correlates of the integration of auditory and visual information for prosodic and semantic processing. This processing was examined in younger adults in Study 1 and in older adults in Study 2. The two studies have been written in manuscript format. At the time of official submission of this dissertation to McGill University, part of the results of Study 1 had been presented at the 16th Annual Meeting of the Cognitive Neuroscience Society in March 2009 and the 17th Annual Meeting of the Cognitive Neuroscience Society in April 2010. Independent analyses conducted on the participants in Study 2, which are not presented in the dissertation, were presented at the Fifth International Conference on Speech Prosody in May 2010.

CONTRIBUTION OF AUTHORS

Abada, S.H., Baum, S.R., Drury, J.E., & Steinhauer, K. (in preparation). When what you see isn't what you get: The influence of visual input on the neural correlates of prosodic processing.

and

Abada, S.H., Steinhauer, K., Drury, J.E., & Baum, S.R. (in preparation). An ERP investigation of older adults' integration of auditory and visual information for prosodic processing.

I (Shani Haviva Abada) was responsible for the development of the ideas, questions and hypotheses that served as the basis for this study. I was responsible for stimuli preparation and study design, participant recruitment, data collection, data analysis, interpretation of results, and manuscript writing. Dr. John E. Drury helped counterbalance the stimuli, thereby strengthening the study design, and provided guidance throughout data analysis, interpretation, and manuscript writing. Dr. Karsten Steinhauer guided me through data analysis using EEProbe and SAS, directed me towards examining eye movement data, and provided guidance throughout data analysis, interpretation, and manuscript writing. Dr. Shari Baum provided guidance and support at all stages of this work, but especially during manuscript writing and revisions.

CHAPTER 1:

General Introduction

As life expectancy increases and as the number of adults older than 65 years approaches record numbers, it becomes increasingly important to understand the process of aging (Hess & Blanchard-Fields, 1996; Birren & Schroots, 2001). This understanding will allow society to ensure that older people remain capable of performing everyday tasks by modifying those tasks when necessary in accordance with changes that may accompany aging (Hess & Blanchard-Fields). For these, among other reasons, the post-World War II period has seen increased interest in aging research (Birren & Schroots). Unfortunately, at this stage, there is little agreement regarding age-related changes that are observed, as well as how to interpret them (Hess & Blanchard-Fields). For example, while virtually all researchers agree that cognitive slowing accompanies aging (Fisher, Fisk, & Duffy, 1995; Hess & Blanchard-Fields), there is little agreement as to the nature and cause of this slowing (Madden, 2001). Slower responses observed in behavioral experiments examining language, for example, may be the result of slower motor abilities (see Ketcham & Stelmach, 2001) or may be related to general cognitive slowing (Cerella, 1985; Salthouse, 1985), task-specific slowing (e.g., Lima, Hale, & Myerson, 1991), or process-specific slowing (e.g., Fisk & Rogers, 1991).

An important goal of research on aging populations is to determine the abilities and processes that change with age and those that are not affected by age. Language is a process of particular interest in aging research, not only because language is vital to remaining an active part of society, but also because there is some variability in linguistic performance in older adults. That is, though aging is associated with difficulties understanding complex structures and linguistic forms that strain working memory (e.g., Byrd, 1993; Kemper, 1992; Light, 1990; Stine & Wingfield, 1990), as well as with decreased sensitivity to temporal auditory cues and decreased ability to interpret certain acoustic cues associated with speech perception and auditory word recognition difficulties (e.g., Tremblay, Piskosz, & Souza, 2002, 2003; Stine-Morrow, Miller, & Nevin, 1999; Wingfield, Aberdeen, & Stine, 1991, among many others), overall linguistic and syntactic knowledge appear to be relatively preserved in older adults (Wingfield, Lindfield, & Goodglass, 2000). Moreover, while many age-related changes are interpreted as signs of decline (see Hess & Blanchard-Fields, 1996), evidence suggesting an increased reliance on contextual information in language processing in older adults (e.g., Cohen & Faulkner, 1983; Abada, Baum, & Titone, 2008, Boothroyd & Nittrouer, 1988; Pichora-Fuller, Schneider, & Daneman, 1995; Sommers & Danielson, 1999; Wingfield, Aberdeen, & Stine, 1991) has been interpreted by some as a reflection of increased efficiency in the language system in older adults. Specifically, an ability to rely on contextual cues for language processing has been taken as an indication that the system has adapted to the requirements for language processing and is therefore able to exploit more cues as language experience increases (see Hess & Blanchard-Fields, 1996).

One domain of language that has received relatively little attention to date in the aging literature is that of prosody (or the melodic properties of speech). Prosody plays a vital role in language, as it interacts with many aspects of language processing (e.g., syllables, lexicon, syntax, emotions, pragmatics), making it an important candidate for investigations of the effects of aging. Understanding any prosodic processing changes related to aging will ultimately provide information regarding many aspects of language processing. If it is found that changes in prosodic processing accompany aging, it may be possible to modify only prosody for older adults and thereby improve comprehension in various aspects of language. To fully understand prosodic processing changes that may be associated with aging, it is necessary to examine the nature of prosodic processing itself, as it occurs in real time, rather than the ultimate outcome of the process as reflected in a behavioral response. To this end, electroencephalography (EEG) is a valuable technology as it provides detailed temporal information regarding language processing in real time. The present research therefore employs EEG to examine the differences and similarities in prosodic processing in younger and older adults. Before turning to the present research, prosodic processing in younger adults and older adults will be reviewed, followed by a review of event-related brain potential (ERP) components relevant to this research and those specifically elicited in response to speech prosody.

Prosodic Processing in Young Adults

Prosody, a primary feature of language (Magne, Schon, & Besson, 2003), refers to nonsegmental aspects of language (Nooteboom, 1997). The acoustic parameters of duration, amplitude, and fundamental frequency (f0) are the primary correlates of prosody (Nooteboom, 1997) and are modulated in every utterance and linguistic unit (Cutler, Dahan, & van Donselaar, 1997). Prosodic cues communicate emotion, convey the intent of an utterance (i.e., whether the utterance is a question or a statement), highlight new or important items in an utterance, disambiguate the meaning of words, and convey syntactic structure (Baum & Pell, 1999; Nooteboom, 1997). While syntactic cues are essential for parsing speech, prosodic cues reinforce syntactic cues and clarify misleading syntax (Cutler et al., 1997), although it is important to note that there is no one-to-one correlation between prosody and syntax. A syntactic boundary can occur without a prosodic boundary, though most prosodic boundaries correspond to syntactic boundaries (Gerken, 1996). Typically, in English, phrase boundaries are signaled in production by an increase in syllable duration in the word preceding the boundary (Lehiste, Olive, & Streeter, 1976), an increase in pause duration at the boundary (Streeter, 1978), and a fall in f0 and increased f0 variability on the word preceding the boundary (Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991), followed by a rise in peak f0 on the syllable following the boundary (Ladd, 1988). As these prosodic cues frequently signal boundaries in production, it is reasonable to assume that listeners are sensitive to them and exploit them to detect boundaries in continuous speech.

Numerous behavioral and electrophysiological studies support the claim that young adult listeners do indeed exploit prosodic cues to understand spoken language (see Cutler et al., 1997 for a review). (As will be seen, far fewer studies address the prosodic abilities of older adults, leaving many unanswered questions.) Despite a wealth of research examining younger (and older) adults' use of prosody to segment speech, understand intention, and identify emotion, the present discussion will focus only on the use of prosody to understand syntax at the phrasal and sentential levels, as this is of greatest relevance to the research presented here. More specifically, as will be explained, both young and older adults exploit phrasal boundaries to interpret utterances, though the nature of the processing mechanisms involved in these interpretations may change with age.

Of particular relevance to the present research are studies demonstrating the ability of adults to exploit prosody to interpret phrase boundaries in sentences with relatively simple syntactic structures. The fundamental issue addressed in these studies can be captured by a simple example. The utterance ‘A plus E times O’ can be produced so as to indicate either of the following interpretations:

1a. (A + E) x O

1b. A + (E x O)

Thus, the same phonetic and lexical content can have two significantly different meanings. In 1978, Streeter (see also e.g., Lehiste, 1973; Lieberman, 1967) made use of just such ambiguous utterances, manipulating f0, duration, and amplitude to create different groupings of phrases of this type and asked young adults to indicate which phrase grouping they heard. She found that young adults reliably used both f0 and duration to successfully interpret these phrases (see also Beach, Katz, & Skowronski, 1996).

Another useful way of exploring disambiguation through prosody is by using stimuli containing temporary syntactic ambiguities. For example, early and late closure ambiguities, such as the following from Speer, Kjelgaard, and Dobroth (1996), have been examined frequently:

2a. Because her grandmother knitted pullovers / Kathy kept warm in the wintertime.

2b. Because her grandmother knitted / pullovers kept Kathy warm in the wintertime.

Lexically, these sentences are identical until the sixth word. However, the underlying syntactic structure differs between the sentences. Ambiguities are resolved either when a disambiguating lexical item is encountered (‘Kathy’ in sentence 2a or ‘kept’ in sentence 2b) or by prosodic cues to phrase boundaries that disambiguate the syntax. For example, the presence or absence of a prosodic (phrase) boundary after ‘knitted’ (e.g., vowel lengthening and drop in f0) provides information supporting early closure of the initial clause of the utterance. In this case, the presence of a prosodic boundary makes it clear that ‘knitted’ in sentence 2b is an intransitive verb and that the following word, ‘pullovers’, is the subject of the next (embedded) clause, rather than the direct object of the verb ‘knitted’ in the first clause. By manipulating the prosody in such utterances, investigators have attempted to determine the degree to which listeners are sensitive to and reliant on prosody when resolving ambiguities, as well as the stage at which prosody is exploited for disambiguation (e.g., Marslen-Wilson Tyler, Warren, Grenier, & Lee, 1992; Nagel, Shapiro, Tuller, & Nawy, 1996; Snedeker & Trueswell, 2003).

Findings for younger adults show that prosody is used for syntactic disambiguation, but its use appears to vary based on the task used (e.g., Speer et al., 1996; Marslen-Wilson et al., 1992; Nagel et al., 1996; Watt & Murray, 1996). For instance, Speer et al. (1996) presented young adult listeners with sentences such as those in (2) with prosodic boundaries manipulated so as to be neutral, conform with, or conflict with the syntax. In both a timed end-of-sentence comprehension task and a cross-modal naming task, results indicated that prosody can facilitate comprehension when it cooperates with syntax and can interfere with comprehension when it conflicts with syntax. In a similar vein, Marslen-Wilson et al. (1992) asked subjects to name a visually-presented probe word and rate its appropriateness to a preceding sentence fragment that was temporarily syntactically ambiguous. Young adults responded faster to the probe when it was consistent with the prosodic context, indicating that prosody was used to disambiguate utterances at early stages in processing (see also Nagel et al., 1996; Watt & Murray, 1996). It is important to note that in each study, participants responded long after the prosodic information of interest was presented. Therefore, while it is possible to conclude that young adults are sensitive to prosodic cues, these studies do not provide insight into how, or how quickly, young adults make use of these cues. This point will be revisited below.

Prosodic Processing in Older Adults

As mentioned above, far fewer investigations have examined prosodic processing in older adults. Based on available studies, it is clear that fundamental sensitivity to prosodic cues is preserved in older adults, though it appears that changes in prosodic processing accompany aging. For example, Taler, Baum, and Saumier (2006) extended studies that have used phrases similar to those used by Streeter (1978) (i.e., ‘A plus E times O’) to an aging population and found that there were no significant differences between younger and older adults when interpreting these types of phrases consisting of a set of conjoined nouns. In addition to these simple phrases, Taler et al. (2006) also used slightly more complex sentences, such as in (3) below, to examine whether younger and older adults were able to use prosody to assign phrase structure.

3a. Madam, Flower is the name of my cat.

3b. Madam Flower is the name of my cat.

Older adults performed significantly *worse* than young adults when answering comprehension questions about the sentences, suggesting that, in this context, older adults were less able to use prosodic cues to disambiguate phrase structure.

In contrast with Taler and colleagues' (2006) findings, in a spontaneous segmentation task using passages taken from popular magazines, participants were told to pause the incoming speech when they wished in order to recall the segment they heard. Both younger and older adults paused the incoming speech stream at major clause boundaries (Wingfield, Kemtes, & Soederberg Miller, 2001), indicating sensitivity to, and exploitation of, prosody (though older adults recalled less than young adults). Similarly, for complex sentences containing temporary syntactic ambiguities, such as in 2, above, the pattern of behavior observed in older adults is largely consistent with the behavior observed in younger adults. For example, when subjects were asked to recall sentences, Wingfield, Wayland, and Stine (1992) found that prosody influenced the recall abilities of younger and older listeners such that sentences containing co-operating prosodic boundaries were easiest to recall, though younger adults were more successful at ignoring conflicting prosody. Nevertheless, consistent with differences observed by Taler et al. (2006), a hint of some age-related differences in prosodic processing were observed. Wingfield et al. (1992) found that, when asked to recall a sentence with a conflicting, ungrammatical prosodic boundary, older adults changed the grammatical structure of the utterances to conform to the prosodic boundary, while still retaining the general meaning of the sentence. Younger adults, on the other hand, changed the prosodic boundary to conform to the grammatical structure, again retaining the meaning of the sentence. This suggests that, while both groups of adults appear sensitive to prosody, the importance given to prosodic cues, or the way prosodic cues are processed, may change with age. Similarly, in a sentence completion task, the syntax of younger and older adults' responses conformed to the prosodic cues to boundaries (Kjelgaard, Titone, & Wingfield, 1999). Here too, however, there were some differences in prosodic processing associated with aging. While both groups of adults were sensitive to prosody and correctly completed early- and late-closure prosody with the appropriate syntax in the majority of cases, older adults were more likely to incorrectly complete early closure prosody sentence onsets with late closure syntax. In these rare cases (6% in the young adults, 14% in the older adults), older adults initiated responses *faster* than did younger adults. The authors interpret this as an indication that younger adults had greater difficulty resolving the conflict between prosody and the preferred late closure response, while older adults had less difficulty resolving the prosodic anomaly in order to choose the less computationally demanding (late closure) grammatical structure. Another possibility, however, is that older adults weigh prosodic cues relative to syntactic structure differently than do younger adults, causing them to rely on one piece of information over the other under certain conditions, which may be task-specific.

In a self-paced Auditory Moving Window (AMW) task, Titone et al. (2006) measured accuracy in repetition (or paraphrasing) of sentences in young and older adults. They found that conflicting prosody hindered understanding in both age groups, though the influence was greater (i.e., more detrimental) for older adults. Interestingly, in this task, cooperating prosody did not facilitate understanding, again measured by repetition accuracy, in either group. The inconsistency between these findings and most of the findings just presented may be associated with the task employed. Specifically, when the AMW task is coupled with the recall task, the demands may potentially alter the attention given to prosody. It is important to point out that the vast majority of studies have relied on off-line tasks, such as recall or sentence completion. As noted earlier, these tasks do not tap processing as it occurs in real time, and thus the behavioral methods used in all these studies may have masked differences and similarities in prosodic processing in younger and older adults that can be illuminated using other methods.

Recently, Steinhauer, Abada, Pauker, Itzhak, and Baum (2010) examined the real-time processing of garden path temporary syntactic ambiguities (such as those in 2) using ERPs. Participants were asked to judge the acceptability of early and late closure sentences with either cooperating or conflicting prosody. Steinhauer et al. (2010) found that older adults accepted conflicting prosody more frequently than younger adults. Despite this difference in behavioral judgments, older adults showed the same neural response to prosody (the Closure Positive Shift [CPS]) as younger adults, but did not display the N400 component (discussed in greater detail below) elicited in younger adults. In this study, the N400 was associated with difficulty interpreting verb argument structure violations in young adults. The presence of the CPS was an indication that in initial stages, prosodic processing is preserved in aging. The absence of the N400 was interpreted as an indication that differences arise at later stages of processing (Steinhauer et al., 2010).

Taken together, the findings in the literature to date reveal that older adults are indeed sensitive to prosodic cues, and that these prosodic cues influence phrasal interpretation. However, it also appears that some changes in prosodic processing may accompany aging. Since conflicting prosody appears to be more difficult for older adults (e.g., Wingfield et al., 1992; Titone et al., 2006; Steinhauer et al., 2010), it has been hypothesized that older adults are more reliant on prosody than are young adults. However, because of the limited information provided by behavioral methods, it is not yet clear whether the age-related differences observed thus far truly reflect an increased reliance on prosody and/or differences in prosodic processing as it occurs in real time, as opposed to differences in performance primarily related to task demands. That is, behavioral examinations mainly provide information regarding the final outcome of a process, not the process itself. Moreover, reaction time is given a great deal of importance in many of these investigations and serves as the primary method to determine how prosody is processed. However, reaction time measures are not an entirely reliable means by which to evaluate language processing in older individuals because of slowing in motor abilities associated with aging (Ketcham & Stelmach, 2001). Therefore, in order to address the questions that remain in the literature (e.g., whether older and younger adults process prosody in the same way, age-related differences in prosodic processing, etc.), a better means of investigating prosodic processing is through the use of electroencephalography (EEG), as this technology provides detailed temporal information about neural activity (Rugg & Coles, 1995). For this reason, the EEG studies presented in this dissertation are crucial to gaining an understanding of the nature of prosodic processing in aging.

Electrophysiological Correlates of Prosodic Processing

There are a number of advantages to using brain-imaging technology, such as EEG, to examine language processing. EEG measures the ongoing processing of language, eliminating the need for any type of behavioral response (other than as a measure of whether subjects attended to individual trials), thereby providing information beyond deliberate, conscious responses. (Although, as demonstrated by Steinhauer et al. [2010], examining both behavioral and EEG responses can illuminate dissociations between different types of processing and provide more insight into language processing overall.) Pannekamp, Toepel, Alter, Hahne, and Friederici (2005) point out that, when examining behavioral responses, a violation condition is normally required in order to draw conclusions regarding the role of prosody in syntactic resolution. These authors further note that the use of electrophysiological measures, specifically event-related potentials (ERPs), and, importantly, the detection of electrophysiological correlates of prosody such as the Closure Positive Shift (CPS; Steinhauer, Alter, & Friederici, 1999) allows for the examination of online prosodic processing at the sentence level, including the time course of prosodic processing, without relying only on accuracy and reaction time measures. While previous studies have demonstrated that prosody is exploited early in sentence comprehension by younger and older adults, the measurements taken were from a point considerably later than the point of prosodic disambiguation (e.g., Marslen-Wilson et al., 1992; Kjelgaard et al., 1999). Thus, it cannot be determined whether prosody was exploited as soon as it became available or later in the process, when it became necessary to use prosody to interpret the syntax. ERPs, on the other hand, can provide information to resolve this issue and thus represent a valuable method to explore prosodic processing.

A growing body of ERP research has examined the neural correlates of prosodic processing, revealing both negative and positive-going brain waves. Thus far, the possible electrophysiological correlates of prosody that have been identified are the CPS, negative-going waveforms (including the right anterior negativity; RAN), and positive-going waveforms (including the P800). The CPS, first discovered by Steinhauer et al. (1999), refers to a large, bilateral, centro-parietal positive-going waveform, most prominent at midline electrodes, that begins between 150 and 200 ms after the offset of a pre-prosodic boundary word and lasts roughly 500 ms. The CPS has been elicited by boundary-related changes in f0, pre-boundary lengthening, and pause duration in German (Steinhauer et al., 1999), Dutch (Kerkhofs, Vonk, Schriefers, & Chwilla, 2007), Japanese (Wolff, Schlesewsky, Hirotani, & Bornkessel-Schlesewsky, 2008), Mandarin Chinese (Li & Yang, 2009) and English (Pauker, Itzhak, Baum, & Steinhauer, under review; Itzhak, Pauker, Drury, Baum, & Steinhauer, 2010) sentences, though not all acoustic boundary cues are required to elicit a CPS response (e.g., see Steinhauer et al., 1999; Steinhauer, 2003 regarding pauses). Further, the CPS has been elicited from both auditory and visual stimuli signaling phrasal boundaries (i.e., commas; Steinhauer & Friederici, 2001) and from boundaries in stimuli with various types of nonlinguistic content (i.e., nonsense speech, humming) (Steinhauer & Friederici, 2001; Pannekamp et al., 2005), though the scalp distribution appears to vary depending on the type of stimulus presented, suggesting that there is an interaction between prosodic processing and other types of information (see Pannekamp et al., 2005). Importantly, though the distribution of the CPS may change based on the type of input, the presence of the CPS itself does not seem to be affected. Rather, whenever a boundary is present, a CPS is elicited. Interestingly, when a prosodic break is expected based on syntactic structure and verb transitivity biases, a CPS is elicited even in the absence of prosodic cues; the CPS in this context does not differ from the CPS elicited by prosodic cues (Itzhak et al., 2010).

In an interesting recent study, Kerkhofs et al. (2007) embedded temporarily syntactically ambiguous Dutch sentences in discourse contexts that created syntactic and prosodic expectations, so that the point at which disambiguating syntactic information and prosodic boundary information became available coincided. They used this design to determine whether the two types of information interact immediately when they become available. The investigators found that prosodic breaks elicited a positivity 400 – 800 ms after pause onset, somewhat later than the typical timing of the CPS. The authors concluded that this positivity was smallest in response to a prosodic break when a syntactic break was expected (i.e., prosodic break in biasing context) and largest when the syntactic break was not expected (i.e., prosodic break in neutral context). However, examination of the waveforms presented in Kerkhofs and colleagues' Figure 3 reveals that prosodic breaks in both neutral and biasing contexts appear to display a positivity. Though the authors do not present statistical comparisons of the presence or absence of the prosodic break within the biasing context (i.e., prosodic break in biasing context versus no-prosodic break in biasing context), it does appear that the CPS is elicited by the expectation of a prosodic break, regardless of the acoustic marker, similar to the findings of Itzhak et al. (2010), and that the discourse context created an expectation for a specific prosodic boundary. Taken together, the findings of Itzhak et al. and Kerkhofs et al. highlight the important role of context in prosodic processing. Despite the consistent elicitation of a CPS by prosodic boundaries under various conditions, relatively little is currently understood about the specific factors that influence and impact the CPS.

Certain negative components have also been hypothesized to be associated with the processing of unexpected and incongruous prosody. When a sentence is expected to end, but instead the sentence-final word contains a prosodic contour that indicates sentence continuation, a right anterior negativity (RAN) has been reported between 300 and 500 ms later (Eckstein & Friederici, 2005, 2006). Similarly, in French, when words in sentence-final position are presented with an incoherent focal accent or incorrect and unexpected stress patterns, negative components have been elicited in the 250 – 450 ms time window after word onset (Magne, Astésano, Lacheret-Dujour, Morel, & Besson, 2005; Magne et al. 2007). In a study examining the perception of infrequent and incorrect prosody compared to frequent prosody in German, Mietz, Toepel, Ischebeck, and Alter (2008) found early negativities elicited both by infrequent and unexpected (incorrect) prosody. They interpreted these negativities as an N400 (see below) with an early onset in response to words that did not prosodically match sentential context and therefore violated prosodic expectations. Magne and colleagues (2007) also concluded that all these components, including the RAN, are part of the “N400 family” and interpreted this as an indication that manipulations of prosody influence lexical access. While interesting, these negativities are not directly relevant to the current research because of the different nature of the stimuli across studies. In particular, the negativities just discussed were elicited by contrasting either metrical stress or prosodic contours at the sentence level, which are both considerably different than the grouping contrast that will be examined here.

Perhaps the most controversial purported electrophysiological correlate of prosody is the P800. In one study that reported a P800, participants listened to statements and questions that were cross-spliced to create prosodically incongruent utterances; that is, statements ended with prosody that would indicate a question. When participants judged whether utterances were prosodically congruous, the P800, a positive waveform with a left lateralized temporal distribution peaking 800 ms after the onset of the prosodically incongruous words (i.e., the first word in the cross-spliced segment), was elicited (Astésano, Besson, & Alter, 2004). The same experiment included utterances that were semantically incongruent, ending with a semantically unrelated word. When participants judged whether the same set of utterances were semantically congruous, the P800 component was not elicited. Based on this, Astésano et al. (2004) claim that the P800 is linked to intonation contour violations. They further claim that this component is distinct from both the CPS, which is linked to prosodic boundaries and has a wider distribution, and the P600, which is linked to syntactic errors and shows bilateral distribution. They do, however, concede that the P600 and P800 may be in the same "family" of ERP components. Interestingly, in the studies above in which a negativity was present, there were also positive components elicited by incongruous sentence-final prosodic manipulations that may be related to the P800. For instance, Eckstein and Friederici (2005) observed a positive peak 800 ms after a word with a prosodic contour incongruently indicating sentence completion. Unlike Astésano and colleagues, these authors interpreted this component as a late P600, because the participants were attending to syntax rather than prosody, and because of the bilateral distribution of the component. Magne at al. (2007) also found a positive component in the 500 – 1200 ms window after the onset of metrically incongruous words, but only in a task in which participants were asked to focus on meter rather than semantic information. They proposed that this positivity is similar to the P800 and concluded that these components may be related to the P300, which marks surprising task-relevant events (Magne et al., 2007). Mietz et al. (2008) found that cross-spliced utterances with incongruous prosody elicited a biphasic N400/P600 component similar to that reported by Steinhauer et al. (1999), which these authors interpreted as a reflection of "effortful integration of the noun into the syntactic sentence structure" (Mietz et al., 2008 p. 167). Taken together with the findings of Kerkhofs et al. (2007), these studies create a strong case for the presence of an additional positivity, distinct from the CPS though not necessarily distinct from the P600 and/or the P300, elicited by prosody-induced anomalies. The prosodic manipulation employed by Astésano and colleagues was one of (illocutionary) intention of an utterance. In the other studies, the P800 was elicited by other types of sentential prosodic incongruities. Thus, although this component may, indeed, reflect a response to anomalous prosody, it is unlikely that the P800 is a response to the type of prosodic boundaries that are the focus of this dissertation.

Thus far, it appears that prosodic boundaries yield a CPS and prosodic incongruities elicit negative and positive components that may or may not be distinct from N400 and P600 components (discussed below). The specific prosodic, acoustic, or linguistic factors that elicit or modulate these components are as yet unclear. Further, whether these same components are elicited and modulated in the same way in older adults is unknown. The current research will begin to address some of these questions by examining prosodic cues as they relate to phrase groupings illustrated in picture stimuli.

Before turning to the present research, a brief mention must be made of the N400 and the P600—two ERP correlates of language processing referred to above that will be particularly relevant to the current investigations. These two components, perhaps the most frequently examined components in language processing, will be discussed in greater detail in the manuscripts presented in Chapters 2 and 3.

The N400, a centro-parietally distributed component with a negative peak 400 ms after the onset of a stimulus (first discovered by Kutas & Hillyard, 1980), shows an increased amplitude after the onset of a semantically or conceptually anomalous stimulus (see Kutas & Federmeier, 2000 for a review) and is also elicited in older adults (e.g., Faustmann, Murdoch, Finnigan, & Copland, 2007). While there is an ongoing debate, with two dominant interpretations, surrounding the processing marked by the N400 (see Lau, Phillips, & Poeppel, 2008 for a discussion), at present the most widely accepted interpretation views the N400 as a reflection of lexical activation/facilitation by context, with the role of context being primary (see also Kutas & Federmeier, 2000 and Kutas, & Federmeier, 2007 for reviews). As will be seen, the findings presented in this work will extend the current interpretation of the N400.

The specific processes marked by the P600, a positive-going waveform with a centro-parietal scalp distribution in young adults (e.g., Osterhoust & Holcomb, 1992) and a more frontal distribution in older adults (e.g., Kemmer, Coulson, de Ochoa, & Kutas, 2004), remain debated. One view that is currently garnering support in the literature is that the P600 marks integration of structured bodies of information (e.g., Kuperberg, 2007; the "generalized mapping" component, Bornkessel & Schlesewsky, 2006, 2008). The findings presented here will address this interpretation.

The Present Research

A large literature supports prosody as a driving force of language acquisition and a central feature of language (e.g., Gleitman & Wanner, 1982; Fernald & Kuhl, 1987; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). If this is indeed the case, it is of great interest to not only understand the processes involved in interpreting prosody, but to also understand how this feature of language changes and whether it is associated with age-related decline. As has been shown, many questions remain regarding prosodic processing, even more so as it relates to aging. Older adults appear sensitive to prosody and seem to exploit prosodic cues to interpret ambiguous utterances (e.g., Taler et al., 2006; Titone et al., 2006). However, while older adults may arrive at the same ultimate response as younger adults, the available data allude to differences in the means by which the two groups reach that response (e.g., Kjelgaard et al., 1999; Titone et al., 2006; Steinhauer et al., 2010). As discussed earlier, one limitation to all these studies (with the exception of Steinhauer et al., 2010) is the use of behavioral tasks which provide the outcome of processing rather than insight into the processing itself. EEG is a useful means to circumvent this issue as it allows for the temporal examination of language processing in real-time without the need for an overt response. Though older adults have displayed some differences in ERP components compared to younger adults (e.g., smaller and later N400, King & Kutas, 1995; Federmeier & Kutas, 1995; more frontal P600, Kemmer et al., 2004; Steinhauer et al., 2010), they nonetheless show similar components that allow for the comparison of language processing between younger and older adults. The goal of this dissertation is to explore age-related changes in prosodic processing and its interaction with visual context. To that end, three main questions are raised. First, what electrophysiological correlates are elicited by auditory-visual prosodic mismatches? Second, how do prosody and visual context interact? Third, do older adults show the same electrophysiological correlates and responses to auditory-visual interactions as do young adults? Chapter 2 addresses the first two questions in a group of 20 English-speaking young adults. Chapter 3 aims to address the third question by examining a group of 16 older adults and comparing them to a subset of the young adults tested in Chapter 2.

To address the questions posed, simple conjoined phrases are presented aurally along with pictures depicting the phrases. Using simple phrases avoids any confounds of working memory limitations as they relate to syntactic processing that may be present in older adults (see e.g., Stine, 1995).

These stimuli create a novel cross-modal prosodic mismatch in which prosodic interpretation only, and not grammaticality or lexical stress, focus or meter, is violated. In the present investigation, a mismatch of prosodic/grouping information results from the cross-modal integration of otherwise acceptable auditory and visual stimuli. That is, if the utterance ‘A plus E times O’ is produced with a late auditory boundary (e.g., ‘A plus E # times O’) and presented at the same time as the equation ‘A + (E x O)’ is seen, the two equations will result in a mismatch because they differ in prosodic interpretation only, without violating syntax per se. Though the syntactic structure and conceptual interpretation differ between the auditory and visual phrases, both interpretations are syntactically correct. The difference lies in which prosodic boundary is appropriate for the context. Thus, the electrophysiological response will be to a mismatch of auditory prosody and information available in the visual context.

One benefit of using such cross-modal stimuli is the ability to examine the interaction between prosody and visual input—a situation akin to what happens in natural language processing. Outside of the laboratory, perceiving language occurs in some type of context, be it a conversation or a location or setting. Research shows that older adults exploit context more than younger adults when perceiving language (e.g., Wingfield et al., 1991; Ganong, 1980; Connine & Clifton, 1987). A paradigm in which this difference is readily evident, and which is most relevant to the proposed research, is cross-modal picture-word interference investigations (Schriefers, Meyer, & Levelt, 1990). In these investigations, participants hear words that are either semantically and/or phonologically related or unrelated to a target picture. Participants are then asked to name the item represented in the picture. In these tasks, visual and auditory information compete, thereby interfering with or priming (i.e., facilitating) word retrieval. Findings show that semantic and phonological associates differentially influence the access of picture names in both young and older adults (Schriefers et al., 1990; Hanauer & Brooks, 2005; Rosinski, 1977; Taylor & Burke, 2002). These findings suggest that a cross-modal interaction with visual information may also be observed when processing prosody.

Using the cross-modal design employed in the present studies will provide insight into the second question addressed by this dissertation: Is there an interaction between prosody and visual information? Identifying such an interaction will illuminate how prosody is processed (and whether that changes with age). In addition to the prosodic mismatch, a semantic mismatch condition was included in which the second pictured item differed from the second spoken noun. Little is known about how semantic information and grouping information available through the visual modality interact with auditory stimuli and the perception of prosodic cues for phrasal interpretation in young adults. There is no evidence of which, if either, of these particular types of cues is weighted more heavily in understanding speech (i.e., semantics vs. prosody and/or visual vs. auditory) or whether there is any difference in the order in which these types of information are processed. Including the semantic mismatch condition also creates a control condition for which the neural response is more easily predicted (as there are more data available regarding semantic anomalies).

While the issue of cue-weighting across modalities is not the primary focus of this dissertation, the studies reported here may provide informative data that address this question. As the picture-word interference examinations above indicate, language processing (in those cases, word retrieval) is mediated by pictures (e.g., Schriefers et al., 1990). As will be discussed in Chapter 2, a number of studies using a variety of methodologies have demonstrated that visual information interacts with and influences the processing of auditory linguistic stimuli (e.g., Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Hanna & Brennan, 2007; McGurk & MacDonald, 1976; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Altmann & Kamide, 2007; Friedrich & Friederici, 2004; Knoeferle, Habets, Crocker, & Münte, 2008; Gleitman, January, Nappa, & Trueswell, 2007; Knoeferle & Crocker, 2007; Farmer, Anderson, & Spivey, 2007, among many others). However, visual stimuli create difficulties when conducting ERP experiments because eye movements are often larger in amplitude than language-related ERP responses. As a result, eye movements may obscure findings and contaminate data. To limit these effects, the majority of ERP studies employ designs that reduce ocular artifacts. Similar precautions to limit eye movement and the loss of data were used here. In the course of data analysis, the issue of eye movement was addressed. As will be seen in Chapter 2, examining this issue provided ground-breaking insight into cross-modal perception using ERPs.

In general, cross-modal processing has remained relatively unexplored in aging in any type of task. Chapter 3 will examine prosodic processing and the integration of auditory and visual input in older adults. Picture-word interference examinations indicate that older adults show roughly the same behavioral pattern as young adults, though some subtle differences have been observed (Taylor &

Burke, 2002). This suggests that there will be some age-related changes in the integration of cross-modal information in the current experiments. For example, it is entirely possible that older adults will attend more to the visual information (the context, in this case), than to the prosodic information, consistent with findings that older adults exploit context more heavily than younger adults when processing other aspects of language (e.g., Cohen & Faulkner, 1983). Thus, the questions posed here are significant in that they address prosodic processing, aging, and modality cue-weighting while providing insight into frequently-elicited ERP components. This research offers essential information into how the brain processes prosody, what influences this processing, whether any changes in these processes accompany aging, and the interpretation of specific ERP components. We turn first to the study of cross-modal interaction in phrase interpretation in young adults.

CHAPTER 2:

When what you see isn't what you get: The influence of visual input on the neural correlates of prosodic processing

Shani Haviva Abada

Shari R. Baum, John E. Drury, Karsten Steinhauer

Prosody refers to the nonsegmental aspects of speech (Nooteboom, 1997) and plays a critical role in conveying phrase boundaries and syntax in utterances (see e.g., Cutler, Dahan, & van Donselaar, 1997). For example, the sentence, 'A woman without her man is nothing' can be produced with two phrase boundaries that create the utterance 'A woman, without her, man is nothing' or with one phrase boundary that creates the utterance 'A woman without her man, is nothing'. While these utterances share identical lexical content, their meanings differ substantially. If a language user heard the first utterance while simultaneously seeing a picture of a woman crying alone in a room, some confusion would likely result. On the other hand, if the same utterance were heard and presented with a picture of a man alone in a room looking upset, the utterance would most likely be easier to interpret. The present investigation uses a related type of cross-modal picture-to-sound matching paradigm to examine the role of visual context in prosodic processing and associated electrophysiological correlates.

There are numerous examples of prosodic boundaries altering the interpretation of sentences. One example that is particularly relevant to the present study is the equation 'one plus two times three' which can be produced with a late boundary or an early boundary, as in the following example (where # symbolizes a prosodic boundary):

1a. 'one # plus two times three' → $1 + (2 \times 3) = 7$ *Early Boundary*

1b. 'one plus two # times three' → $(1 + 2) \times 3 = 9$ *Late Boundary*

While both of these are completely acceptable equations, they contain different prosodic structures and are considerably different in terms of the order of arithmetic operations, such that each equation resolves to a distinct solution. Clearly, prosody plays an important role in distinguishing between these utterances. A number of investigations have employed phrases of this type to create different arithmetic and phrasal groupings and have asked listeners to determine which grouping they heard. Findings show that listeners of all ages successfully exploit prosodic cues to interpret these phrases and render phrase structure judgments (e.g., Streeter, 1978; Lehiste, 1973; Lieberman, 1967; Beach, Katz, & Skowronski, 1996; Taler, Baum, & Saumier, 2006). By manipulating specific prosodic cues (i.e., pre-boundary syllable duration, pause duration, fundamental frequency [f0] and amplitude), these studies have shown that listeners attend to both f0 and duration to render phrasal groupings. Moreover, both duration and f0 allow for the disambiguation of utterances independent of the other cue (Streeter, 1978; Beach et al., 1996). Not only have many studies confirmed the use of prosodic cues to disambiguate more complex utterances (for example, 'Because her grandmother knitted pullovers # Kathy kept warm in the wintertime.' versus 'Because her grandmother knitted # pullovers kept Kathy warm in the wintertime.') (e.g., Nagel, Shapiro, Tuller, & Nawy, 1996; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992, Speer, Kjelgaard, & Dobroth, 1996), there is also evidence from eye-tracking studies that listeners make immediate, on-line use of these cues to interpret sentences (e.g., Snedeker & Trueswell, 2003). Another very sensitive means of determining how quickly and in what way listeners exploit prosodic cues is by using electrophysiological measures, specifically event-related potentials (ERPs), which allow for the examination of the time-course of on-line prosodic processing across the entire utterance, without relying on accuracy and reaction time measures alone (Pannekamp, Toepel, Alter, Hahne, & Friederici, 2005). Moreover, ERPs may provide insights into the nature of the processes underway.

A growing body of ERP research has examined the neural correlates of prosodic processing, revealing a range of both negative and positive-going brain waves. Prosodic boundaries themselves have most often been associated with a positive-going ERP component. For example, Steinhauer, Alter, and Friederici (1999) examined sentences containing temporary syntactic ambiguities, such as the sentence that begins with the fragment in (2) below which can be followed by either of the two endings (a) or (b):

2. When a bear is approaching <u>the people</u> …

a)	… come running	*Early Closure*
b)	… the dogs come running	*Late Closure*

Steinhauer and colleagues (1999) found that prosodic boundaries elicited a large, bilateral, centro-parietal positive-going waveform, most prominent at midline electrodes, that began between 150 and 200 ms after the offset of the word preceding the prosodic boundary and lasted roughly 500 ms, called the Closure Positive Shift (CPS). The CPS has been elicited by boundary-related changes in f0, pre-boundary lengthening, and pause duration in German (Steinhauer et al., 1999), Dutch (Kerkhofs, Vonk, Schriefers, & Chwilla, 2007), Japanese (Wolff, Schlesewsky, Hirotani, & Bornkessel-Schlesewsky, 2008), Mandarin Chinese (Li & Yang, 2009) and English (Pauker, Itzhak, Baum, & Steinhauer, under revision; Itzhak, Pauker, Drury, Baum, & Steinhauer, 2010) sentences, though a pause is not required to elicit a CPS response (Steinhauer et al., 1999; Steinhauer, 2003).

Further, the CPS has been elicited from both auditory and visual stimuli signaling phrasal boundaries (i.e., commas; Steinhauer & Friederici, 2001) and from boundaries in stimuli with various types of nonlinguistic content (i.e., nonsense speech, humming) (Steinhauer & Friederici, 2001; Pannekamp et al., 2005), though the scalp distribution appears to vary depending on the type of stimulus presented (see Pannekamp et al., 2005). This suggests that there is an interaction between prosodic processing and other types of information (Pannekamp et al., 2005). Importantly, though the distribution of the CPS may change based on the type of input, the presence of the CPS itself does not seem to be affected. Rather, whenever a boundary is present, a CPS is elicited. Interestingly, when a prosodic break is expected based on syntactic structure and verb transitivity biases, a CPS is elicited even in the absence of prosodic cues; the CPS in this context does not differ from the CPS elicited by prosodic cues (Itzhak et al., 2010). Kerkhofs et al. (2007) embedded temporarily syntactically ambiguous Dutch sentences in discourse contexts that created syntactic and prosodic expectations, so that the point at which disambiguating syntactic information and prosodic boundary information became available coincided. They used this design to determine whether the two types of information interact immediately when they become available. They found that prosodic breaks elicited a positivity 400 – 800 ms after pause onset, somewhat later than the timing of the CPS. The authors concluded that this positivity was smallest in response to a prosodic break when a syntactic break was expected (i.e., prosodic break in biasing context) and largest when the syntactic break was not expected (i.e., prosodic break in neutral context). However, examination of the waveforms presented in Kerkhofs and colleagues'

Figure 3 reveals that prosodic breaks in both neutral and biasing contexts appear to display a positivity. Though the authors do not present statistical comparisons of the presence or absence of the prosodic break within the biasing context (i.e., prosodic break in biasing context versus no-prosodic break in biasing context), it does appear that the CPS is elicited by the expectation of a prosodic break, regardless of the acoustic marker, similar to the findings of Itzhak et al. (2010), and that the discourse context created an expectation for a specific prosodic boundary. Taken together, the findings of Itzhak et al. and Kerkhofs et al. highlight the important role of context in prosodic processing.

Though Kerkhofs and colleagues (2007) concluded that the positivity elicited by their stimuli was a CPS, they recognized that they could not rule out the interpretation that this positivity was a P600 instead. These two components could not be dissociated in their study, because in the neutral context there was a conflict between prosodic breaks and syntactic structure. In their stimuli, the syntactic structure created an expectation for the sentence to continue without a break (prosodic or syntactic). Therefore, the presence of the prosodic break may have created the perception of a syntactic break that would then require reanalysis, making the elicitation of a P600 rather than a CPS entirely likely. Similar findings were observed by Schmidt-Kassow and Kotz (2009) when they presented listeners with sentences that contained no violations, metrical violations, syntactic violations, or both metrical and syntactic violations. They found that both metrical and syntactic violations elicited a P600 response, which they interpreted as an indication that the P600 reflects the reprocessing associated with violations of rule-based expectancies that is neither specific to syntax nor language. The P600, a positive-going waveform with a centro-parietal scalp distribution, has been linked to syntactic processing (e.g., Osterhout & Holcomb, 1992), revision and reanalysis (e.g., Friederici, 2002; Friederici & Weissenborn, 2007), integration of syntax and semantics (e.g., Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005), and/or general integration processes (e.g., Kaan & Swaab, 2003; Osterhout, Kim, & Kuperberg, 2007; Schmidt-Kassow & Kotz, 2009). While the specific processes marked by the P600 remain debated, one view that is currently garnering support in the literature is that the P600 marks integration of structured bodies of information (e.g., Kuperberg, 2007; the "generalized mapping" component, Bornkessel & Schlesewsky, 2006, 2008). Another (controversial) ERP component—the P800, purported to be an electrophysiological correlate of prosody—has been hypothesized to be related to the P600. In one study that reported a P800 (Astésano, Besson, & Alter, 2004), participants listened to statements and questions that were cross-spliced to create prosodically incongruent utterances; that is, statements ended with prosody that would indicate a question. When participants judged whether utterances were prosodically congruous, the P800, a positive waveform with a left lateralized temporal distribution peaking 800 ms after the onset of the prosodically incongruous words (i.e., the first word in the cross-spliced segment), was elicited, but was absent when participants judged whether the same utterances were semantically congruous. Based on this, Astésano et al. (2004) concluded that the P800 is linked to intonation contour violations and is distinct from both the CPS, which is linked to prosodic boundaries and has a broader distribution, and the P600, which they believe is linked to syntactic errors and shows bilateral distribution. Nonetheless, they did concede that the P600 and P800 may be in the same “family” of ERP components.

Another ERP component that is relevant to the current study and must therefore be addressed is the N400. This well-documented centro-parietally distributed component, first discovered by Kutas & Hillyard (1980), has a negative peak 400 ms after the onset of a stimulus, shows an increased amplitude after the onset of a semantically or conceptually anomalous stimulus, and is elicited by spoken words, written words, pictures, signs, sign language, sounds, and gestures (see Kutas & Federmeier, 2000 for a review). There is an ongoing debate, with two dominant interpretations, surrounding the processing marked by the N400 (see Lau, Phillips, & Poeppel, 2008 for a discussion). According to one interpretation, the N400 reflects the integration of critical semantic or conceptual input with the context and therefore marks post-lexical access mechanisms. According to the other dominant view, the N400 reflects lexical activation, and its relative facilitation, by a context. Lau and colleagues review relevant ERP and fMRI findings and resolve that while there is thus far no conclusive evidence to determine which (or what combination) of these interpretations is reflected by the N400, at present the more widely accepted perspective views the N400 as a reflection of lexical activation/facilitation by context, with the role of context being primary (see also Kutas & Federmeier, 2000 and Kutas & Federmeier, 2007 for reviews).

It should be noted that additional ERP components have been associated with prosodic processing, although the studies that have reported such components have made use of violations of prosody (Eckstein & Friederici, 2005, 2006; Magne, Astésano, Lacheret-Dujour, Morel, & Besson, 2005; Magne et al. 2007; Mietz, Toepel, Ischebeck & Alter, 2008). Moreover, whereas some investigators have purportedly identified unique prosody-specific components, others have suggested that the components are instead part of the N400 or P600 "families" of components (e.g., Magne et al., 2007). It remains unclear whether the components elicited by these prosodic contrasts are distinct markers of prosodic incongruities or are N400- and P600- (or P300-) related components. It is reasonable to conclude, though, that prosodic contours of sentences (Astésano et al., 2004), lexical biases (Itzhak et al., 2010) and discourse context (Kerkhofs et al., 2007) create expectations for upcoming prosodic boundaries, indicating that contextual information plays a role in prosodic processing. In the present study, we extend the notion of context by examining the relationship between visual context and prosody to explore whether visual context also creates expectations for prosodic phrase boundaries.

To date, the majority of examinations of prosody, both behavioral and electrophysiological, have used auditory stimuli alone. However, outside of the laboratory, perceiving language occurs in some type of context, be it a conversation or setting. Many behavioral studies have shown us that language users exploit various contextual cues when perceiving and processing language in both the auditory and written modalities whenever such cues are available (e.g., Ganong, 1980; Connine & Clifton, 1987). Research also shows that what people see influences how they process what they hear (e.g., Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Hanna & Brennan, 2007 among many others). A very good example of this at the perceptual level is the McGurk effect, in which visual input in the form of facial cues, influences phonetic and prosodic perception (e.g., McGurk & MacDonald, 1976). Eye-tracking studies show that a visual scene rapidly influences the perception of linguistic input at a higher level of processing as well (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Altmann & Kamide, 2007; Knoeferle & Crocker, 2007 among many others). In cross-modal picture-word interference investigations, in which participants name target pictures after hearing words that are either semantically and/or phonologically related or unrelated to the picture, visual and auditory information compete, thereby interfering with or facilitating word retrieval, as shown by response time differences in picture naming (Schriefers, Meyer, & Levelt, 1990; Rosinski, 1977; Hanauer & Brooks, 2005). In electrophysiological investigations, pictures have also been shown to influence speech processing. Friedrich and Friederici (2004) presented adults and 19-month-old infants with pictures for 4000 ms. After a delay of 900 ms, but while the picture was still on the screen, an indefinite article followed by a congruous word, incongruous word, pseudoword or nonword, was heard. They found that adults showed a very early negative-going ERP response in temporal electrodes, interpreted as an N400, 100 – 250 ms after the onset of a real word, indicating that word processing is immediately influenced by visual stimuli. Further, incongruous words elicited a larger centro-parietal right-lateralized N400 that was also visible in anterior regions. Importantly, this indicates that a picture was sufficient to modulate the amplitude of the N400 and demonstrates the value of cross-modal ERP examinations. Similarly, Knoeferle, Habets, Crocker, and Münte (2008) presented ambiguous and unambiguous canonical and noncanonical utterances 1 second after the onset of a visual display of three individuals (e.g., a princess, a pirate, and a fencer; these displays are not entirely unlike the displays of three objects used in the present study). They found that visual scenes led to structural revision of ambiguous verbs in canonical utterances, marked by a P600, indicating that non-linguistic visual input can be exploited for syntactic disambiguation just as spoken linguistic cues are. These and other studies show that auditory and visual input are integrated rapidly and that visual input can play an important role in language processing (e.g., Gleitman, January, Nappa, & Trueswell, 2007; Knoeferle & Crocker, 2007; Altmann & Kamide, 2007; Farmer, Anderson, & Spivey, 2007). These investigations support the usefulness of examining prosodic processing within a visual context. The present study builds on these findings to explore how phrase boundaries are exploited by listeners in the presence of a co-operating or conflicting visual context. Specifically, we manipulated spoken prosody and visual grouping displayed in pictures (creating a mismatch between auditory and visual input) and examined the resulting neural responses to determine whether visual input would influence prosodic processing.

The Present Study

This study employs an innovative cross-modal design that creates a novel type of prosodic mismatch in which prosodic interpretation only, and not grammaticality or lexical stress, focus or meter, is violated. The primary goal was to determine whether an utterance that differed from a visual display in prosodic grouping alone would elicit a mismatch response, and, if so, what the neural correlates to the "prosodic mismatch"[1] might be. Further, we examined whether a pause alone was sufficient to cue this difference in grouping. Finally, we examined how visual context would interact with and influence auditory processing. In our analysis of visual processing, a particularly interesting pattern emerged, resulting in an additional innovative goal of examining how eye movements correlate with neural processing. This issue will be discussed at length below.

In this experiment, both the spoken phrase and the visual input were independently grammatically and semantically acceptable, though their cross-modal integration should trigger a prosodic (and perhaps structural/syntactic) mismatch (in relevant conditions), while not themselves yielding a syntactic violation or requiring reanalysis. In previous studies, otherwise acceptable sentences resulted in mismatches because of the inappropriate combination of lexical and prosodic information, each of which suggested a different syntactic structure (e.g., Steinhauer et al., 1999; Pauker et al., under revision). In those studies, however, the mismatch was always between information presented in the same modality. In the present investigation, the mismatch is between information presented in two distinct modalities.

To illustrate the type of mismatch employed, we can return to the 'one plus two times three' example discussed above. Using similar phrases that contained monosyllabic CVC nouns (e.g., 'bag', 'kite', etc.), we explored the

---

[1] Throughout this paper, we refer to the mismatch between an auditory prosodic boundary and visual grouping as a "prosodic mismatch" for ease of exposition and for consistency between condition labels and analysis. It should be noted, however, that this is more accurately described as an auditory-visual grouping mismatch.

neural response to a phrase produced with a “late” auditory boundary (e.g., ‘bag and bed # and cup) during the simultaneous appearance of a visual grouping corresponding to the phrase produced with an “early” auditory boundary (e.g., BAG || BED CUP, where || represents a visual boundary). If the wrong visual representation is seen when the utterance is spoken, a mismatch is created in which there is no syntactic violation. In this study, we used well-known items (see Figure 1 and Appendix 1) to create auditory and visual phrase groupings to examine the electrophysiological responses to this type of contextually-induced prosodic anomaly. One of the benefits of using this cross-modal design is the ability to not only examine the neural response to prosody, but also to examine prosody in a context, which provides the means to explore the interaction between prosody and visual input. In order to ensure that this novel design indeed tapped into the integration of visual context and auditory processing, it was necessary to include another mismatch, the response to which could be more easily predicted. Semantic information was chosen to be the “predictable mismatch” since semantic information is not only easily picturable, salient, and concrete (e.g., Papafragou, 2003; Syrett, Bradley, Kennedy, & Lidz, 2005) in both modalities, but also has a well-documented neural response (e.g., King & Kutas, 1995; Friedrich & Friederici, 2004). Including the semantic mismatch also allowed us to investigate whether prosodic mismatches are detected in the presence of another type of salient violation. As noted above, the prosodic mismatch exploited here was not a violation per se, but was instead a mismatch between otherwise acceptable input across modalities. Since the phrases employed a coordination of concrete nouns as opposed to arithmetic equations, the presence or absence of phrase boundaries results in a difference in abstract concepts of grouping rather than differences in the result of mathematical operations. That is, the conceptual mismatch caused by conflicting visual and auditory groupings was a subtle one, which affected only a rather abstract level of conceptual representation. We expected that the profile of ERP components would allow us to determine whether the human brain processes these kinds of mismatches primarily as a structural conflict (likely to elicit P600-like effects) or as a conceptual mismatch (typically reflected by N400s), or both.

We hypothesized that participants would successfully detect prosodic mismatches, but with less accuracy and increased variability compared to semantic mismatches because, in contrast to semantic information, a prosodic mismatch might be less discernible to listeners. For the semantic mismatch, we predicted a strong N400 response, consistent with Friedrich and Friederici (2004) who found that an aurally presented noun that did not match a concurrently presented picture elicited an N400 response. For prosodic processing, we hypothesized that visual context would be sufficient to create expectations for (or against) a specific prosodic phrasing pattern (auditory boundary), and that these expectations would be reflected in the ERP patterns (discussed in more detail below in Predictions) as soon as the auditory input either matched or mismatched the visual display, given the literature showing that context influences prosodic processing (Itzhak et al., 2010; Kerkhofs et al., 2007; Schmidt-Kassow & Kotz, 2009). Another research question was whether the auditory boundaries would elicit a CPS, given that they were realized only in terms of pauses in absence of both syllable lengthening and boundary tones. The presence of a CPS response to these stimuli would suggest that *every* prosodic boundary elicits a CPS, even when the prosodic boundary reflects phrasal grouping within a simple conjunction as compared to within a more complex utterance. The absence of a CPS would indicate that this component may either only be elicited if acoustic boundary markers are present that were not used in the present study (e.g., preboundary syllable lengthening and boundary tones), under specific task demands, or may only be elicited by certain types of boundaries. It was also predicted that the conditions containing both semantic and prosodic mismatches (discussed below) would display the additive ERP effects of both types of processing and behaviorally this double mismatch would be (at least) as easy to detect as the semantic mismatch. More detailed predictions are discussed following the Methods (below).

Finally, before describing the design, it is important to discuss potential eye movements in the data. As noted, a primary objective of the present investigation was to explore the interaction of visual context and prosodic information using electroencephalography (EEG). Despite the value of this design, it also creates a potential hazard. When examining brain activity through EEG, eye movement is considered more of a liability than an asset. While EEG provides detailed temporal information about neural activity and processing throughout a trial (Rugg & Coles, 1995) and is therefore a valuable means of investigating language, a major challenge in conducting research examining ERPs to linguistic stimuli is avoiding and/or removing artifacts such as eye blinks and saccades, which obscure or conceal the actual EEG signal. These EEG artifacts are often larger in amplitude than most language-related responses. To limit the loss of data due to eye movement artifacts, researchers design experiments that include fixation points and breaks for blinking between stimuli, control the type and size of the visual stimuli presented, and use paradigms and instructions that reduce ocular artifacts or algorithms that correct (rather than reject) EEG signals contaminated with these artifacts, rendering the majority of these studies less than natural. Nonetheless, psycholinguistic ERP research that includes both auditory and visual stimuli without data contamination are beginning to emerge in the literature (e.g., Friedrich & Friederici, 2004; Knoeferle et al., 2008,) as are methods to use the electro-oculogram (EOG, which is being recorded along with the EEG) as a reflection of eye-tracking (e.g., Joyce, Gorodnitsky, King, & Kutas, 2002; Sereno & Rayner, 2003). To limit eye movement, the present study included carefully-controlled visual stimuli that would allow for the examination of the linguistic processes of interest. In particular, all stimuli were of equal size and prominence, the structure of stimuli was universal across conditions, and participants were instructed to fixate the center of the screen because all input required for the task was available while fixating. However, recognizing that eye-movement contamination was a serious issue, we paid careful attention to this. As will be revisited and further discussed in the Methods section, our innovative methodology allowed us to investigate the correlation between eye movements and visual and auditory input while gaining insight into both prosodic and semantic processing.

Methods

Subjects

All subjects were right-handed (as determined by the Edinburgh Handedness Inventory; Oldfield, 1971) native speakers of English with no history of neurological impairments and with normal or corrected-to-normal vision. To ensure that all participants' hearing was within the range of normal, audiometric screening determined that pure tone averages (averaged across 500, 1000, and 2000 Hz) were less than 25 dB HL in the better ear. Data from 20 young adults (ages 18 to 25 years, mean = 21 years, sd = 1.45 years; 11 female) tested at McGill University were included in the analyses. Four additional subjects were tested but their data were excluded from analysis. One female subject was excluded as a result of equipment malfunction; an additional female subject was excluded after falling asleep during testing. Two male subjects were excluded for excessive movement, resulting in too few trials (less than 10% usable data). Written informed consent was obtained prior to testing and subjects were financially compensated for their time. Measurements of auditory working memory (Lehman & Tompkins, 1998), vocabulary (Boston Naming Test [BNT]; Kaplan, Goodglass, & Weintraub, 1978), and cognition (Montreal Cognitive Assessment [MoCA]; Nasreddine et al., 2005) were collected for each participant to serve as potential covariates for ERP and behavioral data.

Stimuli

Stimuli consisted of visual displays and auditory phrases containing three nouns such as 'bag and dog and bed'. In each trial, participants saw a visual display of three nouns with a physical visual boundary, henceforth referred to as ‘wall’, between either nouns one and two or nouns two and three. The visual display was presented simultaneously with the onset of an auditory phrase (e.g., ‘bag and dog and bed’) which included a pause to create either an early or late auditory boundary.

Pictures and auditory recordings of 16 easily picturable monosyllabic CVC nouns that begin and end with a stop consonant (e.g., ‘bike’, ‘pig’, ‘cat’, etc.), were chosen for the visual displays (see Appendix 1). Visual stimuli were pictures taken from the Snodgrass and Vanderwart Picture Inventory (Snodgrass & Vanderwart, 1980). Pictures were resized and positioned equidistantly inside a white rectangle by a trained graphic artist to ensure that all items were equally prominent. As can be seen in Figure 1, this created visual displays of three equidistant objects with a thick vertical black line (the ‘wall’) that served as a physical boundary at one of two possible positions: the ‘Left’ wall between the first and second objects (which corresponds to an early auditory boundary) or the ‘Right’ wall between the second and third objects (which corresponds to the late auditory boundary). Objects were sized and positioned such that it was possible to fixate the center of the screen and attend to the objects without the need to move the eyes to see all the relevant elements. This was done to avoid horizontal eye movements and corresponding ocular artifacts in the EEG signal. Auditory stimuli consisted of multiple repetitions of all CVC nouns as well as the conjunction ‘and’ pronounced with level prosody at a normal speaking rate by an adult female native speaker of English and recorded in a sound-attenuated booth using a portable digital recorder (Marantz Professional PMD670) and a head-mounted microphone (AKG Acoustics C420) and transferred to a computer using an external cardreader (Macally MCR-6U). The mean duration of all items was calculated and the token of each item closest to the overall mean was selected for inclusion in the experiment. Vocalic and voiceless segments of nouns were removed or replicated as needed using Praat software (Boersma & Weenink, 2006) until the duration of all nouns was equivalent within 1 ms (mean = 376.4 ms; sd = .26 ms). The token of 'and' had a duration of 253 ms. Three naïve listeners judged the quality of each item and determined that all words sounded natural and not manipulated.

The 16 monosyllabic CVC nouns were divided into eight sets of four items, or '4-tuples', such that each noun occurred in two 4-tuples. The schema employed to create the 4-tuples is shown in Appendix 1. (The reasoning behind, and advantages of, using these 4-tuples will be discussed below.) Phrases containing three nouns each, similar to those used in previous studies (e.g., Streeter, 1978; Beach et al., 1996), were created from the 4-tuples. For example, as can be seen in Appendix 1, from 4-tuple A (bag dog cup bed) the triplet 'bag - bed - cup' was generated. Four unique phrases of this kind could be generated from each 4-tuple. Each of these unique phrases could be permuted in six ways (e.g., 'bed and bag and cup', 'cup and bed and bag', 'bag and cup and bed', etc.). The six permutations of the four phrases from the eight 4-tuples generated 192 stimuli (6 x 4 x 8). Pictures were arranged to depict each of the auditory stimuli. To create semantic mismatches (discussed below), the fourth noun, which had not been used in the triplet, replaced the middle noun in the visual display. This schema ensured that replacement nouns in phrase-picture pairs were not highly related semantically or phonologically and that every noun could be replaced by two other nouns (one from each 4-tuple), limiting predictability in semantic violation conditions.

Two phrasal groupings (with boundaries marked by ‘#’) are possible with this phrase type: an Early boundary **[EB]** (‘bag # and dog and bed’) and a Late boundary **[LB]** (‘bag and dog # and bed’). The 192 items were randomly divided into the two main boundary conditions (i.e., 96 phrases were presented with an Early boundary and 96 phrases were presented with a Late boundary). A 450 ms interval of silence served as the auditory boundary, since, as discussed above, a pause has been shown to be a sufficient cue to a phrase boundary (Streeter, 1978; Beach et al., 1996; Nagel et al., 1996). As discussed above, the black ‘wall’ separating two of the objects served as the visual boundary, where a ‘left wall’ corresponded to an early boundary and a ‘right wall’ to a late boundary (see Figure 1 for an example). Thus, in both modalities there was only one cue to the phrasal grouping. The crossing of the two auditory and the two visual grouping conditions led to the first four conditions: two matching ***C****ontrol conditions* (EB + left wall, LB + right wall) and two ***P****rosodic mismatch conditions* (EB + right wall, LB + left wall). Moreover, each of these four conditions was also combined with a semantic mismatch where the second object in the visual modality differed from the second noun in the auditory modality. This resulted in four additional conditions, two pure ***S****emantic mismatch* and two ***D****ouble mismatch* (prosodic plus semantic mismatch) conditions. Since the overall EOG and ERP patterns were most strongly driven by the *auditory* input, most analyses will compare conditions separately for (i) early auditory boundaries (EB) and (ii) late auditory boundaries (LB). The labeling of the eight conditions reflects (a) the position of the auditory boundary (E vs. L) and (b) the type of main condition (C, P, S, D). Thus, the condition with an early auditory boundary and a semantic mismatch will be referred to as ES, while a prosodic mismatch in trials with a late auditory boundary will be referred to as LP, and so on (see Figure 1 for examples of the conditions and Figure 2 for condition labels). Each of the 192 auditory items was presented twice: once in a matched Control condition (EC or LC) and once with a visual display that conflicted with the auditory phrase (mismatches). This yielded a total of 384 trials (192 x 2) presented to each participant.

In the matching Control conditions, visual and auditory tokens matched each other both in grouping and in semantics. Thus subjects heard 'bag and dog # and bed' and saw BAG DOG || BED. In the EP and LP conditions, visual stimuli differed from auditory stimuli in phrasal grouping only. For example, in the LP condition, the Early boundary (left wall) visual stimulus was paired with the Late boundary auditory stimulus such that subjects heard 'bag and dog # and bed' but saw BAG || DOG BED). In the ES and LS conditions, visual and auditory stimuli shared the same phrasal boundary, but the *second* noun differed between the modalities. For example, subjects heard 'bag and dog # and bed' and saw BAG CUP || BED. In the ED and LD conditions, visual and auditory stimuli differed both in phrasal grouping and the second noun. For example, in the LD condition, subjects heard 'bag and dog # and bed' and saw BAG || CUP BED. The 4-tuples used to create phrases ensured that each noun appeared an equal number of times in every position in the phrase across the experiment and was replaced by or replaced another noun an equal number of times. This minimized predictability for participants.[2]

The three mismatch conditions in each boundary condition were comprised of 32 trials each. Since half the stimuli were presented in each boundary condition and half were control items, for each trial there was an equal probability of being presented with either boundary or a correct or incorrect item, thus minimizing predictability and eliminating the need for filler items. Each auditory token was paired with four visual tokens. Each participant heard every auditory token twice, once in the Control condition and once in a mismatch condition. The same auditory token was not presented in the same half of the experiment. From these stimuli, three lists were created such that every control token was presented along with only one of the mismatch conditions; as noted, each subject only heard each phrase twice in the experiment, divided across experimental halves. Each phrase was presented six times across the three lists, resulting in 1152 stimuli overall, but only 384 stimuli per list or per participant.

---

[2] While this schema may have allowed for statistical learning, this did not create a problem for our design because our focus was to ensure that differences between ERP responses would not be attenuated due to predictability of a semantic violation. Any statistical learning that may have resulted from the design is orthogonal to the aims of the present study and there is no reason to believe it would modulate a response to prosodic or semantic input.

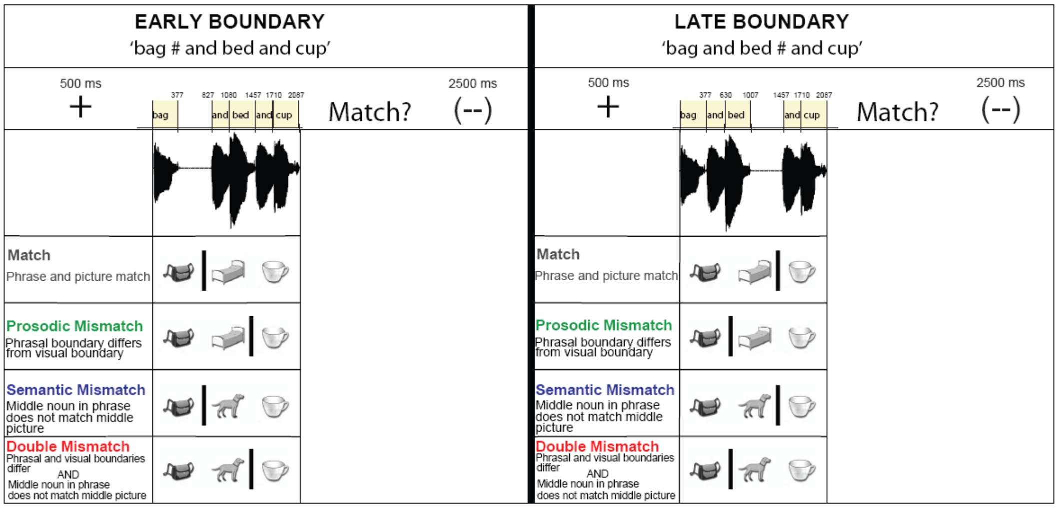

Figure 1. Trial timing and Conditions. Early and Late boundary stimuli (in ms) shown above the auditory input. Trials began with a fixation cross presented for 500 ms followed by auditory and visual stimuli with a simultaneous onset. Twenty ms after the offset of stimuli, the question 'Match?' appeared until participants responded. Participants were then able to blink for 2500 ms until the onset of the subsequent fixation cross. Examples of the visual stimuli presented for each condition are shown beneath the speech waveform.

Procedure

Following the hearing assessment, cognitive and memory testing, and electrode placement, participants were seated in front of a monitor in a sound attenuating chamber, wearing insert earphones (Etymotic Research model ER3-14A). Participants were shown the effects of eye blinks and movement on the EEG signal and were made aware that eye movement would create unwanted artifacts in the data. They were asked to blink and move only when indicated in the experiment and to fixate the center of the screen during stimuli presentation. Participants were told they would see pictures and hear phrases and were asked to indicate whether the two pieces of information presented matched or not.

An initial training session began with the individual aural and visual presentation of each of the 16 nouns. Participants were shown each picture a second time and asked to name each one to ensure they recognized each item. In the rare case that the wrong name was given, the experimenter corrected the participant, who repeated the correct name. The participants were then shown the white rectangle, identified as a shelf, followed by the two ways to divide the shelf with a wall. Participants then heard a phrase and were asked to choose which of two pictures best matched that phrase. If a participant responded incorrectly to any of these items, the experimenter would repeat the phrase, which alerted participants to semantics and phrasal groupings. Finally, one typical trial from the Prosodic Mismatch condition was presented. The phrasal grouping was made explicit to listeners and the Prosodic Mismatch trial was included after pilot subjects who were not specifically made aware of the contrast failed to correctly reject any prosodic mismatches. Participants were not told that Double Mismatch trials could occur. Following this training session, participants saw 30 practice trials to become familiar with the task. No feedback related to accuracy was given once the 30 practice trials began.

The ERP experiment session consisted of one list of 384 items divided into four blocks of 96 items, lasting about 12 minutes each. Two randomization schemas were created for each list and each list and randomization were assigned evenly across groups and male/female participants. Hand assignment was also counterbalanced.

Trial and stimulus timing is presented in Figure 1. Trials began with a fixation cross ("+") presented in the center of the computer screen for 500 ms. Following this, the visual and auditory stimuli presentation began. The picture remained on the screen for the duration of the auditory stimulus (2087 ms). Twenty ms after stimulus offset, the question "Match?" appeared in the center of the screen. Participants had no time limit in which to indicate whether the information from the two modalities was the same or not. After responding, participants were shown the following symbol, which indicated the 2500 ms interval in which they were instructed to move and blink: (--).

Electrophysiological Recording

EEG was continuously recorded (DC mode; 500 Hz sampling rate; Neuroscan Synamps2 amplifier, Neuroscan-Compumedics, Charlotte, North Carolina, USA) from 32 cap-mounted Ag/AgCl electrodes (Electro-Cap International, Eaton, Ohio, USA) referenced to the right mastoid and arranged according to the extended 10-20 system (Jasper, 1958) (impedance <5 kΩ). To measure blinks and eye movements, vertical electro-oculogram (VEOG; from electrodes placed above and below the right eye) and horizontal EOG (HEOG; from electrodes placed at the outer canthus of each eye) were recorded.

Data Analysis

Before beginning EEG data analysis, we addressed the issue of eye movements because it was crucial to ensure that ERP data were not contaminated by these artifacts. Following data collection, EOG activity was examined to see if any confounding artifacts were present. When examining data from individual subjects, a robust and systematic pattern of horizontal eye movement (saccades) was observed in the HEOG electrodes in 13 participants, hereafter referred to as the group of ‘movers’ (represented by solid EOG and ERP lines in Figure 2). The remaining participants (n = 7), hereafter referred to as the group of ‘nonmovers’, did not show movement in these electrodes (represented by dashed EOG and ERP lines in Figure 2). As can be seen in the EEG from the HEOG electrode examples provided in Figure 2 (top), movers’ and nonmovers’ HEOG waveforms were highly distinguishable. On finding that eye movements in the movers group followed exactly the same pattern across individuals, we investigated whether these eye movements reflected language processing differences between groups and could therefore provide more insight than examining ERPs alone. Participants were therefore divided into groups according to whether or not they moved their eyes and these groups were compared to each other behaviorally and electrophysiologically[3].

EEG data were analyzed using EEProbe (ANT, The Netherlands). Single subject averages were computed separately for all conditions following data pre-processing which consisted of filtering (0.16-30 Hz bandpass) and artifact rejection at electrodes Fp1, Fp2, and VEOG (but not HEOG). Averages were based on correct trials only and were computed for 2200 ms epochs beginning at the onset of noun1 of each phrase and ending roughly 100 ms after the offset of noun3. A 200 ms baseline began at the onset of noun1 and ended 200 ms later, 177 ms before the end of noun1. This baseline was chosen because conditions did not differ auditorily within the first 200 ms and thus there was no information to distinguish between conditions. Within this epoch, ERP components and eye movements (from HEOG) were quantified by means of amplitude averages in a series of time windows (TW), detailed below. Figure 2 displays detailed timing of all events in the auditory stimulus as well as the time windows analyzed (labelled 1 – 6 for EOG and A – D for ERPs). Because auditory stimuli were

---

[3] No group differences emerged from t-tests examining measures of working memory, vocabulary or cognitive abilities.

well-controlled and all trials were of equal length, we present the entire epoch (at electrode PZ) or HEOG in all figures.

Eye movements from the HEOG electrode will be presented first. These were analyzed in the following time windows (TWs): initial left saccades from (1) 200 – 300 ms and (2) 300 – 400 ms, early versus late boundary saccades from (3) 400 – 600 ms and (4) 600 – 800 ms, and semantic effects on saccades from (5) 1200 – 1500 ms and (6) 1500 – 1800 ms. Repeated measures ANOVAs were run for the HEOG electrode with the factors **Boundary** (2: Early, Late pause [signaling an auditory prosodic boundary]) x **Wall** (2: Left, Right visual wall [reflecting the position of the visual cue to grouping]) x **Semantics** (2: match, mismatch [of the center noun]) x **Time Window** (6: TWs 1 – 6 above) x **Group** (2: movers, nonmovers). Significant group differences (see below) motivated a detailed exploration to determine the eye movement differences between groups and time windows, as well as whether these differences revealed anything about language processing. Since only the movers showed a main effect of TW, and since this group was, by definition, the only group that showed large eye movements and significant EOG effects (confirmed by significant EOG x Group interactions, reported below), only the findings for the movers, analyzed within each TW, are reported in the eye movements' analyses. When an interaction with Boundary was present, separate ANOVAs were conducted for each boundary with factors Wall and Semantics. It is important to note that in the global ANOVA, a main effect of Wall reflects the position of the visual wall marking the boundary, whereas a Boundary x Wall interaction reflects the auditory-visual grouping mismatch (the prosodic mismatch, see Footnote 1). For an ANOVA within each auditory boundary condition (i.e., in follow-up analyses of eye movement and in ERP analyses), a main effect of Wall reflects the auditory-visual grouping mismatch (prosodic mismatch).

For the ERP effects, following visual inspection, ERPs were examined at midline[4] (Fz/FCz/Cz/CPz/Pz/Oz) electrodes in four TWs: (A) 600 – 800 ms, (B) 1000 – 1200 ms, (C) 1400 – 1650 ms, and (D) 1730 – 1930 ms. Each boundary main condition was analyzed in separate ANOVAs (so that the factor Boundary was not included in these analyses) because effects were elicited in different TWs due to differences in timing of events between boundary conditions. Repeated measures ANOVAs were run within each boundary and included the topographical factor **Anterior/Posterior** (AP; 6). All interactions between distributional factors and Wall were followed up via separate ANOVAs for Wall in each region or electrode. Interactions between Semantics and distributional factors were not decomposed further as these conditions were intended primarily as control conditions and the distributional properties of these components (i.e., the semantic N400 and P600) were not of primary interest for the questions addressed in this study. For all ERP and EOG analyses, Greenhouse-Geisser corrections were employed where applicable.

Finally, behavioral data were analyzed for accuracy and reaction time for correct trials only in a **Boundary** x **Wall** x **Semantics** x **Group** ANOVA to

---

[4] Only midline electrodes were examined to ensure that eye movements were *not* included in analysis. Horizontal eye movements typically affect the EEG at lateral electrodes (especially at frontal sites) whereas midline electrodes are less susceptible to such artifacts (see also Bornkessel & Schlesewsky, 2010). In particular, signals at the parietal midline electrode, Pz, where the majority of the ERP effects were found to be most prominent, did not seem to be influenced by eye movements at all.

maintain consistency between analyses of eye movement, ERPs and behavioral data.

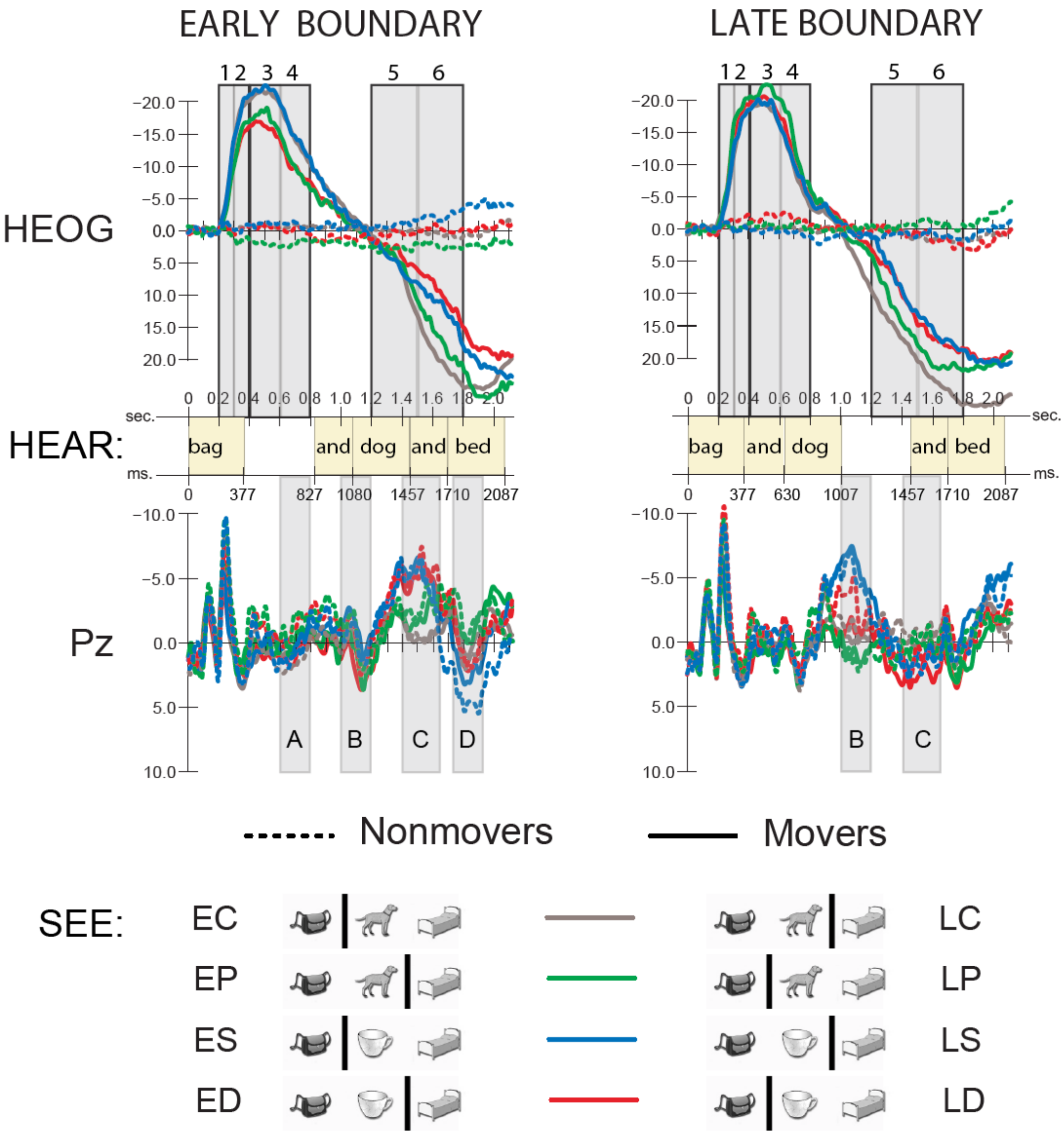

Figure 2. HEOG and ERP for both groups in both conditions. The HEOG electrode (top) and ERP from the PZ electrode (center) time-locked to the onset of the auditory stimulus shown for nonmovers (dashed) and movers (solid). TWs analyzed for EOG (1 – 6 shaded) and ERPs (A – D shaded) are highlighted. Examples of visual information are shown (bottom).

Predictions

The primary questions addressed by the present research are the following:

(i) Is a pause alone sufficient to cue grouping information?

(ii) How will visual context interact with auditory processing?

(iii) Will an utterance that differs from a visual display in prosodic grouping alone elicit a mismatch response? If so, what neural correlates to the prosodic mismatch will emerge?

(iv) Will the processing of grouping information/'prosodic mismatches' be influenced by the co-occurrence of semantic mismatches?

We predicted that the 'wall' in the visual display and a pause in the speech signal would be sufficient to influence the processing of grouping information and elicit a mismatch response. Analyses of the eight conditions in our study were conducted on eye movements recorded from HEOG in a total of six time windows, on electrophysiological responses to both semantics and prosody for both boundary conditions in four time windows, and on behavioral responses (both for accuracy and response time). To aid the navigation of this large quantity of data, more detailed predictions are presented here. We begin with predictions for ERP responses to Semantic, Prosodic and Double mismatches and then discuss predictions for behavioral accuracy and reaction time responses.

For ERP responses to semantic mismatches in both boundary conditions, an N400 was expected approximately 400 ms after the onset of the semantically incongruous word in both the Semantic and Double Mismatch conditions in both boundary conditions, consistent with Friedrich and Friederici (2004). Due to the different timing of events between auditory boundary conditions, this component should occur in different TWs for the Early and Late boundaries.

As to the ERP response for prosodic boundaries, a CPS was predicted for each prosodic boundary, or pause, in the auditory modality, similar to those found in previous investigations (e.g., Steinhauer et al., 1999; Pauker et al., in revision);

these components should also emerge in different TWs for each boundary condition since the onset of the pause differed between the Early and Late boundary conditions. Consistent with findings for other types of stimuli (e.g., Pannekamp et al., 2005), the possibility that the CPS would show a topographical or distributional difference as a result of the difference in acoustic cues, grammatical structure and/or the presence of visual input employed here compared to those employed in previous studies was considered, but there was no basis on which to predict what type of change, if any, might appear. The presence or absence of a CPS will be interpreted in line with what was discussed in the Introduction.

A main goal of the present study was to determine ERP responses to cross-modal auditory-visual grouping mismatches (prosodic mismatches). As this is the first study of its kind, it was less straightforward to predict the response to this type of mismatch compared to semantic mismatches. We expected that the visual input would be sufficient to create expectations for or against an auditory boundary, and that these expectations would be reflected in the ERPs by one of three components (or a combination thereof): (1) a CPS, (2) a P600 and/or P800, or (3) an N400. It was possible that a late positive component, such as the P600 (or possibly the P800), would be elicited by prosodic mismatches in response to a reanalysis of the acoustic signal in relation to the visual signal, consistent with Schmidt-Kassow and Kotz (2009) and Knoeferle et al. (2008). Another possibility was that the prosodic mismatch would elicit an N400 component that would reflect the conceptual mismatch between auditory and visual input and/or the integration of unexpected/anomalous auditory information into the visual context. This last finding would be quite novel as it would clearly differ from most previous N400 findings and could not be accounted for in terms of lexical retrieval (e.g., Lau et al., 2008).

Finally, two alternative predictions were possible for the conditions containing both semantic and prosodic mismatches (Double Mismatch). It was possible that the Double Mismatch would display the additive effects of both types of processing. This would indicate that prosodic and semantic processing recruit distinct neural mechanisms. When both mismatches co-occur, the system would process them (and elicit each ERP effect) independently. Intriguingly, prosodic mismatches could – in principle – become evident prior to semantic mismatches in both conditions. If the Double Mismatch conditions show additive effects of both types of processing, it further indicates that neural processing continues even when it is already clear that there is a mismatch and further processing is not strictly required for successful completion of the task. Alternatively, it is possible that the Double Mismatch condition would show either only one type of processing (perhaps whichever piece of information is encountered first [prosody] or whichever piece of information is more salient and reliable [semantics]) or an interaction between the two types of information, indicating that shared neural mechanisms are recruited for processing both information types.

For behavioral responses analyzed for accuracy, we hypothesized that participants would successfully detect matches and mismatches. More specifically, it was predicted that for co-operating auditory and visual input (Control/match), correct responses would approach ceiling. Similarly, for conditions in which semantic information conflicted between the auditory and visual modalities (Semantic and Double Mismatches), correct ‘mismatch’ responses should approach ceiling since the type of conceptual-semantic mismatch employed was highly salient (e.g., Papafragou, 2003; Syrett et al., 2005). However, it was also predicted that there would be increased variability and less overall accuracy for Prosodic Mismatches because, in comparison to tangible and definitive semantic mismatches, a prosodic mismatch might be less discernible to listeners. Finally, we examined whether an Early versus a Late boundary might influence accuracy because the pause that cues the boundary might be easier to detect later in the utterance (as in the Late boundary condition) compared to earlier in the utterance (this interpretation will be discussed further below).

For response times, no large differences between conditions were predicted because participants were instructed to refrain from responding until presented with a response prompt (delayed response) to avoid motor artifacts in the EEG data. Any differences that did emerge between the conditions should parallel differences observed for accuracy. Specifically, Control, Semantic Mismatch, and Double Mismatch conditions in both boundary conditions should be responded to faster than Prosodic Mismatch conditions because the prosodic mismatch alone would be harder to detect than matches or semantic mismatches. No substantive differences between boundary conditions were predicted.

Results

Eye Movement

Comparison between groups. The Boundary x Wall x Semantics x TW x Group ANOVA, prompted by the robust pattern of eye movement that emerged in many participants (see Data Analysis, above, for description and Figure 2 for examples of group differences in eye movement), revealed significant main effects of both Semantics (F[1,18] = 5.84, p = .0265) and Wall (F[1,18] = 12.78, p = .0022). These main effects were further qualified by multiple interactions with TW and Group (see below). Moreover, the omnibus ANOVA also revealed highly significant differences between time windows (TW F[5,90] = 22.86, p < .0001) and a TW x Group interaction (F[5,90] = 20.81, p < .0001). Follow-up analyses revealed that the TW effect was highly significant in the movers (F[5,60] = 42.37, p < .0001) but failed to reach significance in the nonmovers (F < 1). This supports differences in eye movement at different points in the trial in the movers but not in the nonmovers. Moreover, TW interacted with every linguistic factor (Boundary x TW F[5,90] = 7.22, p = .0028; Semantics x TW F[5,90] = 6.63, p = .0036; Wall x TW F[5,90] = 3.63, p = .0289). Finally, every interaction between TW and an experimental variable further interacted with Group (Boundary x TW x Group F[5,90] = 4.55, p = .0190; Semantics x TW x Group F[5,90] = 6.50, p = .0039; Wall x TW x Group F[5,90] = 3.44, p = .0346; Boundary x Wall x TW x Group F[5,90] = 3.63, p = .0273; Boundary x Wall x Semantics x Group F[1,18] = 5.67, p = .0286). Once again, these interactions were highly significant in the movers (Boundary x TW F[5,60] = 12.12, p = .0003; Semantics x TW F[5,60] = 13.10, p = .0002; Wall x TW F[5,60] = 6.97, p

= .0029; Boundary x Wall x TW F[5,60] = 5.87, p = .0040; Boundary x Wall x Semantics F[1,12] = 5.40, p = .0386) but did not approach significance in the nonmovers (Boundary x TW, F < 1; Semantics x TW, F < 1; TW x Wall F[5,30] = 1.26, p = .3185; Boundary x Wall x TW, F < 1; Boundary x Wall x Semantics F[1,6] = 2.43, p = .1699). The "functional" interpretation of these effects (in terms of eye movement patterns) will be discussed below. These analyses confirm robust differences in eye movement between the groups, as illustrated in Figure 2. The interactions with the condition variables were intriguing; therefore we decided to explore in more detail the patterns of eye movement (for the movers only, since this group alone showed statistically significant movement) and their correspondence with the semantic and prosodic manipulations.

Eye movement pattern in movers. Figure 2 shows the HEOG electrode for both movers and nonmovers. Negative sloping waveforms (negativity plotted upwards) indicate leftward movement and positive sloping waveforms (positivity plotted downwards) indicate rightward movement. Based on visual inspection of the waveforms, the overall pattern of eye movement that can be seen in the movers contains four phases. At the start of the trial, movers fixated the center of the screen where the fixation point was presented. After trial onset, movers looked leftwards towards the first pictured noun. Prosodic boundaries, and whether these boundaries matched or mismatched the visual context, appeared to modulate eye movements. After hearing the second noun, movers fixated the center item and rested there longer if there was a mismatch between auditory and visual information. After this fixation, the saccade towards the third object was reflected by a large positive-going waveform. The first three (most relevant) effects will be described in detail below.

Table 1. Statistical analyses of eye movements from HEOG in movers

	df	200 - 300	300 – 400	400 - 600	600 – 800	1200 – 1500	1500-1800
Boundary	*1,12*	—	—	—	—	13.87**	8.76**
Semantics	*1,12*	—	—	—	—	9.20**	17.84**
Wall	*1,12*	14.02**	19.21**	15.27**	20.59**	—	—
Bound x Sem	*1,12*	—	—	—	—	5.92**	—
Bound x Wall	*1,12*	—	—	9.48**	—	—	5.16**

* p < 0.05
** p < 0.0001

Initial left saccades. Figure 3 shows the HEOG for the movers for an early (left) wall compared to a late (right) visual wall. Across both boundary conditions, within 200 ms after the onset of the visual and auditory stimuli, movers showed a steep negative-going slope, indicating leftward eye movement, for all conditions. Amplitudes were higher at about 250 ms in conditions that contained a left visual wall (i.e., ES, EC, LP, and LD), indicating that participants looked towards the first pictured noun in the utterance and looked further to the left when there was a left wall (yellow line in Figure 3) than when there was no left wall (purple line in Figure 3), though it is not possible to determine the exact fixation point based on the present data. Table 1 reports the statistical analyses of eye movements for the movers and shows significant main effects of Wall found in both the 200 – 300 and 300 – 400 ms TWs, revealing that the effect continued through 400 ms. Importantly, noun1 (which was identical across *all* 8 conditions) ends 377 ms after the onset of the stimulus which means that, prior to 377 ms, the conditions only differ in the presence or absence of a wall. Therefore, these leftward looks are related to the presence of the wall alone and not to differences in the auditory stimuli.

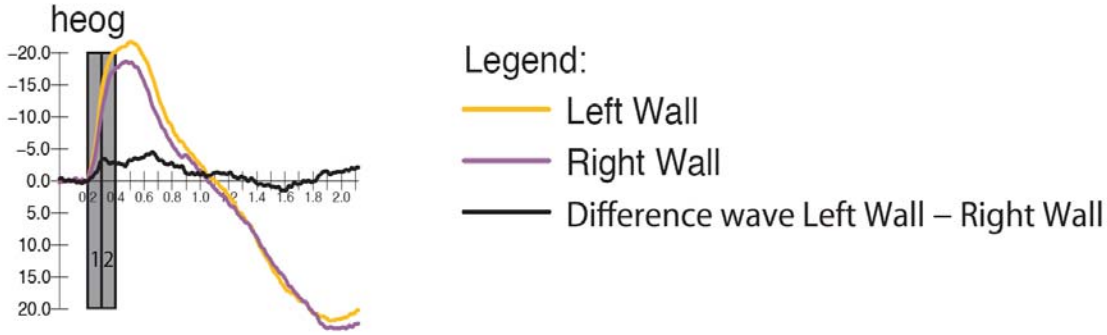

Figure 3. Initial left saccades. Grand average EEG at the HEOG electrode for conditions with a left wall (EC, ES, LP, LD), shown in yellow, compared to a right wall (LC, LS, EP, ED), shown in purple. Both waveforms display large negativities starting at 250 ms, shaded in gray, indicating looks to the left at trial onset. The difference wave (left wall – right wall), shown in black, displays a large negativity indicating larger negativities when a left wall was present compared to a right wall.

Early versus Late boundary saccades. After initially looking towards the first pictured noun, movers initiated rightward saccades to fixate subsequent pictures. By 500 ms (which corresponds to the pause in the Early boundary condition and to the first instance of 'and' in the Late boundary condition, see Figure 1), in all conditions, movers showed positive-going waveforms in the HEOG, indicating rightward eye movement. Since the Late boundary conditions had no early pause ("and noun2" followed immediately after noun1), movements to the right occurred earlier than in the Early boundary conditions. This shows that eye movements were largely guided by the speech stream/prosody, to be discussed in detail in the Discussion. Differences in the positive-going waveform/rightward eye movement were observed both between boundary conditions and between subconditions (see Table 1). Figure 4 displays EOG and difference waves for Early and Late prosodic matches versus mismatches. There is a carry-over effect of the further-left looks (i.e., the initial Wall effects discussed above); specifically, it takes participants longer to fixate noun2 when the initial saccade is further left. This carry-over effect is more persistent in the Early boundary conditions where the rightward saccades (looks to noun2) are initiated even later because the early auditory pause delays these looks. These patterns were confirmed statistically in the 400 – 600 ms and 600 – 800 ms TWs at the HEOG electrode by a significant main effect of Wall and a significant Boundary x Wall interaction. The Boundary x Wall interaction observed in each TW was further examined via separate ANOVAs for each boundary (EB and LB, respectively). The main effect of Wall was significant in the Early boundary conditions in both TWs (400 – 600: F[1,12] = 26.47, p = .0002; 600 – 800: F[1,12] = 18.61, p = .0010). However, in the Late boundary conditions, in the 400 – 600 ms TW, the main effect of Wall did not reach significance (F[1,12] = 2.66, p = .1286). By the 600 – 800 ms TW, it had reached significance, though the effect was never as large as in the Early boundary condition (F[1,12] = 7.76, p = .0165). Figure 4 also illustrates that the HEOG crosses the 0 μV baseline (representing the centre of the screen) approximately 200 ms earlier in the LB conditions (dotted lines) than in the EB conditions (solid lines) where noun2 was presented only after the pause. Taken together, these findings strongly support the notion that the scanning of the visual display was largely synchronized with the pace and grouping of words in the speech stream. That is, visual objects were fixated as soon as they were referred to in the auditory modality, and saccades were already initiated during the pause (otherwise the delay in the crossing of the baseline in EB should have lasted the full pause duration, i.e., 450 rather than 200 ms).

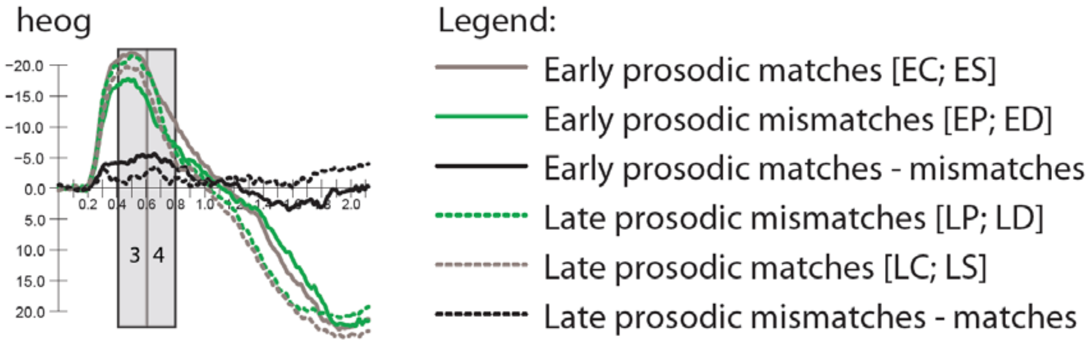

Figure 4. Saccades for Early versus Late auditory boundaries. Prosodic matches (gray) and prosodic mismatches (green) for both boundary conditions (Early = solid; Late = dashed) all display positive-going waveforms indicating rightward movement following the offset of noun1. The difference wave for the Early boundary (solid black line) indicates longer fixations when both a pause and wall are present (prosodic matches) compared to when only a pause is present (prosodic mismatches). The difference wave for the Late boundary (dashed black line) illustrates the same effect for the Late boundary, though it begins later and is not as robust as in the Early boundary.

Semantic effects on saccades. Thus far in the trial, in all conditions, eye movements appeared to be primarily driven by visual and auditory boundaries, based on the morphology of the EOG waveforms. Following the onset of noun2, eye movements appeared to shift to a primarily semantics-driven pattern (illustrated in Figure 5). As mentioned above, HEOG curves in all conditions reached the baseline (center of the screen) once the second (center) noun was heard in the speech signal. As noun2 was heard at 630 ms (offset at 1007 ms) in the LB condition, but only at 1080ms (offset at 1457 ms) in the EB condition (see Figure 1), the crossing point occurred earlier in the LB conditions (around 1000 ms) than in the EB conditions (around 1200 ms). At this point, a semantic mismatch between speech signal and visual display could be detected in Semantic and Double mismatch conditions, which began to pattern together, as did Prosodic mismatch and Control conditions. Interestingly, in both boundary conditions, these semantic mismatches led to very systematically prolonged fixation times (or delayed saccades towards object 3), as illustrated in Figure 5. This delay lasted for several hundred milliseconds in both EB and LB conditions. The latency difference of this effect between LB conditions (peaking around 1300 ms) and EB conditions (peaking around 1700 ms) reflects almost exactly the difference in onset latencies for noun2 in the speech signals (450 ms). The relationship between these semantic effects on *eye movements* and corresponding *neurocognitive ERP measures* of semantic integration will be addressed below. Analyses of the HEOG electrode at TWs 1200 – 1500 ms and 1500 – 1800 ms (see Table 1) showed semantic fixation differences between the boundary conditions (main effect of Boundary), the increased fixation duration of the center pictured noun when there was a mismatch with the auditory stimulus (main effect of Semantics), and the differences in the timing of these effects between boundary conditions (Boundary x Semantics interaction in the 1200 – 1500 ms TW). Follow-up ANOVAs of the Boundary x Semantics interaction in the 1200 – 1500 ms TW revealed a significant main effect of Semantics in the Late boundary condition only (Early: $F[1,12] = 2.07$, $p = .1762$; Late: $F[1,12] = 8.45$, $p = .0132$). While not related to semantic effects, it should be noted that the Boundary x Wall interaction in the 1500 – 1800 ms TW (which, as discussed above, reflects a main effect of prosodic mismatch), did not reveal any significant main effect of Wall in follow-up ANOVAs. Examining Figure 2 reveals a less positive HEOG waveform for prosodic mismatch conditions compared to their controls, indicating either leftward movement or a delay in initiating the saccade towards the third object in the display. Since these data do not provide exact fixation points, follow up examinations are required in order to fully understand this effect.

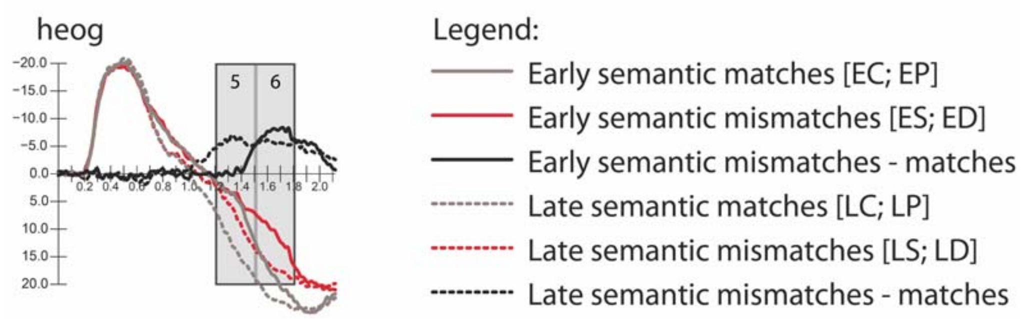

Figure 5. Semantic effects on saccades. All conditions plateau following the onset of noun2, indicating fixation on the center picture. This occurs first in the Late boundary conditions (dashed) because noun2 is heard sooner than it is in the Early boundary conditions (solid). Delayed positivities (rightward movement) in semantic mismatch conditions (red) indicate that participants fixated the center object longer than in match conditions (gray). Difference waves (black) highlight these effects.

ERP Results

After confirming robust group differences in eye movements and finding that those individuals who moved their eyes did so in a systematic way that correlated with both the auditory and visual stimuli (as well as their integration/interaction), ERPs were investigated separately within each group. Figure 6 shows the neural responses to all 4 conditions for the Early and Late boundaries in both groups (i.e., EC, EP, ES, ED, LC, LP, LS, LD). Figure 7 through Figure 13 illustrate the relevant ERP effects, as described below.

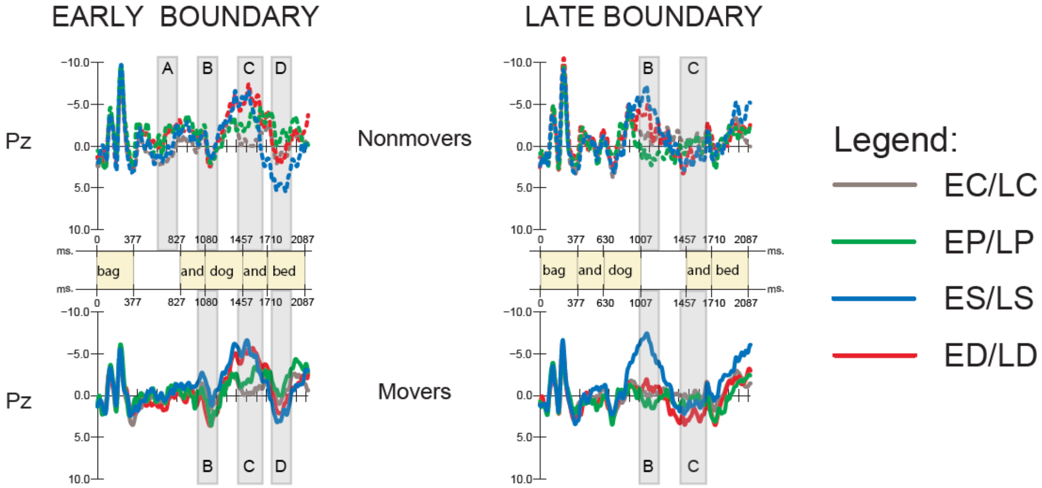

Figure 6. ERPs at electrode PZ. All 4 conditions in both boundary conditions for the nonmovers (dashed, top) and the movers (solid, bottom) for the Early boundary (left) and the Late boundary (right).

In all conditions and in both groups, the onset of the auditory stimulus elicited the typical and expected N1/P2 sequence of components associated with word onsets. Note, however, that due to the concurrent onset of auditory and visual stimuli, the onset components elicited by auditory stimuli were superimposed by an additional N1/P2 pattern evoked by the visual stimuli. Similar (auditory) onset components (although of smaller amplitude) can be seen in each boundary condition elicited by the conjunction 'and' following the respective pause. In addition to these components, semantic and prosodic mismatches elicited various neural responses which are described in turn in what follows.

Table 2. Statistically significant ERP effects in the Early boundary condition

MIDLINE	***df***	600 - 800	1000 – 1200	1400 - 1650	1730 – 1930
Wall	*1,18*	——	——	——	6.71*
Wall x Group	*1,18*	5.10*	——	——	——
Semantics	*1,18*	——	——	24.99**	——
Sem x AP	*5,90*	——	——	4.46*	11.50*

Note: No other effects or interactions reached significance
* p < 0.05
** p < 0.0001

Table 3. Statistically significant ERP effects in the Late boundary condition

MIDLINE	***df***	1000 - 1200	1400 – 1650
Wall	*1,18*	13.13*	——
Semantics	*1,18*	40.55**	——
Sem x Wall x Group	*1,18*	4.74*	——
Sem x AP	*5,90*	11.63*	8.66*

Note: No other effects or interactions reached significance
* p < 0.05
** p < 0.0001

Semantics. For semantic mismatches between auditory and visual input in both boundary conditions, we expected the second noun that did not match the item pictured to elicit an N400. The presence or absence of eye movement was not expected to modulate this response, because even movers had already initiated their saccades towards the corresponding second object during the pause (see above). Figure 7 and Figure 8 illustrate this component in the Early and Late boundary conditions, respectively, using difference waves. The N400 can also be seen in ERP waveforms for all conditions (that is, not difference waves) for both groups in Figure 6.

In addition to the expected N400 response to the semantic mismatch, as illustrated in Figure 7 and Figure 8, a P600 also emerged in both boundary conditions in both groups, suggesting integration and reanalysis of auditory and visual information by participants.

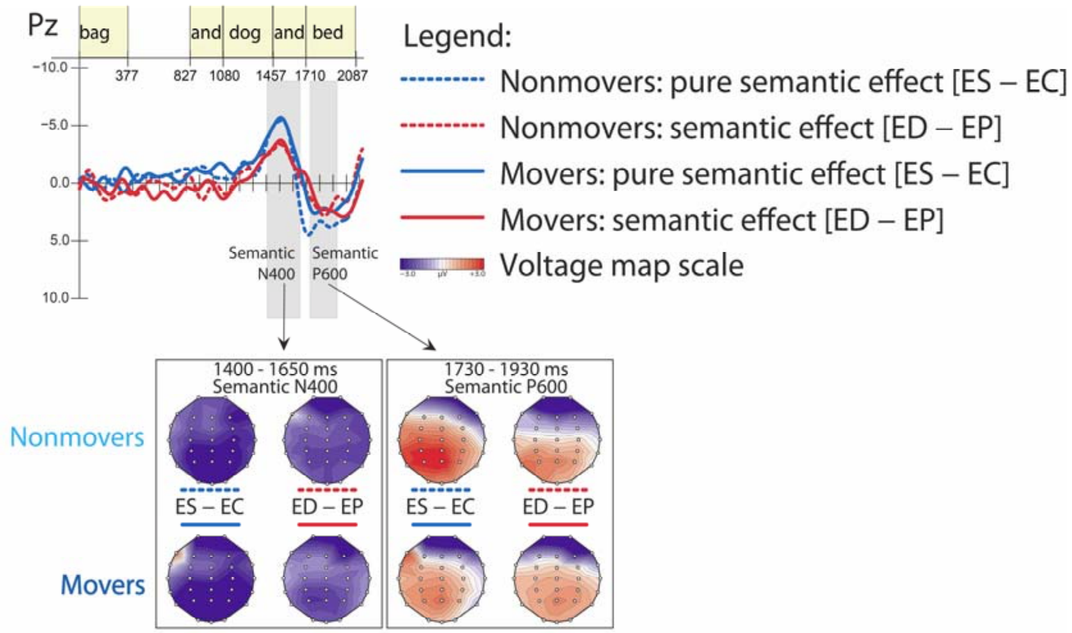

Figure 7. Early boundary semantic effects on ERPs (low pass filtered at 7 Hz) using difference waves for nonmovers and movers. Semantic mismatches elicited N400 and P600 components (shaded) in both ES and ED conditions. Difference waves of semantic mismatch conditions with the corresponding semantic match control condition subtracted illustrate the pure semantic effects in the absence of prosodic effects. The N400 was larger in pure semantic conditions (blue) compared to double mismatches (red). Voltage maps confirm the typical N400 and P600 distributions expected. There were no significant differences between groups.

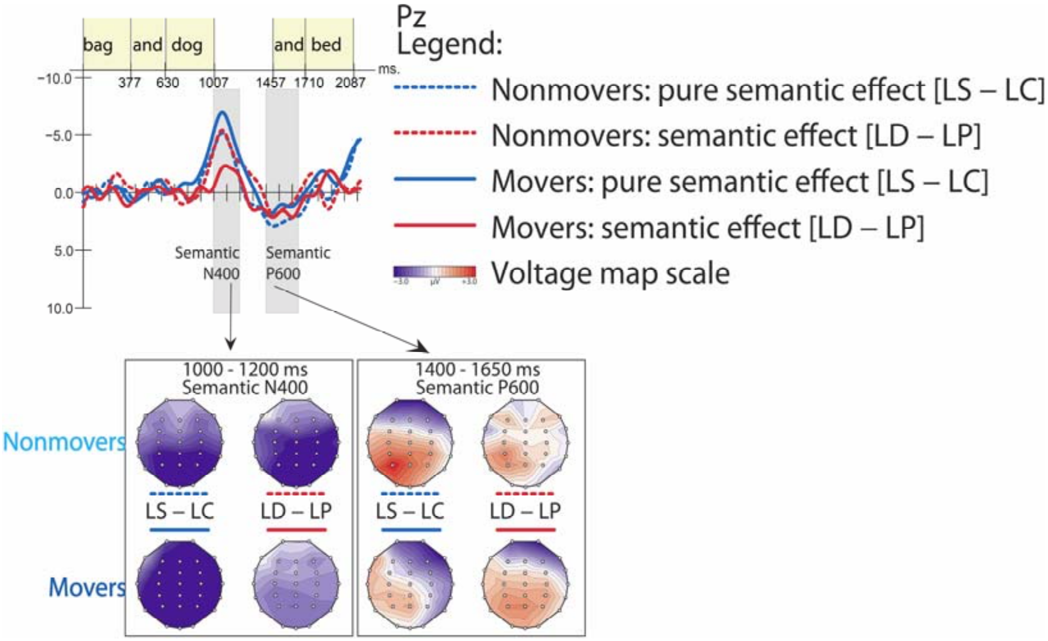

Figure 8. Late boundary semantic effects on ERPs (low pass filtered at 7 Hz) using difference waves for nonmovers and movers. Difference waves of semantic mismatch conditions with the corresponding match control condition subtracted illustrate pure semantic effects. Semantic mismatches elicited an N400 component in the 1000 – 1200 ms TW (shaded) in LS conditions (blue) in nonmovers (dashed) and movers (solid). The LD condition (red) elicited an N400 component in nonmovers but was significantly reduced in movers. Voltage maps confirm this N400 pattern. All conditions elicited a P600 component in both groups. Voltage maps confirm the typical centro-parietal P600 distribution expected.

Semantic N400. In the Early boundary condition, noun2 began at 1080 ms. Both groups showed an increased N400 component in the ES and ED conditions (relative to the control conditions) that began at roughly 1200 ms in the nonmovers and 1230 ms in the movers and peaked roughly 420 ms after the onset of noun2 in both groups. This was confirmed statistically in the 1400 – 1650 ms TW (see Table 2) by a highly significant main effect of Semantics and a Semantics x AP interaction, pointing to a centro-parietal distribution. Voltage maps in Figure 7 show that this effect has the same broad right-lateralized centro-parietal distribution that is typically associated with an N400 (e.g., Kutas & Federmeier, 2000).

In the Late boundary condition, noun2 began at 630 ms and the semantic N400 can be seen in TW 1000 – 1200 ms. In the nonmovers, both the LS and LD conditions showed a negativity that began at 930 ms, continued until 1200 ms and peaked roughly 460 ms after the onset of noun2. In the movers, there was a negativity in the LS condition only that began at 870 ms and continued until 1330 ms with a peak latency of roughly 450 ms after the onset of noun2. The LD condition will be discussed below. These N400 effects in the semantic condition, illustrated in Figure 8, were confirmed statistically by a highly significant main effect of Semantics and a Semantics x AP interaction (see Table 3). Voltage maps in Figure 8 again confirm the typical broad, right-lateralized centro-parietal N400 distribution (e.g., Kutas & Federmeier, 2000), as expected.

Interestingly, Figure 8 (and ERP waveforms in Figure 6) shows that the LD condition had a somewhat different morphology than the LS condition in both groups, though this morphology differed between the groups. In the nonmovers, while the N400 of the LD condition followed the same morphology as the LS condition, it was smaller in amplitude (Figure 6). When these waveforms are examined through difference waves in Figure 8, however, it is evident that the amplitude of the semantic N400 *effect* is consistent across conditions, irrespective of the presence of additional prosodic mismatches. Thus, the lower N400 amplitude for the LD (versus the LS) condition can be attributed to a superimposing positivity related to prosodic processing in *both* the LD and its LP control condition (see Prosodic Positivity in the Early Boundary below). In the movers, a clear N400 was visible only in the LS condition, but not in the LD condition, which displayed a very similar amplitude as the LC condition (Figure

6). However, similar to the pattern in the nonmovers, the small N400 amplitude in LD was partly due to a superimposing (prosodic) positivity (as reflected by the main effects of Wall and Semantics shown in Table 3). The difference waves for movers in Figure 8 reveal that LD did, indeed, elicit a centro-parietal N400 relative to its LP control, resulting in a Semantics main effect and a Semantics x AP interaction across conditions. However, Figure 8 also illustrates that, in striking contrast to the nonmovers, movers showed a substantially reduced semantic N400 effect in the double violation condition (LD). That is, whereas in nonmovers the semantic N400 effect and the prosodic positivity were purely additive, the movers' semantic N400 effect was significantly reduced in the presence of a concurrent prosodic mismatch. Statistically, this observation is reflected by a significant Semantics x Wall x Group interaction (Table 3). Follow-up analyses confirmed that the movers processed prosody and semantics differently when they co-occurred (Semantics x Wall F[1,12] = 11, p = .0061) while the nonmovers did not (F < 1). Voltage maps in Figure 8 further illustrate this interaction in the movers by showing that the N400 response to the LD condition is weaker than the response to the LS condition in that group alone.

Semantic P600. In the Early boundary condition, immediately following the semantic N400, a positive component was elicited by the ES and ED conditions in both the nonmovers and the movers. The P600 component for the ES condition began at 1630 ms in the nonmovers and 1710 ms in the movers and peaked roughly 750 ms after the onset of noun2 in all conditions in both groups. The semantic P600 was supported statistically in TW 1730 – 1930 by a significant Semantics x AP interaction (see Table 2) and the voltage maps in Figure 7 confirmed the broad centro-parietal distribution expected for a P600 component (e.g., Osterhout & Holcomb, 1992).

In the Late boundary condition, the P600 for the semantic mismatch at noun2 began at roughly 1400 ms in the nonmovers and at about 1350 ms in the movers, continued until 1650 ms and peaked roughly 800 ms after the onset of noun2 in both groups. As can be seen in Table 3, this was confirmed statistically in the 1400 – 1650 TW by a significant Semantics x AP interaction. Again, the voltage maps in Figure 8 confirmed the expected P600-like centro-parietal distribution (e.g., Osterhout & Holcomb, 1992).

Prosody. We predicted that auditory boundaries marked by pauses would be reflected in waveforms by the elicitation of the CPS. We therefore begin detailed analyses in this section with an examination of the presence/absence of the CPS and then turn to the response to prosodic mismatches in each boundary condition. Unlike semantic mismatches, for which neural responses were quite similar across boundary conditions and groups, prosodic mismatches elicited different neural responses in each boundary condition and group. In the Early boundary condition, three distinct neural responses to mismatching prosody emerged: (A) a negativity in the 600 – 800 ms TW (and thus roughly 350 ms after the pause) in the nonmovers only, (B) a positivity in the 1000 – 1200 ms TW (700 ms after the onset of the pause) in both groups, and (C) a second negativity in the 1400 – 1650 ms and 1730 – 1930 ms TWs (roughly 600 ms after the onset of the expected, but absent, pause) in both groups. In the Late boundary condition, only a positivity in the 1000 – 1200 ms TW in both groups was observed. (Figure 6 displays the ERP waveforms for each of these conditions and highlights the TWs of interest.) These components are further illustrated for the Early boundary in Figure 10 and Figure 11 and for the Late boundary in Figure 12 and Figure 13. Statistical analyses of these effects for each boundary condition may again be found in Table 2 and Table 3.

Closure Positive Shift. Figure 9 illustrates a direct comparison of ERPs to the EC and LC conditions (thus no violation effects, but a contrast of auditory boundary position only). As this Figure demonstrates, there is no very clear evidence for (or against) a CPS. While there are positive-going shifts after pause onset (prior to the onset components of the first post-boundary word) in both EC and LC conditions, these are relatively small compared to previous auditory CPS findings (i.e., < 2 μV), especially in the LC condition. Moreover, the positive shifts in each condition coincide with the word onset N100 of '*and*' following the boundary in the other condition, making it difficult to determine what drives the apparent differences between conditions (which were significant in a running t-test automatically computed by the analysis program, as can be seen in Figure 9). That is, while the positivities may be CPS responses, it cannot be ruled out that the positivities primarily reflect the absence of the N100. Given the fixed presentation rates employed in this design, time-locking to other events in the utterance (e.g., pause onset) would not change the pattern observed. Thus, even more detailed statistical analyses would result in ambiguous data. Therefore, the current data must be viewed as inconclusive with regard to the CPS, as there is an indication of a positive-going waveform but it cannot be unambiguously isolated from other ERPs. The presence of the CPS in boundaries that signify grouping marked by a pause alone must be further investigated in future follow-up studies.

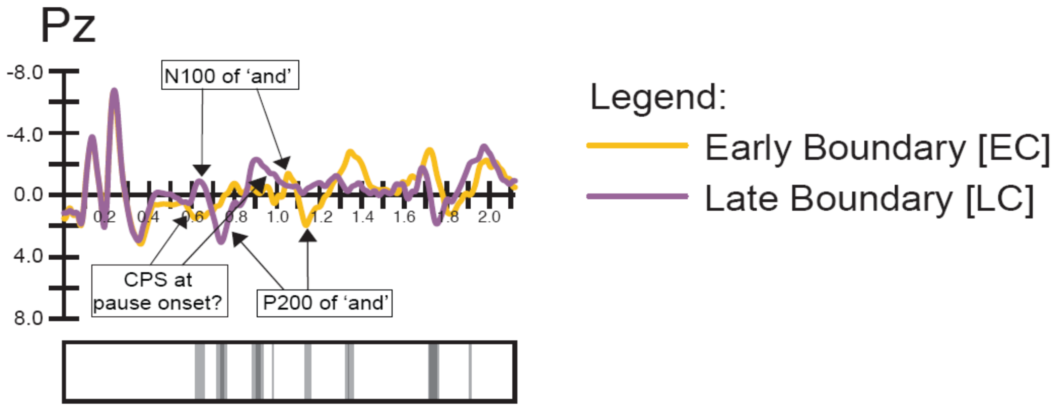

Figure 9. CPS in Early and Late boundary match conditions. EC (yellow) and LC (purple) are compared to examine whether a CPS was elicited by these stimuli. T-tests automatically computed by EEProbe (box at bottom) show that the differences highlighted by the arrows reached significance (significance signified by shaded portions of box). As the arrows indicate, the possible CPS components that emerged overlapped with the N100 to the first word following the pause (‘and’), rendering it difficult to determine whether the positivity is a CPS.

Early Boundary

Negativity for nonmovers. In the Early boundary condition, ERPs showed a negativity for the EP and ED conditions relative to their control conditions that began at 440 ms and was largest between 550 and 780 ms in the *nonmovers* only (see Figure 10). The negativity began only 100 ms after the onset of the unexpected pause, but peaked roughly 350 ms after its onset. Thus, the timing of this component is consistent with what is typically associated with an N400 component (e.g., Kutas & Federmeier, 2000). Analyses in TW 600 – 800 ms, presented in Table 2, revealed a significant main effect of Wall and a significant Wall x Group interaction. Follow-up analyses of each group showed that the main effect of Wall was significant for nonmovers (F[1,6] = 15.04, p = .0082) but not for movers (F[1,12] = .12, p = .7394). The distribution of the negativity is largely consistent with the broad centro-parietal distribution typically observed for the N400 (e.g., Kutas & Federmeier, 2000). However, the voltage map in Figure 10 shows a somewhat more left-lateralized rather than right-lateralized distribution of this component, which would suggest that this may not be a prototypical semantic N400.

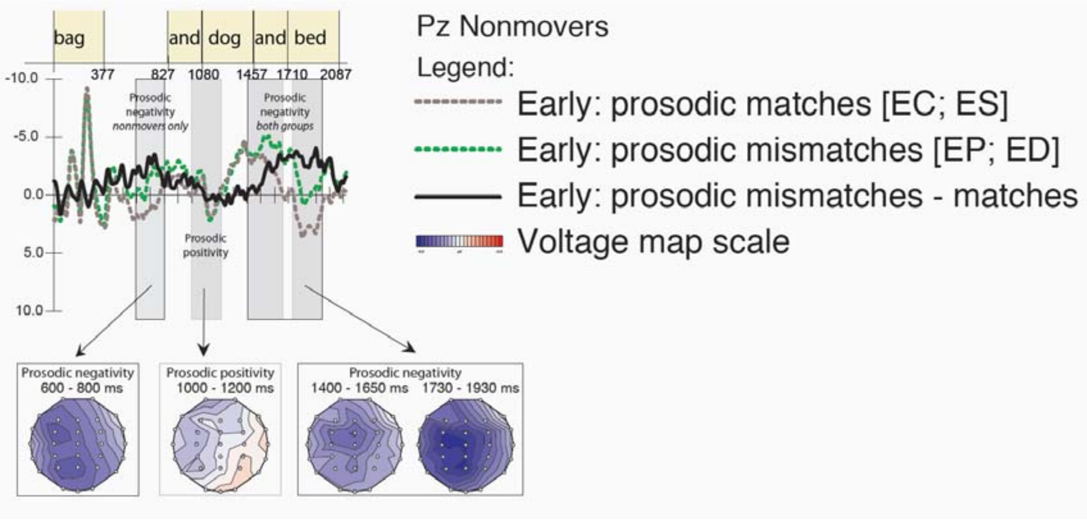

Figure 10. Early boundary prosodic effects on ERPs in nonmovers. Prosodic mismatches (green) & control conditions (gray) differed in 3 ways: (a) in TW 600 - 800 ms prosodic mismatches elicited a negativity not seen in movers (Fig 11), (b) in TW 1000 – 1200 ms prosodic mismatches elicited a positivity in both groups that was stronger in the movers, and (c) in TWs 1400 – 1650 and 1730 – 1930 ms prosodic mismatches elicited a negativity in both groups. The difference wave (black) highlights these effects. Voltage maps for each TW show that the negativities have a distribution typical of an N400. The positivity shows a distribution similar to a P600.

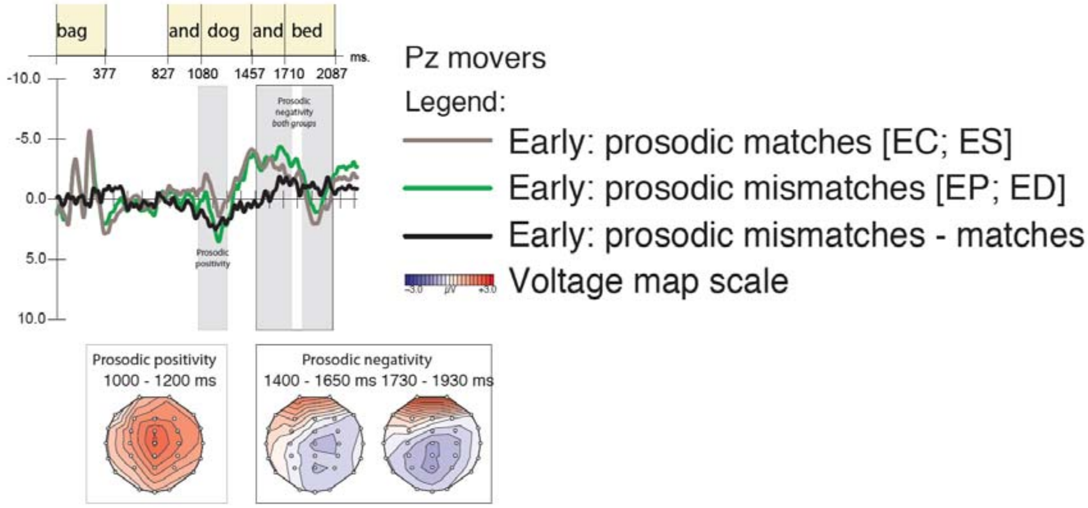

Figure 11. Early boundary prosodic effects on ERPs in movers. Prosodic mismatches (green) and control conditions (gray) differed in two ways: (a) in TW 1000 - 1200 ms prosodic mismatches elicited a positivity in both groups that was stronger in the movers and (b) in TWs 1400 – 1650 and 1730 – 1930 prosodic mismatches elicited a negativity in both groups. The difference wave (black) highlights these effects. Voltage maps show that the positivity is similar in distribution to a P600 and the negativity is similar in distribution to an N400.

Prosodic positivity. An increased positivity in the EP and ED conditions relative to their control conditions was visible in both groups. In the nonmovers, there is some positive-going shift, though it is not large and barely crosses the baseline. In the movers, the positivity began at 1050 ms and persisted until 1240 ms and overlapped with the P200 of the first instance of ‘and’, the first word after the pause (Figure 10 and Figure 11). This positivity occurred not only 200 ms after the onset of noun2, but also began roughly 600 ms after the onset of the unexpected pause (which mismatched with the grouping pattern in the visual display). As can be seen in Table 2, there were no significant main effects or interactions of the factor Wall in this time window in the omnibus ANOVA at the midline (again, in separate analyses for each boundary condition, a main effect of Wall reflects an effect of prosodic mismatch). The difference waves and voltage maps shown in Figure 10 and Figure 11 do not create a strong case for the presence of this component in the nonmovers, but the positivity does seem to be present in the movers. Therefore, although it did not reach significance in the omnibus ANOVA, in order to gain a more complete understanding of this component, we looked at each group independently and found that the positivity did indeed reach significance in the movers (Wall $F[1,12] = 4.88$, $p = .0473$) but not in the nonmovers ($F < 1$). This suggests that the decreased effect in the nonmovers prevented this component from reaching significance in the omnibus ANOVA, but there was no interaction with group because the component may not be completely absent in the nonmovers. Rather, the negativity in the 600 – 800 ms TW, which appears to persist into the 1000 – 1200 ms TW, may have prevented this component from reaching a more positive amplitude (i.e., by canceling out the positivity). Using another baseline might have revealed a significant main effect of Wall; however, as this may have also introduced various artifacts, modification of the baseline was not undertaken.

Prosodic negativity II. In the 1400 – 1650 ms and 1730 – 1930 ms TWs, in addition to the semantic N400 and P600 elicited by the ES and ED conditions, both groups also showed a smaller negativity in the EP condition that began at about 1400 ms in both groups and persisted until roughly 1950 ms. The negativity observed in the EP and ED conditions occurred roughly 400 ms after the visually expected, but absent, pause. In the 1400 – 1650 ms TW, this was supported statistically (see Table 2) by a trend towards a main effect of Wall at the midline ($F[1,18] = 3.25$, $p = .0880$). By the 1730 – 1930 ms TW, the negativity became larger and the difference was supported by a significant main effect of Prosody.

These effects can be seen in Figure 10 and Figure 11. However, since these figures plot the averaged EP and ED waveforms, it is worthwhile to revisit Figure 6 to examine the waveforms for the individual conditions. In the nonmovers, the ED condition exhibited the same morphology as the EC condition. In the movers, the ED condition paralleled the ES condition (a positivity), but reached a smaller positive amplitude. Because in earlier TWs the ED condition exhibited responses to both the semantic and prosodic mismatches, it is clear that participants in both groups detected both errors. This suggests that rather than the movers showing a null effect, in both groups the ED condition showed the additive effect of the prosodic mismatch (a negativity) and the semantic mismatch (a positivity). As a result, for both groups, the ED condition elicited an averaged waveform that was smaller in amplitude than both of the corresponding "pure" mismatch conditions. This is similar to what was seen for semantic effects in the LD condition for the movers (described earlier).

Late Boundary

Prosodic positivity. As shown above, the LS condition elicited an N400 in both the nonmovers and the movers. Interestingly, in both groups, in the same TW, a positivity was elicited by the prosodic mismatch. Figure 12 and Figure 13 show waveforms and difference waves for prosodic effects in the Late boundary condition in the nonmovers and movers. This positivity began at roughly 900 ms, continued until 1300 ms, and peaked at 1100 ms in both groups, 700 ms after the offset of noun1, which co-occurs with the onset of the visually expected, yet absent, auditory boundary. This positivity was confirmed statistically in the 1000 – 1200 ms TW by a significant main effect of Wall (see Table 3).

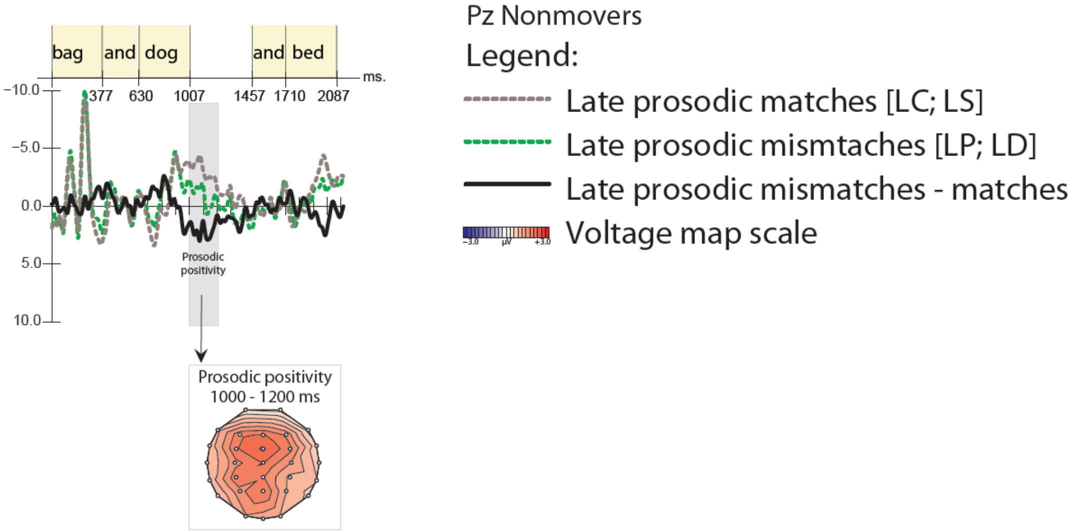

Figure 12. Late boundary prosodic effects on ERPs in nonmovers. Prosodic mismatches (green) and control conditions (gray) differed in TW 1000 – 1200 ms only. The difference wave (black) highlights this effect. Voltage maps show the negativity has a distribution typical of an N400.

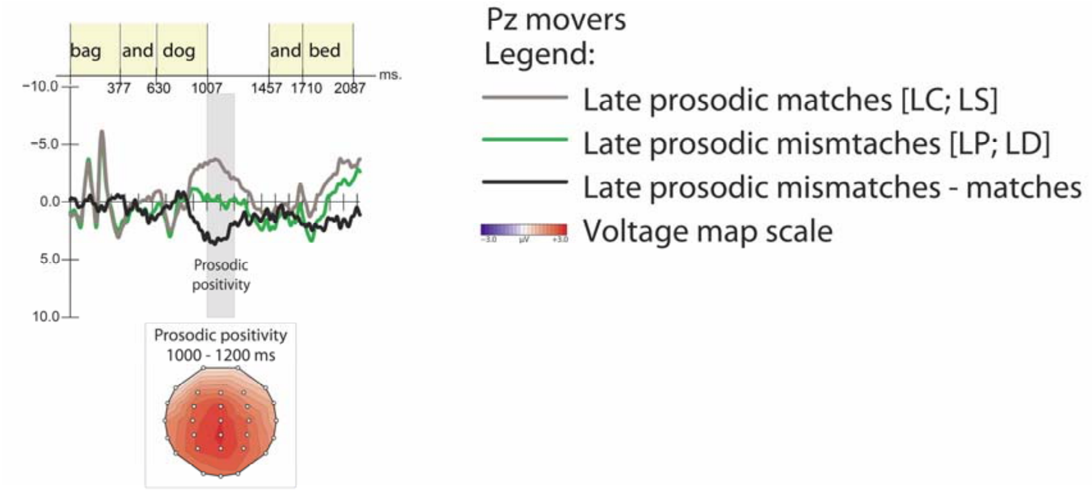

Figure 13. Late boundary prosodic effects on ERPs in movers. Prosodic mismatches (green) and control conditions (gray) differed in TW 1000 – 1200 ms only. Prosodic mismatches elicited a positivity in both groups. The difference wave (black) highlights this effect. Voltage maps show that the positivity has a distribution similar to a P600.

Behavioral Results

Accuracy. Accuracy data for all conditions are illustrated in Figure 14**,** with statistical analyses presented in Table 4. Looking at Figure 14, it is obvious that most conditions reached ceiling levels for accuracy (EC nonmovers = 97.77%, EC movers = 95.83%, LC nonmovers = 97.02%, LC movers = 96.47%, ES nonmovers = 97.77%, ES movers = 97.36%, LS nonmovers = 98.21%, LS movers = 98.32%, ED nonmovers = 99.55%, ED movers = 100%, LD nonmovers = 98.66%, LD movers = 98.56%), with the exception of the pure prosodic mismatch conditions, which are circled in the figure (EP nonmovers = 87.95%, EP movers = 81.25%, LP nonmovers = 90.18%, LP movers = 83.17%). These conditions were responded to significantly less accurately than all other conditions by both groups, reflected by a highly significant Boundary x Wall x Semantics interaction (i.e., main effect of prosodic mismatch in the absence of semantic mismatches). Our 2 x 2 x 2 (Boundary x Wall x Semantics) design resulted in pure prosodic mismatches (EP and LP conditions) being included in the set of semantic match conditions, as well as in the subset of Boundary x Wall interactions, but not in the subsets of either Boundary or Wall conditions (because the EP condition involved an *Early* auditory boundary and a *Right* wall, whereas the LP condition involved a *Late* auditory boundary and a *Left* wall). Therefore, the significant main effect of Semantics as well as the significant Boundary x Wall interaction that may be seen in Table 4 were primarily driven by lower accuracy for the pure prosodic mismatch conditions as well (as evident in Figure 14). Interestingly, though both groups show the same pattern overall, additional group differences also emerged for these conditions. Specifically, in both the EP and LP conditions, the nonmovers showed more accurate responses than the movers. In the most relevant follow-up analysis of the Boundary x Wall x Semantics x Group interaction, although both groups showed a highly significant Boundary x Wall x Semantics interaction (Movers F[1,12] = 127.76, p < .0001; Nonmovers F[1,6] = 36.10, p < .0010), this pattern was particularly strong in the group of movers indicating larger differences between match and mismatch conditions in that group compared to the nonmovers.

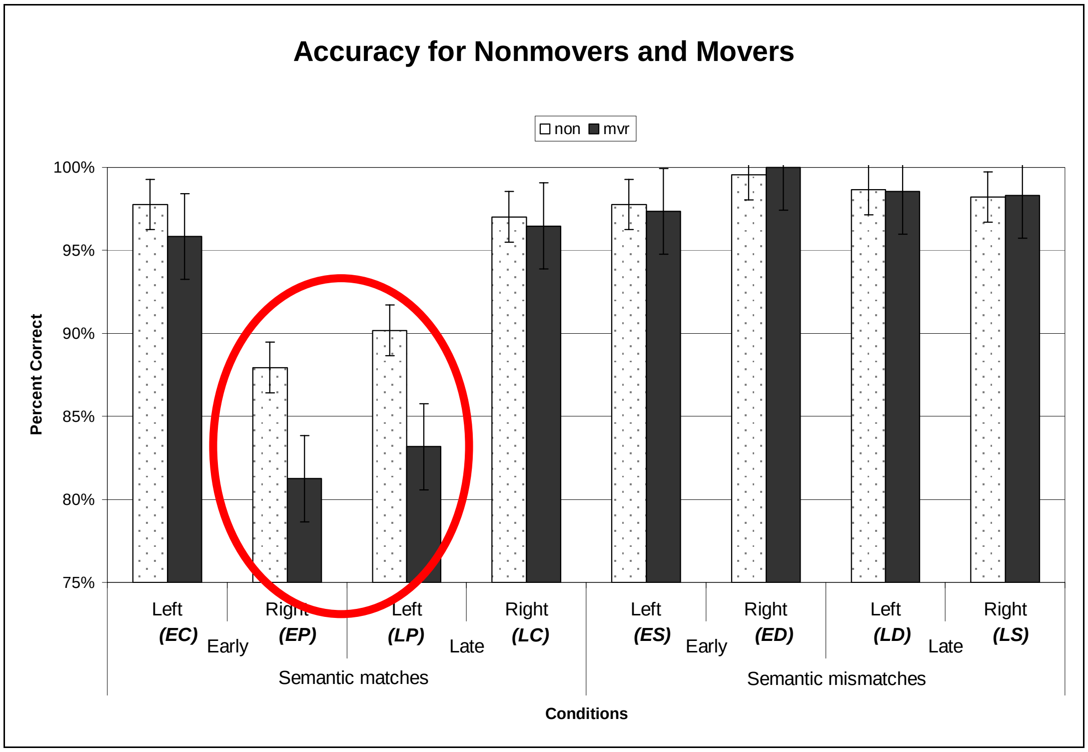

Figure 14. Behavioral accuracy. Accuracy for nonmovers (gray) and movers (black) for all conditions. Responses arranged by factors (semantics, then boundary condition, then wall position) with condition names included in parentheses. Scale begins at 75% in order to illuminate significant differences. Error bars reflect standard error.

Table 4. Statistical analyses of accuracy in behavioral responses

	df	F-value	P-value
Boundary	*1,18*	0.51	0.4837
Boundary x Group	*1,18*	0.06	0.8143
Semantics	***1,18***	**180.98**	**<.0001**
Sem x Group	***1,18***	**13.78**	**0.0016**
Bound x Sem	*1,18*	1.3	0.2694
Bound x Sem x Group	*1,18*	0.06	0.8023
Wall	*1,18*	.01	0.9075
Wall x Group	*1,18*	.40	0.5357
Bound x Wall	***1,18***	**81.46**	**<.0001**
Bound x Wall x Group	***1,18***	**5.85**	**0.0264**
Wall x Sem	*1,18*	3.35	0.0837
Wall x Sem x Group	*1,18*	.02	0.8870
Bound x Wall x Sem	***1,18***	**129.22**	**<.0001**
Bound x Wall x Sem x Group	***1,18***	**7.38**	**0.0141**

Response Times. Reaction times are presented in Figure 15, with statistical analyses presented in Table 5. It should be noted that accuracy, and not speed, was emphasized to participants. Importantly, there was no interaction with group for any factor, showing that the presence or absence of eye movement did not influence the speed of behavioral responses to these stimuli. Nevertheless, at least numerically, the pure prosodic violation conditions again showed a similar pattern to that observed for accuracy data. That is, response times in both groups were slowest in the EP condition, and the group differences between movers and nonmovers were most prominent in the LP condition (with movers responding more slowly than nonmovers; EP nonmovers = 820.87 ms, EP movers = 875.36 ms, LP nonmovers = 665.58 ms, LP movers = 781.15 ms). Although the group difference was not reflected by significant interactions in the global ANOVA for response times, the Boundary x Wall x Semantics interaction, the Boundary x Wall interaction, and the Semantics main effect (across groups) did reach significance again. This pattern of results rules out the possibility that the effects observed for accuracy data may be partly driven by a speed-accuracy trade-off. Rather, the specific processing difficulties in EP and LP conditions (and, to some extent, the group differences in these conditions as well) were reflected by accuracy as well as RT data. As reported above, the only reliable ERP group differences were also observed for prosodic mismatch conditions and thus may be directly linked to the behavioral findings. We will return to these issues in the Discussion section.

It should also be noted that the Semantics main effect (across groups) was by far the most significant effect for RT data ($F > 100$; $p < .0001$). This indicates that response times were consistently faster whenever a semantic mismatch was involved, pointing to the overall ease in detecting this type of cross-modal mismatch independent of concurrent prosodic mismatches in both groups (EC nonmovers = 643.45 ms, EC movers = 680.83 ms, LC nonmovers = 641.68 ms, LC movers = 639.33 ms, ES nonmovers = 582.61 ms, ES movers = 542.86 ms, LS nonmovers = 502.07 ms, LS movers = 506.41 ms, ED nonmovers = 557.25, ED movers = 549.76, LD nonmovers = 512.13, LD movers = 511.25). In contrast to accuracy data, this RT advantage held even in comparison to correct control sentences (cf. left versus right sides of Figure 15).

Finally, reaction time measures revealed that both groups were generally faster at responding to stimuli in the Late auditory boundary condition compared to the Early boundary condition, demonstrated by a main effect of Boundary (Table 5). This is likely because of the nature of the stimuli themselves. The only auditory cue to a boundary was a pause, which can only be perceived relative to other elements in the sentence. Therefore, the presence of the early pause did not preclude a second, longer pause (this will be discussed further in the Discussion).

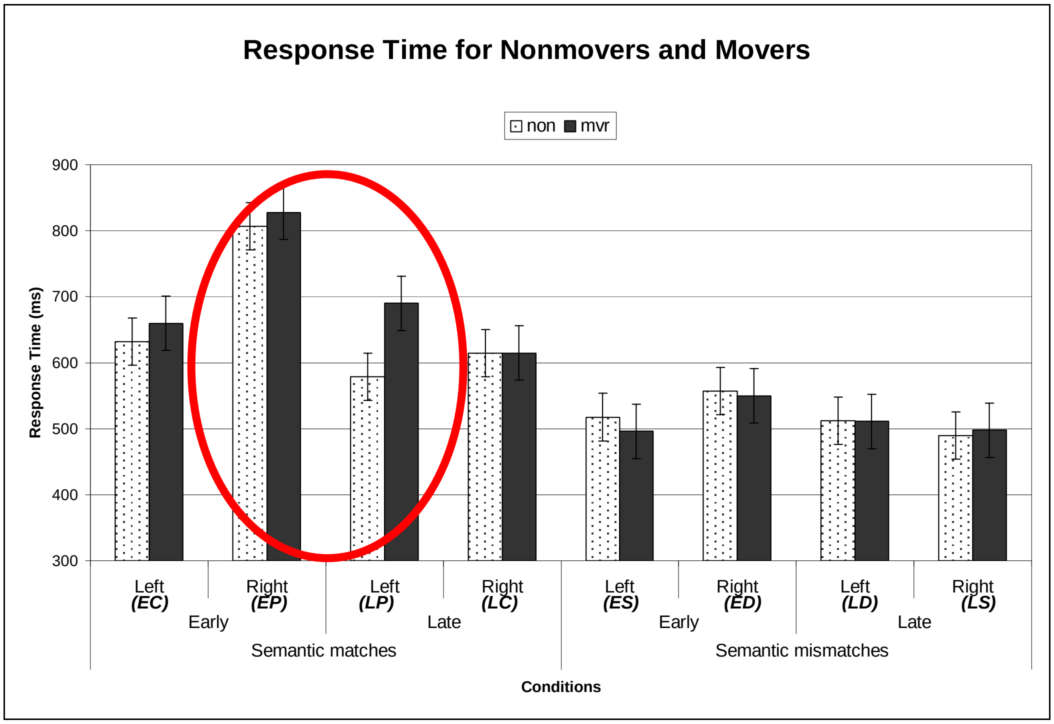

Figure 15. Response time for behavioral data. Response times for nonmovers (gray) and movers (black) for all conditions. Responses arranged by factors (semantics, then boundary

condition, then wall position) with condition names included in parentheses. Scale begins at 300ms in order to illuminate significant differences. Error bars reflect standard error.

Table 5. Statistical analyses of response time for behavioral responses

	df	F-value	P-value
Boundary	*1,18*	**11.34**	**0.0034**
Boundary x Group	*1,18*	0.49	0.4912
Semantics	*1,18*	**110.65**	**<.0001**
Sem x Group	*1,18*	2.72	0.1164
Bound x Sem	*1,18*	0.77	0.3932
Bound x Sem x Group	*1,18*	0.15	0.6987
Wall	*1,18*	1.09	0.3096
Wall x Group	*1,18*	0.41	0.5304
Bound x Wall	*1,18*	**16.84**	**0.0007**
Bound x Wall x Group	*1,18*	1.76	0.2006
Wall x Sem	*1,18*	3.78	0.0678
Wall x Sem x Group	*1,18*	0.67	0.4222
Bound x Wall x Sem	*1,18*	**13.60**	**0.0017**
Bound x Wall x Sem x Group	*1,18*	0.40	0.5375

Discussion

The goals of the present study were to determine whether auditory and visual phrase grouping information are integrated in real-time and interact with one another, and, if so, what electrophysiological components may be elicited by a grammatical and syntactically simple utterance that does not match a visual context either in prosodic grouping or in semantic representation, but is otherwise acceptable. We predicted that a visual display would influence the processing of the acoustic input and that this influence would be reflected in ERP waveforms. On finding that many, but not all, participants exhibited eye movements that followed exactly the same pattern, a second, and particularly innovative, goal emerged to investigate whether patterns of response in the HEOG electrodes could inform our study of prosodic processing. We found a robust and systematic pattern of eye movement that co-varied with differences in behavioral and ERP responses. Our findings supported our initial predictions such that visual context did influence prosodic processing. Based on our findings, it appears that participants either used the visual context to create expectations for the presence of auditory boundaries or searched for visual information that corresponded with auditory input. We will begin our discussion with behavioral findings, then turn to ERP findings, and then discuss the pattern of eye movement observed in this study and how it relates to our other findings.

Behavioral Responses

Behavioral responses showed that individuals were able to detect both semantic and prosodic mismatches between auditory and visual stimuli. One interesting result that emerged in the analysis of response time was that the Late auditory boundary was responded to significantly faster than the Early auditory boundary. It is possible that this is the result of recency effects associated with the timing differences between each boundary condition. As noted earlier, another possibility is that a pause can only be detected relative to other events in an utterance. Therefore, when the pause was heard in the Early boundary, it was still possible that another, longer pause, would be heard later in the utterance. While one could argue that after gaining some experience with the stimuli, participants would expect a pause of a certain duration, it appears that this experience was not sufficient to overcome the possibility that a second pause would exist, resulting in longer response times in the Early boundary condition. (This was also reflected in the ERP responses of the nonmovers by the early prosodic negativity.)

As predicted, participants were best and fastest at correctly rejecting semantic mismatches, as this is the more salient information. While prosodic mismatches were successfully detected above chance levels, participants were least accurate and slowest in their responses to these conditions compared to all other conditions. While the presence of eye movement did not influence the speed of response, it did negatively impact accuracy to some degree, since the movers performed slightly less accurately than the nonmovers, though still with very high accuracy. Thus, both groups showed difficulty detecting prosodic mismatches in both boundary conditions, but the movers showed even more difficulty than the nonmovers. The behavioral data alone do not illuminate when or how this happens. To gain insight into these issues, we turn to the ERP data.

ERPs for Semantics

As expected, in both boundary conditions and in both groups, the semantic mismatch at noun2 elicited an N400 response that was consistent with the timing and distribution typically associated with an N400 to semantically-anomalous stimuli (Kutas & Hillyard, 1980; Kutas & Federmeier, 2000). The presence of the N400 effect replicates Friedrich and Friederici's (2004) findings that a picture that does not match an auditory noun stimulus is sufficient to elicit an N400 response and further extends these findings to the phrase level. Further, the emergence of this component confirms that our design does, indeed, elicit an ERP response that informs our understanding of the interaction between auditory and visual input, as intended.

A less expected, but not surprising, finding was that the semantic mismatch in both boundary conditions and in both groups elicited a P600 component. The timing and distribution of this component is in keeping with what is typically found for a P600 (e.g., Osterhout & Holcomb, 1992). This biphasic N400/P600 pattern has been reported in a number of previous studies, including for verb argument structure violations (e.g., Friederici & Frisch, 2000; Münte, Heinze, Matzke, Wieringa, & Johannes, 1998) and – more relevant in the present context – conceptual-semantic incongruities (e.g., Steinhauer, Drury, Portner, Walenski, & Ullman, 2010). One explanation for the presence of the P600 component following an N400 is that it marks reanalysis and/or integration of the semantic anomaly within the phrase (e.g., Friederici & Weissenborn, 2007), suggesting that participants are not simply recognizing the mismatch, but are also trying to make use of it to interpret the stimuli. Alternatively, this may be an indication that the P600 is elicited whenever there is a mismatch between information from two sources of structured information, rather than to syntactic violations alone (see Münte et al., 1998 for a discussion of this interpretation). This is also in line with a reanalysis/integration interpretation of the P600 (e.g., Kuperberg, 2007; Friederici, 2002; Friederici & Weissenborn, 2007). The positivity to the prosodic anomaly that was elicited in both boundary conditions also supports this interpretation (as discussed below). Importantly, for semantic effects, both groups showed identical timing of both the N400 and P600 components, indicating that eye movement did not influence the speed of processing semantic input.

Prosody

Prosodic negativity. In the Early boundary condition, two negative components were elicited. The first was present for the nonmovers only, while the second was present for both groups. In the first instance, the nonmovers showed a negativity for prosodic mismatches that had a centro-parietal distribution (but appeared more prominent in the left hemisphere) and peaked roughly 350 ms after the onset of the unexpected pause. This component was tentatively interpreted as an N400-like response to the prosodic anomaly because the timing and centro-parietal distribution are overall consistent with what is typically associated with an N400 component (e.g., Kutas & Federmeier, 2000). While the N400 for language is usually right-lateralized (Kutas & Federmeier, 2000), distributional differences for the N400 relative to the presentation of different types of stimuli have been observed (e.g., Ganis, Kutas, & Sereno, 1996; Holcomb & McPherson, 1994; Holcomb & Anderson, 1993; McCallum, Farmer, & Pocock, 1984). It is therefore reasonable to assume that the distribution of the N400 in this study was modulated as a result of the simultaneous auditory and visual input. Therefore, in this instance, the component may be interpreted as an N400 resulting from the mismatch with the conceptual grouping expectation created by the visual input. This suggests that the N400, which is a marker of conceptual meaning and how it relates to context (e.g., Kutas & Federmeier, 2007), is not only not language specific, but also marks violations to contextual expectations for non-semantic conceptual representations across modalities. Crucially, if it is the case that the negativity elicited in the nonmovers is an N400, it is an indication that subjects who did not move their eyes generated abstract conceptual groupings from visual stimuli alone that were then violated by the auditory stimulus. In this study, expectations were not created by linguistic input, but by visual context, supporting the important role of visual information in language processing (e.g., Knoeferle et al., 2008). If this interpretation is correct, this is the first study to find an N400 for (rather abstract) conceptual grouping mismatches between auditory and visual stimuli and would lend support to the interpretation of the N400 as a component reflecting integration of conceptual information with a context.

It is surprising that both groups would show a second response to a prosodic mismatch (in addition to the positivity) in the Early boundary condition only, and not in the Late boundary condition, and even more surprising that the nonmovers would should two negativities. However, participants may have processed the right wall in the absence of the pause as a separate mismatch. Since the only auditory boundary cue was the pause, which can only be detected relative to the other events in the utterance, until the absence of the second pause, there remained a possibility that there could be a longer pause, thereby making the absence of this pause a second prosodic mismatch. This is consistent with differences in accuracy and response time for the Early and Late boundary conditions seen in the behavioral data. The first negativity in the nonmovers also speaks to the group differences seen in the behavioral data. That is, nonmovers were more accurate at detecting prosodic mismatches compared to the movers. It appears that the nonmovers developed expectations earlier than the movers and this type of processing led to higher accuracy in behavioral judgments.

Prosodic positivity. In both the Early and Late boundary conditions, a positive component emerged in the 1000 – 1200 ms TW. This positivity peaked 600 ms after the onset of the unexpected pause in the Early boundary condition (EP and ED conditions) and 600 ms after the expected onset of the pause, which was absent in the Late boundary condition (LP and LD conditions). This was also the earliest time at which the prosodic mismatch could be detected in either boundary condition. This component failed to reach significance as a main effect in the Early boundary condition when both groups were examined together, likely because the effect was cancelled out in the nonmovers because of the earlier negativity. It did, however, reach significance in the movers. In the Late boundary condition, the positivity elicited by prosody superimposed the semantic N400 in the LD condition and significantly reduced the amplitude of the N400 response compared to the corresponding LS condition. This speaks to the strength of the prosodic violation. Specifically, the presence of a semantic violation might be easier to detect behaviorally, but the neural response to prosody, at least in movers, is just as robust as the response to semantics. This pattern will be discussed further below. Interestingly, in the Early boundary condition, a later negativity emerged because the boundary was marked only by a pause and did not preclude the existence of a second pause. This is therefore an indication that prosody is processed automatically as soon as a mismatch between information from different sources is evident, regardless of whether this information will be processed further at a later stage.

The timing and distribution of the prosodic positivity in the Early boundary condition are consistent with a P600 (e.g., Osterhout & Holcomb, 1992). In the Late boundary, this positivity had a widespread scalp distribution, which, while not typical of a P600, is also not completely uncommon (Kemmer, Coulson, de Ochoa, & Kutas, 2004). It is not entirely surprising that the prosodic mismatch elicited a P600, particularly in light of the presence of this component for the *semantic* mismatch. Rather, these findings indicate that the P600 reflects integration between different types of structured information. In this study, the information being integrated came from separate modalities but was nonetheless marked by this component, consistent with Münte et al. (1998).

While it does appear that this positive component is a P600, the possibility that this component is a P800—a component that has been linked to prosodic processing (e.g., Eckstein & Friederici, 2005; Astésano et al., 2004)—must be considered. In the present study, context was created by a visual display which created expectations that were then violated by auditory input. In previous studies that found positivities elicited by prosody, violations were created by prosodic contour mismatching with semantic or syntactic information (i.e., statements concluded with question intonation; Astésano et al., 2004) or violations at the lexical level (focus or metrical stress; Mietz et al., 2008; Magne et al., 2007; Eckstein & Friederici, 2005). While the contrasts previously used and those used here are all included under the domain of prosody, few would argue that the same processes are exploited for these very different types of prosodic processing. It is therefore unlikely that the positivity to the prosodic mismatch elicited by these stimuli is a P800.

Absence of a CPS. While a CPS was expected to be elicited by the phrase boundaries, we could not conclusively find evidence for a CPS in the present study, as the positivity could not be unambiguously isolated from other ERPs. If we could confirm that there really was no CPS in the current investigation, there would be a number of possible explanations for such a finding. One explanation is that the prosodic cues employed did not elicit a CPS.

Previous research showed that a pause was not required to elicit a CPS response (Steinhauer, 2003). Based on the data from this study, one might conclude that a pause alone might not be *sufficient* to elicit this response (but cf. Li & Yang, 2009, who propose that pause can modulate the CPS). A second explanation is that the grammatical structure employed did not elicit a CPS. The utterances used in the present study consisted of simple conjoined noun phrases. It is possible that the CPS is only elicited to phrase boundaries in more complex syntactic structures. At the time of this writing, there is no evidence that either simple phrases of this kind or phrasal groupings, and not other syntactic boundaries, elicit a CPS (but cf. Li & Yang, 2009 who found a CPS elicited by prosodic phrase boundaries). Therefore, it is possible that this grammatical structure simply does not bring forth this component, indicating that the CPS may be linked either to more complex grammar or to prosodic boundaries that mark complex syntax rather than simply groupings. A final possibility is that the task itself prevented the elicitation of the CPS. That is, previous studies required participants to attend to sentences and determine whether they were natural or not (e.g., Steinhauer et al., 1999; Steinhauer, 2003; Pauker et al., under revision). In the present study, the task was not to interpret the stimuli, but rather to determine whether the input from both modalities matched. As a result, the responses elicited marked re-analysis and the realization of prosodic expectations (as discussed above). Therefore, it is possible that the same stimuli presented with a different task might elicit a CPS. Clearly, additional investigations are required to clarify these issues.

Semantics *and* Prosody

Two predictions were made for the conditions containing both semantic and prosodic mismatches (Double Mismatch): (1) semantics and prosody rely on distinct neural correlates or (2) overlapping neural correlates are exploited for semantic and prosodic processing. Examination of the Late boundary condition suggests that a combination of these interpretations may best account for the findings. To review, the LD condition exhibited both the additive effects of semantics (a negativity) and prosody (a positivity) which overlapped and thus resulted in a semantic N400 that showed a significantly smaller amplitude than the semantic N400 in the corresponding LS condition. The main effects (or the additivity) of both semantics and prosody in this TW suggest that there are distinct neural mechanisms recruited for prosodic and semantic processing. There is also no clear evidence that either piece of information was weighted more heavily. In particular, despite being able to determine the presence of the prosodic mismatch prior to the semantic mismatch, and therefore being able to successfully complete the task without further processing semantics, participants did process the semantic mismatch. On the other hand, the response to prosody was so robust that it modulated the N400 such that the amplitude of this component was significantly reduced compared to the pure semantic mismatch. Based on this, it appears that semantics and prosody are processed both distinctly, and concurrently, in the brain. However, in addition to these main effects, an interaction between both types of information was observed in the same TW in the movers. This interaction indicates that, in the movers, while both prosody and semantics were processed concurrently, they were also processed differently when both violations were present as opposed to when a single violation was present. At some point in the analysis of the stimuli, the processing of both types of information has to be integrated. Thus, differences in the nature of prosodic and semantic processing emerged when both types of violations, as compared to one or the other, were present. Some subjects (i.e., the nonmovers) were able to rely on one type of information to render match/mismatch judgments. Other subjects (i.e., the movers), however, seem to integrate and process both pieces of information even when unnecessary to complete the task (because sufficient information had already been received). On the basis of these findings, it again appears that the nonmovers used a strategy that resulted in perhaps more parsimonious processing and, importantly, higher accuracy in behavioral judgments.

Eye movements

Perhaps the most intriguing data to emerge from the present investigation is related to patterns of eye movements and their relationship to the ERP components. The horizontal eye movement pattern that emerged in 13 of the 20 participants showed three stages of processing the auditory and visual input of relevance to the current investigation. In the first stage, at the onset of stimuli in both boundary conditions, movers looked to the left of the visual display towards the first pictured noun. When a left wall (physical boundary) was present, movers looked further to the left than in the absence of the wall. In the second stage, following the offset of noun1, all movers began moving their eyes rightward. When there was an early pause (Early boundary condition), rightward movement was slower or began later than when there was no pause (Late boundary condition). Thus, participants appeared to be following the auditory input while attending to the visual input and initiated the saccade towards the next object only once the corresponding next noun was mentioned (or about to be mentioned) in the speech signal. The later initiation of the saccade in Early (vs. Late) boundary conditions was also reflected by a delayed crossing of the EOG baseline in the Early boundary condition (1200 vs. 1000 ms), reflecting the fact that the target object was reached later.

A parsimonious interpretation of these finding is that the secondary interactions between boundary conditions and prosodic conditions were simply the result of ongoing effects of the wall from the earlier time windows. According to this interpretation, all movers initiated their saccades towards the second object at about the same time, but since the eyes were further left in conditions with a Left wall (ES and EC; LP and LD), the saccades had to be longer. As a consequence, the eye position in these conditions was always slightly further to the left compared to conditions with a Right wall. In this latter case, one would expect that (a) the difference in eye position should decrease over time and converge at the landing point and (b) the landing point (center of the screen represented by the EOG baseline level) should be reached at approximately the same point in time. Given that this was overall the pattern found (see Figure 2 and Figure 5 as well as corresponding analyses, but note that in both boundary conditions there was a small difference in the time at which the landing point was reached), a spill-over of the Wall effect due to the differences in initial fixation points seems to be supported by the data.

In the third stage, after noun2 in both boundary conditions, eye movements showed a crossover from prosody-driven movement to semantic-driven movement. In semantic match conditions, participants fixated the center pictured noun and then continued moving their eyes rightward to subsequent items. In conditions containing a semantic mismatch, movers continued to fixate the center item until they had heard most of the pause (Early boundary) or the second instance of ‘and’ (Late boundary). This pattern shows that language users rapidly consider visual context (within 250 ms of stimulus onset) and again supports ongoing integration between auditory and visual input throughout processing. There is a great deal of interaction between information from the two modalities as participants integrate input from both modalities and continuously modify the input they exploit most heavily relative to what input is most informative at any given time. This finding is in line with previous audiovisual language processing research showing a rapid use of depicted events (Knoeferle & Crocker, 2007), simultaneous parallel and integrative processing of visual information and linguistic computations (Gleitman et al., 2007), and language-mediated eye movements influenced by conceptual structures (Altmann & Kamide, 2007).

Implications of eye movement. Significant differences in neural responses between groups suggest that the movers and nonmovers processed the stimuli differently. High behavioral accuracy indicates that both groups were successful at this task, though nonmovers were more accurate. Measurements of hearing, working memory, cognition and vocabulary revealed no differences between the groups that could explain the differing patterns of performance.

Since the only identified difference between these groups lies in whether or not they moved their eyes, eye movement patterns suggest that the groups may have used distinct strategies to complete the task.

The first question that must be asked is: what are the different strategies utilized by each group? One possibility is that each group is weighting input from each modality differently. The nonmovers, who fixated on the visual input and created expectations/predictions which were then violated (or not) by the auditory input (reflected by the first prosodic negativity), were holding the visual input constant and weighing it more heavily. They were, in essence, asking, "Does the auditory input match the visual input?" The movers, on the other hand, were letting the auditory input guide their attention and drive their visual search (reflected by the robust P600 to the prosodic mismatch in boundary conditions). Their approach may have been, "Does the visual input match the auditory input?" It is possible that the movers were attending so much to the auditory input that the result was an inability to inhibit their eye movement. To summarize this idea, the nonmovers used the visual input to create expectations and then modified this based on auditory input, whereas the movers refrained from creating these expectations by allowing the auditory input to direct their attention.

A second possibility is that the different neural components could have been a product of other differences rather than a difference in strategy alone. The nonmovers stayed fixated on the center of the screen and could therefore observe both boundaries. The movers attended to the left side of the picture and could therefore only see one boundary. Since the visual attention of participants in each group was directed at different places in the same time window, there were –

unsurprisingly and in-line with eye-tracking data and examinations of visual attention (e.g., Tanenhaus et al., 1995; Gleitman et al., 2007) – differences in perception. While these hypotheses may explain the findings, they do not take into account potential reasons for why visual attention was directed differently in each group or why different strategies were employed.

Thus, rather obvious questions that arise from these findings and interpretations are *why* there are differences between these groups and what drives different patterns of eye movement to begin with. The present study alone cannot answer these questions. There were no differences between these groups in the independent measures taken, other than the presence or absence of eye movement. Nonetheless, it is possible that another kind of measurement would illuminate group differences. It is also possible that the nonmovers were better at following instructions or that the nonmovers simply paid better attention to the instructions. Perhaps one strategy was easier, less taxing on resources, or even more effective when used outside of the laboratory, than the other, and individuals chose one or the other based on this. When asked about the task at the end of the study, many participants reported difficulty preventing their eyes from moving. Unfortunately, as these eye movement findings were not expected, it is not known whether those who claimed to have found this difficult were those who moved their eyes. Therefore, while interesting, it remains an anecdotal observation. It would, however, be difficult to argue that preventing all eye movement is a natural means of processing language. It is clear that language users continuously move their eyes and attend to visual information, as has been shown by a wealth of eye-tracking studies and that visual attention is influenced by auditory input (e.g., Gleitman et al., 2007). Likewise, where visual attention is directed influences the perception of auditory linguistic input (e.g., Georgiades & Harris, 1997). Moreover, much research has shown that visual input from facial cues plays an important role in speech perception (McGurk & MacDonald, 1976). Indeed, some very minimal eye movement remained in the nonmovers, which can be seen by very small positivities in HEOG displayed in Figure 2. The present investigation expanded our knowledge on the process of integration of auditory and visual information. However, since this was not a very natural task, it remains only an important first step in an area that requires much more investigation.

From this 'naturalness' explanation, further questions arise. If eye movement is the more natural method of processing language, is it a better way to process language? Were these participants at an advantage over those who did not move their eyes? Behaviorally, in this task, this was not the case. On the contrary, if anything, those who did not move their eyes appeared to have a small advantage in terms of accuracy in the present task. In terms of neural correlates, however, it is difficult to discern whether either group showed any processing advantage. For prosodic processing, the movers might have displayed an advantage because the P600 responses to prosody were larger in the movers than in the nonmovers. The effect of prosody was stronger in the LD condition for the movers than for the nonmovers. On the other hand, the nonmovers showed an additional prosodic mismatch component in the Early boundary condition that was absent in the movers and occurred prior to any neural response to prosody in the movers. Therefore, at this point it is difficult to conclude that one pattern of eye movement is preferable to the other.

It is important to remember that these data do not determine exact fixation points and further analyses are required to advance our understanding of eye movement patterns during auditory-visual integration of this nature. Further, the present study employed the simultaneous onset of auditory and visual information. The result was that participants received all the visual information at once, whereas auditory input was sequential. Future investigations are required that include a manipulation of the timing at which input from each modality is received in order to increase our understanding of precisely how individuals process and integrate visual and auditory information.

Conclusions

This study showed that visual context influences the neural correlates of prosodic and semantic processing. A mismatch between auditory and visual information elicited both N400-like and P600-like components both for prosodic and semantic mismatches. Based on these findings, we argued that the N400 should be interpreted as a marker of relatively abstract conceptual integration beyond the lexico-semantic level (see Lau et al., 2008); moreover, the findings are in keeping with the hypothesis that the P600 marks general integration of structured information, even across modalities. Of particular interest was the finding that mismatches between auditory and visual prosodic information elicited different components in individuals who moved their eyes while attending to the stimuli compared to those who did not. For those who did not move their eyes, we concluded that participants created expectations for conceptual groupings based on the visual stimuli and that pauses in the auditory input were sufficient to violate those expectations. Those individuals who did move their eyes appeared to do so as driven by the auditory input. While it is not possible to make firm conclusions regarding why eye movement and corresponding neural differences emerged from these stimuli, it is clear that the HEOG data presented here inform our understanding of the processes under investigation. Time windows that contained substantial eye movement differences in the absence of ERP differences, and vice versa, confirm that the HEOG data do indeed provide different, and complementary, information than the ERP data and are a useful means of gaining insight into the processing and integration of auditory and visual information. Differences in electrophysiological, but not behavioral, responses to the stimuli indicate that, unsurprisingly, eye movement strongly influences the integration of auditory and visual input. Crucially, these data show that using well-controlled visual stimuli allows for the measurement of neural responses to audiovisual stimuli without contamination of ERP responses. Information from these two independent sources (HEOG and ERP) can be synergistically combined to advance our understanding of language processing in context. Using this type of methodology will allow us to gain increased insight into the temporal ordering of language processing effects and to take another step towards more ecologically valid research.

PREFACE TO CHAPTER 3

Chapter 2 examined how visual context interacted with and influenced auditory processing in young adults. Through behavioral, ERP, and HEOG data, we found that young adults integrated input from the auditory and visual modalities and continuously modified the input they exploited most heavily relative to what input was most informative at any given time. Specifically, both prosodic and semantic mismatches elicited N400 and P600 components in the young adults. Interestingly, we found that the young adults did not behave entirely homogeneously. Instead, two groups could be defined on the basis of horizontal eye movement data: those who did not move their eyes during stimulus presentation (the nonmovers) and those who used the auditory input to guide their eye movements (the movers). Somewhat surprisingly, we found that these two groups showed different neural responses to stimuli.

As noted in the introduction, one of the primary aims of this dissertation was to examine prosodic processing in an aging population, given somewhat equivocal evidence in the literature showing age-related changes associated with prosodic processing (e.g., Wingfield, Wayland, & Stine, 1992; Kjelgaard, Titone, & Wingfield, 1999; Steinhauer, Abada, Pauker, Itzhak, & Baum, 2010). In particular, we wanted to make use of the exquisite temporal precision afforded by ERP analyses to determine whether some of the conflicting findings in the literature could be accounted for by differences in the stage of processing to which the various tasks used were sensitive. Therefore, our second study used the same design and stimuli as in the first experiment to examine the influence of visual context on auditory processing in older adults aged 65 to 80. As will be seen, the data from these older adults was compared to the data from the young adult movers alone because all older adults showed eye movement during the task.

CHAPTER 3:

An ERP investigation of older adults' integration of auditory and visual information for prosodic processing

Shani Haviva Abada

Karsten Steinhauer, John E. Drury, Shari R. Baum

Aging is associated with losses to certain cognitive and motor functions (Spencer & Raz, 1995; Jonides et al., 2000; Ketcham & Stelmach, 2001). In the domain of language, older adults have shown difficulty understanding complex structures and linguistic forms that strain working memory (e.g., Byrd, 1993; Kemper, 1992; Light, 1990; Stine & Wingfield, 1990). Aging is also associated with decreased sensitivity to temporal auditory cues and decreased ability to interpret certain acoustic cues associated with speech perception and auditory word recognition difficulties (e.g., Tremblay, Piskosz, & Souza, 2002, 2003; Stine-Morrow, Miller, & Nevin, 1999; Wingfield, Aberdeen, & Stine, 1991, among many others). On the other hand, overall linguistic and syntactic knowledge appear to be relatively preserved (Wingfield, Lindfield, & Goodglass, 2000). Older adults perform well, and perhaps even better than younger adults, on measures of vocabulary (see Verhaegen, 2003 for a review) and they are able to successfully exploit linguistic context in spoken word recognition (Pichora-Fuller, Schneider, & Daneman, 1995; Wingfield et al., 1991; Wingfield, Alexander, & Cavigelli, 1994; Abada, Baum, & Titone, 2008). Evidence also suggests there may be an increased reliance on contextual information in language processing in older adults (e.g., Cohen & Faulkner, 1983; Abada et al., 2008).

Examinations of the electrophysiological correlates of language processing (which reflect real-time processing mechanisms) have revealed that, while there are some changes in event-related potential (ERP) components associated with aging, the same components are typically elicited under similar conditions in both young and older adults. In general, exogenous ERP components (e.g., N1 and P2) show little or no change associated with aging (Iragui, Kutas, Mitchiner, &

Hillyard, 1993) while endogenous ERP components (e.g., N2, P3, N400, P600) show changes in amplitude, latency and/or distribution (e.g., Iragui et al., 1993; Faustmann, Murdoch, Finnigan, & Copland, 2007; Kemmer, Coulson, de Ochoa, & Kutas, 2004). For example, investigations of the N400—a component known to be modulated by semantic or contextual relatedness (Kutas & Hillyard, 1980)—have consistently shown reduced amplitudes for this component in older adults compared to younger adults (e.g., King & Kutas, 1995; Federmeier & Kutas, 2005; Faustmann et al., 2007). It has been suggested that this change is related to older adults' difficulty inhibiting the activation of words that are weakly related to the context (King & Kutas, 1995). Latency changes to the N400 as a result of aging are more controversial. In studies using word lists (Harbin, Marsh, & Harvey, 1984), written stimuli (Gunter, Jackson, & Mulder, 1992, 1995, 1998), and delays between context and the critical word (Ford et al., 1996, Woodward, Ford, & Hammet, 1993), delays in latency of the N400 ranging from 35 ms to 120 ms have been observed in older adults relative to their younger peers. However, in studies using natural connected speech, no such delays have been reported (Federmeier, McLennan, & De Ochoa, 2002; Federmeier, Van Petten, Schwartz, & Kutas, 2003; Faustmann et al., 2007). It is possible that those studies which reported latency delays in the N400 made use of less-natural stimuli, which may have increased task demands and therefore slowed processing in older adults.

Though very little data are available regarding the effects of aging on the P600 — a component sensitive to syntactic parsing difficulty (e.g., Osterhout & Holcomb, 1992) — there is evidence that simple syntactic violations (Kemmer et al., 2004) and garden path sentences (Steinhauer, Abada, Pauker, Itzhak, & Baum,

2010) elicit a P600 in older adults, as in young adults. Unlike the N400, no changes in amplitude or latency have been reported as compared to younger adults. Rather, only the distribution of the P600 has been shown to be modulated. For instance, Kemmer et al. (2004) found a more frontal, bilaterally symmetric distribution of the P600 in older adults compared to young adults. Similarly, Steinhauer et al. (2010) found a broader, less posterior distribution of the P600, especially over the right hemisphere, in older compared to younger subjects. Importantly, both of these studies also found decreased accuracy (and slower response times, Kemmer et al.) on the behavioral task included in the study.

One domain of language that has received comparatively little attention to date in the aging literature is that of prosody, a primary feature of language (Magne, Schon, & Besson, 2003). Prosody refers to nonsegmental aspects of language (Nooteboom, 1997) that communicate emotion, convey the intent of an utterance (i.e., whether the utterance is a question or a statement), highlight new or important items in an utterance, disambiguate the meaning of words, and convey syntactic structure (Baum & Pell, 1999; Nooteboom, 1997). Prosody's vital role in language makes it an important candidate for aging investigations. Since prosody interacts with so many aspects of language processing (e.g., syllables, lexicon, syntax, emotions, pragmatics), understanding prosodic processing changes related to aging will ultimately provide information regarding many aspects of language processing.

The limited but growing body of research on prosodic processing in aging has generally documented sensitivity to speech prosody information in aging adults. Older adults are able to exploit prosodic contours to parse syntax and facilitate sentence recall (Cohen & Faulkner, 1983; Kjelgaard, Titone, & Wingfield, 1999; Wingfield, Wayland, & Stine, 1992). In simple phrases, such as the phrases ‘pink and black # and green’ and ‘pink # and black and green’ (where # signals a phrase boundary), previous studies have reported that, like young adults, older adults are able to exploit prosodic cues to assign phrase structure and accurately interpret the stimuli (Taler, Baum, & Saumier, 2006). Examinations of more complex sentences containing temporary syntactic ambiguities also show that older adults exploit prosody in self-paced listening (Wingfield, Kemtes, & Soederberg Miller, 2001), recall (Wingfield et al., 1992), and sentence completion (Kjelgaard et al., 1999) tasks.

While these studies show that older adults are sensitive to prosodic cues, there are indications of age-related differences in prosodic processing as well. For example, when asked comprehension questions about utterances whose interpretation hinged on differences in prosodic boundary only (e.g., ‘Madam, Flower is the name of my cat.’ versus ‘Madam Flower is the name of my cat.’), older adults performed significantly worse than did their younger counterparts (Taler et al., 2006). In a sentence completion task, Kjelgaard and colleagues (1999) found that older adults were as sensitive as younger adults to prosody signaling early and late closure structures and showed a similar late-closure parsing bias when presented with sentence fragments containing neutral prosody. However, older adults were more likely to complete sentence fragments that were presented with early closure prosody with late closure syntax. In the few trials in which this error occurred (6% in the young, 14% in the older adults), older adults were also *faster* at initiating responses than were younger adults. The authors interpret these findings as an indication that the elderly participants were more quickly able to resolve the conflict between prosody and late closure syntax in favor of late closure syntax since this structure is less resource taxing than early closure syntax. These findings also indicate that there are age-related differences associated with the weighting of prosodic cues relative to syntactic structure that require further exploration.

Along a similar line of thinking, there is also evidence in the literature that older adults rely on prosody more heavily than do younger adults. For instance, in a self-paced auditory moving window task, conflicting prosody hindered understanding to a greater extent in older adults than in younger adults, suggesting that older adults may in fact be *more* reliant on prosody than are young adults (Titone et al., 2006). Wingfield and colleagues (1992) presented elderly subjects with sentences containing major clause boundaries with prosodic cues that either conformed or conflicted with those boundaries and asked participants to recall the sentences. They found that older adults were not only more reliant on prosody, reflected by an increased error rate in prosody-syntax mismatches, but were also more likely to change the original syntax to conform with the prosodic boundary. In contrast, younger adults were more likely to change conflicting prosody to conform to the original syntax of the utterance.

Unfortunately, behavioral paradigms have thus far been unable to conclusively determine whether older adults process prosody and integrate prosody and syntax differently from young adults in real-time. To further address this issue, Steinhauer and colleagues (2010) examined behavioral and electrophysiological responses to garden path sentences (with temporary syntactic ambiguities) in older adults (for examinations of young adults using behavioral tasks see Beach, 1991; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Speer, Kjelgaard, & Dobroth, 1996, among others). For example, a sentence that begins with the fragment in (1) below can be followed by either of the two endings (A) or (B):

(1) When a bear is approaching <u>the people</u> …

c) … the dogs come running. *Late Closure*

d) … come running. *Early Closure*

In a previous study examining young adults, Pauker, Itzhak, Baum, and Steinhauer (under revision) replicated the closure positive shift (CPS) in response to prosodic boundaries (Steinhauer, Alter, & Friederici, 1999) and demonstrated P600 and biphasic N400-P600 garden path effects in response to two types of sentences containing prosody-syntax mismatches. Both missing and superfluous prosodic boundaries showed an immediate influence on parsing and overrode typical preferences, inducing garden path effects. In their follow-up investigation with older adults, participants were asked to judge the acceptability of early and late closure sentences with either cooperating or conflicting prosody; Steinhauer et al. (2010) found that older adults accepted conflicting prosody more frequently than younger adults. Interestingly, a CPS was elicited by prosodic boundaries and showed similar timing and distribution in both groups. For prosody–syntax mismatches, older adults demonstrated P600 components with the same latency as those observed in younger adults (Pauker et al., under revision), though the distribution of the P600 was more anterior in the older adults. The older adults did not, however, show the N400 response that young adults showed. This was interpreted as an indication that there are differences related to the integration of conceptual knowledge associated with aging. This study showed that older adults undergo the same initial stages of prosodic processing as young adults, as reflected in the CPS and P600 components. There were, however, differences in later stages of prosodic processing, in particular with respect to integration and (re)analysis of linguistic input, demonstrated by differences in behavioral responses and the absence of the N400 component. In these later stages, older adults appear to modify lexical and syntactic structure such that they conform to prosodic structure, again suggesting that older adults are more reliant on prosody than are younger adults (Wingfield et al., 1992).

As should be evident, the extent to which prosodic processing changes with advancing age remains equivocal, especially in real-time language processing (with only a very small number of investigations to date). The present investigation attempts to advance our understanding of this important issue while controlling a number of critical variables. All of the prior studies which showed an *increased* reliance on prosody in older adults made use of sentences containing temporary syntactic ambiguities, which are difficult to understand because they challenge memory and attentional resources (Abney & Johnson, 1991). When investigations employed less complex syntax, as in the ‘Madam flower’ sentences described above or phrases of the type ‘pink and black and green’, older adults performed equivalent to or worse than young adults (Taler et al., 2006). The question then arises: Are older adults more reliant on prosody only when faced with demanding syntax? A second question arises from Steinhauer and colleagues’ (2010) findings that older adults show differences in the integration of prosody with conceptual knowledge. Is this differential pattern limited to the integration of prosody and sentence comprehension or does it include other areas of language processing and the processing of input from other modalities?

The Present Study

In the present study, we employed the same methodology used to investigate young adults in Chapter 2 to address the two questions raised above. In order to investigate prosodic processing in the absence of complex syntactic violations that accompany the examination of garden path sentences, we collected behavioral and ERP responses from older adults to simple auditory phrases (e.g., ‘bag and bed and cup’) containing either an early or late auditory phrase boundary and compared those responses to a subset of the younger adults examined in Chapter 2. In order to examine the real-time integration of prosodic boundary information with visual input signaling boundaries, auditory phrases were presented simultaneously with visual stimuli consisting of a horizontal array of three pictures corresponding to the three words in the phrase, arranged with a visual boundary between two of the three. The three objects were visually grouped in ways that were either congruent or not with the corresponding prosodic grouping (see Figure 1). Phrases and pictures either shared the same semantic and prosodic/grouping content, or differed in one or both of these domains.

When we examined younger adults, we found that participants were able to detect both grouping and semantic mismatches between auditory and visual information with high accuracy. Both prosodic and semantic mismatches elicited N400 and P600 components that marked the integration of this information between the two modalities. In particular, when a conceptual expectation for either the presence or absence of an auditory boundary was violated (either by the presence of an unexpected auditory boundary or by the absence of an expected boundary), an N400 was elicited. At the point where all the auditory and visual information required to render a match/mismatch judgment became available, which occurred at the same point in both auditory boundary conditions (see Figure 1), a P600 was elicited. These findings were particularly compelling because they extended our understanding of what each of these components marked. In particular, the dominant view of the N400 is that it is a reflection of lexical activation, and its relative facilitation, by a context (see Lau, Phillips, & Poeppel, 2008). Another view is that the N400 reflects the integration of critical semantic or conceptual input with a context (see Lau et al. for a discussion of both interpretations). Our findings lend support to this second interpretation and further extend this view to consider the N400 as a component reflecting integration of conceptual information—broadly construed—with a context. With respect to the P600, one current interpretation in the literature is that the P600 reflects difficulties during the integration of structured bodies of information (e.g., Kuperberg, 2007; Bornkessel & Schlesewsky, 2006, 2008). Our findings supported this view and further showed that the P600 reflects integration between different types of information, even when the information being integrated is from separate modalities (a view consistent with Münte, 2008).

One somewhat surprising ERP finding was that a CPS component was *not* elicited by these stimuli in young adults. The CPS—a large, bilateral, centro-parietal positive-going waveform, most prominent at midline electrodes, that begins 150 and 200 ms after the onset of a prosodic boundary and lasts roughly 500 ms—is elicited by auditory and visual stimuli signaling phrasal boundaries (Steinhauer et al., 1999; Kerkhofs, Vonk, Schriefers, & Chwilla, 2007; Wolff, Schlesewsky, Hirotani, & Bornkessel-Schlesewsky, 2008; Li & Yang, 2009; Pauker et al., under revision; Itzhak, Pauker, Drury, Baum, & Steinhauer, 2010; Steinhauer, 2003; Steinhauer & Friederici, 2001) and from boundaries in stimuli with various types of nonlinguistic content (i.e., nonsense speech, humming) (Steinhauer & Friederici, 2001; Pannekamp, Toepel, Alter, Hahne, & Friederici, 2005). The absence of a CPS in Chapter 2 may be due to the prosodic cues employed (i.e., the pause alone was not sufficient to elicit a CPS), the grammatical structure employed (i.e., the grammar was either not complex enough or a CPS is not elicited by grouping alone), or to the task itself (i.e., the task did not require attention to grammaticality or naturalness, but rather to the integration of input from two modalities). It should be noted, however, that this investigation did not include an analysis of time windows that would potentially show a CPS as this component did not emerge in visual inspection of the waveforms.

In addition to these ERP findings, an additional, and particularly exciting, finding emerged in the examination of young adults (Chapter 2). Though participants were instructed to fixate the center of the screen, where all necessary visual information would be available, 13 of the 20 young adults tested were unable to inhibit horizontal eye movement (referred to in Chapter 2 as the group of 'movers'). Instead, these individuals demonstrated a systematic pattern of eye movement, beginning at the left side of the visual display and continuing rightward, guided by the auditory input. Intriguingly, while there were no cognitive or hearing differences between the ‘movers’ and the other participants, the presence or absence of eye movement modulated the neural response in each group. This suggests that the two groups employed different strategies for this task and that those strategies were reflected in the ERPs (for a full discussion, see Chapter 2). Since eye movement played an important role in the neural response elicited by these stimuli, we paid particular attention to this issue when examining older adults, in order to ensure that any differences observed were related to aging and not to eye movement.

We predicted that behaviorally, older adults would be able to detect both auditory-visual semantic and prosodic/grouping mismatches (henceforth referred to as ‘prosodic mismatches’), though we also expected increased variability and lower overall accuracy in the detection of prosodic mismatches. Despite lower accuracy at detecting prosodic mismatches, we anticipated the same neural correlates (i.e., the N400 and P600) in older adults as those seen in younger adults and did not anticipate a CPS, given its absence in the data for young adults (and the analyses conducted, as noted above). Therefore, there would be some dissociation between the offline behavioral response and the online neural response. However, in keeping with previous ERP findings for the elderly, we expected that the N400 components elicited both by semantic mismatches and prosodic mismatches would be smaller in amplitude (e.g., King & Kutas, 1995; Federmeier & Kutas, 2005; Faustmann et al., 2007). We could not rule out the possibility that these components would also be delayed in latency as a result of the increased task demands of this cross-modal method. We expected that P600 components would be more frontally distributed (Kemmer et al., 2004; Steinhauer et al., 2010) and did not expect any latency or amplitude differences.

Methods

Subjects

All subjects were right-handed (as determined by the Edinburgh Handedness Inventory; Oldfield, 1971) native speakers of English with no history of neurological impairments and with normal or corrected-to-normal vision. To confirm that all participants' hearing was within the range of normal, audiometric screening ensured that pure tone averages (averaged across 500, 1000, and 2000 Hz) were less than 25 dB HL in the better ear. Data from 11 older adults (ages 65 to 80 years, mean = 70, sd = 3.5 years; 6 female) were acquired at McGill University and compared to data from 13 young adults (ages 18 to 25 years, mean = 22 years, sd = 1.5 years; 7 female), a subset of those presented in Chapter 2. This subset of young participants was chosen because they were comparable to the older adults in that every subject moved their eyes when observing the stimuli (discussed below). Five additional older subjects were tested but their data were excluded from analysis for the following reasons: one male subject was excluded as a result of equipment malfunction; three female and one male subject were excluded because they did not achieve a minimum score of 25 on the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005). Written informed consent was obtained prior to testing and subjects were financially compensated for their time. Prior to testing, subjects were also screened on tests of auditory working memory (Lehman & Tompkins, 1998), vocabulary (Boston Naming Test

[BNT]; Kaplan, Goodglass, & Weintraub, 1978), and cognition (MoCA; Nasreddine et al., 2005) (see Appendix 2).

Stimuli

Stimuli consisted of visual displays and auditory phrases containing three nouns. In each trial, participants saw a visual display of three nouns with a physical visual boundary, henceforth referred to as ‘wall’, between either nouns one and two (Left wall) or nouns two and three (Right wall). The visual display was presented simultaneously with the onset of an auditory phrase (e.g., ‘bag and bed and cup’) which included a pause to create either an early or late auditory boundary.

Pictures and auditory recordings of 16 easily picturable monosyllabic CVC nouns that begin and end with a stop consonant (e.g., ‘bike’, ‘pig’, ‘cat’, etc.), were chosen for the visual displays (see Figure 1). Visual stimuli were pictures taken from the Snodgrass and Vanderwart Picture Inventory (Snodgrass & Vanderwart, 1980). Pictures were resized and positioned equidistantly inside a white rectangle by a trained graphic artist to ensure that all items were equally prominent. As can be seen in Figure 1, this created visual displays of three equidistant objects with a vertical thick black line (the ‘wall’) at one of two possible positions: the ‘Left’ wall between the first and second objects (which is suggestive of an early auditory boundary) or the ‘Right’ wall between the second and third objects (which is suggestive of a late auditory boundary). Objects were sized and positioned such that it was possible to fixate the center of the screen and attend to the objects without the need to move the eyes to see all the relevant elements. This was done to avoid horizontal eye movements and corresponding ocular artifacts in the EEG signal. Auditory stimuli consisted of multiple repetitions of all CVC nouns as well as the conjunction 'and' pronounced with level prosody at a normal speaking rate by an adult female native speaker of English and recorded in a sound-attenuated booth using a portable digital recorder (Marantz Professional PMD670) and a head-mounted microphone (AKG Acoustics C420) and transferred to a computer using an external cardreader (Macally MCR-6U). The mean duration of all items was calculated and the token of each item closest to the overall mean was selected for inclusion in the experiment. Vocalic and voiceless segments of nouns were removed or replicated as needed using Praat software (Boersma & Weenink, 2006) until the duration of all nouns was equivalent within 1 ms (mean = 376.4 ms; sd = .26 ms). The token of 'and' had a duration of 253 ms. Three naïve listeners judged the quality of each item and determined that all words sounded natural and not manipulated.

The 16 monosyllabic CVC nouns were divided into eight sets of four items, or '4-tuples', such that each noun occurred in two 4-tuples. The schema employed to create the 4-tuples is shown in Appendix 1. (The reasoning behind, and advantages of, using these 4-tuples will be discussed below.) Phrases containing three nouns each, similar to those used in previous studies (e.g., Streeter, 1978; Taler et al., 2006), were created from the 4-tuples. For example, as can be seen in Appendix 1, from 4-tuple A (bag dog cup bed) the triplet 'bag - bed - cup' was generated. Four unique phrases of this kind could be generated from each 4-tuple. Each of these unique phrases could be permuted in six ways (e.g., 'bed and bag and cup', 'cup and bed and bag', 'bag and cup and bed', etc.). The six permutations of the four phrases from the eight 4-tuples generated 192 stimuli (6 x 4 x 8). Pictures were arranged to depict each of the auditory stimuli. To create semantic mismatches (discussed below), the fourth noun, which had not been used in the triplet, replaced the middle noun in the visual display. This schema ensured that replacement nouns in phrase-picture pairs were not highly related semantically or phonologically and that every noun could be replaced by two other nouns (one from each 4-tuple), limiting predictability in semantic violation conditions.

Two phrasal groupings (with boundaries marked by '#') are possible with this phrase type: an Early boundary **[EB]** ('bag # and dog and bed') and a Late boundary **[LB]** ('bag and dog # and bed'). The 192 items were randomly divided into the two main boundary conditions (i.e., 96 phrases were presented with an Early boundary and 96 phrases were presented with a Late boundary). A 450 ms interval of silence served as the auditory boundary, since, as discussed above, a pause has been shown to be a sufficient cue to a phrase boundary (Streeter, 1978; Nagel et al., 1996). As discussed above, the black 'wall' separating two of the objects served as the visual boundary, where a 'left wall' corresponded to an early boundary and a 'right wall' to a late boundary (see Figure 1 for an example). Thus, in both modalities there was only one cue to the phrasal grouping. The crossing of the two auditory and the two visual grouping conditions led to the first four conditions: two matching ***C**ontrol conditions* (EB + left wall, LB + right wall) and two ***P**rosodic mismatch conditions* (EB + right wall, LB + left wall). Moreover, each of these four conditions was also combined with a semantic mismatch where the second object in the visual modality differed from the second noun in the auditory modality. This resulted in four additional conditions, two pure ***S****emantic mismatch* and two ***D****ouble mismatch* (prosodic plus semantic mismatch) conditions. Since the overall EOG and ERP patterns were most strongly driven by the *auditory* input, most analyses will compare conditions separately for (i) early auditory boundaries (EB) and (ii) late auditory boundaries (LB). The labeling of the eight conditions reflects (a) the position of the auditory boundary (E vs. L) and (b) the type of main condition (C, P, S, D). Thus, the condition with an early auditory boundary and a semantic mismatch will be referred to as ES, while a prosodic mismatch in trials with a late auditory boundary will be referred to as LP, and so on (see Figure 1 for examples of the conditions and Figure 2 for condition labels). Each of the 192 auditory items was presented twice: once in a matched Control condition (EC or LC) and once with a visual display that conflicted with the auditory phrase (mismatches). This yielded a total of 384 trials (192 x 2) presented to each participant.

In the matching Control conditions, visual and auditory tokens matched each other both in grouping and in semantics. Thus subjects heard ‘bag and dog # and bed’ and saw BAG DOG || BED. In the EP and LP subconditions, visual stimuli differed from auditory stimuli in phrasal grouping only. For example, in the LP subcondition, the Left wall visual stimulus was paired with the Late boundary auditory stimulus such that subjects heard ‘bag and dog # and bed’ but saw BAG || DOG BED). In the ES and LS subconditions, visual and auditory stimuli shared the same phrasal boundary, but the *second* noun differed between the modalities. For example, subjects heard ‘bag and dog # and bed’ and saw BAG CUP || BED. In the ED and LD subconditions, visual and auditory stimuli differed both in phrasal grouping and the second noun. For example, in the LD subcondition, subjects heard ‘bag and dog # and bed’ and saw BAG || CUP BED. The 4-tuples used to create phrases ensured that each noun appeared an equal number of times in every position in the phrase across the experiment and was replaced by or replaced another noun an equal number of times. This minimized predictability for participants.[1]

The three mismatch conditions in each (Early and Late) boundary condition were comprised of 32 trials each. Since half the stimuli were presented in each boundary condition and half were control items, for each trial there was an equal probability of being presented with either boundary or a correct or incorrect item, thus minimizing predictability and eliminating the need for filler items. Each auditory token was paired with four visual tokens. Each participant heard every auditory token twice, once in the Control subcondition and once in a mismatch subcondition. The same auditory token was not presented in the same half of the experiment. From these stimuli, three lists were created such that every control token was presented along with only one of the mismatch subconditions; as noted, each subject only heard each phrase twice in the experiment, divided across experimental halves. Each phrase was presented six times across the three lists, resulting in 1152 stimuli overall, but only 384 stimuli per list or per participant.

---

[1] While this schema may have allowed for statistical learning, this did not create a problem for our design because our focus was to ensure that differences between ERP responses would not be attenuated due to predictability of a semantic violation. Any statistical learning that may have resulted from the design is orthogonal to the aims of the present study and there is no reason to believe it would modulate a response to prosodic or semantic input.

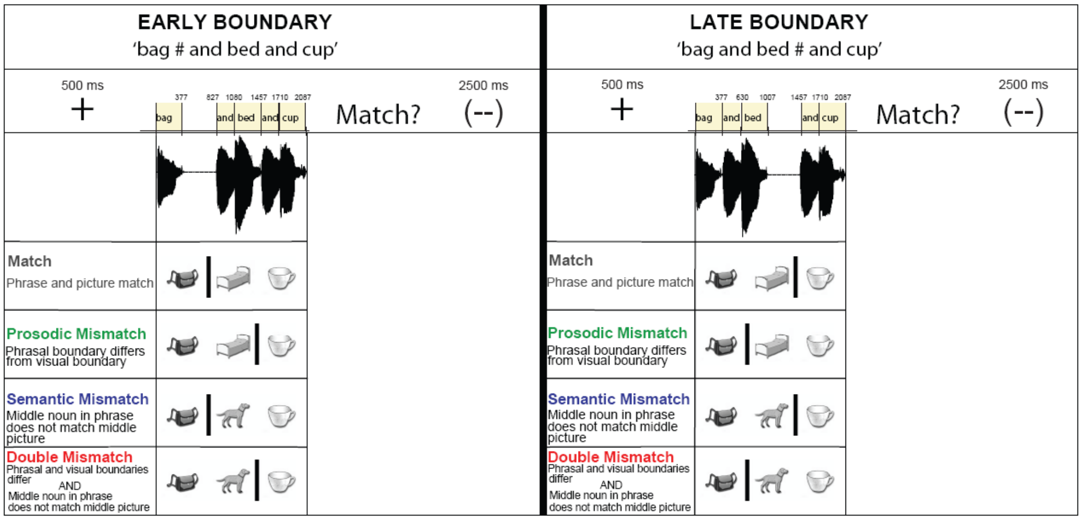

Figure 1. Trial timing and Conditions. Early and Late boundary stimuli (in ms) shown above the auditory input. Trials began with a fixation cross presented for 500 ms followed by auditory and visual stimuli with a simultaneous onset. Twenty ms after the offset of stimuli, the question 'Match?' appeared until participants responded. Participants were then able to blink for 2500 ms until the onset of the subsequent fixation cross. Examples of the visual stimuli presented for each condition are shown beneath the speech waveform.

Procedure

Following the hearing assessment, cognitive and memory testing, and electrode placement, participants were seated in front of a monitor in a sound-attenuating chamber, wearing insert earphones (Etymotic Research model ER3-14A). Participants were shown the effects of eye blinks and movement on the EEG signal and were aware that eye movement would create unwanted artifacts in the data. They were asked to blink and move only when indicated in the experiment and to fixate the center of the screen during stimuli presentation. Participants were told they would see pictures and hear phrases and were asked to indicate whether or not the two pieces of information presented matched.

An initial training session began with the individual aural and visual presentation of each of the 16 nouns. Participants were shown each picture a second time and asked to name each one to ensure they recognized each item. In the rare case that the wrong name was given, the experimenter corrected the participant, who repeated the correct name. The participants were then shown the white rectangle, identified as a shelf, followed by the two ways to divide the shelf with a wall. Participants then heard a phrase and were asked to choose which of two pictures best matched that phrase. If a participant responded incorrectly to any of these items, the experimenter would repeat the phrase, which alerted participants to semantics and phrasal groupings. Finally, one typical trial from the Prosodic Mismatch subcondition was presented. The phrasal grouping was made explicit to listeners subsequent to analyses of data from pilot subjects who were not specifically made aware of the contrast, which showed that these individuals failed to correctly reject any prosodic mismatches. Participants were not told that Double Mismatch trials could occur. Following this training session, participants saw 30 practice trials to become familiar with the task. No feedback related to accuracy was given once the 30 practice trials began.

The ERP experiment session consisted of one list of 384 items divided into four blocks of 96 items, lasting about 12 minutes each. Two randomization schemas were created for each list and each list and randomization was assigned evenly across groups and male/female participants. Hand assignment was also counterbalanced.

Trial and stimulus timing is presented in Figure 1. Trials began with a fixation cross (“+”) presented in the center of the computer screen for 500 ms. Following this, the visual and auditory stimuli presentation began. The picture remained on the screen for the duration of the auditory stimulus (2087 ms). Twenty ms after stimulus offset, the question “Match?” appeared in the center of the screen. Participants had no time limit to indicate whether the information from the two modalities was the same or not. After responding, participants were shown the following symbol, which indicated the 2500 ms interval in which they were instructed to move and blink: (--).

Electrophysiological Recording

EEG was continuously recorded (DC mode; 500 Hz sampling rate; Neuroscan Synamps2 amplifier, Neuroscan-Compumedics, Charlotte, North Carolina, USA) from 32 cap-mounted Ag/AgCl electrodes (Electro-Cap International, Eaton, Ohio, USA) referenced to the right mastoid and arranged according to the extended 10-20 system (Jasper, 1958) (impedance <5 kΩ). To measure blinks and eye movements, vertical electro-oculogram (VEOG; from electrodes placed above and below the right eye) and horizontal EOG (HEOG; from electrodes placed at the outer canthus of each eye) were recorded.

Data Analysis

In Chapter 2, before beginning EEG data analysis, we addressed the issue of eye movements because it was critical to ensure that ERP data were not contaminated by these artifacts. We found a robust and systematic pattern of horizontal eye movement (saccades) in the HEOG electrodes in 13 young adult participants, while the remaining participants (n = 7, the ‘nonmovers’) did not show movement in these electrodes. Importantly, in Chapter 2, we found that the presence or absence of eye movement played a crucial role in the processing of prosodic and semantic information in this task. Specifically, the neural correlates elicited by grouping mismatches differed between those who moved their eyes and those who did not move their eyes, suggesting that a different strategy was employed by each group. To ensure that any differences observed in the present study were related to age rather than eye movement, it was necessary to examine the pattern of eye movements in the older adults so that they could be compared to a similar group of young adults with regard to eye movement. Unlike in young adults, older participants did not differ with regard to their eye movement patterns. Instead, every older adult exhibited eye movement. As can be seen in Figure 2, which displays the HEOG electrode for young movers and older adults, the eye movements of the older adults displayed the same overall pattern as did the eye movements of the young movers. It should be noted, however, that peak amplitudes of the HEOG waveforms for the older adults appeared much larger than for the young adults because older adults sat closer to the screen so that they could easily see the visual stimuli and therefore their saccades were larger. Since all the older adults were movers, they were compared to the young adult movers alone in order to ensure that any differences found are related to aging as opposed to eye movement.

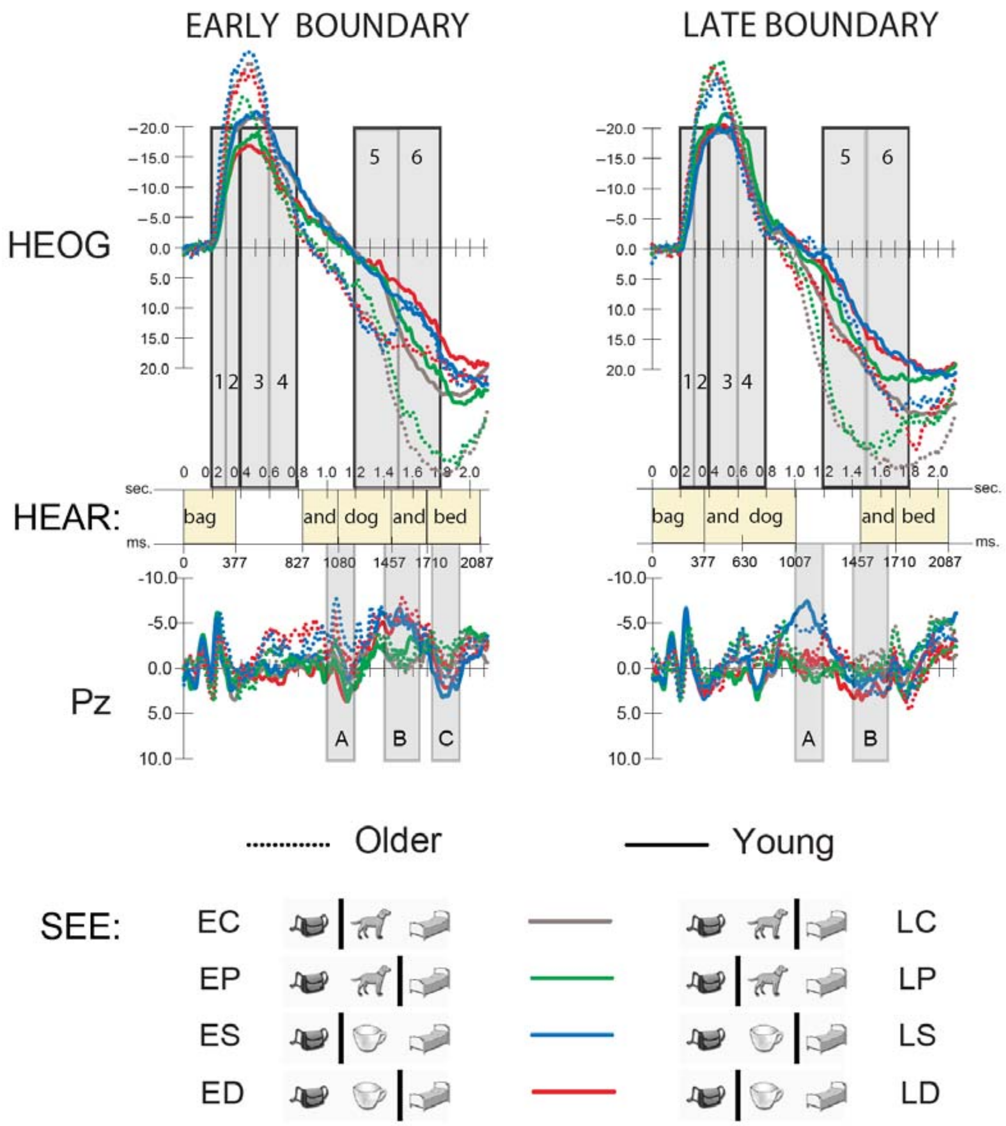

Figure 2. ERP and HEOG for young and older adults. ERP waveforms at electrode Pz (middle) and HEOG waveforms (top) are presented for young (solid) and older (dotted) adults for all conditions in both the Early and Late boundary conditions.

Behavioral data were analyzed for accuracy using repeated measures analyses of variance (ANOVA) with factors **Boundary** (2: Early, Late pause) x **Wall** (2: Left, Right visual wall) x **Semantics** (2: match, mismatch) x **Group** (2: older, young) to maintain consistency between analyses of ERP data and eye movement in the previous investigation. ERP data were analyzed using EEProbe (ANT, The Netherlands). Single subject averages were computed separately for all conditions following data pre-processing, which consisted of filtering (0.16-30 Hz bandpass) and artifact rejection at electrodes Fp1, Fp2, and VEOG (but not HEOG). Averages were based on correct trials only[2] and were computed for 2200 ms epochs beginning at the onset of noun1 of each phrase and ending roughly 100 ms after the offset of noun3. A 200 ms baseline began at the onset of noun1 and ended 200 ms later, 177 ms before the end of noun1. This baseline was chosen because conditions did not differ auditorily within the first 200 ms and thus there was no information to distinguish between conditions. Within this epoch, ERP components were quantified by means of amplitude averages in a series of time windows (TW), detailed below. Figure 2 displays detailed timing of all events in the auditory stimulus as well as the time windows analyzed (labelled A – C). Because auditory stimuli were well-controlled and all trials were of equal length, we present the entire epoch (at electrode Pz) in all figures.

Following visual inspection, ERPs were examined at midline[3] (Fz/FCz/Cz/CPz/Pz/Oz) electrodes in three TWs: (A) 1000 – 1200 ms, (B) 1400 – 1650 ms, and (C) 1730 – 1930 ms. The Early and Late boundary conditions were

---

[2] There was one exception to this. For the EP and LP conditions, 4 older adults consistently responded incorrectly (i.e., responded ‘match’ to mismatch conditions) suggesting that they did not detect prosodic mismatches. For these 4 individuals, in the EP and LP conditions only, averages were based on the incorrect trials to avoid having to exclude their data completely. The option of splitting the older adults into two groups (i.e., those that detected prosody and those that did not) was considered but it was decided that this was not a reasonable option due to the extremely small group sizes that would emerge (n = 7 and n = 4, respectively). These groups would not provide enough power to analyze ERP data and would result in increased noise in the data. Therefore, the group was analyzed as a whole to examine whether older adults on average perform similarly to younger adults. The issue of variability within older adults will be addressed in the Discussion and explored further in future investigations.

[3] Only midline electrodes were examined to ensure that eye movements were *not* included in analysis. Horizontal eye movements typically affect the EEG at lateral electrodes (especially at frontal sites) whereas midline electrodes are less susceptible to such artifacts (see also Bornkessel & Schlesewsky, 2010). In particular, signals at the parietal midline electrode, Pz, where the majority of the ERP effects were found to be most prominent, did not seem to be influenced by eye movements at all.

analyzed in separate repeated measures ANOVAs because effects were elicited in different TWs due to differences in events between the boundary conditions. An additional topographical factor **Anterior/Posterior** (AP; 6) was included in the analyses. All interactions between distributional factors and experimental factors were followed up via separate ANOVAs for experimental factor at each electrode. Greenhouse-Geisser corrections were employed where applicable.

Results

Behavioral Accuracy Results

Behavioral accuracy data for all conditions are illustrated in Figure 3, with statistical analyses presented in Table 1. Looking at Figure 3, it is obvious that most conditions reached ceiling levels for accuracy (EC young = 95.83%, LC young = 96.47%, EC elderly = 99.05%, LC elderly = 98.11%, ES young = 97.36%, LS young = 98.32%, ES elderly = 96.88%, LS elderly = 98.30%, ED young = 100%, LD young = 98.56%, ED elderly = 98.86%, LD elderly = 99.43%), with the exception of the pure prosodic mismatch conditions, which are circled in the figure (EP young = 81.25%, LP young = 83.17%, EP elderly = 57.10%, LP elderly = 62.22%). These conditions were responded to significantly less accurately than all other conditions by both groups, reflected by a highly significant Boundary x Wall x Semantics interaction (i.e., main effect of prosodic mismatch in the absence of semantic mismatches). Our 2 x 2 x 2 (Boundary x Wall x Semantics) design resulted in pure prosodic mismatches (EP and LP conditions) being included in the set of semantic match conditions, as well as in the subset of Boundary x Wall interactions, but not in the subsets of either Boundary or Wall conditions (because the EP condition involved an *Early* auditory boundary and a *Right* wall, whereas the LP condition involved a *Late* auditory boundary and a *Left* wall). Therefore, the significant main effect of Semantics as well as the significant Boundary x Wall interaction that may be seen in Table 1 were primarily driven by lower accuracy for the pure prosodic mismatch conditions as well (as evident in Figure 3). Interestingly, though both groups show the same pattern overall, additional group differences also emerged for these conditions. Specifically, in both the EP and LP conditions, the young adults were far more accurate than the older adults. In the most relevant follow-up analysis of the Boundary x Wall x Semantics x Group interaction, although both groups showed a highly significant Boundary x Wall x Semantics interaction (Young F[1,12] = 127.76, p < .0001; Elderly F[1,10] = 739.96, p < .0001), this pattern was particularly strong in the older adults indicating larger differences between match and mismatch conditions in that group compared to the younger adults.

Table 1. Statistical analysis of behavioral accuracy for both groups

Effect	df	F-value	P-value
Boundary	***1,22***	**4.18**	**0.0532**
Boundary x Group	*1,22*	1.02	0.3235
Semantics	***1,22***	**806.42**	**<.0001**
Semantics x Group	***1,22***	**95.94**	**<.0001**
Boundary x Semantics	*1,22*	1.68	0.2086
Bound x Sem x Group	*1,22*	0.05	0.8317
Wall	*1,22*	1.03	0.3218
Wall x Group	*1,22*	2.47	0.1307
Boundary x Wall	***1,22***	**611.74**	**<.0001**
Bound x Wall x Group	***1,22***	**152.06**	**<.0001**
Wall x Semantics	***1,22***	**6.91**	**0.0153**
Wall x Semantics x Group	*1,22*	0.64	0.4320
Bound x Wall x Sem	***1,22***	**768.10**	**<.0001**
Bound x Wall x Sem x Group	***1,22***	**155.02**	**<.0001**

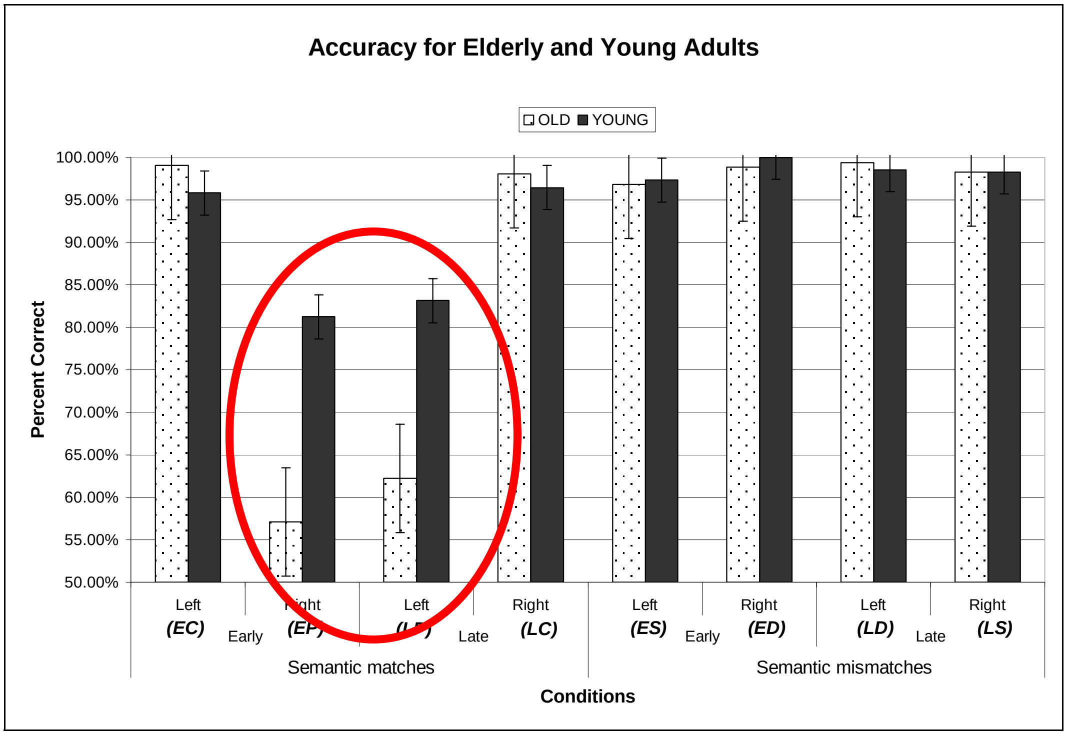

Figure 3. Behavioral accuracy. Accuracy for older adults (gray) and young adults (black) for all conditions. Responses arranged by factors (semantics, then boundary condition, then wall position) with condition names included in parentheses. Scale begins at 50% in order to illuminate significant differences. Error bars reflect standard error.

ERP Results

In all conditions and in both groups, the onset of the auditory stimulus elicited the typical and expected N1/P2 sequence of components associated with word onsets. Note, however, that due to the concurrent onset of auditory and visual stimuli, the onset components elicited by auditory stimuli were superimposed by an additional N1/P2 pattern evoked by the visual stimuli. Similar (auditory) onset components (although of smaller amplitude) can be seen in each boundary condition elicited by the conjunction ‘and’ following the respective pause. In addition to these components, semantic and prosodic mismatches elicited various neural responses which are described in turn in what follows. Figure 4 shows ERP waveforms for all conditions for both groups.

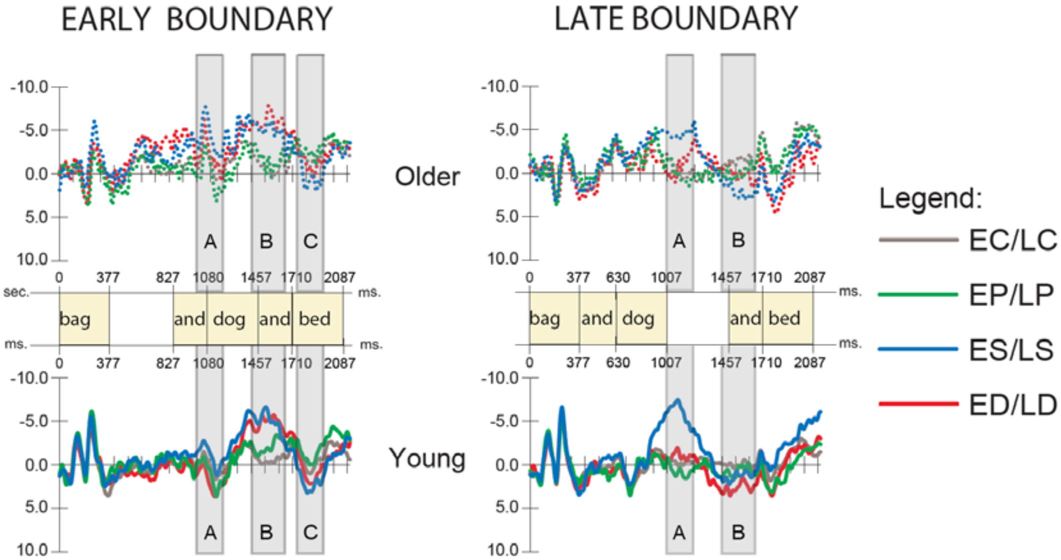

Figure 4. ERPs at electrode PZ. All 4 conditions in both boundary conditions for the older adults (dotted, top) and the younger adults (solid, bottom) for the Early boundary (left) and the Late boundary (right).

Table 2. Statistical analyses of ERP effects in the Early boundary condition

MIDLINE	***df***	**600 - 800**	**1000 - 1200**	**1400 – 1650**	**1730 – 1930**
Wall	*1,22*	——	11.88*	——	——
Semantics	*1,22*	——	4.39*	23.15**	6.42*
Wall x AP	*5,110*	——	7.58*	——	——
Sem x AP	*5,110*	——	——	23.72**	——
Sem x AP x Group	*5,110*	——	——	5.14*	6.15*
Wall x Sem x AP	*5,110*	——	——	4.51*	——

Note: No other effects or interactions reached significance

* p < 0.05

** p < 0.0001

Table 3. Statistical analyses of ERP effects in the Late boundary condition

MIDLINE	***df***	**1000 - 1200**	**1400 - 1650**	**1730 - 1930**
Wall	*1,18*	16.02**	——	5.65**
Semantics	*1,18*	27.00***	——	——
Sem x Group	*1,18*	——	——	21.18**
Wall x Sem	*1,18*	12.19**	——	——
Wall x AP	*5,90*	4.80**	——	——
Wall x AP x Group	*5,90*	3.06*	——	3.41**
Sem x AP	*5,90*	6.81**	——	2.86*
Wall x Sem x AP	*5,90*	——	2.99*	——

Note: No other effects or interactions reached significance

* p < 0.05

** p < 0.0001

Semantics

We expected semantic mismatches between auditory and visual input in both Early and Late boundary conditions to elicit the same components in the older adults as in the younger adults. Specifically, noun2 in the ES/LS and ED/LD conditions, which did not match the picture, should have elicited both an N400 and P600 response. In keeping with previous research (e.g., Iragui et al., 1993; Faustmann et al., 2007; Kemmer et al., 2004), we anticipated that in older adults the N400 would have a smaller amplitude and possibly be delayed in latency and the P600 to be more frontally distributed. Figure 5 and Figure 6 illustrate these components in the Early and Late boundary conditions, respectively, using difference waves. The N400 can also be seen in ERP waveforms for both groups in Figure 2.

Semantic N400. Noun2 began at 1080 ms in the Early boundary condition. Both groups showed an N400 component in the ES and ED conditions that began at roughly 1280 ms in the old and 1230 ms in the young and peaked at roughly 1530 ms in the old and 1500 ms in the young, approximately 400+ ms after the onset of noun2. This was confirmed statistically in the 1400 – 1650 ms TW (see Table 2) by a highly significant main effect of Semantics and a Semantics x AP interaction. The additional Semantics x AP x Group interaction points to distributional differences of the N400 related to aging. The young adults showed a broader N400 distribution than the older adults, who showed a slightly more posterior distribution of the N400 (see Table 4 for follow up analyses at each electrode). These distributional differences are also illustrated in the voltage maps displayed in Figure 5. Surprisingly, there were no significant differences in amplitude between the groups.

In the Late boundary condition, noun2 began at 630 ms and the semantic N400 can be seen in TW 1000 – 1200 ms. The LS condition will be addressed before turning to the LD condition, as the effects in these conditions differed. In the young movers, there was a negativity in the LS condition that began at 870 ms and continued until 1330 ms with a peak latency of roughly 450 ms after the onset of noun2. In the older adults, the negativity in the LS condition appeared to begin later, at 930 ms, and peaked later as well, at 1200 ms, or, roughly 550 ms after the onset of noun2. In the LS condition, this effect continued until 1350 ms in the older adults, whereas it lasted until 1500 ms in the LD condition. The N400 in both groups was confirmed by a highly significant main effect of Semantics and a Semantics x AP interaction in the 1000 – 1200 ms TW. Follow-up analyses of the Semantics x AP interaction showed a broad distribution of the N400 that was strongest over parietal and occipital electrodes (see Table 4), which can also be seen in the voltage maps displayed in Figure 6. Again, there were no group differences in amplitude of the N400 component.

As reported in Chapter 2, in the young movers, the N400 in the LD (double mismatch) condition was significantly lower in amplitude than the N400 in the LS (semantic mismatch) condition because of the simultaneous semantic and prosodic processing. That is, the semantic N400 was superimposed by a concurrent prosodic positivity (see below). As these two main effects were opposite in polarity, they partly cancelled each other out in the double mismatch condition. In addition to their co-occurrence, the two types of mismatches also interacted with one another. In the older adults, the negativity in the LD condition began at 1080 ms, later than the negativity in the LS condition, but peaked at the same time as in the LS condition (1200 ms), though the peak in the LD condition was significantly lower in amplitude than in the LS condition (as in the young adults). These effects were confirmed by a significant Wall x Semantics interaction and can be seen in Figure 6**.** Importantly, the young and older adults showed similar interactions, though it appears that the N400 for the older adults peaked at a later latency than the N400 for younger adults.

Table 4. Follow-up analyses of interactions at each electrode (in each group when factors interacted with group). Follow-up analyses presented for Semantic effects then Wall effects in each auditory boundary condition.

			Early Boundary			**Late Boundary**	
			1400 – 1650	**1730 - 1930**	**1000 - 1200**	**1000 – 1200**	
Group	Electrode	*df*	**Sem x AP x Group**	**Sem x AP x Group**	**Wall x AP**	**Sem x AP**	**Wall x AP x Group**
elderly	Fz	*1,10*	——	13.13*			11.80*
	FCz	*1,10*	——	6.21*			9.56*
	Cz	*1,10*	——	5.31*			7.64*
	CPz	*1,10*	9.07*	6.63*			5.19*
	Pz	*1,10*	21.45*	8.94*			——
	Oz	*1,10*	62.84**	——			——
young	Fz	*1,12*	——	——			——
	FCz	*1,12*	——	——			8.03*
	Cz	*1,12*	6.32*	——			9.08*
	CPz	*1,12*	19.16*	——			11.04*
	Pz	*1,12*	26.84*	——			11.27*
	Oz	*1,12*	27.54*	——			9.31*
omnibus	Fz	*1,22*			12.41*	8.69*	
	FCz	*1,22*			14.03*	10.48*	
	Cz	*1,22*			13.51*	16.81*	
	CPz	*1,22*			11.52*	28.21**	
	Pz	*1,22*			7.69*	38.72**	
	Oz	*1,22*			——	52.08**	

Semantic P600. In the young adults in the Early boundary condition, immediately following the semantic N400, a positive component – the P600 – was elicited by the ES and ED conditions. The P600 began at 1720 ms in the young and at 1750 ms in the older adults and peaked at roughly 1840 ms and 1860 ms in the young and the old, respectively. In the older adults, the P600 in the ED condition showed the same timing as in the ES condition, but appeared to be smaller in amplitude (see Figure 5). The semantic P600 was supported statistically in TW 1730 – 1930 ms by a significant main effect of Semantics. Group differences in distribution are evident from a significant Semantics x AP x Group interaction which was followed up in each electrode in each group (see Table 4). Older adults show a broad, but more frontal, P600. The interaction did not resolve to a significant main effect of semantics in any electrode in the young adults.

In the Late boundary condition, the P600 for the semantic mismatch at noun2 in the LS and LD conditions began at roughly 1350 ms in the young adults, continued until 1650 ms and peaked roughly 800 ms after the onset of noun2. In the older adults, the P600 in the LS condition began at 1340 ms, continued until 1850 ms and peaked at 1600 ms, or 970 ms after the onset of noun2, considerably later than the peak in the young adults (see Figure 6). Even more surprisingly, the P600 in the LD condition in the older adults began at 1600 ms (towards the end of the effect in the LS condition), peaked at 1780 ms, and then followed the same morphology as in the LS condition. In Chapter 2, we examined the semantic P600 for the Late boundary in the 1400 – 1650 ms and 1730 – 1930 ms TWs. As can be seen in Table 3, when the young and old groups were examined together, there was no significant main effect of semantics in the 1400 – 1650 ms TW. Rather surprisingly, there was only a Wall x Semantics x AP interaction which did not resolve to a significant Wall x Semantics interaction at any electrode. In the 1730 – 1930 ms TW, however, there was a significant Semantics x Group interaction. While the main effect of Semantics was significant in both groups (young: F[1,12] = 3.67, p = .0171; old: F[1,10] = 12.66, p = .0052), it is clear from Figure 6 that this component is both larger and later in the elderly.

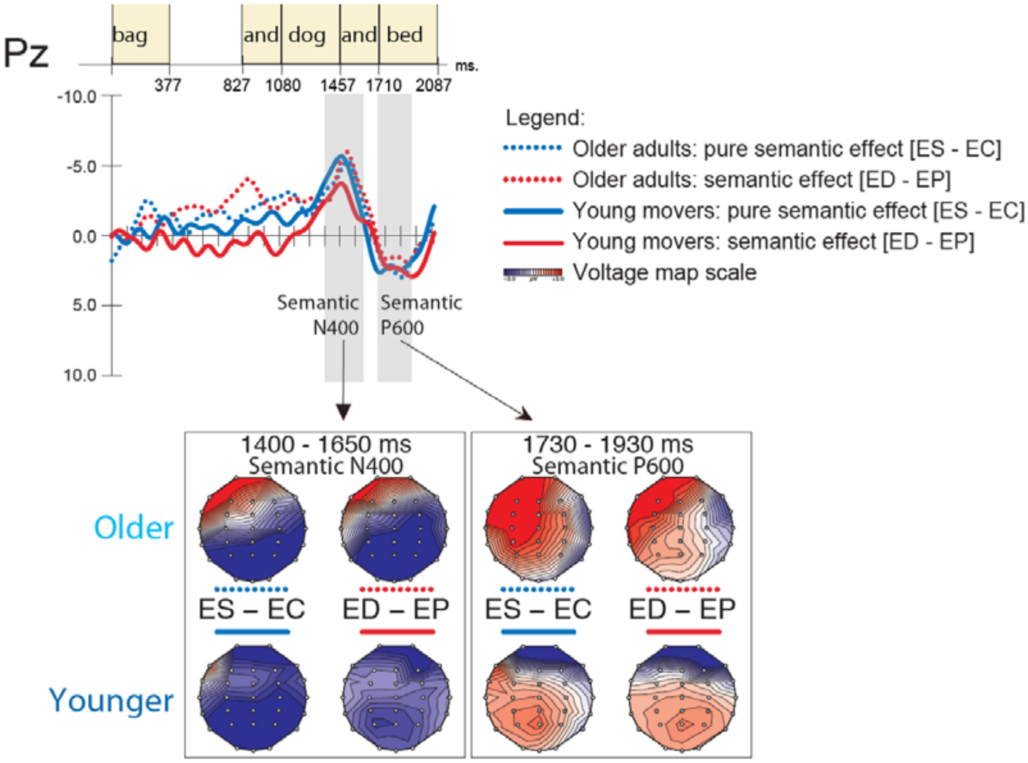

Figure 5. Early boundary semantic effects on ERP (low pass filtered at 7 Hz) using difference waves for old and young adults. Difference waves of semantic mismatch conditions with the corresponding semantic match control condition subtracted for the older adults (dotted) and younger adults (solid) illustrate the pure semantic effects in the absence of prosodic effects. Semantic mismatches elicited N400 and P600 components (shaded) in both ES and ED conditions. The N400 was larger in pure semantic conditions (blue) compared to double mismatches (red). Voltage maps confirm the typical N400 and P600 distributions expected in young adults and more frontal distribution of the components in older adults.

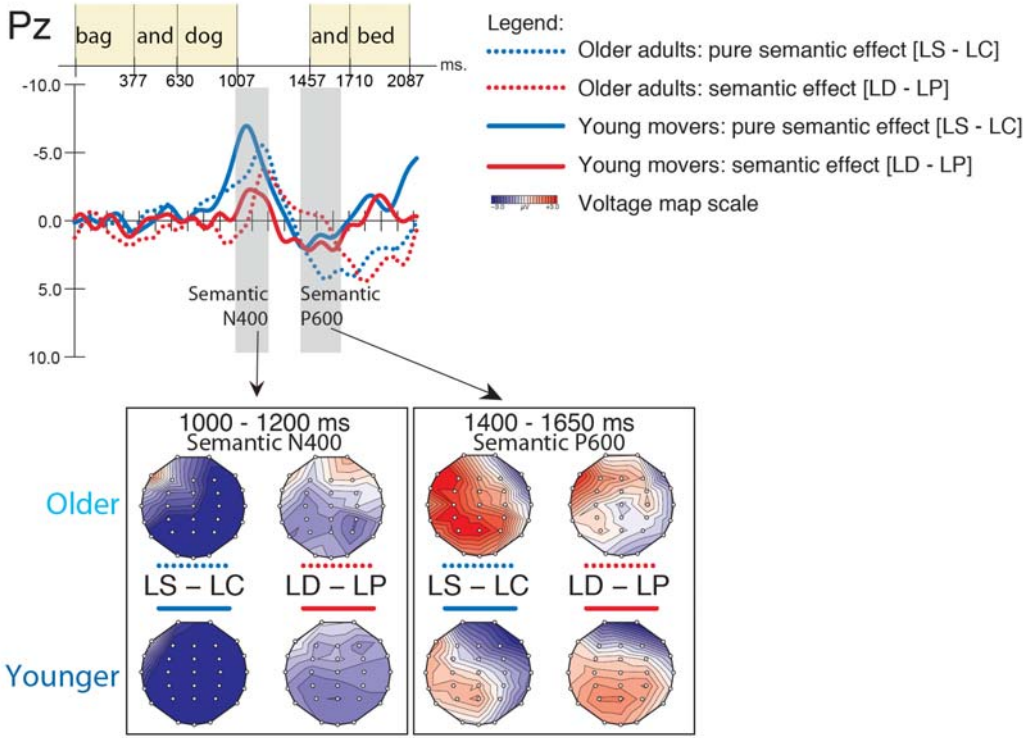

Figure 6. Late boundary semantic effects on ERPs (low pass filtered at 7 Hz) using difference waves for old and young adults. Difference waves of semantic mismatch conditions with the corresponding match control condition subtracted illustrate pure semantic effects in the old (dotted) and young (solid) adult groups. Semantic mismatches elicited an N400 component in the 1000 – 1200 ms TW (shaded) in LS (blue) in both groups. LD (red) elicited an N400 that was significantly reduced compared to LS. Voltage maps confirm this N400 pattern. Both LS and LD conditions elicited a P600 component in both groups in the 1400 – 1650 ms TW (shaded). Voltage maps confirm the typical P600 distribution expected in young adults and a more broad distribution in older adults.

Prosody

We expected prosodic mismatches between auditory and visual input in both Early and Late boundary conditions to elicit the same components in the older adults as in the younger adult movers. Specifically, the onset of the first instance of ‘and’ in the Late boundary should have elicited a positivity in the 1000 – 1200 ms TW. While this positivity did not reach significance when time-locked to the onset of the pause in the Early boundary condition in young adults in Chapter 2, we nonetheless examined the 1000 – 1200 ms TW here because there was an indication that the positivity was present in the group of movers. The onset of the second instance of ‘and’ in the Early boundary should have elicited a negativity in the 1400 – 1650 ms and 1730 – 1930 ms TWs. Since these components can be interpreted as P600-like and N400-like responses, we anticipated the same differences in distribution and/or amplitude for these components between the younger and older adults described above. Figure 7 and Figure 8 illustrate these components in the Early and Late boundary conditions, respectively, using difference waves; ERP waveforms for both groups can be seen in Figure 2.

Early Boundary Prosodic Positivity. Both groups displayed an increased positivity in the EP and ED conditions relative to their control conditions, as illustrated in Figure 7. Difference waves in Figure 7 show that the positivity in the older adults began at 900 ms, peaked at roughly 1150 ms, and continued until 1400 ms. In the young movers, the positivity began at 1050 ms and persisted until 1240 ms. The presence of this component was confirmed by a significant main effect of Wall and a Wall x AP interaction in the 1000 – 1200 ms TW. Follow-up analyses presented in Table 4 showed that this component had a broad fronto-central distribution in both groups.

In addition, and somewhat surprisingly, a main effect of semantics emerged in this TW. In Figure 5 it appears that there is some negativity in the ES and ED conditions in the older adults and in the ES condition only in the younger adults, which begins prior to the onset of noun2, and therefore prior to the onset of the relevant semantic information. It is unclear how to interpret this small but significant effect as the second noun was not presented until 1080 ms; nonetheless, it appears unrelated to the prosodic positivity found in the same TW.

Early Boundary Prosodic Negativity. In the young adults, in the 1400 – 1650 ms and 1730 – 1930 ms TWs, the prosodic mismatch elicited a negativity that began at 1440 ms and persisted until roughly 1950 ms. This negativity occurred in the same TWs as the semantic N400 and P600 elicited by the ES and ED conditions. As can be seen in Figure 7, in the older adults, there was some negativity that began at 1300 ms, peaked at roughly 1800 ms and persisted until 2000 ms. However, this failed to reach significance (i.e., there was no effect of Wall and no Wall x Group interaction) in the ANOVA examining both groups.

The emergence of a Wall x Semantics x AP interaction in the 1400 – 1650 ms TW (the same TW as the semantic N400) seemed to suggest that when the prosodic and semantic mismatches co-occurred, each type of information was processed differently than when only one type of mismatch was present. However, subsequent follow-up analyses showed that the interaction failed to resolve to a significant Wall x Semantics interaction at any electrode (see Table 4).

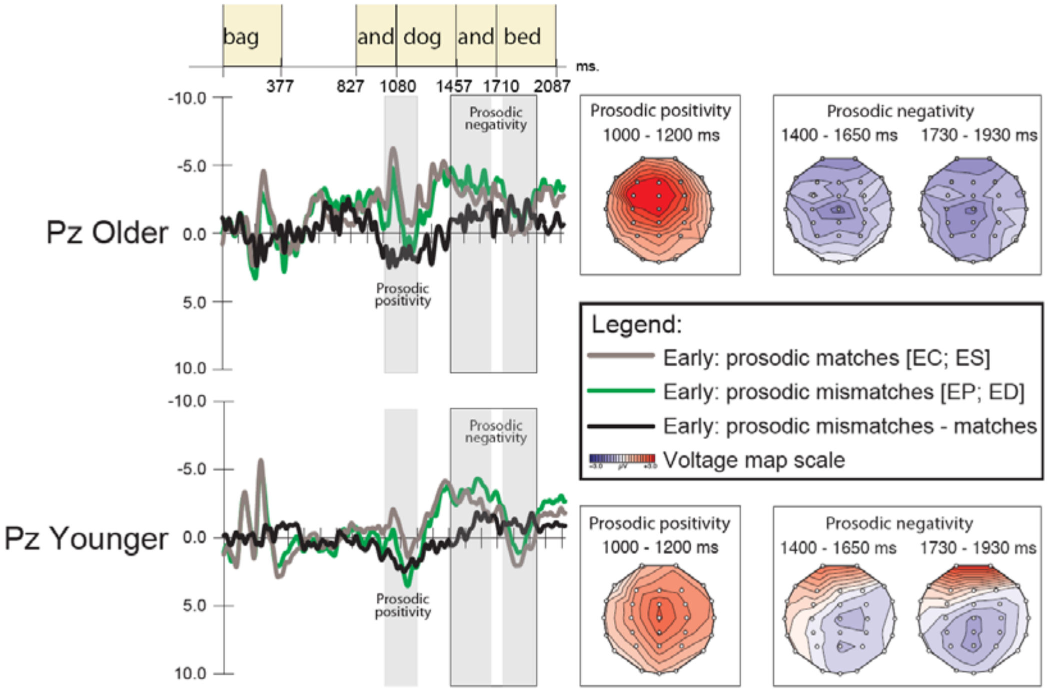

Figure 7. Prosodic mismatches in the Early boundary. Prosodic mismatches (green) and control mismatches (gray) are presented for older adults (top) and younger adults (bottom) along with voltage maps for regions of interest. The difference wave (black) shows that prosodic mismatches elicit a positivity in both groups and a negativity that was significant in the young adults in Chapter 2 but failed to reach significance here. Voltage maps illustrate a broad distribution of the positivity in young adults and more frontal distribution in older adults.

Late boundary Prosodic Positivity. As shown above, the LS condition elicited an N400 for the semantic anomaly in both age groups. Interestingly, in both groups, in the same TW, a positivity was elicited by the prosodic mismatch. Figure 8 shows waveforms and difference waves for prosodic effects in the Late boundary condition in the young and older adults. In the young adults, this positivity began at roughly 900 ms, continued until 1300 ms, and peaked at 1100 ms, 700 ms after the offset of noun1, which co-occurred with the onset of the visually expected, yet absent, auditory boundary. In the older adults, the positivity also began at 900 ms, peaked at roughly 1000 ms, and persisted until 1200 ms. A significant main effect of Wall in the 1000 – 1200 ms TW (see Table 3) confirmed this positive component in both groups. However, significant Wall x AP and Wall x AP x Group interactions showed that there were some differences in the distribution of this component between the groups. Specifically, follow-up ANOVAs revealed that the positivity was more frontally distributed in the older adults compared to the younger adults, who showed a centro-posterior distribution (see Table 4).

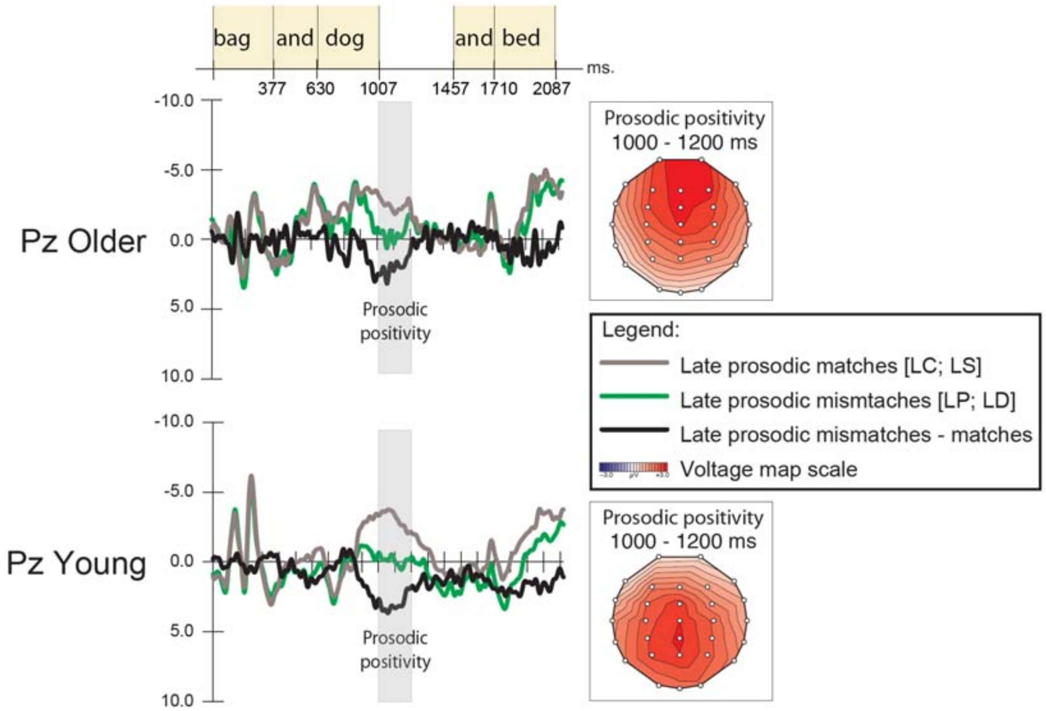

Figure 8. Prosodic mismatches in the Late boundary. Prosodic mismatches (green) and control condition (gray) are presented for older adults (top) and younger adults (bottom) along with voltage maps. Difference waves (black) illustrate that 700 ms after the absence of a pause that is expected based on visual input elicits a positive component (in the 1000 – 1200 ms TW) in both groups. Voltage maps illustrate the more frontal distribution of the positivity in the older adults compared to younger adults.

Discussion

The present study employed ERPs in an auditory-visual match judgment task in order to examine the integration of prosodic information with visual input in real time in older adults, whose responses were compared to those of a subset of younger adults reported in Chapter 2. In the previous study, two groups of young adults emerged: those who moved their eyes ('movers') and those who fixated the center of the visual display ('nonmovers'). The presence or absence of horizontal eye movement in the young adults was linked to differences in the patterns of ERP responses. In the present study, eye movements were briefly examined in older adults to ensure they were being compared to a comparable group of young adults. In so doing, it became evident that every older adult displayed horizontal eye movement consistent with the pattern of eye movements observed in the young adult movers. Before discussing relevant ERP and behavioral findings, we will address the implications of eye movement.

Implications of Eye Movement

In Chapter 2, we discussed possible explanations for the presence/absence of eye movement in different participants. It was hypothesized that the group differences in eye movement were either a product of independent differences in the participants (e.g., differences in visual attention) or the result of using different strategies (for example, weighting information from each modality differently, with the nonmovers being guided by visual input and the movers being guided by auditory input). As there were no group differences related to any of the independent measures taken, it was not possible to tease apart these hypotheses. As discussed in Chapter 2, it would be difficult to argue that inhibiting all eye movement, especially in the presence of visual input, would be a natural behavior. It is unclear why two distinct groups of young adults emerged and what ultimately contributed to these different ocular patterns.

In the present study, every older adult moved their eyes in a similar pattern to the one observed in the young adult movers. The presence of eye movement in all of the older adults may contribute to our understanding of these patterns in the young adults as well. It is generally the case that hearing loss accompanies normal aging (e.g., Schneider & Hamstra, 1999; Snell, 1997; van Rooij & Plomp, 1992, among many others), which would probably lead elderly individuals to be less reliant on auditory input, and therefore more heavily reliant on visual input (Chapter 2). However, it appeared that the young participants who did *not* move their eyes weighted visual input more heavily than auditory input. Therefore, the interpretation that older adults moved their eyes as a result of a lack of confidence in the auditory input is not entirely supported. It is also generally assumed that older adults use less resource-taxing strategies and make use of any contextual cues or external resources available to them (e.g. Cohen & Faulkner, 1983; Boothroyd & Nittrouer, 1988; Sommers & Danielson, 1999). Based on this, it would seem that moving the eyes during this task would serve as a less demanding strategy and aid in the successful completion of the task. This hypothesis corresponds well with the patterns observed in the young adults in that it accounts for why the majority of young adults moved their eyes despite explicit instruction to the contrary.

Semantics

Turning to the experimental effects of interest, we first discuss our findings for the processing of semantic mismatches in aging. For semantic mismatches, we found that older adults detected semantic and double mismatches with high accuracy, as did young adults, although older adults were slightly more accurate in identifying matches (compared to mismatches), whereas young adults' identification of mismatches was slightly better. This is in keeping with examinations of semantic priming and semantic processing of incongruities in sentences and word lists that have shown that semantic effects remain relatively stable with advancing age (e.g., Bowles & Poon, 1985; Burke, White, & Diaz, 1987; Byrd, 1984), though the effects may happen slower (reflected by response times) or be larger in magnitude than in young adults (e.g., Nebes, Boller, & Holland, 1986; Petros, Zehr, & Chabot, 1983; Chapman, Chapman, Curran, & Miller, 1994). In the current study, ERP responses revealed that older adults showed a biphasic N400 – P600 response to auditory-visual semantic mismatches, just as the young adults did. While we expected that the N400 would be smaller in amplitude and perhaps later in onset in the older adults (following e.g., Faustmann et al., 1997), we found that the semantic N400 in the Early auditory boundary condition showed no amplitude or latency differences between groups but was more posteriorly distributed in older adults compared to younger adults. In contrast, the semantic N400 in the Late auditory boundary condition showed no distributional or amplitude differences, but there did appear to be a difference in latency. The N400 for the semantic mismatch began later (60 ms later in the pure semantic mismatch and 200 ms later in the double mismatch) and peaked 130 ms later in the older adults than the same component in the young adults.

Based on the present findings, as well as previous investigations of the N400 in semantic processing, it appears that task demands and the type of stimuli used modulate the morphology of the resulting electrophysiological components in older adults. Previous tasks examining the N400 found latency delays in tasks that used less natural speech—and were therefore likely more difficult (Harbin et al., 1984; Gunter et al., 1992, 1995, 1998; Ford et al., 1996; Woodward et al., 1993)—but not in those using natural speech (e.g., Federmeier et al., 2002; Federmeier et al., 2003; Faustmann et al., 2007). In the present study, despite accuracy scores approaching ceiling, delays to the semantic N400 in at least one condition (as well as the presence of eye movement) suggest that integrating auditory and visual input was demanding. These latency changes observed in the N400 are also in keeping with theories of general cognitive slowing (Cerella, 1985; Salthouse, 1985) and task-specific slowing (e.g., Lima, Hale, & Myerson, 1991) in aging, although it should be borne in mind that the latency differences did not emerge across-the-board.

Another interesting finding is that when a prosodic mismatch co-occurred with a semantic mismatch (Double Mismatch) in the Late auditory boundary condition, semantic processing (reflected by the N400) was modulated in each population. While both groups showed additive and interactive processing, resulting in a waveform that shows the average of overlapping positive (prosodic P600) and negative (semantic N400) components that cancel each other out, these effects begin later in older adults as a result of N400 delays, especially in the LD condition. This is further evidence that task demands (i.e., the increased processing load of simultaneous prosodic and semantic mismatch processing) delay the N400 in older adults. Nonetheless, the finding that older adults process the double mismatch in the same manner as young adults, despite latency delays, indicates that similar neural structures and processes are recruited for this task in the aging brain.

In keeping with Kemmer et al. (2004) and Steinhauer et al. (2010), the P600 in response to the semantic mismatch (thought to reflect the conflict between structured sources of information across the two modalities) was expected to be more frontally distributed in the older adults. For the Early auditory boundary only, this distributional pattern was realized. In the Late auditory boundary, the P600 was larger in amplitude and peaked later (170 ms in the pure semantic mismatch and 350 ms in double mismatch) in the older adults compared to the younger adults. These latency delays are the result of the N400 delays (both in peak latency and component offset) discussed above. While one delay does not necessitate delays in other components, in this case the waveform was biphasic and the subsequent phase (the P600) could not occur until the offset of the first phase (the N400). In general, the P600 findings are in keeping with previous research on the P600 in aging (Kemmer et al., 2004; Steinhauer et al., 2010) and indicate that, despite potential differences in the neural generators underlying the component, older adults integrate semantic (and prosodic) information across modalities in much the same way as do younger adults.

Prosody

Turning to the primary focus of the current investigation – the processing of prosody – the behavioral judgments indicated that both groups responded more accurately to the Late auditory boundary condition compared to the Early auditory boundary condition. This may be due to recency effects, but also may be due to the fact that the only boundary cue was a pause, which can only be detected relative to other events in an utterance. Therefore, the pause in the Early boundary condition did not preclude a second longer pause, whereas the absence of an early pause in the Late boundary condition was more easily detected (see Chapter 2, for further discussion).

Of more interest, with respect to judgments of prosodic mismatches, older adults correctly rejected these conditions with far less accuracy and with much more variability than did younger adults. At first glance, these findings would suggest that older adults were not as sensitive to prosody as the young adults. (Indeed, a subset of older adults were not as sensitive to prosody as they were unable to identify prosodic mismatches in the behavioral task, indicating that older adults form a heterogeneous group with respect to prosodic processing.) However, the ERP data demonstrate that older adults showed on-line sensitivity to the prosodic cues, although they also reveal differences between the age groups. The point at which auditory-visual grouping mismatches can be detected is the same across boundary conditions (i.e., when a boundary is heard but no wall is seen in the Early boundary condition; when a boundary is not heard but a wall is seen in the Late boundary condition). In both auditory boundary conditions, roughly 700 ms after this point, both older and younger adults showed a positivity. This positivity is interpreted as a P600 elicited as a function of the integration of mismatching auditory and visual information. In the Early boundary condition, both groups showed a fronto-central distribution of this prosodic P600, consistent with Kemmer et al. (2004) and Steinhauer et al. (2010). In the Late boundary condition, the prosodic P600 was more frontally distributed in the older adults than in the younger adults, where it had a more central distribution (again consistent with Kemmer et al. and Steinhauer et al.). The presence of this component further indicates that, despite the behavioral judgments, older adults are sensitive to grouping information and are able to integrate prosodic and visual information in some way. This indicates that the processes marked by the P600 (integration of structured bodies of information) are relatively intact in advanced age. An absence of latency delays suggests that age-related cognitive slowing cannot entirely account for changes in language processing associated with aging. Distributional differences in the P600 may point to differences in the neural generators employed in older compared to younger adults; however, as discussed by Kemmer and colleagues, since the waveforms observed in particular electrodes can be generated by entirely different neural regions, it is difficult to conclude with certainty that a more frontal distribution of the P600 is the result of increased activity in frontal electrodes as opposed to processing differences in a distinct brain region. Dissociations between behavioral responses and the elicitation of the P600 found in the present study and by both Kemmer et al. and Steinhauer et al. point to differences in response-related processes associated with aging (see also, e.g., Bashore & Smulders, 1995; Madden, Pierce, & Allen, 1993). They further support Kemmer and colleagues' hypothesis that memory, motor, and strategic processes, rather than decreased language abilities, are likely responsible for poor performance on behavioral tasks.

In addition to the P600-like response to the prosodic mismatch in the young adults, a negativity (interpreted as an N400 to the violation of conceptual expectations) was also observed in response to the *absence* of a pause between the second and third items in the Early auditory boundary condition when a wall was seen (see Figure and Chapter 2). This same component did not reach significance in the elderly group. Older adults, therefore, elicited only one of the two electrophysiological responses to the prosodic mismatch observed in younger adults (i.e., the P600 and not the N400). At the earliest stage at which the prosodic mismatch can be detected in the Early boundary condition (when a pause is heard but no wall is seen), a P600 is elicited (recall, the P600 therefore occurs prior to the N400 in this case). The differences arise primarily at a later stage at which, in this study, expectations and realizations for prosodic and visual information are *integrated*; in this case, the older adults fail to show an N400 response. That is, when older adults first encounter the prosodic mismatch (again, when the pause is heard), they show the same P600 response as do younger adults (albeit with a different distribution). At later stages, when expectations that might have been created earlier in the trial were violated (when a right wall is seen, creating the expectation to hear a Late auditory boundary, and no pause is heard), older adults do not show an N400 response, suggesting that they are not processing information from the two modalities in the same manner as young adults. It is possible that older adults are not creating expectations for auditory input/generating abstract conceptual groupings from visual stimuli and are instead entirely guided by auditory input throughout the trial. (In Chapter 2 this was offered as one possible strategy used by those who moved their eyes, though those young adults still appeared to generate expectations for auditory stimuli.) Therefore, when the older adults see the left wall but do not hear a pause, no expectations are violated. Then, between the time at which the relevant prosodic information is heard and the time at which participants must render a judgment, the visual display remains on the screen. In this time (after they have heard auditory input and are no longer guided by it exclusively), older adults perhaps change the perceived auditory boundary location to fit the visual input, resulting in fewer rejections in the prosodic mismatch condition and a dissociation between behavioral and ERP results (at least in terms of the P600). While previous studies have reported an *increased* reliance on prosody relative to syntactic information, the hypothesis that older adults modify the perception of the auditory boundary to fit the visual input would suggest that in this study older adults rely more heavily on (or recall more accurately) visual information relative to prosodic information. This is not surprising, as visual information is more concrete and not fleeting, as is speech. This interpretation would be in line with working memory difficulties associated with aging (e.g., Kemper, 1992; Light, 1990) and with Kemmer and colleagues' (2004) account of low response accuracy being related to non-language cognitive processing difficulties, rather than decreased prosodic processing abilities, because it would suggest that older adults have difficulty remembering the location of a pause and instead weight the concrete information—received for both a longer time and more recently—more heavily than the fleeting auditory input.

An alternative interpretation of the absence of the N400 is that older adults processed the prosodic mismatch earlier in the trial (i.e., the P600 response) and therefore did not continue to process information related to these mismatches. In other words, once the initial processing is completed and they are able to render a judgment, the P600 is elicited and no further processing of pauses is initiated. This could be support for increased language processing efficiency in aging (Hess & Blanchard-Fields, 1996) since it would suggest a more parsimonious pattern of language processing in older adults, resulting in only one stage of processing rather than two. However, given the low accuracy scores for prosodic mismatches, it is more likely indicative of either an inability or a difficulty integrating this information.

In sum, initially, older adults detected prosodic mismatches similarly to young adults and were successful at integrating auditory and visual information, reflected by the presence of the P600 component. Difficulties in prosodic processing appear to arise at later stages of processing, reflected by the absence of the N400 component and low accuracy scores for the prosodic mismatch condition in the behavioral task. This dissociation between ERP and behavioral responses is in keeping with findings by both Kemmer et al. (2004) and Steinhauer et al. (2010) and is likely the result of decreased abilities related to memory, motor, or strategic abilities.

Conclusions

This study showed that visual context influences the neural correlates of prosodic and semantic processing in older adults, though not in exactly the same manner as observed in younger adults. A mismatch between auditory and visual semantic information elicited a biphasic N400 and P600 waveform for semantic mismatches in both groups and a P600 (or P600-like) component for prosodic mismatches. The presence of a P600 component for prosodic mismatches that differed primarily in distribution in older adults indicates that initial stages of prosodic processing are relatively intact in the aging brain. Latency delays in the N400 semantic responses support a theory of cognitive slowing in aging adults (e.g., Cerella, 1985; Salthouse, 1985; Lima et al., 1991) and further speak to the role of task demands in modulating the morphology of this neural component in older adults. Specifically, increased task demands slow neural processing (at least the processing reflected by the semantic N400). Our interpretation of the dissociation between behavioral and ERP findings for prosodic mismatches supports the hypothesis that older adults exhibit on-line sensitivity to prosodic cues in speech, but may either integrate the prosodic information with syntactic and semantic information at later stages of language processing in a manner different from young adults (in keeping with Steinhauer et al., 2010) or have lower accuracy scores as a result of decreased working memory or cognitive abilities (in keeping with Kemmer et al., 2004). It should also be recalled that since a subset of older adults were not able to detect prosodic mismatches in the behavioral task, these findings represent how older adults process prosody *on average* rather than offering detailed insight into different abilities represented in this heterogeneous group (for more on this topic, see Abada, Steinhauer, Drury, & Baum, 2010).

With respect to prosodic processing abilities in older adults in general, while it may at first appear that our findings are inconsistent with an increased reliance on prosody in the elderly (e.g., Wingfield et al., 1992; Kjelgaard et al., 1999), it is critical to consider differences in task demands across the investigations. For example, in previous studies (e.g., Wingfield et al., 1992; Kjelgaard et al., 1999), prosody served as an additional cue to aid in completion of the task in that prosody helped older adults recall or complete sentences. In the present study, the task was quite different. Participants were asked to detect and integrate prosodic and visual cues. Further, rather than employing complex transient syntactic ambiguities, we made use of simple phrases where prosody marked grouping and did not disambiguate syntactic structures. Older adults showed no difficulty *detecting* and *integrating* prosodic cues and mismatches, reflected by the elicitation of the P600, but did show difficulty re-assessing conceptual expectations created by auditory and visual grouping cues, reflected by the absence of the N400 and low accuracy scores on the behavioral task.

The present study represents an important step towards gaining a more accurate understanding of prosodic processing in older adults because it used a more ecological task and examined not only sensitivity to prosodic cues, but integration of prosody and visual information. We found that prosodic and semantic abilities are largely preserved in older adults, though it appears that increased task demands, integration, and non-language-related motor and memory abilities negatively impact accuracy and speed in older adults. We further find that models of cognitive slowing do not account for all language differences associated with aging and instead suggest that there may be some differences in the specific neural generators responsible for processing semantic and prosodic information, though currently it is not possible to make any specific claims in that regard. Crucially, our findings show that behavioral examinations of prosodic processing in older adults offer a very limited view into prosodic processing abilities. Future additional investigations of real-time prosodic processing are required in order to gain a more complete understanding of how this ability changes with age. Moreover, future examinations must include different syntactic structures and tasks so that the full nature of prosodic processing under different settings may be illuminated.

CHAPTER 4:

General Discussion

The goal of this dissertation was to explore age-related changes in prosodic processing and its interaction with visual context. To that end, three main questions were examined. First, what electrophysiological correlates are elicited by auditory-visual ('prosodic') grouping mismatches? Second, how do speech prosody and visual context interact? Third, do older adults show the same electrophysiological correlates and responses to auditory-visual integration as do young adults? The first study (Chapter 2) addressed the first two questions in a group of 20 English-speaking young adults. The second study (Chapter 3) addressed the third question by examining a group of 11 older adults and comparing them to a subset of the young adults discussed in Chapter 2.

Both studies used the same cross-modal design that used simple conjoined phrases (e.g., 'bag and bed and cup') presented aurally along with pictures depicting the phrases. Using simple phrases avoided any confounds of working memory limitations related to syntactic processing that may be present in older adults (see e.g., Stine, 1995). Visual displays either matched the auditory display, differed from the auditory phrase in visual grouping (prosodic mismatch), differed from the auditory phrase by showing a different item in the second position (semantic mismatch), or differed both prosodically and semantically (double mismatch). Behavioral judgment responses and electrophysiological responses were recorded to determine how individuals integrated what they heard and what they saw.

When young adults were examined (Chapter 2), an intriguing pattern emerged. The horizontal electro-oculogram (HEOG; recorded from electrodes placed at the outer canthus of each eye) showed a pattern of eye movement that was consistent across 13 of the 20 participants (the ‘movers’). The remaining 7 participants (the ‘nonmovers’), however, fixated the center of the screen as they were instructed. Horizontal eye movement in the movers followed a systematic pattern across all group members. Specifically, eye movement in this group was guided by the auditory input. Behaviorally, both groups showed high accuracy scores for all conditions, indicating that all young adults were successful at detecting both prosodic and semantic mismatches. However, in both groups accuracy was lower for prosodic mismatches. Interestingly, the nonmovers were overall more accurate than the movers, suggesting that eye movement did not aid in detecting mismatches. Reaction time measurements did not reveal any group differences, but, as expected, responses to prosodic mismatches were slower than all other conditions.

ERP responses provided more insight into how auditory and visual input were integrated. Both prosodic and semantic mismatches elicited N400 and P600 responses in both groups. Group differences, however, demonstrated that the two groups used different strategies to complete the task. N400 responses showed that nonmovers were creating expectations based on visual input that were then violated by the auditory input. P600 responses were elicited by the integration of the mismatching information presented aurally and visually. As will be discussed below, these findings can inform current interpretations of N400 and P600 components.

Older adults, as described in Chapter 3, showed similar neural correlates overall to young adults, although an N400 to prosodic mismatches was not elicited in this group. Behaviorally, older adults also responded similarly to younger adults, although they had even more difficulty detecting prosodic mismatches, resulting in significantly lower accuracy scores in the pure prosodic mismatch conditions. Most strikingly, every older adult displayed a horizontal eye movement pattern similar to the movers, though not every young adult moved their eyes. This may suggest that eye movement during cross-modal stimulus presentation is a strategic option for young adults, but cannot be inhibited in older adults.

Findings for both young and older adults have a number of important implications when examined in the context of the current literature. Our findings contribute to the interpretation of certain ERP components, as well as to our understanding of prosodic (and semantic) processing, and auditory-visual integration in both young and older adults. We will address each of these issues, beginning with a discussion of the N400 and P600 ERP components.

As summarized in the Introduction in Chapter 2, both the N400 and P600 components have received a great deal of attention in the literature and various interpretations have been offered to explain each component. The present findings extend the current views on each component. There are currently two dominant views regarding the N400 (discussed in Lau, Phillips, & Poeppel, 2008). One view holds that the N400 reflects the integration of critical semantic or conceptual input with context and therefore marks post-lexical access mechanisms; the alternative view holds that the N400 reflects lexical activation, and its relative facilitation, by a context. Lau and colleagues review relevant ERP and fMRI findings and resolve that while there is thus far no conclusive evidence to determine which (or what combination) of these interpretations is reflected by the N400, at present the more widely accepted perspective views the N400 as a reflection of lexical activation/facilitation by context, with the role of context being primary (see also Kutas & Federmeier, 2000 and Kutas & Federmeier, 2007 for reviews). In the present studies, an N400 was elicited when an expectation for or against hearing a boundary (marked by pause) was violated. That is, an N400 was elicited to a stimulus that was not lexical, but did carry conceptual meaning since the pause and wall marked conceptual grouping. Once an expectation for or against a grouping of items was created, the violation of that expectation resulted in an N400 response. This is evidence *against* an interpretation of the N400 as solely a reflection of lexical activation by a context. Rather, our findings speak in favor of the N400 as a reflection of the integration of conceptual input with a context. Even more intriguing, our findings extend this interpretation to include the integration of auditory input (both prosodic and semantic) with a visual context.

While there have been two commonly held interpretations of the N400, the interpretation of the P600 is much more debated and has changed over time. Currently, an interpretation that is garnering increased support in the literature holds that the P600 marks integration of structured bodies of information (e.g., Kuperberg, 2007; the "generalized mapping" component, Bornkessel & Schlesewsky, 2006, 2008). Compatible with this view is the hypothesis that the P600 is elicited whenever there is a mismatch between information from two sources of structured information, rather than to syntactic violations alone (see Münte, Heinze, Matzke, Wieringa, & Johannes, 1998 for a discussion of this interpretation). The current findings not only support these interpretations, but extend them to include a mismatch between structured bodies of information from *two different* modalities.

In terms of language processing, our findings confirm that participants integrate auditory and visual semantic information as soon as they become available (e.g., Friedrich & Friederici, 2004) and further extend this to the phrase level as opposed to the single word level. Whereas conceptual-semantic integration across modalities had been demonstrated before (although in a much simpler picture/word matching paradigm), our present data are the first to elucidate the real-time integration of auditory and visual *grouping* information. Interestingly, while behavioral responses showed increased difficulty detecting these prosodic mismatches, particularly in older adults (which will be revisited below), ERP analyses revealed that all participants in both age groups showed sensitivity to the prosodic mismatch. In line with Steinhauer and colleagues (2010), this indicates that off-line behavioral responses do not provide sufficiently detailed information regarding prosodic processing in real time. These findings not only open the door to the broader examination of prosodic processing within a context but also create a new method of examining auditory-visual integration using ERPs.

In that regard, the novel methodology used in these studies illuminated the influence of eye movement on language processing in a ground-breaking way. While a wealth of studies have demonstrated that eye movement patterns can inform our understanding of language processing (e.g., Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Hanna & Brennan, 2007; McGurk & MacDonald, 1957; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Friedrich &

Friederici, 2004; Knoeferle, Habets, Crocker, & Münte, 2008; Gleitman, January, Nappa, & Trueswell, 2007; Knoeferle & Crocker, 2007; Altmann & Kamide, 2007; Farmer, Anderson, & Spivey, 2007, among many others) and that visual attention is influenced by linguistic input (e.g., Gleitman et al., 2007), the present findings demonstrate that the respective processing strategy as reflected by (the presence versus absence of) eye movement also modulates the neural response elicited by critical stimuli. We interpreted the modulation of ERP waveforms by eye movements to result from strategic differences in task completion between the two groups of young adults, as no other differences between the groups were evident. It is unclear at this time why two distinct strategies emerged and what contributed to the use of one or the other. It is possible that one strategy was easier, less taxing on resources, or even more effective when used outside of the laboratory, prompting individuals to choose one strategy over the other. In Chapter 2, we argued that coordinating one's eye movement with the incoming speech input may reflect a more natural processing 'strategy' when simultaneously attending to both modalities. However, in our task, the more natural strategy did not offer an advantage in terms of accuracy, as the nonmovers were more accurate on behavioral measurements than were the movers. In Chapter 3, we determined that every older adult displayed eye movement and suggested that this strategy may be less resource taxing than inhibiting one's eye movements (e.g., Boothroyd & Nittrouer, 1988). It is clear that investigating eye movement (here, from HEOG sources) simultaneously with neural processing from ERP sources provides increased insight into language processing than when either source is examined independently. Importantly, our methodology did not compromise the quality of the ERP data, indicating that by carefully controlling stimuli, it is possible to design more ecological studies that can advance our understanding of language processing in context.

One of the main goals of this work was to examine prosodic processing in aging. In Chapter 3 we found that older adults performed the task similarly to young adults overall. That is, older adults were very successful at detecting matches and semantic mismatches and elicited largely the same neural correlates to semantic mismatches as did young adults (though there were some distributional and timing differences). With respect to prosodic mismatches, however, differences emerged between the age groups. Overall, older adults were far less successful at detecting prosodic mismatches than were younger adults. Our findings indicate, nonetheless, that older adults are sensitive to prosodic mismatches as reflected in on-line ERP responses, and are able to integrate auditory and visual grouping information. In initial stages (i.e., when information is first presented), older adults process grouping information similarly to younger adults, supported by the elicitation of the P600 component at the point at which the mismatch becomes evident in both auditory boundary conditions. Differences arise at later stages at which expectations for boundaries/walls are either confirmed or violated. When those expectations are violated, older adults do not elicit the same neural components (the N400) as do young adults. At even later stages, as reflected in reduced off-line accuracy measures for prosodic mismatch conditions, there are differences in response-related processes, in line with Kemmer, Coulson, de Ochoa, and Kutas (2004).

While these findings do demonstrate group differences, it is possible that these differences are not the result of prosodic processing alone. That is, behavioral differences and dissociations between on-line and off-line measures have been reported in previous studies (e.g., Kemmer et al., 2004; Steinhauer, Abada, Pauker, Itzhak, & Baum, 2010) and may be associated with decreases in memory, motor, and strategic processes, rather than decreased language processing abilities per se (Kemmer et al). The absence of the N400 may not be the result of decreased abilities detecting prosodic mismatches. Instead, it may be the case that older adults are entirely guided by auditory input and are not creating expectations throughout the trial as are young adults. At the offset of the auditory events, the visual information remains on the screen and is a more stable cue. Therefore, when asked to render judgments, they recall the visual input better rather than integrating the auditory and visual information off-line. The factors contributing to these differences are not specific to prosody. Rather, the elicitation of the P600 supports previous findings that older adults do not differ from young adults in initial stages of prosodic processing (Steinhauer et al., 2010) and further support the need for on-line measurements in order to gain more accurate insight into prosodic processing. Nonetheless, it is important to note that there were no substantive differences between older and younger participants with respect to semantic processing, indicating that there is something unique about the treatment of prosody in aging that merits further investigation.

Limitations and Future Directions

The studies presented here are a strong first step towards more ecologically valid research examining the integration of auditory and visual information in both young and older adults and provide insight into the processes reflected by well-documented neural components (the N400 and the P600). These studies are the first to present simultaneous ERP and HEOG data elicited by picture stimuli (as opposed to reading) and are also the first to examine eye movement in prosodic processing in older adults. Thus, despite some important initial contributions, more questions are being raised than answered, and limitations of the present work must be kept in mind.

The carefully constructed stimulus design employed provided new insights into the on-line integration of auditory and visual information. However, as discussed in Chapter 2, because of the fixed presentation, the design could not offer unequivocal support for or against the elicitation of the Closure Positive Shift (CPS) in response to these stimuli. That is, while there is some evidence of a positivity following the onset of the pause in both the Early and Late boundary conditions, the positivity overlaps with the negative-going N100 elicited by the word '*and*' in the *control* condition. Therefore, it is not clear whether this is a positive shift elicited by the pause or primarily reflects the absence of an N100. Since every trial contained the exact same timing of events (with fixed pause durations), even time-locking the ERPs to the offset of the pause would not help dissociate the CPS from the N100 (as is typically possible in less controlled natural speech).

Another limitation of our design is that it cannot speak to the on-line processing of prosody in the absence of mismatch or violation judgment requirements. Therefore, these findings do not contribute to the question of what neural correlates are elicited by prosodic boundaries marking grouping when individuals attend to auditory boundaries alone. The stimuli themselves, however, can address this issue in follow-up studies using only the match conditions using a method similar to that used by Streeter (1978). That is, participants could be presented with an auditory phrase and asked to place the boundary in the appropriate location following the offset of the phrase. Moreover, the auditory boundary could be marked by a pause in some trials and by differences in fundamental frequency in other trials in order to better determine which acoustic parameters are required to elicit an ERP response to such a prosodic boundary. Such a study would be the first to measure on-line prosodic processing in the absence of any potential mismatch or violation in the entire experiment.

One problematic issue (raised in Footnote 2 in Chapter 3) arose in the examination of older adults. Of the older adults who were not excluded from the study based on MoCA scores, 4 of them incorrectly accepted prosodic mismatches throughout the study. Eliminating these individuals from the study would have resulted in too few older adults to conduct ERP analyses (n = 7). The small number of individuals also prevented a reliable direct comparison of the two groups of older adults. Since the goal of this study was to examine prosodic processing in older adults, it was decided that all participants should be included in analyses and follow-up investigations should examine this variability in older adults in more detail. The group of "less prosodically sensitive" older adults indicates that prosodic processing may not be as homogeneous an ability as some other aspects of language processing in older adults. Rather, older adults seem to show a great deal of variability when processing prosody. Interestingly, in an investigation using syntactically complex and working memory taxing sentences, Grossman and colleagues found that some older adults were successful at the task while others performed poorly when syntax became very complex and resource demanding (Grossman et al., 2002a, 2002b). The authors interpret these findings as an indication that older adults use a compensatory approach to sentence processing, though the specific strategy is modified by individual biological and cognitive factors. Thus, the findings presented here offer insight into how older adults *on average* perform in comparison with young adults. However, it is clear that this issue must be examined in more detail in future studies. To that end, data from additional older adults will be collected to determine whether this heterogeneity in performance continues or whether the 4 participants in this study form an anomalous group.

The methodology and findings presented here open the door to many future investigations. First, as just discussed, further examinations of older adults are required to clarify prosodic processing in aging populations. Second, the issue of eye movement and visual attention must be more fully investigated using this methodology and different linguistic constructions. Another interesting finding that emerged in piloting was that participants who were not informed of the potential for prosodic mismatches did not detect the mismatch behaviorally. Similarly, when preparing stimuli, individuals who were asked to review the stimuli and were aware of the focus on prosody in the study did not detect the semantic mismatch. ERP data were either not analyzed or not recorded for these few individuals so it remains unclear whether these attentional differences were reflected in neural responses. An examination in which participants are instructed to attend to semantics, prosody, or given no instruction, would illuminate the role attention plays in language processing and the associated neural correlates in novel ways. Another important study to undertake to address the question of weighting auditory and visual information would use the same stimuli and manipulate the time at which information from each modality is presented. These findings would also likely speak to the processes reflected by the N400 and P600 because of differences in how participants would create expectations and integrate information between the present study and that study.

Conclusions

The present investigations examined the neural correlates of the integration of auditory and visual prosodic/grouping and semantic information in young and older adults. These studies showed that auditory and visual information are successfully integrated by both age groups, illuminated different eye movement strategies used by both groups, and provided important insight into the mechanisms reflected by the N400 and P600 components. Our findings revealed that eye movement (and therefore visual attention) can play a critical role in language processing. Through this research, the interpretation of the N400 can be extended to include the integration of conceptual information into a context, even if that context is visual. The interpretation of the P600 as reflecting the integration of structured bodies of information was supported and extended to include the integration of information across auditory and visual modalities. Finally, we concluded that while older adults process prosodic information similarly to younger adults in initial stages, some differences emerge at later stages of processing. As is well known, older adults are a far more heterogeneous population in terms of prosodic processing than are young adults. The studies presented herein are a critical first step in examining the on-line response to prosody and its role in language interpretation within a visual context using ERP methodologies. The findings raise intriguing questions about eye movement and visual attention that must be examined in future studies.

REFERENCES

Abada, S.H., Baum, S.R., & Titone, D. (2008). The effects of central- vs. peripheral-feature contexts on phonetic ambiguity resolution in younger and older listeners. *Experimental Aging Research, 34 (3),* 232-250.

Abada, S.H., Steinhauer, K., Drury, J.E., & Baum, S.R. (2010). Age differences in electrophysiological correlates of cross-modal phrasal interpretation. Poster presented at the Fifth International Conference on Speech Prosody, Chicago, IL.

Abney, S.P., & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research, 20,* 233-250.

Altmann, G.T.M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57,* 502-518.

Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research, 18*, 172-184.

Bashore, T.R., & Smulders, F. (1995). Do general slowing functions mask local slowing effects? A chronopsychophysiological perspective. In P. A. Allen & T. R. Bashore (Eds.), *Age differences in word and language processing* (pp. 390–425). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.

Baum, S.R., & Pell, M.D. (1999). The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology, 13*, 581-608.

Beach, C.M. (1991). The interpretation of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language, 30,* 644-663.

Beach, C.M., Katz, W.F., & Skowronski, A. (1996). Children's processing of prosodic cues for phrasal interpretation. *Journal of the Acoustical Society of America, 99*, 1148-1160.

Birren, J.E., & Schroots, J.J.F. (2001). The History of Geropsychology. In J. E. Birren & K. W. Schaie (Eds.), *The Handbook of the Psychology of Aging* (Fifth ed.). San Diego: Academic Press.

Boersma, P., & Weenink, D. (2006). Praat: doing phonetics by computer (Version 4.4.16) [Computer program]. Retrieved April 1, 2006, from http://www.praat.org/

Boothroyd, A., and Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America, 84,* 101-114.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review, 113,* 787-821.

Bornkessel, I., & Schlesewsky, M. (2010). *"Mind the gap!" On the interdependence between eye movements and neurophysiological responses during real time language comprehension.* Talk presented at the 23rd Annual CUNY Conference on Human Sentence Processing, New York, NY.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews, 59,* 55-73.

Bowles, N.L., & Poon, L.W. (1985). Aging and retrieval of words in semantic memory. *Journal of Gerontology, 40,* 71–77.

Burke, D.M., White, H., & Diaz, D. L. (1987). Semantic priming in young and older adults: Evidence for age constancy in automatic and attentional processes. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 79–88.

Byrd, M. (1984). Age differences in the retrieval of information from semantic memory. *Experimental Aging Research, 10,* 29–33.

Byrd, M. (1993). Adult age differences in the ability to write prose passages. *Educational Gerontology, 19,* 375–396.

Cerella, J. (1985). Information processing in the elderly. *Psychological Bulletin, 98*, 67-83.

Chapman, L.J., Chapman, J.P., Curran, T.E., & Miller, M.B. (1994). Do children and the elderly show heightened semantic priming? How to answer the question. *Developmental Review, 14,* 159–185.

Cohen, G., & Faulkner, D. (1986). Word recognition: Age differences in contextual facilitation effects. *British Journal of Psychology, 74,* 239-251.

Connine, C.M., & Clifton, C.C., Jr. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 291-299.

Cutler, A., Dahan, D., and van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40,* 141-201.

Eckstein, K., & Friederici, A.D. (2005). Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. *Cognitive Brain Research, 25*, 130-143.

Eckstein, K., & Friederici, A.D. (2006). It's early: Event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience, 18(10),* 1696-1711.

Farmer, T.A., Anderson, S.E., & Spivey, M.J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language, 57,* 570-595.

Faustmann, A., Murdoch, B.E., Finnigan, S.P., & Copland, D.A. (2007). Effects of advancing age on the processing of semantic anomalies in adults: Evidence from event-related brain potentials. *Experimental Aging Research, 33,* 439-460.

Federmeier, K.D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology, 42*, 133-141.

Federmeier, K.D., McLennan, D.B., & de Ochoa, E. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39,* 133–146.

Federmeier, K.D., Van Petten, C., Schwartz, T.J., & Kutas, M. (2003). Sounds, words, sentences: Age-related changes across levels of language processing. *Psychology and Aging, 18,* 858–872.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10,* 279-293.

Fisher, D.L., Fisk, A.D., & Duffy, S.A. (1995). Why latent models are needed to test hypotheses about the slowing of word and language processes in older adults. In P. A. Allen & T. R. Bashore (Eds.), *Age differences in word and language processing*. Amsterdam: Elsevier.

Fisk, A.D., & Rogers, W.A. (1991). Toward an understanding of age-related visual search effects. *Journal of Experimental Psychology: General, 120*, 131-149.

Ford, J.M., Woodward, S.H., Sullivan, E.V., Isaacks, B.G., Tinklenberg, J.R., Yesavage, J.A., & Roth, W.T. (1996). N400 evidence of abnormal responses to speech in Alzheimer's disease. *Electroencephalography and Clinical Neurophysiology, 99,* 235-246.

Friederici, A.D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*, 78-84.

Friederici, A.D., & Frisch, S. (2000). Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language, 43*, 476-507.

Friederici, A.D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax-semantic interface. *Brain Research, 1146*, 50-58.

Friedrich, M., & Friederici, A.D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of*

Cognitive Neuroscience, 16 (8), 1465-1477.

Ganis, G., Kutas, M., & Sereno, M.I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8,* 89-106.

Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110-125.

Georgiades, M.S., & Harris, J.P. (1997). Biasing effects in ambiguous figures: removal or fixation of critical features can affect perception. *Visual Cognition, 4*, 383-408.

Gerken, L.A. (1996). Prosody's role in language acquisition and adult parsing. *Journal of Psycholinguistic Research, 25:2,* 345-356.

Gleitman, L.R., January, D., Nappa, R., & Trueswell, J.C. (2007). On the *give* and *take* between event apprehension and utterance formulation. *Journal of Memory and Language, 57,* 544-569.

Gleitman, L.R., & Wanner, E. (1982). Language acquisition: The state of the art. In E. Wanner & L.R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3 – 48). Cambridge, England: Cambridge University Press.

Grossman, M., Cooke, A., DeVita, C., Alsop, D., Detre, J., Chen, W., & Gee, J. (2002a). Age-related changed in working memory during sentence comprehension: an fMRI study. *NeuroImage, 15,* 302-317.

Grossman, M., Cooke, A., DeVita, C., Chen, W., Moore, P., Detre, J., Alsop, D., & Gee, J. (2002b) Sentence processing strategies in healthy seniors with poor comprehension: an fMRI study. *Brain and Language, 80,* 296-313.

Gunter, T.C., Jackson, J.L., & Mulder, G. (1992). An electrophysiological study of semantic processing in young and middle- aged academics. *Psychophysiology, 29,* 38-54.

Gunter, T.C., Jackson, J.L., & Mulder, G. (1995). Language, memory, and aging: An electrophysiological exploration of the N400 during reading of memory-demanding sentences. *Psychophysiology, 32,* 215-229.

Gunter, T.C., Jackson, J.L., & Mulder, G. (1998). Priming and aging: An electrophysiological investigation of N400 and recall. *Brain and Language, 65,* 333-355.

Hanna, J.E., & Brennan, S.E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*, 596-615.

Hanauer, J.B., & Brooks, P.J. (2005). Contributions of response set and semantic relatedness to cross-modal Stroop-like picture-word interference in children and adults. *Journal of Experimental Child Psychology, 90*, 21-47.

Harbin, T.J., Marsh, G.R., & Harvey, T. (1984). Differences in the late components of he event-related potential due to age and to semantic and non-semantic tasks. *Electroencephalography and Clinical Neurophysiology, 59,* 489-496.

Hess, T.M., & Blanchard-Fields, F. (1996). Introduction to the study of cognitive change in adulthood. In F. Blanchard-Fields & T. M. Hess (Eds.), *Perspectives on Cognitive Change in Adulthood and Aging* (pp. 3-24). New York: The McGraw-Hill Companies, Inc.

Holcomb, P.J., & Anderson, J.E. (1993) Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes, 8(4),* 379-411.

Holcomb, P.J., & McPherson, W.B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition, 24(2),* 259-276.

Iragui, V., Kutas, M., Mitchiner, M., & Hillyard, S. (1993). Effects of aging on event-related brain potentials and reaction times in an auditory oddball task. *Psychophysiology, 30,* 10-22.

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.

Jasper, H.A. (1958). The ten-twenty system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10,* 371-375.

Jonides, J., Schumacher, E.H., Smith, E.E., Koeppe, R., Awh, E., Reuter-Lorenz, P.A., Marshuetz, C., & Willis, C. (1998). The role of parietal cortex in verbal working memory. *Journal of Neuroscience, 9*, 462–475.

Joyce, C.A., Gorodnitsky, I.F., King, J.W., & Kutas, M. (2002). Tracking eye fixations with electroocular and electroencephalographic recordings. *Psychophysiology, 39,* 607-618.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*, 159-201.

Kaan, E., & Swaab, T.Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive*

Neuroscience, 15, 98-110.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). Boston Naming Test. Philadelphia: Lea & Febiger.

Kemmer, L., Coulson, S., de Ochoa, E., & Kutas, M. (2004). Syntactic processing with aging: An event-related potential study. *Psychophysiology, 41*, 372-384.

Kemper, S. (1992). Language and aging. In F.I.M. Craik & T.A. Salthouse (Eds.), *Handbook of aging and cognition* (pp. 213-270). Hillsdale, NJ: Erlbaum.

Kerkhofs, R., Vonk, W., Schriefers, H. & Chwilla, D.J. (2007). Discourse, syntax, and prosody: The brain reveals and immediate interaction. *Journal of Cognitive Neuroscience, 19:9*, 1421-1434.

Ketcham, C.J., & Stelmach, G.E. (2001). Age-Related Declines in Motor Control. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the Psychology of Aging*. San Diego: Academic Press.

King, J.W., & Kutas, M. (1995). Do the waves begin to waver? ERP studies of language processing in the elderly. In P. A. Allen & T. R. Bashore (Eds.), *Age Differences in Word and Language Processing* (pp. 314-344). Amsterdam: Elsevier.

Kjelgaard, M.M., Titone, D.A., & Wingfield, A. (1999). The influence of prosodic structure on the interpretation of temporary syntactic ambiguity by young and elderly listeners. *Experimental Aging Research, 25*, 187-207.

Knoeferle, P., & Crocker, M.W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of*

Memory and Language, 57, 519-543.

Knoeferle, P., Habets, B., Crocker M.W., & Münte, T.F. (2008). Visual scenes trigger immediate syntactic reanalysis: Evidence from ERPs during situated spoken comprehension. *Cerebral Cortex, 18,* 789-795.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *NeuroImage, 37,* 1445-1456.

Kuperberg, G.R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23-49.

Kutas, M., & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 4(12)*, 463-470.

Kutas, M., & Federmeier, K.D. (2007) Event-related brain potential (ERP) studies of sentence processing. In Gaskell, M.G. (Ed.), *The Oxford handbook of psycholinguistics* (385 – 406). Oxford University Press: Oxford.

Kutas, M., & Hillyard, S.A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology,11*, 99-116.

Ladd, D.R. (1988). Declination “reset” and the hierarchical organization of utterances. *Journal of the Acoustical Society of America, 84,* 530-544.

Lau, E.F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Neuroscience, 9,* 920-933.

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America, 54(5)*, 1228-1234.

Lehiste, I., Olive, J.P., & Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America, 60,* 1199-1202.

Lehman, M.T., & Tompkins, C.A. (1998). Reliability and validity of an auditory working memory measure: data from elderly and right-hemisphere damaged adults. *Aphasiology, 12: 7/8,* 771–785.

Li, W., & Yang, Y. (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience, 158,* 1416-1425.

Lieberman, P. (1967). *Intonation, Perception, and Language.* MIT Research Monograph No. 38, MIT, Cambridge, MA.

Light, L.L. (1990). Interactions between memory and language in old age. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 275-290). San Diego, CA: Academic Press.

Lima, S. D., Hale, S., & Myerson, J. (1991). How general is general slowing? Evidence from the lexical domain. *Psychology and Aging, 98,* 615-618.

Madden, D. J. (2001). Speed and Timing of Behavioral Processes. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the Psychology of Aging*. San Diego: Academic Press.

Madden, D. J., Pierce, T. W., & Allen, P. A. (1993). Age-related slowing and the time course of semantic priming in visual word identification. *Psychology and Aging, 8,* 490–507.

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. *Cerebral Cortex, 17*, 2659-2668.

Magne C., Astésano C., Lacheret-Dujour A., Morel M., & Besson M. (2005). Online processing of ‘‘pop-out’’ words in spoken French dialogues. *Journal of Cognitive Neuroscience, 17,* 740-756.

Magne, C., Schon, D., & Besson, M. (2003) Prosodic and melodic processing in adults and children. *Annals New York Academy of Sciences, 999,* 461-476.

Marslen-Wilson, W.D., Tyler, L.K., Warren, P., Grenier, P., & Lee, C.S. (1992). Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology, 45A,* 73-87.

Mietz, A., Toepel, U., Ischebeck, A., & Alter, K. (2008). Inadequate and infrequent are not alike: ERPs to deviant prosodic patterns in spoken sentence comprehension. *Brain and Language, 104*, 159-169.

McCallum, W.C., Farmer, S.F., & Pocock, P.V. (1984). The effects of physical and semantic incongruites on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology: Evoked Potentials, 59(6),* 477-488.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

Münte, T.F., Heinze, H.J., Matzke, M., Wieringa, B.M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*, 217-226.

Nagel, H.N., Shapiro, L.P., Tuller, B., & Nawy, R. (1996). Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. *Journal of Psycholinguistic Research*, *25*, 319-344.

Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society, 53(4)*, 695-699.

Nebes, R.D., Boller, F., & Holland, A. (1986). Use of semantic context by patients with Alzheimer's Disease. *Psychology and Aging, 1,* 261-269.

Nooteboom, S. (1997). The prosody of speech: Melody and Rhythm. In W.J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 640-673). Cambridge, MA: Blackwell Publishers Ltd.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9:1,* 97-113.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potential elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785-806.

Osterhout, L., Kim. A., & Kuperberg, G. (2007). The neurobiology of sentence comprehension. In M. Spivey, M. Joanaisse, & K. McRae (Eds), *The Cambridge Handbook of Psycholinguistics*. Cambridge: Cambridge University Press.

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(3)*, 407-421.

Papafragou, A. (2003) Aspectuality and Scalar Structure. Proceedings from the 27th Annual Boston University Conference on Language Development, Cascadilla Press, Somerville, Massachusetts.

Pauker, E., Itzhak, I., Baum, S.R., & Steinhauer, K. (under review). Co-operating and conflicting prosody in spoken English garden path sentences: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience,* 48pp.

Petros, T.V., Zehr, D.H., & Chabot, R.J. (1983). Adult age differences in accessing and retrieving information from long-term memory. *Journal of Gerontology, 38,* 589–592.

Phillips, C., Kazanina, N., & Abada, S.H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research, 22*, 407-428.

Pichora-Fuller, M.K., Schneider, B.A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97,* 593-608.

Price, P.J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90,* 2956-2970.

Rosinski, R.R. (1977). Picture-word interference in semantically based. *Child Development, 38*, 643-647.

Rugg, M.D., & Coles, M.G.H. (1995). The ERP and cognitive psychology: Conceptual issues. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (Vol. 25, pp. 27-39). Oxford: Oxford University Press.

Salthouse, T.A. (1985). Speed of behavior and its implications for cognition. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: van Nostrand Reinhold Co., Inc.

Schmidt-Kassow, M., & Kotz, S.A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience, 21:9,* 1693-1708.

Schneider, B.A., Daneman, M., & Pichora-Fuller, M.K. (2002). Listening in aging adults: from discourse comprehension to psychoacoustics. *Canadian Journal of Experimental Psychology, 56(3)*, 139-152.

Schneider, B., & Hamstra, S. (1999). Gap detection thresholds as a function of tonal duration for younger and older listeners. *Journal of the Acoustical Society of America, 106,* 371-380.

Schriefers, H., Meyer, A.S., & Levelt, W.J.M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*, 86-102.

Sereno, S.C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences, 7:11,* 489-493.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*, 103-130.

Snell, K. (1997). Age-related changes in temporal gap detection. *Journal of the*

Acoustical Society of America, 101, 2214-2220.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6 (2)*, 174-215.

Soderstrom, M., Seidl, A., Kemler Nelson, D.G., & Jusczyk, P.W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, *49*, 249-267.

Sommers, M.S., & Danielson, S.M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging*, *14:3*, 458-472.

Speer, S.R., Kjelgaard, M.M., & Dobroth, K.M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, *25*, 249-271.

Spencer, W.D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging, 10 (4),* 527-539.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language, 86*, 142-164.

Steinhauer, K., Abada, S.H., Pauker, E., Itzhak, I., & Baum, S.R. (2010). Prosody-syntax interactions in aging: Event-related potentials reveal dissociations between online and offline measures. *Neuroscience Letters*, *472(2)*, 133-138.

Steinhauer, K., Alter, K., & Friederici, A.D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*, 191-196.

Steinhauer, K., Drury, J.E., Portner, P., Walenski, M., & Ullman, M.T. (2010). Syntax, concepts, and logic in the temporal dynamics of language comprehension: Evidence from event-related potentials. *Neuropsychologia, 48 (6),* 1525-1542.

Steinhauer, K., & Friederici, A. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research, 30(3)*, 267-295.

Stine, E.A.L. (1995). Aging and the distribution of resources in working memory. In P. A. Allen & T. R. Bashore (Eds.), *Age Differences in Word and Language Processing* (pp. 171-186). Amsterdam: Elsevier.

Stine, E.A.L., & Wingfield, A. (1990). How much do working memory deficits contribute to age differences in discourse memory? *European Journal of Cognitive Psychology, 2,* 289-304.

Stine-Morrow, E., Miller, L., & Nevin, J. (1999). The effects of context and feedback on age differences in spoken word recognition. *Journal of Gerontology: Psychological Science, B 54,* P125-P134.

Streeter, L.A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America*, 64, 1582-1592.

Syrett, K., Bradley E., Kennedy C., and Lidz, J. (2005). *Shifting Standards: Childrens' Understanding of Gradable Adjectives.* Paper presented at the 2005 Annual Meeting of the Linguistics Society of America, San

Francisco, CA.

Taler, V., Baum, S.R., & Saumier, D. (2006). *Perception of linguistic and affective prosody in younger and older adults.* Poster presented at the The 28th Annual Conference of the Cognitive Science Society, Vancouver, Canada.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science. 268,* 632-634.

Taylor, J.K., & Burke, D.M. (2002). Asymmetric aging effects on semantic and phonological processes: Naming in the picture-word interference task. *Psychology and Aging, 17(4)*, 662-676.

Titone, D.A., Koh, C.K., Kjelgaard, M.M., Bruce, S., Speer, S.R., & Wingfield, A. (2006). Age-related impairments in the revision of syntactic misanalyses: Effects of Prosody. *Language and Speech, 49(1)*, 75-99.

Tremblay, K.L., Piskosz, M., & Souza, P. (2002). Aging alters the neural representation of speech cues. *NeuroReport, 13,* 1865–1870.

Tremblay, K.L., Piskosz, M., & Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology, 114,* 1332–1343.

van Rooij, J.C.G.M., & Plomp, R. (1992). Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion. *Journal of the Acoustical Society of America, 91,* 1028-1033.

Verhaegen, P. (2003). Aging and vocabulary scores: A meta-analysis. *Psychology and Aging, 18,* 332–339.

Watt, S.M., & Murray, W.S. (1996). Prosodic form and parsing commitments. *Journal of Psycholinguistic Research*, *25*, 291-318.

Wingfield, A., Aberdeen, J.S., & Stine, E.A.L. (1991). Word onset gating and linguistic context in spoken word recognition by young and elderly adults. *Journal of Gerontology: Psychological Sciences*, *46*, 127-129.

Wingfield, A., Alexander, A.H., & Cavigelli, S. (1994). Does memory constrain utilization of top-down information in spoken word recognition? Evidence from normal aging. *Language and Speech, 37 (3),* 221-235.

Wingfield, A., Kemtes, K.A., & Soederberg Miller, L. (2001). Adult aging and listening patterns for spoken prose: Spontaneous segmentation versus self-paced listening. *Experimental Aging Research, 27*, 229-239.

Wingfield, A., Lindfield, K., & Goodglass, H. (2000). Effects of age and hearing sensitivity on the use of prosodic information in spoken word recognition. *Journal of Speech, Language, Hearing Research, 43,* 915-925.

Wingfield, A., Wayland, S.C., & Stine, E.A.L. (1992). Adult age differences in the use of prosody for syntactic parsing and recall of spoken sentences. *Journal of Gerontology, 47(5)*, P350-P356.

Wolff, S., Schlesewsky, M., Hirotani, M., & Bornkessel-Schlesewsky, I. (2008). The neural mechanisms of word order processing revisited: Electrophysiological evidence from Japanese. *Brain and Language, 107,* 133-157.

Woodward, S.H., Ford, J.M., & Hammett, S.C. (1993). N4 to spoken sentences in young and older subjects. *Electroencephalography and Clinical Neurophysiology, 87,* 306–320.

APPENDIX 1. All nouns used in stimulus creation. Nouns were divided into 4-tuples and each word was used in 2 4-tuples. Triplets were created from the 4-tuples, permuted, and the middle word was replaced by whatever word was not included in the triplet. Each word could therefore be replaced by 4 alternate words.

A>>>	bag	dog	cup	bed
B>>>	cap	bug	book	kite
C>>>	pig	cat	bike	top
D>>>	duck	goat	cake	boot
	E^^^	**F^^^**	**G^^^**	**H^^^**

HEAR:			**SEE:**		
cup	bag	bed	CUP	BAG	BED
cup	dog	bed	CUP	DOG	BED
cup	**bag**	bed	CUP	**DOG**	BED
cup	**dog**	bed	CUP	**BAG**	BED
bag	dog	bed	BAG	DOG	BED
bag	cup	bed	BAG	CUP	BED
bag	**dog**	bed	BAG	**CUP**	BED
bag	**cup**	bed	BAG	**DOG**	BED
bag	cup	dog	BAG	CUP	DOG
bag	bed	dog	BAG	BED	DOG
bag	**cup**	dog	BAG	**BED**	DOG
bag	**bed**	dog	BAG	**CUP**	DOG
dog	bag	cup	DOG	BAG	CUP
dog	bed	cup	DOG	BED	CUP
dog	**bag**	cup	DOG	**BED**	CUP
dog	**bed**	cup	DOG	**BAG**	CUP
bed	bag	cup	BED	BAG	CUP
bed	dog	cup	BED	DOG	CUP
bed	**bag**	cup	BED	**DOG**	CUP
bed	**dog**	cup	BED	**BAG**	CUP
bed	dog	bag	BED	DOG	BAG
bed	cup	bag	BED	CUP	BAG
bed	**dog**	bag	BED	**CUP**	BAG
bed	**cup**	bag	BED	**DOG**	BAG
dog	cup	bag	DOG	CUP	BAG
dog	bed	bag	DOG	BED	BAG
dog	**cup**	bag	DOG	**BED**	BAG
dog	**bed**	bag	DOG	**CUP**	BAG
cup	bag	dog	CUP	BAG	DOG
cup	bed	dog	CUP	BED	DOG
cup	**bag**	dog	CUP	**BED**	DOG
cup	**bed**	dog	CUP	**BAG**	DOG

Appendix 2. Background cognitive data scores for older adults. Excluded subjects scores are shaded in gray (low MoCA scores) or blue (equipment failure).

Gender	Age	Years of Ed	MoCA	BNT	AWM recall	AWM t/f
M	71	11	25	56	36	42
M	66	11	27	58	24	42
F	75	12	29	52	27	42
F	68	18	27	47	38	42
F	73	13	24	52	26	41
F	72	13	23	52	27	42
F	68	14	26	58	31	42
M	74	18	27	58	31	42
M	64	17	22	59	27	41
F	73	10	26	57	26	42
M	66	11	25	60	29	42
F	66	20	25	55	30	42
F	72	16	23	58	28	42
F	69	20	29	55	26	42
M	72	16	26	57	29	42
M	73	11	29	56	34	41
mean	**70**	**15**	**27**	**56**	**30**	**42**
sd	**3.46**	**3.88**	**1.51**	**3.58**	**3.98**	**0.30**

MoCA: Montreal Cognitive Assessment
BNT: Boston Naming Test
AWM recall: Auditory Working Memory recall score
AWM t/f: Auditory Working Memory true/false score