

PERCEPTION OF JAPANESE TEMPORALLY-CUED PHONETIC CONTRASTS BY
JAPANESE AND AMERICAN ENGLISH LISTENERS:
BEHAVIORAL AND ELECTROPHYSIOLOGICAL MEASURES

by

MIWAKO HISAGI

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences
in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York

2007

UMI Number: 3283197

Copyright 2007 by
Hisagi, Miwako

All rights reserved.

UMI[®]

UMI Microform 3283197

Copyright 2007 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

©2007

Miwako Hisagi

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty
in Speech and Hearing Sciences in satisfaction of the dissertation requirement
for the degree of Doctor of philosophy.

Date

Valerie L. Shafer, Ph.D.
Co-Chair (adviser) of Examining Committee

Date

Winifred Strange, Ph.D.
Co-Chair of Examining Committee

Date

Martin R. Gitterman, Ph.D.
Executive Officer

Elise S. Sussman, Ph.D.
Supervisory Committee

Jason D. Zevin, Ph.D.
Outside Reader

The City University of New York

Abstract

PERCEPTION OF JAPANESE TEMPORALLY-CUED PHONETIC
CONTRASTS BY JAPANESE AND AMERICAN ENGLISH LISTENERS:
BEHAVIORAL AND ELECTROPHYSIOLOGICAL MEASURES

by

Miwako Hisagi

Co-chairs: Professor Valerie L. Shafer and Professor Winifred Strange

This study examined American listeners' perception of Japanese temporally-cued contrasts of vowels (*kado* vs. *kaado*) and consonants (*nifi* vs. *niffi*). Japanese (JP) phonology includes distinctive contrasts of consonant and vowel "length" that are cued almost exclusively by duration differences at the word level. American English (AE) phonology, on the other hand, uses such temporal cues only as secondary phonetic features. Previous cross-language behavioral studies of temporally-cued contrasts produced mixed results, in part due to choice of stimuli and tasks. The present study examines the question by considering the implications of first-language (L1) perception as an "over-learned" automatic process of selective perception in adults. The notion of "highly over-learned processes" interfering with second-language (L2) speech perception can be related to models of L1 acquisition in which infants are claimed to learn to automatically focus on language-specific features of speech sounds. In this study, the event-related potential component, Mismatch Negativity (MMN), was used to index "pre-attentive" discrimination of temporally-cued vowel and consonant contrasts in a categorical oddball task by two group of listeners: AE learners with no experience in

learning JP and native JP controls. The role of attention was examined in two experiments: Exp 1 Auditory-Attend, in which listeners attended to the auditory input (implicitly counted target deviants), and Exp 2 Visual-Attend, in which listeners attended to (implicitly counted) deviants in a simultaneously presented visual categorial oddball shape discrimination task. Follow-up behavioral tests indicated that AE listeners had more difficulty on the consonant contrast. Results showed effects of attentional focus for both language groups. For the easier vowel contrast, MMN amplitudes were smaller for AE listeners in the Visual-Attend experiment than JP controls and AE listeners in the Auditory-Attend experiment. For the more difficult consonant contrast, AE listeners produced very small MMNs in both attention experiments, while JP listeners produced larger MMNs in the Auditory-Attend experiment. This pattern of results supports the conclusion that native listeners show more robust indices of “pre-attentive” discrimination than do non-native listeners on both vowel and consonant contrasts; however, attentional resources may be required to process difficult consonant contrasts even in a listener’s native language.

Acknowledgements

I would like to express my appreciation for the National Science Foundation (NSF) Dissertation Enhancement Award (Division of Behavioral and Cognitive Sciences), which permitted me to focus on my dissertation for 18 months. I am also grateful for the National Institute of Health research grant to Dr. Valerie L. Shafer and Dr. Winifred Strange, which allowed me to sharpen my skills as a research assistant, as well as for the generous support from the City University of New York during my doctoral career: the CUNY Science Fellowship (2002-2004), the CUNY University Fellowship (2004-2007), and the CUNY Student Senator Merit Award (2003).

My most sincere thanks and appreciation go to the faculty of the Speech and Hearing Sciences Department at the City University of New York-Graduate Center.

Thanks to Dr. Valerie L. Shafer, my co-chair, for introducing me to the electrophysiological method and allowing me to use the LAB intensively. She pointed out a number of inconsistencies and shortcomings in mid-later versions of this thesis and helped me develop some of its more user-friendly features. Her ever-present sense of humor helped me enormously in coping with the pressures inherent in the work at hand. Her unending patience and welcomed mentoring made all the difference to me. From start to finish, she demanded my best, and her enthusiastic interest and generous advice made my work more focused, professional and polished than I ever could have achieved alone.

Thanks to Dr. Winifred Strange, also my co-chair, for accepting me into her laboratory and giving me an enormous amount of training especially during my earlier years at CUNY. I appreciate her teaching me the most rigorous approach to experimental

design while still insisting on capturing “real world” phenomena. She helped me enormously with the very earliest version of this thesis. I am fortunate to have had the opportunity to absorb some of her knowledge, skills, ethics, and fascination with cross-language speech perception. Special thanks for allowing me to use the LAB and for her thorough critique of each and every chapter.

My thanks go as well to Dr. Elise S. Sussman, Associate Professor, Dominick P. Purpura Department of Neuroscience, Albert Einstein College of Medicine, for introducing me to a different approach in the electrophysiological method and whose helpful comments, support and encouragement were invaluable. I am indebted to her for guiding me in the right direction and for teaching me how to think and write about MMN-related studies.

Without the strong support and guidance of these three fine scientists and dedicated teachers, this work would not have been completed. To all of them I am genuinely grateful.

I thank Dr. Jason D. Zevin, Assistant Professor of Psychology in Psychiatry, Sackler Institute for Developmental Psychobiology, Weill-Cornell Medical College, my outside reader, who also inspired much of my research.

I am grateful to Dr. Martin R. Gitterman, Executive Officer, for his consistent support of my teaching and my academic career, and for his open door at the Graduate Center.

Many thanks to Dr. James J. Jenkins, who was of enormous help, from picking apart the details of my statistical analyses on behavioral data to assessing the big picture. His support, encouragement, and concrete suggestions significantly improved my work.

Special thanks to Dr. Loraine K. Obler, who was a part of my second exam committee and who so effectively supported my academic work.

Special mention needs to be made of those who helped me behind the scenes and lightened my load throughout. Warm thanks to Gary Chant for his invaluable technical assistance, Linda Ashour and Loretta Walker for their administrative assistance, and the CUNY-GC security officers for their kind help and consideration.

I am also grateful to the “mass mind” of the Developmental Neurolinguistic Laboratory (DNL) and Speech Perception and Acoustic Laboratory (SAPL), with special thanks to its local and long-distance members (“passenger power rangers” for me), Dr. Erika Levy, Dr. Valeriy Shafiro, and Dr. Kanae Nishi.

And a very hearty thank-you to Dr. Arild Hestvik for teaching me EEGLAB.

It would be impossible to say enough about my special friends and colleagues for their forbearance and heart-warming encouragement: Many, many thanks to Hia Datta (coffee buddy as well), Karen Garrido-Nag, Marcin Wroblewski, and Dr. Nathan Maxwell, to name just a few.

I owe a particular debt of gratitude to Jane and Mitch Kirsch, who have always been there for me throughout my time in the United States. It has been their thoughtful and generous support that have enabled me to make it this far.

Finally, enormous thanks to my family and other relatives in Japan for helping me through my doctoral life with their loving support, patience, and tenderness -- and countless care packages.

Of course, any errors or omissions of research, theory or presentation are wholly my own.

Table of Contents

	<u>Page</u>
Abstract	iv
Acknowledgments	vi
Table of Contents	ix
List of Tables	xiv
List of Figures	xv
List of Appendices	xix
Chapter 1. Introduction	1
1.1 Statement of the Problem	1
1.2 Attention and Automatic Selective Perception (ASP)	1
1.3 Attention in First Language Acquisition	4
1.4 Electrophysiological Measures of Auditory Processing	5
1.5 Perceptual Assimilation Model (PAM)	9
1.6 Japanese Temporal Cues and Behavioral Measures	10
1.7 Temporal Cues and Electrophysiological Measures	13
1.8 Summary	15
1.9 The Present Study	16
1.10 Overview	19
Chapter 2. Experiment 1: Auditory-Attend Experimental Group	21
2.1 Method	21
2.1.1 Stimulus Materials	21

<u>Auditory stimuli:</u>	21
<u>Familiarization task materials:</u>	24
<u>Visual stimuli:</u>	24
2.1.2 Participants	25
2.1.3 Paradigms and Procedures	26
<u>ERP task:</u>	26
<u>Behavioral Task:</u>	27
2.1.4 Electroencephalogram (EEG) Recording and Data Analysis	27
<u>Electrode Placements:</u>	27
<u>Data Recording:</u>	27
<u>ERP Data Post-Processing:</u>	28
<u>Analysis- presence of MMN, N2b and P3b:</u>	29
<u>Behavioral Data:</u>	32
2.2. Results	32
2.2.1 Behavioral Results	32
<u>Behavior on the counting task (during ERP recording):</u>	32
<u>Behavior on the button-press task (after ERP recording):</u>	33
2.2.2 Electrophysiological Results	34
2.2.2.1 Mismatch Negativity (MMN), N2b and P3b	34
<u>Vowel Contrast:</u>	35
<u>[tado]-as-standard condition:</u>	35
<u>[taado]-as-standard condition:</u>	38
<u>[tado vs taado]:</u>	40

<u>Consonant Contrast:</u>	41
<u><i>[miʃi]-as-standard condition:</i></u>	41
<u><i>[miʃʃi]-as-standard condition:</i></u>	43
<u><i>[miʃi vs. miʃʃi]:</i></u>	45
<u>Vowel vs. Consonant Contrasts:</u>	47
2.2.2.2 Late Negativity (LN)	48
<u>Vowel Contrast:</u>	49
<u><i>[tado]-as-standard condition:</i></u>	49
<u><i>[taado]-as-standard condition:</i></u>	51
<u><i>[tado vs taado]:</i></u>	52
<u>Consonant Contrast:</u>	53
<u><i>[miʃi]-as-standard condition:</i></u>	53
<u><i>[miʃʃi]-as-standard condition:</i></u>	55
<u><i>[miʃi vs. miʃʃi]:</i></u>	56
2.2.3 Discussion	57
<u>Vowel versus Consonant Contrasts</u>	59
<u>Short as standard and Long as standard</u>	60
Chapter 3. Experiment 2: Visual-Attention	61
3.1 Overview	61
3.2 Methods	61
3.2.1 Stimulus Materials	61
<u>Auditory and Visual stimuli:</u>	61

<u>Familiarization task materials:</u>	62
<u>Behavioral Task:</u>	62
3.1.2 Participants	62
3.2.3 Paradigms and Procedures	62
3.2.4 Electroencephalogram (EEG) Recording and Data Analysis	63
3.3 Results	63
3.3.1 Behavioral Results	63
<u>Behavior on the counting task (during ERP recording):</u>	63
<u>Behavior on the button-press task (after ERP recording):</u>	64
3.3.2 Electrophysiological Results	65
3.3.2.1 MMN	65
<u>Vowel Contrast:</u>	66
<u><i>[tado]-as-standard condition:</i></u>	66
<u><i>[taado]-as-standard condition:</i></u>	67
<u><i>[tado vs taado]:</i></u>	69
<u>Consonant Contrast:</u>	69
<u><i>[miʃi]-as-standard condition:</i></u>	69
<u><i>[miʃʃi]-as-standard condition:</i></u>	70
<u><i>[miʃi vs. miʃʃi]:</i></u>	72
<u>Vowel vs. Consonant Contrasts:</u>	72
3.3.2.2 Late Negativity (LN)	73
<u><i>[tado]-as-standard condition:</i></u>	73

3.3.3 Discussion	74
<u>Vowel versus Consonant Contrasts</u>	76
Chapter 4. Experiment 1 vs. Experiment 2	79
4.1 Vowel Contrast	79
4.2 Consonant Contrast	81
4.3 Vowel vs. Consonant Contrasts	82
4.4 Summary	83
Chapter 5. General Discussion	85
5.1 Automaticity of Processing	89
5.2 Vowel vs. Consonant Duration Contrasts	93
5.3 Topography of Language Group Differences	95
Chapter 6. Conclusion	97
7. Tables	99
8. Figures	104
9. Appendices	164
10. References	180

List of tables

	<u>Page</u>
Table 1. Target Stimuli: real word models and the nonsense words that served as the experimental stimuli	99
Table 2. Summary of acoustic analysis of target nonsense stimulus materials: VOT and vowel consonant durations (ms) and pitch of two syllables (Hz) of selected four stimuli of each phonetic category. Standard deviation (SD) is shown for the contrasting target segments.	100
Table 3. An example of the test procedures and time for each task. The order of contrast type (vowel vs. consonant) and standard order (short-as-standard vs. long-as-standard) were counter-balanced.	101
Table 4. Auditory-Attend Experiment: Behavioral data The results of Counting auditory targets during ERP recording and Button-press auditory targets after the ERP recording showed the mean of the number correct, standard error of mean, median, mode and standard deviation. One of AE participant for <i>[miʃi]-as-standard condition</i> in Reaction time was not included due to that the participant missed the target.	102
Table 5. Visual-Attend Experiment: Behavioral data The results of Counting visual targets during ERP recording and Button-press auditory targets after the ERP recording showed the mean of the number correct, standard error of mean, median, mode and standard deviation.	103

List of figures

	<u>Page</u>
Figure 1. Design of Study	104
Figure 2. Categorical oddball discrimination paradigm for Auditory-Attend Experiment	105
Figure 3. Categorical oddball discrimination paradigm for Visual-Attend Experiment	106
Figure 4(a-d). Global Field Power (GFP)	107
Figure 5. AAE [tado] topographical voltage maps of the standard, deviant, and MMN	111
Figure 6. Counting of auditory deviants in the Auditory-Attend Experiment (14 blocks)	112
Figure 7. Button-press behavioral data to the auditory task for the Auditory-Attend Experiment (15 deviants)	113
Figure 8. Auditory-Attend Grand Average data (Fz)	114
Figure 9. Auditory-Attend Grand Average data (Pz)	115
Figure 10. [tado] in the Auditory-Attend Experiment - Language Group effect	116
Figure 11. [tado] in the Auditory-Attend Experiment – Site and Hemisphere effect	117
Figure 12. [tado] in the Auditory-Attend Experiment - Inversion effect	118
Figure 13. [taado] in the Auditory-Attend Experiment - Language Group effect	119
Figure 14. [taado] in the Auditory-Attend Experiment - Site and Hemisphere effect	120
Figure 15. [taado] for the Auditory-Attend Experiment – Inversion effect	121
Figure 16. [tado vs. taado] in the Auditory-Attend Experiment – Order effect	122

Figure 17. [tado vs. taado] in the Auditory-Attend Experiment – Hemisphere effect	123
Figure 18. [tado vs. taado] in the Auditory-Attend Experiment – P3b effect	124
Figure 19. [miʃi] in the Auditory-Attend Experiment – Site effect	125
Figure 20. [miʃi] in the Auditory-Attend Experiment – Inversion effect	126
Figure 21. [miʃʃi] in the Auditory-Attend Experiment – Language Group effect	127
Figure 22. [miʃʃi] in the Auditory-Attend Experiment – Site and Hemisphere effect	128
Figure 23. [miʃʃi] in the Auditory-Attend Experiment – Inversion effect	129
Figure 24. [miʃi vs. miʃʃi] in the Auditory-Attend Experiment – Order effect	130
Figure 25. [miʃi vs. miʃʃi] in the Auditory-Attend Experiment – Site effect	131
Figure 26. [miʃi vs. miʃʃi] in the Auditory-Attend Experiment – Hemisphere effect	132
Figure 27. Vowel vs. Consonant in the Auditory-Attend Experiment	133
Figure 28. Vowel vs. Consonant in the Auditory-Attend Experiment	134
Figure 29. Vowel vs. Consonant in the Auditory-Attend Experiment	135
Figure 30. Late Negativity (LN) [tado] in the Auditory-Attend Experiment – Site and Hemisphere effect	136
Figure 31. Late Negativity (LN) [tado] in the Auditory-Attend Experiment – Language Group effect	137
Figure 32. Late Negativity (LN) [tado] in the Auditory-Attend Experiment – Site and Hemisphere effect	138
Figure 33. Late Negativity (LN) [taado] in the Auditory-Attend Experiment – Language Group effect	139
Figure 34. Late Negativity (LN) [taado] in the Auditory-Attend Experiment – Site and Hemisphere effect	140

Figure 35. Late Negativity (LN) [tado vs. taado] in the Auditory-Attend Experiment – Site effect	141
Figure 36. Late Negativity (LN) [miʃi] in the Auditory-Attend Experiment - Site and Hemisphere effect	142
Figure 37. Late Negativity (LN) [miʃʃi] in the Auditory-Attend Experiment – Site and Hemisphere effect	143
Figure 38. Late Negativity (LN) [miʃi vs. miʃʃi] in the Auditory-Attend Experiment – Site effect	144
Figure 39. Late Negativity (LN) [miʃi vs. miʃʃi] in the Auditory-Attend Experiment – Hemisphere effect	145
Figure 40. Counting of visual deviants in the Visual-Attend Experiment (14 blocks)	146
Figure 41. Button-press behavioral data to the auditory task for the Visual-Attend Experiment (15 deviants)	147
Figure 42. Visual-Attention Grand average data (Fz)	148
Figure 43. [tado] in the Visual-Attend Experiment – Site and Hemisphere effect	149
Figure 44. [taado] in the Visual-Attend Experiment – Language Group effect among three hemispheres	150
Figure 45. [tado vs. taado] in the Visual-Attend Experiment – Hemisphere effect	151
Figure 46. [miʃi] in the Visual-Attend Experiment – Hemisphere effect	152
Figure 47. [miʃʃi] in the Visual-Attend Experiment – Hemisphere effect	153
Figure 48. [miʃi vs. miʃʃi] in the Visual-Attend Experiment – Hemisphere effect	154
Figure 49. AE group [tado] in the Auditory- and Visual- Attend Experiments	155
Figure 50. Vowel in the Auditory- and Visual- Attend Experiments – Language group effect	156

Figure 51. Vowel in the Auditory- and Visual- Attend Experiments – Experimental Group effect	157
Figure 52. JP group [miʃi] in the Auditory- and Visual- Attend Experiments	158
Figure 53. JP group [miʃʃi] in the Auditory and Visual- Attend Experiments	159
Figure 54. Consonant in the Auditory- and Visual- Attend Experiments – Experimental Group effect	160
Figure 55. Vowel vs. Consonant in the Auditory- and Visual- Attend Experiments – Site effect	161
Figure 56(a-b). Vowel vs. Consonant in the Auditory- and Visual- Attend Experiments – Time effect	162

List of appendixes

	<u>Page</u>
Appendix A. Protocol for Speakers	164
Appendix B. Acoustic analysis of Target real words	167
Appendix C. Acoustic analysis of Real words of previous study	168
Appendix D. Consent Forms	169
Appendix E. Language background questionnaire	171
Appendix F. The number of stimuli per block	173
Appendix G. Instructions for ERPs (auditory)	175
Appendix H. ERP Data Post-Processing	176
Appendix I. Auditory-Attend Experiment: All nine sites with LM and RM: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4	177
Appendix J. Instructions for ERPs (visual)	178
Appendix K. Visual-Attend Experiment: All nine sites with LM and RM: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4	179

Chapter 1. Introduction

1.1 Statement of the Problem

Past cross-language research has shown that phonetic segments that are not contrasted in a listener's native language (L1) but are contrasted in another language (L2) can cause significant perceptual difficulty if the listener has not had considerable experience with the L2 (see Best, 1995; Strange, 1995). For example, Japanese listeners have difficulty perceiving the English r/l phonetic contrast (e.g., rock vs. lock) although the Japanese can produce such sounds, which exist in Japanese as phonetic variants of one contrast. One interpretation of this finding is that the highly over-learned processes by which adult listeners differentiate and recognize native phonemic categories somehow "interfere with" the perception of phonemic categories that are not present in the native language. The role of attention was discussed by Gordon, Eberhardt, and Rueckl (1993), who suggest that the amount of attention makes a difference in recognizing the important acoustic information for phonetic segments/contrasts. The present study examined the question from a novel direction by considering the implications of first language perception as an "over-learned" process, and by exploring the effect of attention in speech perception, using neurophysiological indices of discrimination, event-related potentials (ERPs).

1.2 Attention and Automatic Selective Perception (ASP)

Researchers in speech perception often make references to the notion that speech processing is "automatic" and that the development of speech perception requires attention, but these views have generally been mentioned in passing without presenting a detailed model of their role in speech perception. Theoretical treatments of attention (and

in particular auditory attention) have defined selective or focused attention in the following way. Pugh, Shaywitz, Shaywitz, Fulbright, Byrd, Skudlarski, Shankweiler, Katz, Constable, Fletcher, Lacadie, Marchione, and Gore, (1996: 160) states that it is necessary to have “the organism target some sensory dimension for focal processing while effectively inhibiting others” and this is selective attention. They suggest that cognitive systems are active in controlling the attention shift and the engage/disengage operations when task demands stress any component operations. In addition, within such controlling operations, processing information that relates to a particular sensory dimension should affect the responsiveness or activation state of cortical and/or subcortical regions. Researchers have argued over whether there is early or late selection of stimuli to be processed (see Lavie, 2005). In Lavie’s model, high perceptual load leads to early selection of relevant stimuli and less processing of distracter stimuli. In contrast, high cognitive load leads to late selection of relevant stimuli and, therefore, greater processing of irrelevant stimuli because fewer resources are available for controlling processing. Irrespective of which model is followed, more attentional resources are needed for information processing that is less automatic. As Näätänen (1988) points out, selective attention and automaticity are inversely related. These theoretical models of auditory attention have not been extended to address native-language speech perception, which is a highly over-learned process. Research on visual selective attention suggests that highly familiar information, such as famous faces, are more difficult to ignore in selective attention tasks, suggesting that they may be automatically processed (see, Lavie & DeFockert, 2003). These data suggest that the role of attention in processing speech, which is highly familiar, differs from processing novel auditory information.

A recent theoretical framework presented by Strange & Shafer (2006) directly addresses the role of selective perception/attention processes in native and non-native speech perception. Strange's Automatic Selective Perception (ASP) model (Strange & Shafer, 2006), characterizes first-language speech processing in adults as reflecting automatic selective perceptual routines for the detection of language-specific (phonologically relevant) information in acoustic signals. In the ASP model, selective perceptual processing of L1 phonetic segments/sequences is automatic due to extensive L1 experience. In contrast, non-native listeners may require attentional resources to discriminate the same phonetic contrasts. Strange and Shafer (2006) consider perceptual difficulty from the viewpoint of the L2 learning processes. Their focus is to predict levels of difficulty in cross-language speech perception as a function of stimulus and task factors, as well as L1/L2 phonological and phonetic similarities and differences. They agree with other theorists (e.g. Flege, 1995, Best, 1995) who suggest that perceptual difficulties partially underlie production difficulties of L2 learners and explain why L2 learners can neither produce non-accented speech nor discriminate phonetic differences in the L2 at native levels of accuracy. To illustrate this issue, for native speakers of American English (AE), detecting the acoustic information needed to differentiate AE /æ/ in "pat" and /ɑ/ in "pot" is automatic, requiring few attentional resources. In contrast, for native Japanese speakers, /æ/ and /ɑ/ are both perceived as very similar to Japanese /a/, although acoustic differences between these segments may be detected under some listening conditions. Thus, when listening to AE speech, Japanese L2 learners of AE may have difficulty categorizing these segments as different phonemes unless the stimulus and task conditions allow them to use attentional resources.

The ASP model predicts success by non-native listeners in a test of phonetic perception to be dependent on the degree of attention (i.e., cognitive effort) that can be used in the task. Non-native listeners can recover the phonetically-relevant information specified by multiple spectral and temporal parameters, such as formants and voice onset time (VOT), but they require additional cognitive effort to access that phonetically-relevant information to discriminate or identify non-native phonemic contrasts. Without focused attention on the relevant cues differentiating these phones, the non-native listeners may not categorize stimuli appropriately. Differences between L1 and L2 phonology in how phonetic information specifies contrasts predict differences in the amount of attention needed in processing.

1.3 Attention in First Language Acquisition

The notion of “highly over-learned processes” interfering with L2 speech perception can be related to models of first language acquisition, in which selective attention is hypothesized to play an important role in learning phonological and lexical information (Jusczyk, 1997; Werker & Curtin, 2005). Jusczyk (1997) claimed that infants learn to focus attention automatically on phonologically relevant features of speech sounds (see Word Recognition and Phonetic Structure Acquisition (WRAPSA), Jusczyk, 1997). Werker and Curtin (2005) claim that the level of automaticity in language processing depends on the developmental stage and that attention plays a role in selecting particular information (see Processing Rich Information from Multi-dimensional Interactive Representations (PRIMIR), Werker & Curtin, 2005). In summary, a fundamental claim in both WRAPSA and PRIMIR is that infants use selective attentional processes to detect and integrate phonologically-relevant acoustic cues in their native

language. An additional claim of WRAPSA is that the process of selecting the relevant cues becomes automatic with experience in L1.

1.4 Electrophysiological Measures of Auditory Processing

A large amount of research on human perception focuses on automaticity of selective perception/attention processes. Näätänen (1988) suggested that perceptual processing is not completely automatic but depends on the level of attention. Näätänen (1988:119) states, “The greater the extent to which information processing is automatic, the less there is to be explained by attention.” Thus, according to Näätänen, there is a strong relationship between automaticity and the level of attention. Exploring automaticity of information processing can be a challenge using behavioral methods, since it is difficult to evaluate directly the processing of a particular stimulus without a behavioral response to that stimulus. Over the last decade, studies have used neurophysiological tools, such as ERPs and functional Magnetic Resonance Imaging (fMRI), to examine “preattentive” perceptual processing of auditory stimuli (Näätänen 1988; Kahneman & Treisman, 1984; Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Yucel, Petty, McCarthy & Belger, 2005; Polich, 2003). In particular, ERP measures can provide information regarding processes preceding behavioral output and in the absence of focused attention. Specifically, the discriminative ERP component Mismatch Negativity (MMN) serves as an index of pre-attentive discrimination of a stimulus or of a pattern change in a sequence of auditory stimuli (Näätänen, 1990; Näätänen & Winkler, 1999, Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler., 2001 for review). Some studies have indicated that the MMN also reflects the subject’s behavioral responses (Tervaniemi, Rytönen, Schröger, Ilmoniemi, & Näätänen, 2001; Tiitinen, May,

Reinikainen, & Näätänen., 1994), although there is not always a direct relationship between the two (e.g., see Shafer, Schwartz & Kurtzberg, 2004).

The MMN to a change in acoustic properties can be elicited both with and without attention to the auditory modality (e.g., Shafer et al, 2004; Sussman, Kujala, Halmetoja, Lyytinen, Alku, and Näätänen, 2004; Szymanski, Yund, and Woods, 1999) or to the visual modality (Alho, Woods, & Algazi, 1994). For example, Sussman et al. (2004) examined attentional effects on speech processing of a Finnish vowel duration contrast, “tuli (fire) vs. tuuli (wind)”. A second goal was to determine whether detecting these vowel duration changes is dependent on their position within a word (beginning, middle, or end of three-syllable pseudo-words). They found an effect of attention in which the MMN amplitude was larger at the mastoid electrode sites (but not frontocentral electrode sites) when the listeners attended to the target stimuli compared with when they ignored such sounds. The first position had the largest MMN followed by the second and the third had the smallest MMN.

The results of another study using MMN in a dichotic listening paradigm (Szymanski et al., 1999) also suggested that speech processing is affected by attention. Participants were asked to respond to targets in one ear while ignoring the information to the other ear. There were two types of deviant features: 1) intensity, where the intensity of /b/ was increased /B/ or 2) phoneme, where /b/ was changed to /g/. These deviants could occur in three different positions 1) VC formant transitions (/iBbi/ or /igbi/); 2) CV formant transitions (/ibBi/ or /ibgi/), or 3) both, where VC and CV formant transitions changed (igi/ or /iBi/). The listener’s native phonological system was reflected at the automatic level, but attention altered the relevance of acoustic-phonetic cues. Attention

enhanced the MMN and the MMN amplitude was equal or larger for greater differences between word forms in terms of intensity or formant transitions, when the forms were behavioral targets. However, when phoneme change (formant differences) was not the target (intensity was), a change in the formant transitions of [ibi] both from the first vowel and into the second vowel to yield [igi] (condition 3 above) yielded a smaller MMN than only changing the first or second transitions (which resulted in [igbi] and [ibgi], respectively)(conditions 1 and 2 above) even though these single changes were acoustically less different than the double change. In other words, when asked to identify intensity deviants and not phonemic deviants, listeners processed the non-target phonemic deviants automatically on the basis of phonological salience, with the [ibi-igbi] and [ibi-ibgi] being phonologically more salient than [ibi-igi] change.

The status of a speech contrast in a listener's L1 has been shown to affect the amplitude and/or latency of MMN (e.g., Näätänen, Lehtokoski, Lennes, Cheour, Huotilainen, Iivonen, Vainio, Alku, Ilmoniemi, Luuk, Allik, Sinkkonen, & Alho, 1997; Shafer, et al., 2004; Nenonen, Shestakova, Huotilainen, & Näätänen, 2003). Specifically, several studies have shown that MMN is larger in amplitude or earlier in latency in adult listeners for whom a speech contrast is distinctive in their L1 as compared to listeners for whom the contrast is non-distinctive (Menning, Imaizumi, Zwitterlood, & Pantev, 2002; Nenonen et al, 2003). This finding supports the notion that L1 learning early in life leads to automatic pre-attentive processing of phonetic contrasts. In general, the presence of an MMN means that a change in the stimulus was discriminated by the brain at a pre-attentive stage of processing, whereas the absence of an MMN means that a change in the stimulus was not discriminated by the brain at this early stage of processing. Furthermore,

a smaller and later MMN indicates a more difficult discrimination (Menning et al., 2002; Nenonen et al 2003; Nenonen, Shestakova, Huotilainen, & Naatanen, 2005; Dehaena-Lambertz, Dupoux, & Gout, 2000; Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006). A smaller or later MMN suggests that listeners cannot discriminate non-native speech sounds as well as native contrasts. Thus, the amplitude and latency of MMN reflects the level of difficulty of a speech contrast that is a function of the listeners' past experience with the contrast.

There are also attention-related ERP components, N2b and P3b (see Sussman et al., 2004). The N2b is elicited by stimulus deviance when a listener attends to an infrequent stimulus within a sequence of frequent stimuli. The N2b usually occurs around 200 and 300 ms from the stimulus onset and follows the MMN. The P3b is also elicited by attended infrequent stimuli, but unlike MMN, the deviant target has to be the focus of attention. The processes indexed by these components are not the focus of this dissertation, but N2b and P3b are expected to be elicited, along with MMN, when the stimulus of interest is the attended target.

In summary, processes indexed by MMN are largely considered to reflect pre-attentive or automatic discrimination processes (Sussman, Winkler, & Wang., 2003). Both N2b and P3b components are seen only under conditions of focused attention. Listeners show smaller or later MMNs to non-native contrasts. These characteristics of MMN demonstrate that it can be used to explore automaticity of speech processing in native versus non-native listeners.

1.5 Perceptual Assimilation Model (PAM)

Some non-native contrasts are more difficult to discriminate and categorize than others. The level of difficulty can be predicted on the basis of similarity across languages and from the phonology of L1. Best's Perceptual Assimilation Model (PAM) (Best, 1995, 2003; Best & McRoberts, 2003) attempts to explain patterns of difficulty of L2 contrasts in relation to the L1 phoneme inventory. Three different perceptual assimilation patterns are introduced here. In the Two Category (TC) pattern, the contrasting non-native segments are assimilated into two native categories, and thus are the least difficult for non-native listeners to discriminate. In the Single Category (SC) pattern, the two non-native segments are assimilated as equally good (or poor) exemplars of a single native category: these segments are hardest for non-native listeners to discriminate. Another pattern of assimilation, called the Category Goodness (CG) pattern is seen when two non-native segments are assimilated into a single native category, but one is perceived as a better instance of the native category than the other. The difficulty in discriminating this type of contrast ranges from moderate to very good, depending on the level of perceived difference in goodness.

According to this model differences in phonology between two languages such as English, a stress-timed language, and Japanese, a mora-timed language, can be used to predict differences in perception of non-native contrasts. For instance, the two American English (AE) vowels /æ:/ and /ɑ:/ are assimilated into one native phoneme category of /ɑ:/ by Japanese, often as equally good exemplars of that category; thus, they are very difficult for non-native listeners to discriminate (Hisagi, Strange, Akahane-Yamada, Kubo, 2005). In comparison, for the AE /ɑ:-Λ/ contrast, if Japanese assimilate /ɑ:/ as a

good exemplar of long Japanese /a:/ and /Λ/ as a poorer exemplar of /a:/ or as short /a/, discrimination would be much better.

Patterns of perceptual assimilation of Japanese vowel length and consonant length contrasts by AE listeners might predict higher discrimination accuracy on vowel duration contrasts than on consonant duration contrasts by AE listeners. AE listeners use durational differences in vowels as a secondary cue to differentiate vowel phonemes (e.g., “bade” versus “bed”) and consonant phonemes (e.g., “pat” vs. “pad”) (see Borden, Harris & Raphael, 2003). Consonant duration differences do not serve as a secondary phoneme cue in English, although they do serve as cues for a word boundary in English (e.g., night train vs. night rain). Thus, we can assume that Japanese long and short consonants assimilate into the same category for English speakers, with neither member of the pair being considered a better exemplar of the English consonant duration (SC pattern in PAM), while Japanese long and short vowels may be assimilated in a CG (short /a/ as a poorer exemplar of AE /ɑ:/, long /a:/ as a better exemplar of AE /ɑ:/) or TC pattern (short /a/ to /Λ/ and long /a:/ to /ɑ:/).

In addition, there may be psychoacoustic reasons why certain contrasts are easier than others, and some studies suggest that duration differences are easier than spectral differences (Werker & Tees, 1984; Bohn & Flege, 1990). However, to this author’s knowledge, there are no claims about whether consonant duration differences are more or less salient than vowel duration differences from a psychoacoustic viewpoint.

1.6 Japanese Temporal Cues and Behavioral Measures

The present study focused on Japanese (JP) temporally-cued differences because Japanese phonology includes distinctive contrasts in both consonant and vowel

“quantity” (or “length”) that are cued almost exclusively by duration differences at the word level. AE listeners, on the other hand, use such temporal cues only as secondary cues. The prosodic structure of JP, within which the phonemic length contrasts and syllable-structure contrasts are embedded, differs considerably from the prosodic patterns which English speakers experience. For example, the minimal pair [obasan] “aunt” versus [obaasan] “grandmother” differentiates meaning by vowel duration. The minimal pair [oto] “sound” versus [otto] “husband” differentiates meaning by consonant duration. As described earlier, AE listeners use vowel duration as a secondary cue to final consonant voicing (e.g., the vowel is longer in “bid” compared to “bit”), but consonant duration is not a cue at the word level (Borden et al., 2003). Even here, other cues, such as aspiration, can be used to segment the words appropriately. AE phonetic experience with consonant duration is, therefore, much more limited than it is with vowel duration. If phonetic experience with temporal differences in segments and syllables were predictive of discrimination of non-native phonological contrasts of quantity, as suggested by some studies (Werker, Gilbert, Humphrey, & Tees, 1981; Tees & Werker, 1984; Polka, 1991), then JP consonant length contrasts should be more difficult than vowel length contrasts for naïve AE listeners.

Past investigations of AE speakers’ perception of JP duration contrasts have presented mixed results. Several studies support the notion that AE phonology predicts performance of AE listeners on JP temporal cues. Specifically, two studies found that performance on vowel duration (e.g., [kado] vs. [kaado]) was better than on consonant duration (e.g., [ni/i] vs. [niʃ/i]) for AE speakers (Enomoto, 1992; Muroi, 1995). In contrast, other studies of AE listeners’ perception of JP duration contrasts have shown no

differences in performance for vowels versus consonants (Tajima, Kato, Rothwell, & Munhall, 2003; Hisagi & Strange, under revision). In general, performance was well above chance for both contrast types in all of these studies. However, within a contrast type (e.g., consonant duration), performance varied depending on the particular word-pair (e.g., Enomoto, 1992; Hisagi & Strange, under revision).

Results from these JP studies were inconsistent. Two of the four studies support the claim that L1 predicts L2 performance on temporally-cued distinctions, but all suggest that these distinctions are *relatively* easy for naïve or L2 listeners. However, performance on these studies varied from a low of 76% to a high of 100% accuracy. Under ideal listening conditions in which stimulus variability was low, with synthetic speech stimuli presented in citation form, performance was high and often at ceiling. Under more difficult conditions with more complex stimuli, performance by AE listeners was considerably worse (Hirata, 1990; Tajima et al., 2003; Hisagi & Strange, under revision). Manipulation of instructions (directing attention to the duration differences) also changed performance (Hisagi & Strange, under revision). A second goal of this dissertation was to examine further whether LI predicts performance on non-native contrasts by including non-native vowel- and consonant-duration contrasts that have different status in the L1.

The results from these studies leave us in doubt concerning the degree to which perception of temporally-cued distinctions in an L2 are affected by a listener's L1. However, we know that identifying word meaning on the basis of duration is difficult for L2 learners of JP who do not use duration as a primary distinctive cue in their L1 in real life situations (Toda, 1994; Muroi, 1995; Hirata, 2004a, 2004b). The considerable

differences across tasks and stimuli probably account for some of the variability in performance across studies. Additionally, the behavioral response is the output of a series of processes, and tasks and stimuli may affect these processes differently. Participants may respond based on different criteria. For example, one participant may focus more on acoustic differences, whereas another may focus on phonological differences.

Examination of speech processing at a stage preceding the behavioral response using a measure such as MMN could help explain the different pattern of results seen across studies.

1.7 Temporal Cues and Electrophysiological Measures

Recently, several studies have compared L1 versus L2 processing of temporally-cued contrasts and found larger MMNs for native than non-native listeners (Menning et al., 2002; Nenonen et al. 2003; Nenonen, et al., 2005; Dehaena-Lambertz et al., 2000; Ylinen, et al., 2006). Advanced Russian learners of Finnish (participant age: 10-14 years old) and native Finnish speakers as control subjects were examined on a vowel duration difference (*[ka:]* 200 ms (standard) versus *[ka]* 150 ms (deviant) or *[kæ:]* to *[kæ]*) that is phonemic in Finnish but not Russian (Nenonen et al., 2003; Nenonen et al., 2005). The first study also examined duration differences of complex tones as a non-speech control. For the non-speech stimuli, there was no difference between Russian and Finnish listeners in amplitude of MMN. For the speech stimuli, both groups showed significant MMN responses, but the amplitude of MMN was larger for Finnish (L1) than for Russian (L2) listeners.

Similar to the study by Nenonen, et al. (2003), Ylinen et al. (2006) examined the Finnish phoneme length difference *[tu:ku]* versus *[tuku]* by Russian L2 learners of

Finnish and by native Russian speakers who had no knowledge of Finnish, with a Finnish control group. MMNs were elicited by all three groups in both across-category (across-category deviant which was 62 ms shorter than standard, CVCV with a short vowel) and within-category change (within-category deviant which was 62 ms longer than standard CVVCV with a long vowel). The standard was categorized as CVVCV with a long vowel. A larger MMN was elicited in native Finnish speakers compared with the other two groups. However, there was no difference between the two non-native groups. There was also an interesting finding in that the response pattern to across- and within-category changes was the same in all three groups. If the MMN reflects the phoneme boundary, then the across-category change should elicit a larger MMN than that elicited by the within-category change for the Finnish speakers, but not for the non-native speakers of Finnish. Therefore, the researchers concluded that the MMN was enhanced because of activation of native language phonetic prototypes, but did not reflect phoneme boundaries.

In a study using a different design, Dehaene-Lambertz and colleagues (2000) examined a contrast of syllable structure (e.g., [igʷmo] vs. [igmo]) using multiple tokens to study French (L1) and JP (L2) perception of these syllable structures using ERPs. They found a discriminative ERP response for the L1 but not the L2 listeners. The component reported was probably the MMN or the MMN overlapping with an N2b-like component, which is elicited in tasks requiring discrimination of targets (Näätänen, 1990).

Menning, et al., (2002) showed that, even with training on a particular duration contrast, non-native listeners show worse performance and a smaller MMN than native listeners. They trained German participants (unfamiliar with JP) on durational changes in a JP consonant contrast, leading to improved behavioral accuracy and a larger and earlier

MMN. However, the JP L1 listeners still performed better on the behavioral task and had a significantly larger MMN than did the German group.

These findings are consistent with the notion that non-native groups are less automatic than native groups at processing durational contrasts. In sum, these studies showed that MMN amplitude reflected L1-L2 phonological differences. Native speakers, in general, exhibited larger and earlier MMNs than non-native speakers. The studies demonstrated that MMN can serve as a useful measure in exploring speech perception in a non-native language. Previous studies have not compared speech processing under both attention-independent and attention-dependent conditions. Most studies, except the Dehaene-Lambertz study, were conducted with a passive task where the listener's attention was directed away from the auditory stimuli by instructions to watch a video movie or to read a book, and this task does not carefully control for attention. Moreover, most of these ERP studies have examined discrimination of a pair of stimuli that can be discriminated easily using acoustic-phonetic cues (but see Dehaene-Lambertz et al., 2000).

1.8 Summary

Behavioral research has provided mixed results regarding the effect of L1 on the perception of L2 temporally-cued distinctions. These mixed results are partly due to differences in the complexity of the stimuli and task demands. No electrophysiological study directly comparing discrimination of vowel and consonant contrasts in native (JP) versus non-native (AE) listeners has been undertaken. Investigations of speech perception of non-native contrasts using MMN illustrate that this electrophysiological component allows comparison of L1/L2 effects under different attention conditions.

1.9 The Present Study

The purpose of this study was to extend this research on the effects of phonological functionality (L1 experience) and attentional variables on the perception of temporally-cued distinctions by examining the perception of vowel (*tado* vs. *taado*) and consonant (*miɸi* vs. *miɸɸi*) duration contrasts by native and non-native listeners in attend (auditory mode) and non-attend (visual mode) conditions. A primary goal was to examine whether attention to the auditory versus visual mode affects auditory perception in different ways in listeners for whom the contrasts are native versus non-native. Specifically, an auditory-mode task was designed to reveal non-native versus native performance under focused attention to the contrasts of interest, whereas a visual-mode task was designed to examine automaticity of processing the contrasts of interest when attention was focused away from them. The ERP components, MMN, N2b, and P3b, and behavioral discrimination measures were used to examine these questions. The groups tested for this study were native speakers of JP and native speakers of AE who had never studied JP and who had had limited experience with any second languages. AE and JP listeners are likely to use different cues to recognize JP words, because JP, but not AE, listeners rely heavily on temporal cues to distinguish meaning at the word level. The performance of naïve AE listeners provides a baseline indicating the perceptual difficulty of these JP contrasts for learners.

Discrimination of stimuli can be achieved on the basis of acoustic-phonetic differences and does not require categorical judgments (e.g., Alho et al., 1994; Nenonen et al., 2003; Sussman, 2004; Inouchi, Kubota, Ferrari, & Roberts., 2003; Winkler & Czigler., 1998; Arnott & Alain, 2002; Sussman et al., 2004; Garrido, Hisagi & Shafer, in

preparation). One way to lead participants' attention to phonemic differences, not acoustic ones, is the categorial discrimination task. A behavioral categorial (name identity) discrimination task requires listeners to decide whether two physically different segments belong to the same or different phonetic categories without requiring the use of unfamiliar response labels. Very few ERP studies have been done using stimuli in which more than one token of each phonetic category are presented (but see, Dehaene-Lamberts et al., 2000; Menning et al., 2002; Winkler, Kujala, Alku, & Näätänen, 2003). This dissertation used a "categorial oddball" discrimination task in which multiple natural-speech exemplars of each phonetic segment were presented as the "standard" and "deviant" stimuli. This technique required participants to rely on categorial information in discrimination. Nonsense words were chosen in order to prevent the JP listeners from using lexical information to facilitate discrimination.

There were two experiments. In the Auditory-Attend Experiment (Exp 1), ERP indices of auditory discrimination of vowel and consonant duration contrasts were examined in AE and JP listeners in a task requiring attention to the speech stimuli of interest via the counting of the deviant word forms. In the Visual-Attend Experiment (Exp 2), these contrasts were examined in a visual-attention task requiring counting of visual deviants. The visual-attention task was used to direct attention away from the auditory stimuli in a more controlled fashion than in the typical "passive" visual task (e.g., watching a video) used in most studies. For each participant in each experiment, two stimulus types (vowel and consonant) were presented on the same day.

This research addressed the following questions:

Do AE and JP listeners show robust discrimination of JP temporally-cued contrasts when attention is on the *auditory* channel (Exp 1), as indexed by MMN and by the attention-dependent N2b and P3b, as opposed to when attention is directed away from the auditory channel to the visual channel (Exp 2), as indexed by MMN?

a. If so, are there differences between JP versus AE groups in the amplitude and latency of MMN indicating differences in processing as a function of phonological experience?

b. Do AE or JP listeners show differences in amplitude and latency of MMN between vowel and consonant contrasts reflecting possible differences in psychoacoustic salience?

In further analyses, the results of both experiments (Exp 1 and 2) were examined together, asking the question, “Are there differences between Exp 1 versus Exp 2 groups in amplitude and latency of MMNs?” Specifically, this comparison determined the role of attentional focus on brain indices of discrimination by native and non-native listeners. Specifically do non-native listeners reflect less automatic processing of these contrasts and do native listeners show any effects of attention on processing these duration contrasts?

We predicted that native speakers’ perception patterns in discriminating phonetic contrasts would reflect automatic processing, while non-native speakers’ brain responses to these contrasts would reflect the effects of attentional manipulations. The AE listeners might not show evidence of discrimination of the duration differences without attention (Exp 2), whereas the native speakers of JP would show automatic pre-attentive discrimination. Specifically, in the Visual-Attend Experiment (Exp 2), AE listeners

would show absent or small MMNs because processing of non-native contrasts is less automatic. In the Auditory-Attend Experiment (Exp 1), the AE listeners would show MMNs because they could make use of attentional resources, but the MMNs might be smaller in amplitude and/or later in latency than for the JP listeners. In both experiments, it was predicted that MMNs would be larger for vowel contrasts than for consonant contrasts for the AE listeners because English uses duration as a secondary cue for categorizing vowels but not consonants. We also predicted that the behavioral task requiring target identification with a button press would produce better than chance-level performance because previous studies, including studies that involved harder tasks and more complex stimuli, showed high performance by non-native listeners. Behavioral discrimination of the contrasts would provide further information regarding the relationship between MMN and behavioral measures.

1.10 Overview

This dissertation employed a between subjects design. American English (AE) was the target and Japanese (JP) the control, with two different experiments manipulating attention: an Auditory-Attend Experiment (Exp 1) and a Visual-Attend Experiment (Exp 2). Figure 1 describes the design of the dissertation. In Exp 1 for the ERP task, participants silently counted the deviant auditory targets. In Exp 2 for the ERP task, participants silently counted the deviant targets in a visual sequence and ignored the auditory stimuli. Within each experiment, there were two independent variables, vowel and consonant contrast types. Within each contrast type, there were two nested orders of testing: a) short as standard, long as deviant and b) long as standard, short as deviant. The study followed a mixed design with two between groups variables (language and attend

conditions) and two within subjects variables (contrast type and order within each contrast type). After the ERP experiments, a button-press behavioral discrimination task requiring discrimination of the auditory targets was given.

Chapter 2. Experiment 1: Auditory-Attend Experimental Group

2.1 Method

2.1.1 Stimulus Materials

Auditory stimuli: The stimuli were created in sentence context to maintain the naturalness of each token. The stimulus materials were initially recorded by four native female speakers of JP (Tokyo dialect), in an Industrial Acoustics Co (IAC) chamber. The output from a dynamic microphone (SHURE SM-48) was pre-amplified (Earthworks microphone preamp Lab 101), digitized (sampling rate: 22.05kHz and 16 bits resolution and mono channel) via a Soundblaster Live Wave DF80 sound card, and stored as wave files on a PC computer using SoundForge 4.5 (Sonic Foundry, Inc) software. Calibration was performed such that the direct current (DC) offset was adjusted, and autosnap to zero was selected. The input level was set such that target words varied from approximately -12 dB to -3 dB, with no clipping. Root mean squared (RMS) power and DC offset were periodically monitored by means of SoundForge's "statistics" function to ensure appropriate levels (-3 ~ -12 dB range) during recording. The experimenter and the speaker communicated with each other through a wired intercom, and the experimenter monitored the speaker's productions during the recording session.

The Japanese target words were disyllables consisting of contrasting vowels (in bold) (CVCV vs. CVVCV) and contrasting consonants (in bold) (CVCV vs. CVCCV). Table 1 shows both target real words and nonsense words. Based on the study of Hisagi and Strange (under revision), the most difficult real word pair for each contrast type was selected to create analogous nonsense words. The most difficult pairs were [kado vs. kaado] for the vowel length contrast and [niʃi vs. niʃʃi] for the consonant length contrast.

The pitch pattern of [kado] and [kaado] was high-low for both words, and the pitch pattern of [niʃi] and [niʃʃi] was low-low for both words. Thus, within each contrast type, there was no pitch information to serve as a secondary cue to discriminate the two words. Only durational differences differentiated the members of each minimal pair. Based on these words, the nonsense words, [tado vs. taado] for the vowel contrast and [miʃi vs. miʃʃi] for the consonant contrast were created by substituting the initial consonant. To avoid speakers' awareness of the target stimuli, 86 additional real and nonsense filler words (minimal pairs and non-minimal pairs) were included in the word list.

All words were produced in carrier sentences, "*Kore-wa (Target) desune*", spoken eight times at the speaker's self-selected normal speaking rate (8 sentences x 8 tokens = 64 sentences). Target and filler items were randomized within a block. Instructions were given as follows: "Please speak as if you are talking to a Japanese speaker" to avoid using over-careful pronunciation. Any mispronounced or disfluent utterances were either self-corrected by the speaker or the experimenter asked for a repetition at the end of each block (see Appendix A for protocol for speakers). The speaker who had the most consistent production across all stimuli was chosen as a speaker.

A new set of two native speakers of JP listened to all of the target stimuli randomly and transcribed each word to ensure that none of the stimuli was ambiguous. Four of the eight tokens of each nonsense word were selected based on their acoustic data using Multi-Speech (Kay Elemetrics Corp, Model 3700 Version 2.2). The first and eighth stimuli were discarded because they generally were not the best tokens. The participants were too casual for the first and too practiced for the last. Basic criteria were: choose the

stimuli which have the same pitch pattern and then choose the shortest four stimuli. After the selected tokens were analyzed to ensure that there were no recording artifacts and that the signal was clean, the target words were extracted from the sentences, using SoundForge 4.5 software. The quality and intelligibility of the extracted words were evaluated by a new set of two native speakers of JP on the computer-sound system used in the ERP study. The evaluators transcribed the words first and then rated the “goodness” on an 11-point scale (-5 as bad to +5 as good). Only the words that scored better than +3 on this scale were used for the experiment.

Acoustic analysis of the extracted target nonsense words for final four-stimulus set was conducted (see Table 2). Durational measurements were obtained to observe the difference between long and short words of each contrast type. For the vowel contrast, the VOT, the target vowel /a/, the closure for the consonant /d/, and the second vowel /o/ were measured. For the consonant contrast, the duration of the nasal /m/, the initial vowel /i/, the target consonant /ʃ/, and the second vowel /i/ were measured. The ratio on average of the long to short target vowel /a/ in /tado/ versus /aa/ in /taado/ was 1.61, with variability in the range from 83 ms to 88 ms for /a/ and from 128 ms to 148 ms for /aa/. The ratio on average of the long to short target consonant /ʃ/ in /miʃi/ versus /ʃʃ/ in /miʃʃi/ was 1.54, with variability in the range from 84 ms to 95 ms for /ʃ/ and from 128 ms to 152 ms for /ʃʃ/. The ratio of long to short in word duration was 1.21 for the vowel contrast and 1.11 for the consonant contrast. The mean pitch and pitch range for the segments are presented in Table 2. In general, these differed minimally and/or overlapped for the short and long stimuli. To ensure that the nonsense words had similar acoustic qualities to the target real words, acoustic measurements were made for the vowel length

contrast [kado vs. kaado] and for the consonant length contrast [niʃi vs. niʃʃi] from the same speaker. Similar duration, duration ratios and pitch patterns were found for the real words, as shown in Appendix B, and were comparable to the real words in Hisagi and Strange (under revision) (see Appendix C).

Familiarization task materials: The familiarization tasks were designed to focus the participants' attention on vowel properties for the vowel task and on consonant properties for the consonant task. To encourage participants to adjust to nonsense words, only nonsense words were presented for this task. The same procedures used for creating the experimental stimuli were used here. The familiarization task materials for AE participants were, for the vowel contrast, nonsense words, /kæpi/ and /kepi/ and for the consonant contrast /bozə˘/ and /bosə˘/. A native female speaker of AE (New York dialect) recorded the words in carrier sentences, "*The word is (Target) this time.*"

The familiarization task materials for the JP participants were nonsense words, /kuto/ and /keto/ for the vowel contrast and /pepu/ and /pekku/ for the consonant contrast. The stimuli for the JP familiarization task were the filler nonsense words included in the original protocol, but the speaker was a different speaker than used in the main experiment. The familiarization words were produced in carrier sentences, "*Korewa (Target) desune ('This is target'),*"

Visual stimuli: Visual stimuli were created in a Word document using drawing tools. The goal was to equate the Auditory-Attend (Experiment 1) and Visual-Attend (Experiment 2) tasks for difficulty, creating an identical environment for the two tasks. Four different sizes of pentagon and hexagon shapes (color: orange with black background) and four different sizes of circle and oval shapes (color: green with black background) were

created. Three researchers judged the level of difficulty of these contrasts. The pentagon versus hexagon contrast was easier than the circle versus oval contrast. Based on the prediction that the vowel contrast would be easier than the consonant contrast, the pentagon versus hexagon contrast was presented with vowel sounds, and the circle versus oval contrast was presented with consonant sounds. The visual materials for the Auditory-Attend Experiment group were used simply to have the same environment as for the Visual-Attend Experiment group, and to fix participants' eyes to the screen.

2.1.2 Participants

Twelve native speakers of JP (two males and ten females), who had lived in the United States less than 36 months, served as the native-language group. Twelve native speakers of AE, who had never studied JP and had no strong second language background (i.e., not fluent in any L2 and never used an L2 regularly after formal study of the language), participated as non-native listeners (four males and eight females). An additional three participants were tested, but excluded from the final sample because of noisy data (see below). For the AE participants, foreign language experience with Spanish or French in junior-high, high school and/or college were accepted because these languages do not use temporal cues to distinguish vowels. The JP participants were native speakers of the Tokyo (Kanto) dialect, who had had minimal (i.e., only the standard) English education in Japan. All participants were between 21 and 40 years of age with normal hearing (at 500, 1000, 2000, 4000 Hz at 20 dB HL). The participants were recruited by advertisement on the CUNY college campus, at the CUNY-Graduate Center, at a Japanese bookstore, or on the internet through *Craigslist* and *marugoto* web pages in

New York City (web pages for AE and JP readers, respectively). All participants were paid to participate in the experiment.

2.1.3 Paradigms and Procedures

All testing took place in a 9' x 10' soundproof, electrically-shielded booth. Using E-prime to control the experiment, stimuli were presented over speakers (Realistic Minimus -7) at 75 dB SPL. Participants were seated in comfortable chairs. Each participant was asked to fill out a consent form (see Appendix D) and a language background questionnaire before the experiment (see Appendix E). Each participant was tested for approximately 3.5 hours with frequent breaks (see Table 3 for greater detail regarding the procedure).

ERP task: A categorial oddball paradigm was used with standards presented on 85% and deviants on 15% of the trials (see Figure 2). Words occurred with inter-stimulus intervals (ISI) of 800 ms resulting in stimulus onset asynchronies (SOAs) of 1035 ms to 1150 ms. The total number of trials was 1400 (1190 standard + 210 deviant). Trains of stimuli were presented in fourteen blocks for each condition (14 x 4 = 56 blocks in total). Each block had a different number of standards and deviants (average: 100 trials per block) (see Appendix F). Participants received a randomized order of blocks for each condition. The four conditions (vowel and consonant by order: short as standard and short as deviant) were counter-balanced. Each block lasted approximately 1-2 minutes. Each condition took approximately 30 minutes.

At the beginning of the experiment, detailed oral instructions with sample words were given to make participants aware of what they needed to attend to (i.e., deviant stimuli). The participants were instructed to listen to randomly ordered JP words, count

the deviant sounds silently, and report to an investigator at the end of each block (see Appendix G for the exact instructions). During the ERP recording, the participants were asked to look at the computer screen used for the Visual-Attend Experiment group (Exp 2) (see Figure 3).

Behavioral Task: After the end of the ERP session, a ten-minute behavioral discrimination task using the same auditory stimuli was given to judge participants' behavioral performance. Participants were asked to press a button when they heard a deviant stimulus. In this task one block of 100 trials (85 standard and 15 deviant) was presented for each order and contrast. The order of contrast type and short versus long stimuli as deviants was counter-balanced. The stimulus orders, ISI, and SOA were identical to those of the electrophysiological Auditory-Attend Experiment.

2.1.4 Electroencephalogram (EEG) Recording and Data Analysis

Electrode Placements: A Geodesic net of 65 electrodes wrapped in sponges was placed on the participant's scalp, after soaking in saline solution for five minutes. The Vertex served as the reference during data collection. Vertical and horizontal eye movements were monitored from frontal electrodes Fp1 (left) and Fp2 (right) and electrodes placed below each eye. Impedance was maintained below 40 k Ω , which is acceptable for the high-impedance Geodesic amplifiers (200 M Ω ; Picton, Alain, Otten, Ritter, & Achim 2000; Ferree, Luu, Russell & Tucker, 2001).

Data Recording: The EEG was amplified with a bandpass of 0.1 to 30 Hz, using Geodesic Amplifiers. A Geodesic software system (NetStation version 3.0) in continuous mode was used to acquire the data at a sampling rate of 250 Hz per channel for later off-line processing. During data acquisition, all channels were observed by an experimenter

to monitor each participant's state, artifacts due to electrical interference, defective electrode contacts, and/or excessive muscle movement.

ERP Data Post-Processing (see Appendix H): The continuous EEG was processed off-line, using a lowpass filter of 20 Hz. The EEG was segmented into epochs with an analysis time of 1000 ms post-stimulus and a 100 ms pre-stimulus baseline. The data were transferred into an EEGLAB program (Matlab-based) to perform an Independent Component Analysis (ICA) (Bell & Sejnowski, 1995; Glass, Frishkoff, Frank, Davery, Dien & Maloney, 2004) for eye blink correction. The algorithm decomposes the signal into components and removes those components highly correlated with an eye-blink component ($r > .9$), and then recomposes the signal minus these components. The corrected data were then baseline corrected and examined for artifacts. An epoch for a single channel was marked bad if electrophysiological activity exceeded $\pm 70 \mu\text{V}$. A channel was marked as bad if greater than 20 % of the total epochs were marked as bad. An epoch was rejected if more than 10 channels for that epoch were marked as bad. The data had no more than 10% of trials lost on average due to eye blink activity. Only one participant was discarded at this point due to high artifact rate (greater than 50% of trials rejected). Channels with high artifact rates on greater than 20% of trials were replaced by interpolation using adjacent sites. ERP averages were calculated for each stimulus type (standard, deviant) and baseline corrected using the 100-ms pre-stimulus activity. Two participants were discarded because their data were noisy as determined by the absence of clear N1 and P2 obligatory components. ERP averages were re-referenced to an average-reference. Finally, the subtraction waveforms were created by subtracting the standard from the deviant to see the presence or absence of MMN.

Analysis- presence of MMN, N2b and P3b: The time intervals used in the ANOVAs were determined by calculating global field power (GFP) from the grand-average data (Lehmann & Skrandies, 1980). GFP is the standard deviation calculated from the mean of all 65 channels as a function of time. A peak of GFP may reflect a maximum of underlying dipolar brain activity that contributes to the surface potential field and is useful for identifying time intervals for analyzing a component of interest (Shafer, Ponton, Datta, Morr., & Schwartz, in press). Figure 4 displays the GFP for all four conditions for the AE and JP groups. The onset and offset of the analysis time interval period corresponded to the beginning and the ending of the GFP peaks, respectively.

For each stimulus condition (tado, taado, miʃi, miʃʃi), five-way analyses of variances (ANOVAs) with language group (AE vs. JP), stimulus (standard vs. deviant), site (frontal: F3, Fz, F4 vs. central: C3, Cz, C4), hemisphere (left vs. midline vs. right), and time in 40 ms time windows (vowels: 124-160; 164-200; 204-240; 244-280; 284-320; 324-360; consonants: 204-240; 244-280; 284-320; 324-360; 364-400; 404-440; 444-480) were carried out to establish whether there were significant differences in stimulus conditions that were consistent with MMN timing and topography. Frontal and central sites were chosen because the MMN component is typically largest there; this decision was confirmed by topographical voltage maps of the subtraction waveforms (Figure 5). Hemisphere was included as a factor in the analyses because previous research suggests that the greatest language-related differences might be seen at left hemisphere sites (e.g., Alho, Winkler, Escera, Huotilainen, Virtanen, Jaaskelainen, Pekkonen, & Ilmoniemi, 1998; Näätänen et al., 1997; Shestakova, Brattico, Huotilainen, Galunov, Soloviev, Sams, Ilmoniemi, & Näätänen, 2002; Shtyrov, Kujala, Ahveninen, Tervaniemi, Alku,

Ilmoniemi, & Näätänen, 1998, Shtyrov, Kujala, Palva, Ilmoniemi, & Näätänen, 2000; Shafer, et al., 2004; Binder, Westbury, McKiernan, Possing, & Medler, 2005). Time was included to determine whether the MMN began earlier in one language group than in another.

If the five-way ANOVA revealed significant main effects or interactions including Stimulus and Language Group, then within-subject ANOVAs were used to determine whether there were any significant differences between deviant and standard stimuli for each language group. To follow-up language group differences that interacted with stimulus, additional ANOVAs using subtraction waveforms (deviant – standard) were undertaken and included site, hemisphere and time where appropriate.

An N2b discriminative component might be observed in the Auditory-Attend task following or partially overlapping the MMN. The N2b was expected to be seen as greater negativity of the deviant than standard; it is largest at central sites, does not invert in polarity at the mastoids, and thus can be separated from the MMN by analyses examining the mastoids relative to frontal and central sites. Four-way ANOVAs of language group (AE vs. JP), site {central: C3, Cz, C4; mastoids: LM, LRM [(LM+RM)/2], RM}, hemisphere (left vs. midline vs. right), and time in 40 ms time windows (vowels: 124-160; 164-200; 204-240; 244-280; 284-320; 324-360; consonants: 204-240; 244-280; 284-320; 324-360; 364-400; 404-440; 444-480) were conducted on the subtraction waveforms so that it could be determined whether MMNs were overlapped by N2b. In particular, site was included in the analysis so that inversion (or lack of inversion) in conjunction with central negativity could be used to identify MMN versus N2b.

The presence of P3b, which indicates attentional allocation (Polich, 2003), was also measured from the subtraction waveforms for the Auditory-Attend Experiment. To determine whether P3b was present, four-way ANOVAs of language group (AE vs. JP), contrast types (short vs. long), hemisphere (P3 vs. Pz vs. P4), and time in 40 ms time windows between 400-800 ms (vowels: 404-440; 444-480; 484-520; 524-560; 564-600; consonant: 484-520; 524-560; 564-600; 604-640; 644-680; 684-720; 724-760) were conducted.

A later negativity was also observed and the same series of analyses used for the earlier negativity were carried out for this late component, except that the time intervals extend from 444 to 720 ms for both vowels and consonants. The outcome of these analyses were interpreted more strictly ($p < 0.01$) because we had no specific predictions for this late negativity.

A second question was whether the contrast types (vowel vs. consonant) were different in difficulty. For this reason, one set of analyses included contrast type as a factor. Five-way ANOVAs of contrast type (vowel vs. consonant), language group (AE vs. JP), site (frontal: F3, Fz, F4 vs. central: C3, Cz, C4), hemisphere (left vs. midline vs. right), and time in 40 ms time windows (vowels: 204-240; 244-280; consonants: 324-360; 364-400) were conducted using the subtraction waveforms to examine differences between vowels and consonants.

We also noticed differences in the latency of the discriminative responses related to which stimulus served as the standard for a contrast pair (e.g., short as standard versus long as standard). Analyses were also carried out comparing these order differences. The

results from these analyses were interpreted more strictly ($p < 0.01$) because we had no specific predictions for order differences.

Stepdown analyses followed up on significant interactions, and Tukey's Honestly Significantly Different (HSD) post-hoc tests were used for post-hoc pairwise comparisons. The Greenhouse-Geisser correction was used for adjustment when necessary (i.e., more than two levels of site were compared). The specific details of the analyses are given at the beginning of each section, below, for greater clarity.

Behavioral Data: For the counting and button-press behavioral tasks, the number of correct and incorrect responses was calculated for the analyses. The distribution of scores in descriptive statistics was inspected first to see whether the data should be analyzed using parametric or nonparametric statistics. Median and Wilcoxin Matched Pairs tests were used to compare groups in the cases where nonparametric tests were appropriate (Siegel & Castellan, 1988). For the button-press task, reaction time (RT), measured from the onset of the deviant stimuli, was also examined, using repeated measures ANOVAs. RT (in milliseconds) was analyzed using the correct responses only.

2.2. Results

2.2.1 Behavioral Results

Behavior on the counting task (during ERP recording): Conditions are named according to which stimulus served as the standard (tado, taado, miʃi, miʃʃi). For the behavioral responses, 0-2 errors per block counted as “no error.” The mean scores in Figure 6 describe the number of “no error blocks” per condition out of 14 possible. Because the data were not normally distributed, Median tests were conducted to determine an overall group difference and differences in each condition. The JP group performed better than

the AE group overall; however, Median tests showed no significant language difference for any condition ($p < 0.05$). Wilcoxin Matched Pairs non-parametric tests were used to assess whether the vowel duration contrast was easier than the consonant duration contrast. The [tado]-as-standard condition was easier than [miʃi] ($p = 0.000$) and [miʃʃi] ($p = 0.003$). The [taado]-as-standard condition was also easier than [miʃi] ($p = 0.000$) and [miʃʃi] ($p = 0.000$). The same test was conducted with language group (AE vs. JP) separately. In the AE group, the [tado]-as-standard condition was easier than [miʃi] ($p = 0.037$), but not [miʃʃi] ($p = 0.10$). The [taado]-as-standard condition was also easier than [miʃi] ($p = 0.0058$) and [miʃʃi] ($p = 0.020$). In the JP group, the [tado]-as-standard condition was easier than [miʃi] ($p = 0.003$) and [miʃʃi] ($p = 0.015$). The [taado]-as-standard condition was also easier than [miʃi] ($p = 0.0033$) and [miʃʃi] ($p = 0.0058$). Thus, for both groups, overall performance was better for the vowel contrast than for the consonant contrast. While the JP group performed more accurately overall, the difference between language groups for each standard condition were small.

Behavior on the button-press task (after ERP recording): Table 4 shows the descriptive statistics for performance on the task and Figure 7 displays the mean correct detections of the deviant stimuli (out of 15 possible). Median tests revealed that there were significant differences between the AE and JP groups for all conditions (tado: $p < 0.0001$; taado: $p < 0.05$; miʃi: $p < 0.001$; miʃʃi: $p < 0.001$). In all comparisons, the JP group was more accurate than the AE group. Wilcoxin Matched-Pairs tests were used to assess whether the vowel contrast was easier than the consonant contrast. The [tado]-as-standard

condition was easier than [miʃi] ($p = 0.001$) and [miʃʃi] ($p = 0.05$). The [taado]-as-standard condition was also easier than [miʃi] ($p = 0.000$) and [miʃʃi] ($p = 0.008$). The same test was conducted with language group (AE vs. JP) separately. In the AE group, the [tado]-as-standard condition was easier than [miʃi] ($p = 0.041$), but not [miʃʃi] ($p = 0.10$). The [taado]-as-standard condition was also easier than [miʃi] ($p = 0.0041$) and [miʃʃi] ($p = 0.017$). In the JP group, the [tado]-as-standard condition was easier than [miʃi] ($p = 0.007$), but not [miʃʃi] ($p = 0.155$). The [taado]-as-standard condition was also easier than [miʃi] ($p = 0.0033$), but not [miʃʃi] ($p = 0.1688$). Thus, there was a language group difference where the JP group showed better performance than the AE group on both contrasts in both short as standard and long as standard orders. For both language groups, overall performance was better for the vowel contrast than for the consonant contrast. Reaction Time (RT) was also examined using Median tests and ANOVA. There was no overall significant difference in RT between the two groups.

2.2.2 Electrophysiological results

2.2.2.1 Mismatch Negativity (MMN), N2b and P3b

The grand averages of the ERPs to the standard, deviant and subtraction (deviant – standard) waveforms are shown at Fz and LM for each language group in Figure 8 (see Appendix I for all nine sites with LM and RM: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4).

Examination of the subtraction waveform revealed three negative deflections between 150 and 700 ms for both language groups. The onset of the duration difference between the long and short vowel was approximately 98 ms and between the long and short consonant was 210 ms. MMN was expected no earlier than 90 ms later. The first two

negative deflections between 160 and 400 ms for the vowels and 200 and 480 ms for the consonants probably reflect N1 recovery and MMN activity and were examined using ANOVAs and are presented in section 2.2.2.1. Negativity from 240 to 440 ms at Cz may be N2b (because there was less inversion at the mastoids). Later negativity began approximately 400 ms after the onset of stimulus difference. The late negativity (LN) was examined using ANOVAs and are presented in section 2.2.2.2. Figure 9 shows subtraction waveforms (deviant - standard) at Pz. Positivity from 400 to 800 ms at Pz is probably the P3b. Statistical analyses are presented below for each contrast type and standard-deviant order (e.g., [tado]-as-standard condition) to determine whether these apparent differences between the deviant and the standard were significant. Only effects including Stimulus as a factor are reported for the five-way and four-way ANOVAs.

Vowel Contrast:

[tado]-as-standard condition: The five-way ANOVAs comparing the ERPs for the standard and the deviant revealed a significant main effect of Stimulus [$F(1, 22) = 23.53, p = 0.000$] and significant interactions of Stimulus x Site x Language Group [$F(1, 22) = 5.79, p = 0.025$], Stimulus x Hemisphere x Language Group [$F(2, 44) = 3.50, p = 0.04$], Stimulus x Time [$F(5, 110) = 21.38, p = 0.000$], Stimulus x Site x Hemisphere [$F(2, 44) = 6.59, p = 0.004$], Stimulus x Site x Time [$F(5, 110) = 3.89, p = 0.013$], Stimulus x Hemisphere x Time [$F(10, 220) = 4.03, p = 0.002$], and Stimulus x Site x Hemisphere x Time [$F(10, 220) = 2.59, p = 0.048$]. These significant findings indicate that the deviant was more negative than the standard in some cases and that its magnitude was affected by the factors of Language Group, Site, Hemisphere and Time.

Four-way ANOVAs were carried out to see whether there were differences in the MMN related to language group (AE vs. JP) separately that interacted with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was a significant main effect of Stimulus [$F(1, 11) = 19.20, p = 0.001$]; there were also significant interactions of Stimulus x Time [$F(5, 55) = 4.50, p = 0.001$], Stimulus x Site x Hemisphere [$F(2, 22) = 5.06, p = 0.02$], and Stimulus x Site x Time [$F(5, 55) = 3.11, p = 0.05$]. In the JP group, there was a significant main effect of Stimulus [$F(1, 11) = 7.12, p = 0.022$]; there were also significant interactions of Stimulus x Time [$F(5, 55) = 21.26, p = 0.000$] and Stimulus x Hemisphere x Time [$F(10, 110) = 3.32, p = 0.016$]. These significant findings confirm that the deviant was more negative than the standard in both language groups.

To simplify interpretation of the interactions, four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (124-160; 164-200; 204-240; 244-280; 284-320; 324-360 ms) were carried out using the subtraction waveforms (deviant – standard). Significant interactions of Site x Language Group [$F(1, 22) = 5.79, p = 0.025$] and Hemisphere x Language Group [$F(2, 44) = 3.50, p = 0.039$] revealed that the AE group showed the greatest negativity at central sites while the JP group showed the greatest negativity at frontal sites and the JP group generally showed greater negativity at the left than the midline or right sites, as shown in Figure 10. However, post-hoc pairwise comparisons were not significant. Significant two-way interactions of Site x Hemisphere [$F(2, 44) = 6.59, p = 0.003$] and Site x Time [$F(5, 110) = 3.89, p = 0.003$] were also found. A significant interaction of Site x Hemisphere x Time showed that the negativity was larger at the central than frontal

region for the left sites [$F(10, 220) = 2.59, p = 0.05$] (see Figure 11). An analysis of Site x Hemisphere for each time interval followed up on these results. Significant interactions of Site x Hemisphere were seen in the following five time intervals (ms): 164-200 (time 2) [$F(2, 44) = 7.52, p = 0.02$], 204-240 (time 3) [$F(2, 44) = 6.36, p = 0.004$], 244-280 (time 4) [$F(2, 44) = 5.65, p = 0.007$], 284-320 (time 5) [$F(2, 44) = 5.79, p = 0.006$], 324-360 (time 6) [$F(2, 44) = 5.06, p = 0.01$]. Post-hoc tests ($p < 0.05$) showed that the following four time intervals: 164-200 (time 2), 204-240 (time 3), 244-280 (time 4), and 284-320 (time 5) ms showed a significant difference between the left-frontal and central sites. The 284-320 ms (time 5) interval showed a significant difference between the right frontal and central sites and the 324-360 ms (time 6) interval for the midline frontal and central sites. The frontal sites showed a larger negativity at the midline (i.e., Fz) *before* the peak of MMN while the central sites showed a larger negativity, particularly at the left site (C3), *after* the peak of MMN.

Four-way ANOVAs with language group (AE vs. JP), hemisphere (left vs. midline vs. right), site (central sites vs. mastoids), and time (124-160; 164-200; 204-240; 244-280; 284-320; 324-360 ms) were carried out to determine whether a portion of the negativity could be considered the N2b. A Site x Time interaction [$F(5, 110) = 21.70, p = 0.000$] was found (see Figure 12). Post-hoc tests ($p < 0.05$) showed significant differences between the central and mastoid sites for the intervals 204-240 (time 3), 244-280 (time 4) and 284-320 (time 5) ms. Examination of Figure 12 suggests that there may be N2b activity in times 4 and 5 because it does not differ for central and mastoid sites.

In sum, both language groups had a robust MMN for the vowel contrast with the short vowel as the standard and long vowel as the deviant. The AE group showed a larger

MMN at central sites while the JP group showed it at frontal sites. The JP group also showed a larger MMN at the left than the midline or right sites. There was a weak indication of N2b, but the large central negativity may have indicated overlap of N2b with the MMN.

[taado]-as-standard condition: Figure 4 displays GFP for the subtraction waveform when [taado] was the standard. The MMN appears to extend from 220 to 350 ms. The five-way ANOVAs comparing the ERPs for the standard and the deviant revealed a significant main effect of Stimulus [$F(1, 22) = 17.50, p = 0.000$] and significant interactions of Stimulus x Time [$F(5, 110) = 31.76, p = 0.000$], Stimulus x Time x Language Group [$F(5, 110) = 2.50, p = 0.07$], Stimulus x Site x Time [$F(5, 110) = 3.86, p = 0.019$], Stimulus x Hemisphere x Time [$F(10, 220) = 2.48, p = 0.05$], and Stimulus x Site x Hemisphere x Time [$F(10, 220) = 6.01, p = 0.000$]. These significant findings indicate that the deviant was more negative than the standard and that its magnitude was affected by the factors of Language Group, Site, Hemisphere and Time.

Four-way ANOVAs were carried out to see whether there were differences in the MMN related to language group (AE vs. JP) separately that interact with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was a significant main effect of Stimulus [$F(1, 11) = 8.78, p = 0.013$]; there were also significant interactions of Stimulus x Time [$F(5, 55) = 17.24, p = 0.000$] and Stimulus x Site x Hemisphere x Time [$F(10, 110) = 3.08, p = 0.05$]. In the JP group, there was a significant main effect of Stimulus [$F(1, 11) = 9.05, p = 0.012$]; there were also significant interactions of Stimulus x Time [$F(5, 55) = 16.96, p = 0.000$], Stimulus x Site x Time [$F(5, 55) = 4.40, p = 0.011$], and Stimulus x Site x Hemisphere x Time [$F(10,$

110) = 3.34, $p = 0.021$]. These significant findings confirm that the deviant was more negative than the standard in both language groups.

To simplify interpretation of the interactions, four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (124-160; 164-200; 204-240; 244-280; 284-320; 324-360 ms) were carried out using the subtraction waveforms (deviant – standard). A significant interaction of Time x Language Group [$F(5, 110) = 2.51, p = 0.03$] was observed (see Figure 13): the AE group showed a slightly larger MMN than the JP group in the intervals 284-320 (time 5) ms and 324-360 (time 6) ms although post-hoc comparisons of the language groups for each time-interval were not significant. Significant interactions of Site x Time [$F(5, 110) = 3.86, p = 0.003$] and Hemisphere x Time [$F(10, 220) = 2.48, p = 0.0078$] (see Figure 14) were seen. Post-hoc tests ($p < 0.05$) showed significantly larger MMN for the frontal compared to the central sites from 204-240 (time 3) ms. The midline of the hemisphere from 244-280 (time 4) ms showed a larger MMN than the left and right hemisphere sites, which was supported by post-hoc tests ($p = 0.03$).

Four-way ANOVAs with language group (AE vs. JP), hemisphere (left vs. midline vs. right), site (central sites vs. mastoids), and time (124-160; 164-200; 204-240; 244-280; 284-320; 324-360 ms) were carried out to determine whether a portion of the negativity could be considered the N2b. There were main effects of Site [$F(1, 22) = 22.46, p = 0.0000$] and a Site x Time interaction [$F(5, 110) = 27.96, p = 0.0000$] (see Figure 15). Post-hoc tests ($p < 0.05$) showed that the mastoids were significantly more positive than the central sites in intervals 244-280 (time 4), 284-320 (time 5) and 324-360 (time 6) ms, consistent with MMN. However, there was no evidence of N2b in these time intervals.

That is to say, there was no central negativity without mastoid inversion. There was no significant interaction with Language Group.

In sum, both language groups had the MMN for the vowel contrast with the long vowel as standard and the short vowel as deviant. The AE group showed a slightly larger MMN than the JP group in the later time intervals. The mid-time intervals showed the largest MMN amplitude at the midline-frontal site. There was no clear indication of N2b.

[tado vs taado]: Five-way ANOVAs with language group (AE vs. JP), standard order (tado vs. taado), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (124-160; 164-200; 204-240; 244-280; 284-320; 324-360 ms) were carried out using the subtraction waveforms (deviant – standard). There was no main effect of standard order, but there were some interactions. A significant Order x Time interaction was found [$F(5, 110) = 36.95, p = 0.0000$] (see Figure 16). The MMN to the [tado]-as-standard condition showed an earlier latency than the MMN to the [taado]-as-standard condition. Post-hoc tests ($p < 0.01$) supported that the [tado]-as-standard condition was larger from 164-200 and 204-240 ms, while the [taado]-as-standard condition was larger from 244-280, 284-320, and 324-360 ms. A significant interaction of Order x Hemisphere x Time [$F(10, 220) = 3.63, p = 0.007$] was also found (see Figure 17). Examination of Figure 17 reveals that this difference between contrasts is primarily one of time. Specifically, the only clear hemisphere difference is in the final time interval where the negativity ends earlier for [tado]-as standard condition than [taado]-as standard condition.

For the posterior sites, four-way ANOVAs of language group (AE vs. JP), standard order (short vs. long), hemisphere (left as P3 vs. midline as Pz vs. right as P4), and time in 40 ms time windows between 400 and 800 ms (404-440; 444-480; 484-520;

524-560; 564-600 ms) were carried out using the subtraction waveforms (deviant – standard). Significant interactions of Order x Time [$F(4, 88) = 10.514, p = 0.0000$] and Order x Hemisphere x Time [$F(8, 176) = 3.52, p = 0.006$] were seen. Figure 18 reveals that the Pz site showed the largest amplitude P3b for both contrast types, but peaked earlier for [tado] than [taado]-as standard conditions. Post-hoc tests ($p < 0.01$) showed that the [tado]-as-standard condition had the larger P3b for the 444-480 (time 2) and 484-520 (time 3) ms intervals for all three sites compared to the [taado]-as-standard condition.

In sum, the MMN to the [tado]-as-standard condition showed an earlier latency than the MMN to the [taado]-as-standard condition. The Pz site showed the largest amplitude P3b for both standard orders and the [tado]-as-standard condition had a larger P3b than the [taado]-as-standard condition.

Consonant Contrast:

[mi\i]-as-standard condition: The five-way ANOVAs comparing the ERPs for the standard and the deviant revealed a significant main effect of Stimulus [$F(1, 22) = 15.29, p = 0.000$] and significant interactions of Stimulus x Time [$F(6, 132) = 4.16, p = 0.005$], Stimulus x Site x Time [$F(6, 132) = 3.28, p = 0.04$], and a marginal interaction of Stimulus x Site x Hemisphere x Time [$F(12, 264) = 2.12, p = 0.08$]. These significant findings indicate that the deviant was more negative than the standard and that its magnitude was affected by the factors of Site, Time and marginally by Hemisphere.

Four-way ANOVAs were carried out to see whether there were differences in the MMN related to language group (AE vs. JP) separately that interact with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was a significant main effect of Stimulus [$F(1, 11) = 4.83, p = 0.05$]; there were also significant

interactions of Stimulus x Time [$F(6, 66) = 2.96, p = 0.013$] and Stimulus x Site x Time [$F(6, 66) = 3.46, p = 0.05$]. In the JP group, there was a significant main effect of Stimulus [$F(1, 11) = 15.64, p = 0.002$], but there were no significant interactions. These significant findings confirm that the deviant was more negative than the standard for both language groups.

To simplify interpretation of the interactions, three-way ANOVAs with site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (204-240; 244-280; 284-320; 324-360; 364-400; 404-440; 444-480 ms) were carried out using the subtraction waveforms (deviant – standard). Language Group was not included because there were no significant interactions in the five-way ANOVAs above. A significant interaction was found of Site x Time [$F(6, 138) = 3.19, p = 0.006$] (see Figure 19). The central sites showed a larger MMN than the frontal sites, and the largest peak amplitude was shown in time 324-360 (time 4) ms. Post-hoc tests ($p < 0.05$) showed significant differences for both 364-400 (time 5) ms and 404-440 (time 6) ms between the frontal and central sites. A significant interaction was also found of Hemisphere x Time [$F(12, 276) = 1.82, p = 0.04$] although the post-hoc tests ($p < 0.05$) showed no difference across the left-midline and right sites.

Four-way ANOVAs with language group (AE vs. JP), hemisphere (left vs. midline vs. right), site (central sites vs. mastoids), and time (204-240; 244-280; 284-320; 324-360; 364-400; 404-440; 444-480 ms) were carried out to determine whether a portion of the negativity could be considered the N2b. There was a main effect of Site [$F(1, 22) = 11.12, p = 0.003$], and a significant interaction of a Site x Time [$F(6, 132) = 2.90, p = 0.01$]. Post-hoc tests ($p < 0.05$) showed significant difference between the central and

mastoid sites for 204-240 (time 1); 284-320 (time 3); 324-360 (time 4); 404-440 (time 6); and 444-480 (time 7) ms. Figure 20 shows these negative peaks with inversions at the mastoids, but no clear evidence of N2b. That is to say, there was no central negativity without corresponding absence of inversion at the mastoids.

In sum, both language groups had the MMN for the consonant contrast with the short consonant as the standard, with no overall significant difference between groups. The central sites showed larger MMNs than the frontal sites and there was no clear indication of N2b.

[miʃi]-as-standard condition: The five-way ANOVAs comparing the ERPs for the standard and the deviant revealed a significant main effect of Stimulus [$F(1, 22) = 15.48, p = 0.000$] and significant interactions of Stimulus x Time [$F(6, 132) = 9.85, p = 0.000$], Stimulus x Site x Hemisphere [$F(2, 44) = 4.11, p = 0.03$], Stimulus x Hemisphere x Time [$F(12, 264) = 6.96, p = 0.000$], and Stimulus x Site x Hemisphere x Time [$F(12, 264) = 3.33, p = 0.009$]. A Stimulus x Hemisphere x Time x Language Group interaction approached significance [$F(12, 264) = 1.88, p = 0.09$]. These significant findings indicate that the deviant was more negative than the standard and that its magnitude was affected by the factors of Language Group, Site, Hemisphere and Time.

Four-way ANOVAs were carried out to see whether there were differences in the MMN related to language group (AE vs. JP) separately that interact with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was a significant main effect of Stimulus [$F(1, 11) = 8.45, p = 0.014$]; there were also significant interactions of Stimulus x Time [$F(6, 66) = 4.67, p = 0.001$] and a marginal interaction of Stimulus x Site x Hemisphere x Time [$F(12, 132) = 2.50, p = 0.07$]. In the

JP group, there was a significant main effect of Stimulus [$F(1, 11) = 7.58, p = 0.019$]; there were also significant interactions of Stimulus x Time [$F(6, 66) = 8.14, p = 0.000$] and Stimulus x Hemisphere x Time [$F(12, 132) = 7.18, p = 0.000$]. These significant findings confirm that the deviant was more negative than the standard for both language groups.

To simplify interpretation of the interactions, four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (204-240; 244-280; 284-320; 324-360; 364-400; 404-440; 444-480 ms) were carried out using the subtraction waveforms (deviant – standard). A significant interaction was seen of Site x Hemisphere [$F(2, 44) = 4.11, p = 0.023$]. The negativity was somewhat larger at the left for central sites and at the midline for frontal sites, although post-hoc tests were not significant. A Hemisphere x Time interaction was seen as well [$F(12, 264) = 6.96, p = 0.000$]. The negativity was largest in amplitude in the 364-400 ms interval times across left, right and midline sites. In addition, it appears that there was a midline negativity from 244-280 ms time. A marginal significant interaction of Hemisphere x Time x Language Group was also seen [$F(12, 264) = 1.88, p = 0.09$]. Figure 21 shows greater negativity from 404 to 480 ms (times 6 and 7) at left and midline sites for the AE compared to JP group. Post-hoc tests ($p < 0.05$) showed that there was a difference between the AE and JP groups in the 404-440 ms interval (time 6) at the midline sites, where the AE group showed a larger negativity than the JP group. A significant interaction of Site x Hemisphere x Time was also seen [$F(12, 264) = 3.33, p = 0.009$] (see Figure 22). Each time bin was analyzed separately, and there was an interaction of Site x Hemisphere in the following time bins ($p < 0.05$): 244-280 (time 2), 364-400 (time 5),

404-440 (time 6), and 444-480 (time 7) ms. In general, central sites showed a larger negativity than frontal sites, particularly in the left hemisphere.

Four-way ANOVAs with language group (AE vs. JP), hemisphere (left vs. midline vs. right), site (central sites vs. mastoids), and time (204-240; 244-280; 284-320; 324-360; 364-400; 404-440; 444-480 ms) were carried out to determine whether a portion of the negativity could be considered the N2b. There was a main effect of Site [$F(1, 22) = 13.14, p = 0.002$] and a Site x Time interaction [$F(6, 132) = 11.61, p = 0.000$] (Figure 23). Post-hoc tests ($p < 0.05$) showed significant difference between the central and mastoid sites for 244-280 (time 2), 364-400 (time 5), and 404-440 (time 6) ms. There was no clear evidence of N2b.

In sum, both language groups produced significant MMN components for the consonant contrast with the long consonant as standard. The AE group yielded a larger MMN in left and midline sites than the JP group in later time intervals. The central sites showed a larger MMN than frontal sites particularly in the left hemisphere. The AE group had a larger MMN than the JP group in the later time intervals. There was no clear indication of N2b.

[miʃi vs. miʃfi]: Five-way ANOVAs with language group (AE vs. JP), contrast type (miʃi vs. miʃfi), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (204-240; 244-280; 284-320; 324-360; 364-400; 404-440; 444-480 ms) were carried out using the subtraction waveforms (deviant – standard). There was no main effect of standard order, but there was an interaction of Order x Time [$F(6, 132) = 8.76, p = 0.000$] as shown in Figure 24. The [miʃi]-as-standard condition showed earlier negativity than the [miʃfi]-as-standard condition, but there was also an earlier negativity in interval 244-280

ms (time 2) for the [miʃfi] contrast. Post-hoc tests ($p < 0.01$) showed that there was a significant difference between the two standard orders in the intervals 364-400 (time 5) and 444-480 (time 7) ms. A significant interaction of Order x Site x Time [$F(6, 132) = 3.76, p = 0.009$] was also found. As reported above, the negativity for the [miʃi]-as-standard condition was larger at the central than frontal sites, while there was no difference in site for the [miʃfi]-as-standard condition which can be seen in Figure 25. A significant interaction of Order x Hemisphere x Time [$F(12, 264) = 5.14, p = 0.000$] was also observed and is shown in Figure 26. This difference is largely an artifact of the time difference in the negativities.

Four-way ANOVAs were undertaken to examine the P3b with language group (AE vs. JP), contrast types (short vs. long), hemisphere (left as P3 vs. midline as Pz vs. right as P4), and time in 40 ms time windows between 400-800 ms (consonant: 484-520; 524-560; 564-600; 604-640; 644-680; 684-720; 724-760 ms) using the subtraction waveforms (deviant – standard). A main effect of standard order indicated that the positivity to the [miʃi]-as-standard condition was larger than the positivity to the [miʃfi]-as-standard condition [$F(1, 22) = 7.36, p = 0.013$]. There was also a significant interaction of Order x Time [$F(6, 132) = 3.92, p = 0.001$]. Post-hoc tests ($p < 0.01$) showed that the [miʃi]-as-standard condition had a greater positivity than did the [miʃfi]-as-standard condition in all time bins except 724-760 ms (time 7).

In sum, the [miʃi]-as-standard condition produced earlier negativities than the [miʃfi]-as-standard condition. There was an earlier negativity for the [miʃfi]-as-standard condition, but too early to be MMN to the consonant duration contrast. The [miʃi]-as-

standard condition produced a larger MMN at the central than frontal sites, while there was no difference in site for the [miʃʃi]-as-standard condition. The [miʃi]-as-standard condition showed a larger P3b than the [miʃʃi]-as-standard condition.

Vowel vs. Consonant Contrasts:

The discriminative responses in the ERPs to the vowel-length and consonant-length contrasts were found at different latencies because the acoustic differences began at different latencies from the onset of the words. Thus, to remove this difference and focus on amplitude and topography differences between the vowel and consonant contrasts, two time bins which framed the peak of the greatest negativity for each contrast type and standard order (vowels: [tado]: 204-240; [taado]: 244-280; consonants: [miʃi]: 324-360; [miʃʃi]: 364-400) were compared. Each time bin was an average of the short and long contrast (e.g., time bin 204-240 ms: [tado] + [taado] /2; time bin 244-280 ms: [tado] + [taado] /2). Five-way ANOVAs with language group (AE vs. JP), contrast type (vowel vs. consonant), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (204-240; 244-280 for vowels) and (324-360; 364-400 for consonants) were carried out using the subtraction waveforms (deviant – standard). A significant main effect of the vowel versus consonant contrast was found [$F(1, 22) = 11.82, p = 0.002$]. The vowel showed a larger negativity than the consonant. There was a significant interaction of Contrast x Site [$F(1, 22) = 6.86, p = 0.016$] as shown in Figure 27. A significant interaction of Contrast x Site x Time [$F(1, 22) = 6.24, p = 0.02$] (see Figure 28) indicated that the frontal sites for both time bins and the central sites in time 244-280 ms showed the stronger effect for vowels; these results were supported by post-hoc tests ($p < 0.05$). A significant interaction of Contrast x Hemisphere x Time was observed [$F(2, 44) = 4.53,$

$p = 0.02$] as shown in Figure 29. Post-hoc tests ($p < 0.05$) indicated that the negativity to the vowel contrasts was larger than that to the consonant contrasts at all hemispheres and in both time bins. In the vowel contrast, the right hemisphere showed equivalent negativity for the 204-240 ms and the 244-280 ms interval while both the left and midline of regions showed greater negativity for the later compared to earlier time bins.

In sum, for both language groups, the vowel contrast elicited much larger MMNs than the consonant contrast.

2.2.2.2 Late Negativity (LN)

The deviant also showed greater negativity than the standard later than 400 ms. We did not predict this later negativity and thus present analyses in a separate section from MMN and were stricter in the criterion for concluding a significant effect ($p < 0.01$). To examine whether this late negativity (LN) was significant and whether language group, site, hemisphere or time interacted with this stimulus difference, ANOVAs comparing stimulus (standard, deviant) were carried out for each contrast and standard-order condition separately (tado, taado, miʃi, miʃʃi). For each condition, five-way analyses of variances (ANOVAs) with language group (AE vs. JP), stimulus (standard vs. deviant), site (frontal: F3, Fz, F4 vs. central: C3, Cz, C4), hemisphere (left vs. midline vs. right), and time in 40 ms time windows (444-480; 484-520; 524-560; 564-600; 604-640; 644-680; 684-720 ms) were carried out. The time intervals used for these ANOVAs were determined from the GFP figures and topography maps.

The grand averages of the ERPs to the standard, deviant and subtraction waveforms are shown at Fz and LM for each language group in Figure 8 (see Appendix I for all nine sites with LM and RM: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4). Figure 8 reveals a

greater late negativity to the deviant compared to the standard for both stimulus contrasts and for both language groups, although the difference appears to be larger in some cases. Statistical analyses are presented below for each contrast type and standard order (e.g., [tado]-as-standard condition) to determine whether these apparent differences between the deviant and the standard are significant. Only effects including Stimulus as a factor will be reported for the five-way ANOVAs.

Vowel Contrast:

[tado]-as-standard condition: The five-way ANOVAs comparing the ERPs for the standard and the deviant revealed a significant main effect of Stimulus [$F(1, 22) = 17.05$, $p = 0.000$] and significant interactions of Stimulus x Site [$F(1, 22) = 30.57$, $p = 0.000$], Stimulus x Time [$F(6, 132) = 16.85$, $p = 0.000$], Stimulus x Time x Language Group [$F(6, 132) = 3.74$, $p = 0.024$], Stimulus x Site x Hemisphere [$F(2, 44) = 11.32$, $p = 0.000$], Stimulus x Hemisphere x Time [$F(12, 264) = 8.23$, $p = 0.000$] and Stimulus x Site x Hemisphere x Time [$F(12, 264) = 6.68$, $p = 0.000$]. These significant findings indicate that the deviant was more negative than the standard and that its magnitude was affected by the factors of Language Group, Site, Hemisphere and Time.

Four-way ANOVAs were carried out to see whether there were differences in the LN related to language group (AE vs. JP) separately that interacted with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was a significant main effect of Stimulus [$F(1, 11) = 23.24$, $p = 0.0005$]; there were also significant interactions of Stimulus x Site [$F(1, 11) = 47.06$, $p = 0.000$], Stimulus x Time [$F(6, 66) = 13.24$, $p = 0.000$], Stimulus x Site x Hemisphere [$F(2, 22) = 11.06$, $p = 0.002$] and Stimulus x Hemisphere x Time [$F(12, 132) = 6.87$, $p = 0.000$]. In the JP

group, there were significant interactions of Stimulus x Time [$F(6, 66) = 4.41, p = 0.001$] and Stimulus x Site x Hemisphere x Time [$F(12, 132) = 4.57, p = 0.006$]. These significant findings confirm that a LN was present for both language groups.

To simplify interpretation of the interactions, four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (444-480; 484-520; 524-560; 564-600; 604-640; 644-680; 684-720 ms) were carried out using the subtraction waveforms (deviant – standard). A significant interaction of Site x Hemisphere [$F(2, 44) = 11.32; p = 0.000$] (see Figure 30) revealed that the frontal sites showed larger LNs than the central sites across left, midline and right sites but more so for the midline and the right sites, and post-hoc tests ($p < 0.05$) supported this result. A Time x Language Group interaction was significant [$F(6, 132) = 3.74, p = 0.0018$]. Figure 31 suggests that the AE group had a larger LN than the JP group, especially in time intervals 644-680 (time 6) ms, although pairwise post-hoc comparisons did not show a significant difference between groups for any time-interval. Hemisphere x Time [$F(12, 264) = 8.23, p = 0.000$] and Site x Hemisphere x Time interactions [$F(12, 264) = 6.68, p = 0.000$] (see Figure 32) were found. Two-way analyses examining site and hemisphere for each time separately reveal that the frontal sites elicited larger LN than the central sites at left-midline and right sites across all the times bins (time 1: [$F(2, 44) = 3.60, p = 0.04$]; time 2: [$F(2, 44) = 5.92, p = 0.005$]; time 3: [$F(2, 44) = 8.82, p = 0.0006$]; time 4: [$F(2, 44) = 11.51, p = 0.000$]; time 5: [$F(2, 44) = 11.90, p = 0.000$]; time 6: [$F(2, 44) = 15.89, p = 0.000$]; time 7: [$F(2, 44) = 16.16, p = 0.000$]), which was supported by post-hoc tests except the left hemisphere in time 7.

In sum, both language groups had the LN for the vowel contrast with the short vowel as standard. The frontal sites showed a larger LN than the central sites especially at the midline and the right sites. The AE group had a larger LN than the JP group.

[taado]-as-standard condition: The five-way ANOVAs comparing the ERPs for the standard and the deviant revealed a significant main effect of Stimulus [$F(1, 22) = 7.51, p = 0.012$] and significant interactions of Stimulus x Site [$F(1, 22) = 39.86, p = 0.000$], Stimulus x Hemisphere x Language Group [$F(2, 44) = 4.85, p = 0.019$], Stimulus x Time [$F(6, 132) = 12.01, p = 0.000$], Stimulus x Site x Time [$F(6, 132) = 2.73, p = 0.08$], Stimulus x Hemisphere x Time [$F(12, 264) = 5.33, p = 0.004$] and Stimulus x Site x Hemisphere x Time [$F(12, 264) = 6.87, p = 0.000$]. These significant findings indicate that the LN is present and that its magnitude is affected by the factors of Language Group, Site, Hemisphere and Time.

Four-way ANOVAs were carried out to see whether there were differences in the LN related to language group (AE vs. JP) separately that interact with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. For the AE group, there were significant interactions of Stimulus x Site [$F(1, 11) = 28.96, p = 0.000$] and Stimulus x Time [$F(6, 66) = 6.94, p = 0.000$]. For the JP group, there were significant interactions of Stimulus x Site [$F(1, 11) = 11.78, p = 0.006$], Stimulus x Time [$F(6, 66) = 6.69, p = 0.000$] and Stimulus x Site x Hemisphere x Time [$F(12, 132) = 5.23, p = 0.006$]. These significant findings confirm that the LN is present in both language groups, and that it interacts with Site, Hemisphere and Time.

To simplify interpretation of the interactions, four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and

time (444-480; 484-520; 524-560; 564-600; 604-640; 644-680; 684-720 ms) were carried out using the subtraction waveforms (deviant – standard). An interaction of Site x Time [$F(6, 132) = 2.73, p = 0.016$] was seen. Post-hoc tests showed that the LN was larger at frontal than central sites ($p < 0.05$), as it was for the [tado]-as-standard condition. A significant interaction of Hemisphere x Language Group was also found [$F(2, 44) = 4.85, p = 0.013$]. Figure 33 shows that the AE group had the larger negativity at the right than midline or left sites, while the JP group showed the smallest LN at the right site, although pairwise post-hoc tests do not show significant differences between these sites. A significant interaction of Hemisphere x Time [$F(12, 264) = 5.33, p = 0.0000$] showed that the LN was seen strongly at the midline sites. A significant interaction of Site x Hemisphere x Time were found. The LN was seen more strongly at the frontal sites than at the central sites [$F(12, 264) = 6.87, p = 0.000$] (see Figure 34), which was supported by post-hoc tests ($p < 0.05$). Two-way analyses examining site and hemisphere for each time separately reveal that the frontal sites elicited larger MMN than the central sites in all hemispheres at time interval 684-720 ms (time 7): [$F(2, 44) = 5.30, p = 0.009$].

In sum, both language groups produced a LN in the frontal sites for the vowel contrast with the long vowel as standard. There was a little negativity at the central site.

The AE group had the largest and the JP had the smallest LN at the right site.

[tado vs. taado]: Four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (444-480; 484-520; 524-560; 564-600; 604-640; 644-680; 684-720 ms) were carried out using the subtraction waveforms (deviant – standard). There was no main effect of standard order, but a significant interaction of Order x Site x Time [$F(6, 132) = 4.42, p = 0.0089$] (see Figure

35) was found. Two-way analyses examining standard order and site for each time separately reveal that there was no significant interaction for any time bin. Thus, the overall pattern led to the difference, but there was no particular time bin which reflected the large difference. Figure 35 suggests that the difference between [tado]-as-standard condition and [taado]-as-standard condition was greater for central than the frontal sites, but probably less so for the time interval 564-600 (time 4) ms. Post-hoc tests also showed that the [tado]-as-standard condition showed a slightly larger LN than did the [taado]-as-standard condition in time 524-560 (time 3) and 564-600 (time 4) ms.

In sum, there was no significant difference in the magnitude of the LN between two standard orders although [tado]-as-standard condition tended to produce a slightly larger LN than did the [taado]-as-standard condition.

Consonant Contrast:

[mi\i]-as-standard condition: The five-way ANOVAs comparing the ERPs for the standard and the deviant revealed a significant main effect of Stimulus [$F(1, 22) = 7.50$, $p = 0.012$] and significant interactions of Stimulus x Site [$F(1, 22) = 23.28$, $p = 0.000$], Stimulus x Hemisphere [$F(2, 44) = 7.33$, $p = 0.002$], Stimulus x Time [$F(6, 132) = 12.08$, $p = 0.000$], Stimulus x Site x Time [$F(6, 132) = 42.94$, $p = 0.000$], Stimulus x Hemisphere x Time [$F(12, 264) = 5.18$, $p = 0.001$], and Stimulus x Site x Hemisphere x Time [$F(12, 264) = 2.56$, $p = 0.004$]. These significant findings indicate that a late negativity was present and that its magnitude was affected by the factors of Site, Hemisphere and Time.

Four-way ANOVAs were carried out to see whether there were differences in the LN related to language group (AE vs. JP) separately that interact with site (frontal vs.

central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was no main effect of Stimulus [$F(1, 11) = 2.83, p = 0.121$], but there were significant interactions of Stimulus x Time [$F(6, 66) = 6.14, p = 0.000$], Stimulus x Site x Time [$F(6, 66) = 24.08, p = 0.000$], and Stimulus x Hemisphere x Time [$F(12, 132) = 4.30, p = 0.008$]. In the JP group, there was a significant main effect of Stimulus [$F(1, 11) = 4.70, p = 0.05$]; there were also significant interactions of Stimulus x Site [$F(1, 11) = 16.60, p = 0.002$], Stimulus x Hemisphere [$F(2, 22) = 4.88, p = 0.018$], Stimulus x Time [$F(6, 66) = 6.32, p = 0.000$], and Stimulus x Hemisphere x Time [$F(6, 66) = 19.06, p = 0.000$]. These significant findings confirm that the LN was present for both groups.

To simplify interpretation of the interactions, four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (444-480; 484-520; 524-560; 564-600; 604-640; 644-680; 684-720 ms) were carried out using the subtraction waveforms (deviant – standard). Significant interactions of Site x Time [$F(6, 132) = 42.94, p = 0.000$] and Hemisphere x Time [$F(12, 264) = 5.18, p = 0.000$] were found (see Figure 36). Frontal sites showed larger LNs than the central sites in all time intervals except the 444-480 ms time interval (time 1) supported by post-hoc tests ($p < 0.05$) and the left site showed the largest LN followed by the right and then midline sites. Post-hoc tests ($p < 0.05$) showed that the left and right sites had significantly larger LNs than the midline site except the 444-480 ms time interval (time 1).

In sum, both language groups had the LN at frontal sites for the consonant contrast with the short consonant as the standard, but negligible LNs at central sites. The left-frontal site showed the largest LN.

[missi]-as-standard condition: The five-way ANOVAs comparing the ERPs for the standard and the deviant did not reveal a significant main effect of Stimulus [$F(1, 22) = 0.27, p = 0.61$]. There was, however, a significant interaction of Stimulus x Site x Time [$F(6, 132) = 15.10, p = 0.000$]. This finding indicates the presence of the late negativity.

Four-way ANOVAs were carried out to see whether there were differences in the LN related to language group (AE vs. JP) separately that interact with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was no main effect of Stimulus [$F(1, 11) = 0.13, p = 0.72$], but there was a significant interaction of Stimulus x Site x Time [$F(6, 66) = 16.71, p = 0.000$]. In the JP group, there was no main effect of Stimulus [$F(1, 11) = 0.14, p = 0.71$], but there was a significant interaction of Stimulus x Time [$F(6, 66) = 4.91, p = 0.000$]. These findings confirm that the LN was present for both language groups.

To simplify interpretation of the interactions, four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (444-480; 484-520; 524-560; 564-600; 604-640; 644-680; 684-720 ms) were carried out using the subtraction waveforms (deviant – standard). A significant interaction of Site x Time [$F(6, 132) = 15.10, p = 0.000$] was observed (see Figure 37 top). Frontal sites showed larger LNs than the central sites in time intervals 564-600 (time 4), 604-640 (time 5), 644-680 (time 6) and 684-720 (time 7) ms. Post-hoc comparisons ($p < 0.05$) indicated a larger negativity at the frontal than central sites, in the time intervals 604-640 (time 5), 644-680 (time 6), and 684-720 (time 7) ms. A significant interaction of Hemisphere x Time [$F(12, 264) = 2.96, p = 0.000$] was found (see Figure 37 bottom) and the left site showed the largest LN except the time interval 444-480 (time 1) ms. Post-hoc

tests showed a larger negativity at left sites than midline or right sites, in time intervals 564-600 (time 4), 604-640 (time 5), and 644-680 (time 6) ms ($p < 0.05$).

In sum, both language groups had the LN for the consonant contrast with the long vowel as the standard. The frontal sites showed a larger LN than the central sites and the left site showed the largest LN.

[miʃi vs. miʃfi]: Four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (444-480; 484-520; 524-560; 564-600; 604-640; 644-680; 684-720 ms) were carried out using the subtraction waveforms (deviant – standard). There was no main effect of standard order, but there were significant interactions of Order x Site [$F(1, 22) = 10.93, p = 0.003$], Order x Time [$F(6, 132) = 5.40, p = 0.000$], Site x Time [$F(6, 132) = 52.64, p = 0.000$] and Order x Site x Time [$F(6, 132) = 5.91, p = 0.007$] (see Figure 38). The LN was seen at frontal sites from time interval 564-600 (time 4) ms for [miʃi]-as-standard condition, but was later and smaller for [miʃfi]-as-standard condition. Two-way analyses examining standard order and site for each time separately reveal that there was significant interaction in the following time intervals: 524-560 (time 3): [$F(1, 22) = 12.57, p = 0.002$]; 564-600 (time 4): [$F(1, 22) = 15.89, p = 0.001$]; 604-640 (time 5): [$F(1, 22) = 13.84, p = 0.001$] and 644-680 (time 6): [$F(1, 22) = 9.89, p = 0.005$]. Significant interactions were also found in Hemisphere x Time: [$F(12, 264) = 3.10, p = 0.0004$] and Order x Hemisphere x Time: [$F(12, 264) = 5.30, p = 0.0002$]. Two-way analyses examining contrast and hemisphere for each time separately reveal that there was a marginal significant interaction only in time interval 524-560 (time 3) ms: [$F(2, 44) = 3.94, p = 0.03$]. Figure 39 shows that the left sites were more negative for all the time

bins except time intervals 444-480 (time 1) and 484-520 (time 2) ms, and that there was a greater difference between [miʃi] and [miʃʃi] at lateral than midline sites. Examination of Figure 39 suggests that the interaction is the result of greater negativity for the [miʃi]-as-standard than the [miʃʃi]-as-standard condition at the lateral than midline sites, and that the peak negativities occurred in different time intervals.

In sum, the LN for [miʃʃi]-as-standard condition was later and smaller than for the [miʃi]-as-standard condition for both language groups. There was greater negativity of [miʃi] than [miʃʃi] at the lateral compared to midline sites.

2.2.3 Discussion

AE and JP listeners both showed robust discrimination of JP temporally-cued vowel and consonant contrasts in this Auditory-Attend experiment, as evidenced by both behavioral and electrophysiological measures. During ERP recording, performance on the implicit counting task by the JP group was better than the AE group overall, but there was no significant group difference on any one of the four contrast-standard-order conditions. Thus, the two groups appeared to be attending to the deviant stimuli at approximately equal levels during this part of the experiment. In the subsequent button-press behavioral task, however, the JP group performed significantly better than the AE group on both contrasts in both standard orders. Thus, there was an effect of language experience on this behavioral measure of discrimination. However, overall performance for both groups was above chance for both vowel and consonant contrast conditions. As for the contrast type, the vowel contrast was significantly easier to discriminate behaviorally than the consonant contrast for both groups.

On electrophysiological tasks, both groups showed significant MMNs for both contrasts in both standard-order conditions. The only language group differences in ERP data were in topography and lateralization. The early negativity peaking at frontal-central sites between 200 and 230 ms for the short vowel, and 250-280 ms for the long vowel as standard, and the second negativity, peaking between 300 and 330 ms for the short consonant and between 370 and 400 ms for the long consonant as standard are consistent with the timing and topography of MMN. Significant negativity was found at central sites for the AE group and at frontal sites for the JP group for the short vowel as standard from 204 to 240 ms. For the long vowel as standard, the largest negativity was found at the midline-frontal site (i.e., Fz) for both groups from 244 to 280 ms. For the short consonant as standard, the largest negativity was found at central sites for both groups from 324 to 440 ms. For the long consonant as standard, the largest negativity was found for both groups at the left-central (i.e., C3) and the midline-frontal (i.e., Fz) sites from 364 to 400 ms. These negativities have the timing and topography of the MMN (Näätänen, 1990). Specifically, the negativity peaks between 100 and 300 ms after the time range where the short and long stimuli first can be identified as different (98 ms for vowels and 211 ms for consonants). There was no clear evidence of N2b. That is, the central sites did not show an absence of inversion at the mastoids in the time range of a central negativity. However, the AE group did show larger central than frontal negativity in later time intervals (with some mastoid inversion) for the [miʃfi]-as-standard condition, which may indicate overlap of the MMN and N2b. Both groups also showed a late positivity at posterior sites (i.e., P3b) peaking between 400 and 800 ms (largest time bins: [tado]: 484-520 ms; [taado]: 524-560 ms; [miʃi]: 644-680 ms; [miʃfi]: 684-720 ms).

Both groups also showed a later negativity (LN). This LN is consistent with the LN reported in several studies (Shafer, Morr, Datta, Kurtzberg & Schwartz, 2005a; Zachau, Rinker, Körner, Kohls, Maas, Hennighausen and Schecker, 2005; Shestakova, Huotilainen, Čeponiene and Cheour, 2003). The frontal sites showed a much larger LN than did the central sites. In addition, the LN was largest at the frontal sites of the left hemisphere (i.e., F3). The AE group elicited a larger LN than did the JP group, which may also indicate the level of attention (Shafer et al. 2005a). The AE group might have been using more attentional resources than the JP group if the LN is considered some measure of attention. Thus, it appears that the topographical difference also hints at the level/degree of attention.

Vowel versus Consonant Contrasts – Behavioral data supported the conclusion that the vowel contrast was easier than the consonant contrast for both group. This may reflect a difference in psychoacoustic salience for the JP listeners, since both contrasts are phonologically functional in their language. As for the AE group, in addition to this psychoacoustic salience, language experience might also be a factor in the language group difference because AE group showed significantly poorer performance than the JP group especially on the consonant contrast.

With respect to the ERP data, there was also evidence that the vowel contrast was easier than the consonant contrast, as evidenced by the higher amplitude MMN peak components produced in the vowel contrast conditions. Note that the difference in duration of the consonant segments comes in the middle of the word; thus the MMN and other measures are expected to be seen later. The finding of later latency MMNs to the consonants compared to the vowels was expected because the consonant duration

difference occurs approximately 100 ms later than the vowel duration difference. However, the negativity that is presumably the MMN was also smaller in magnitude for the consonants than the vowels, suggesting that the consonant contrast was psychoacoustically less salient than the vowel contrast as suggested by the behavioral data. This will be addressed further in the general discussion after Experiment II.

Short as standard and Long as standard – There were also differences in the ERPs dependent on which stimulus of a contrast pair served as the standard. The major finding was that the latency of the discriminative components, MMN, LN, P3b, were later when the long stimulus served as the standard. Such asymmetries have been observed in other studies including Shafer et al. (2004) and Friedrich, Weber, and Friederici (2004). There were also some interactions of time with hemisphere and site, but these appear to be primarily an artifact of the time difference, and will not be discussed further. This finding of an asymmetry in the component latencies may be the result of psychoacoustic factors and will be addressed further in the general discussion.

In summary, the MMN and P3b were present for both language groups, contrast orders and contrast types. A significant LN was also found. Evidence for the N2b was less strong. Overall, the presence of a robust MMN for the AE group similar in amplitude to that of the JP group suggests that, with attention, non-native listeners can fairly easily discriminate non-phonemic vowel and consonant duration differences. However, subtle differences in topography, hemisphere and latency suggest that they may be engaging resources differently to arrive at a similar behavioral outcome. These findings will be discussed in greater detail in chapter 5 after presenting the results of the visual task.

Chapter 3. Experiment 2: Visual-Attend Experimental Group

3.1 Overview

In the Auditory-Attend Experiment, the discrimination task required *direct* attention to the target speech stimuli via the counting of the deviant word forms. In the Visual-Attend Experiment, using the same paradigm and same environment as in the Exp 1, discriminability of target speech stimuli was examined when attention was focused away from the auditory input. We predicted that the AE listeners might not show evidence of discrimination of the duration differences in vowels and consonants without attention, whereas the native speakers of JP would show automatic pre-attentive discrimination of both vowel and consonant contrasts. That is, AE listeners were predicted to show absent or small MMNs because processing of these speech contrasts by AE listeners is less automatic. The results of both experiments (Exp 1 and 2) were examined together later to ask whether the attention manipulation produced differences in discriminative responses for AE versus JP groups (i.e., whether non-native listeners were less automatic at discriminating such contrasts, and whether native listeners showed any effects of attention on processing such native speech contrasts). Four-way ANOVAs with experimental group (Auditory vs. Visual), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (124-160; 164-200; 204-240, 244-280; 284-320; 324-360 ms) were carried out for each language group and each contrast separately.

3.2 Methods

3.2.1 Stimulus Materials

Auditory and Visual stimuli: The same stimuli as in Exp 1 were used.

Familiarization task materials: The same visual stimuli used for the experiment were used for the familiarization task.

Behavioral Task: The same stimuli as in Exp 1 were used.

3.2.2 Participants

Twelve native speakers of JP (two males and ten females) served as a control group, and twelve native speakers of AE participated as target listeners (four males and eight females). The criteria for the selection were the same as in Exp 1.

3.2.3 Paradigms and Procedures

The overall procedures were the same as for Exp 1. As in the Exp 1, the pentagon versus hexagon series and the circle versus oval series were created. The pentagon versus hexagon contrast was presented with vowel stimuli (color: orange with black background), while the circle versus oval contrast was presented with consonant stimuli (color: green with black background). All four shapes were used as the standard and deviant, and the order was counter-balanced (see Figure 3 for the paradigm of the visual-attention condition). The only differences compared to the Auditory-Attend task were that a visual oddball task was presented and that the ISI was 780 ms for the shapes presented along with the vowel stimuli (800 ms ISI) and 820 ms for the shapes presented along with the consonant stimuli (800 ms ISI), in order to make sure that the presentation of auditory and visual stimuli would not overlap. SOAs were 1015 ms to 1170 ms (235 & 290 ms + 780 ms for vowels; 300 & 350 ms + 820 ms for consonants). At the beginning of the experiment, detailed oral instructions (see Appendix J) about the visual shapes (same as the actual shapes) were given to make the participants aware of what they needed to do to. They were asked to ignore the auditory stimuli and watch the computer

screen. Participants were asked to count a target shape (i.e., deviant shape) silently and to report the number to an investigator at the end of each block. Each block lasted approximately 1-2 minutes. For the button-press behavioral task at the end of the session, participants were given the same instructions as in Exp 1 with sample words (about two minutes). Because they were told to ignore the auditory stimuli during the ERP recording, they needed to be aware of what they needed to attend to in the behavioral task (i.e., deviant stimuli).

3.2.4 Electroencephalogram (EEG) Recording and Data Analysis

The analyses were identical to experiment 1 for the MMN and LN. The N2b and P3b were not expected in this paradigm because attention was directed away from the auditory stimuli.

3.3 Results

3.3.1 Behavioral Results

Behavior on the counting task (during ERP recording): Conditions are named according to which stimulus served as the standard, for example, pentagon with [tado], hexagon with [taado], circle with [miʃi], and oval with [miʃʃi]. The mean scores presented in Figure 40 describes the number of “no error blocks” reported in each condition. The same analysis as in Exp 1 was used. Because the data were not normally distributed, a Median test was conducted to determine the significance of group differences. There was no group difference for any of the conditions ($p < 0.05$) except for [tado]-as-standard condition (i.e., pentagon as the standard and hexagon as the deviant). Arguably this result was spurious, in light of the fact that the other seven (ok as is) comparisons were non-significant. Thus, overall, the task was equally difficult for both the JP and the AE groups,

as was seen in the Exp 1. A Wilcoxin Matched-Pairs test was used to assess whether performance on pentagon versus hexagon shapes with vowel auditory input was more accurate than performance on the circle versus oval shapes with consonant contrast input. Results showed that performance on the pentagon versus hexagon task with [tado]-as-standard condition was better than on the circle versus oval task with [mifi] ($p = 0.019$) and [miffi] ($p = 0.0011$). The [taado]-as-standard condition was also easier than [miffi] ($p = 0.002$) and approached significance for [mifi] ($p = 0.06$). The same test was conducted with language group (AE vs. JP) separately. In the AE group, pentagon versus hexagon shape discrimination with [tado]-as-standard condition was not easier than [mifi] ($p = 0.09$) and [miffi] ($p = 0.07$). The [taado]-as-standard condition was also not easier than [mifi] ($p = 0.153$), but was easier in [miffi] ($p = 0.05$). In the JP group performance on the visual discrimination task in the [tado]-as-standard condition was not easier than [mifi] ($p = 0.09$), but was easier in [miffi] ($p = 0.0077$). Performance in the [taado]-as-standard condition was also not easier than [mifi] ($p = 0.176$), but was easier than [miffi] ($p = 0.0077$). Thus, for both groups, overall performance was better in pentagon versus hexagon task presented with the vowel stimuli than in the circle versus oval task presented with the consonant stimuli.

Behavior on the button-press task (after ERP recording): The same behavioral test as in Exp 1 was conducted at the end of the ERP session in this experiment. A Median test revealed significant differences between the AE and JP groups for all conditions except the [taado]-as-standard condition (tado: $p < 0.001$; mifi: $p < 0.01$; miffi: $p < 0.01$) (see

Table 5 and Figure 41). In all comparisons, the JP group was more accurate than the AE group. A Wilcoxin Matched-Pairs test was used to assess whether the vowel contrast was easier than the consonant contrast. The [tado]-as-standard condition was easier than [miʃi] ($p = 0.000$) and [miʃfi] ($p = 0.014$). The [taado]-as-standard condition was also easier than [miʃi] ($p = 0.000$) and [miʃfi] ($p = 0.000$). The same test was conducted with language group (AE vs. JP) separately. In the AE group, [tado]-as-standard condition was easier than [miʃi] ($p = 0.018$), but not [miʃfi] ($p = 0.126$). The [taado]-as-standard condition was also easier than [miʃi] ($p = 0.005$) and [miʃfi] ($p = 0.003$). In the JP group, [tado]-as-standard condition was easier than [miʃi] ($p = 0.0033$) and [miʃfi] ($p = 0.012$). [taado]-as-standard condition was also easier than [miʃi] ($p = 0.0033$) and [miʃfi] ($p = 0.005$). Thus, there was a language group difference where the JP group showed a better performance than the AE group for three out of four contrast types and for both groups the overall performance was better in vowel than the consonant contrast. Reaction Time (RT) was also examined by using a Median test and ANOVA. There was no significant overall difference between the two groups (at $p < 0.05$).

3.3.2 Electrophysiological results

3.3.2.1 MMN

The same factors as in Exp 1 were used for the ANOVAs. The grand averages of the ERPs to the standard, deviant and subtraction waveforms are shown at Fz and LM for each language group in Figure 42 {see Appendix K for all 12 sites: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, LM, LRM [(LM+RM)/2], RM}. This figure reveals a greater negativity to the deviant compared to the standard for both stimulus contrasts and for both language

groups, although the difference appears to be larger in some cases. Statistical analyses are presented below for each contrast-standard order condition (e.g., vowel-[tado] as the standard) to determine whether these apparent differences between the deviant and the standard were significant. Only effects including Stimulus as a factor are reported for the five-way ANOVAs.

Vowel Contrast:

[tado]-as-standard condition: The five-way ANOVAs comparing the ERPs for the standard and the deviant revealed a significant main effect of Stimulus [$F(1, 22) = 15.10, p = 0.000$] and significant interactions of Stimulus x Time [$F(5, 110) = 17.61, p = 0.000$], Stimulus x Hemisphere x Time [$F(10, 220) = 3.87, p = 0.002$], and Stimulus x Site x Hemisphere x Time [$F(10, 220) = 3.30, p = 0.008$]. These significant findings indicate that the deviant was more negative than the standard and that its magnitude was affected by the factors of Site, Hemisphere and Time.

Four-way ANOVAs were carried out to see whether there were differences in the MMN related to language group (AE vs. JP) separately that interact with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was no main effect of Stimulus [$F(1, 11) = 3.02, p = 0.11$], but there was a significant interaction of Stimulus x Time [$F(5, 55) = 5.69, p = 0.000$] and marginally for Stimulus x Site x Hemisphere x Time [$F(10, 110) = 2.63, p = 0.06$]. In the JP group, there was a significant main effect of Stimulus [$F(1, 11) = 16.10, p = 0.002$]; there was also a significant interaction of Stimulus x Time [$F(5, 55) = 14.12, p = 0.000$] and marginally for Stimulus x Hemisphere x Time [$F(10, 110) = 2.45, p = 0.07$]. These findings confirm that the deviant was more negative than the standard for both groups.

To simplify interpretation of the interactions, four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (124-160; 164-200; 204-240; 244-280; 284-320; 324-360 ms) were carried out using the subtraction waveforms (deviant – standard). Significant interactions of Hemisphere x Time [$F(10, 220) = 3.87, p = 0.0008$] and Site x Hemisphere x Time [$F(10, 220) = 3.30, p = 0.008$] revealed that the frontal sites were larger than the central, but only at the midline from 204 to 240 (time 3) ms (see Figure 43). Two-way analyses examining site and hemisphere for each time separately reveal that there was only marginal significance at time interval 204-240 ms [$F(2, 44) = 2.86, p = 0.07$].

In sum, both language groups showed MMNs for the vowel contrast in the short vowel standard condition, but they appeared larger for the JP group (although this did not reach significance). MMNs at the frontal sites were larger than at the central sites, but only at the midline.

[taado]-as-standard condition: The five-way ANOVAs comparing the ERPs for the standard and the deviant revealed a significant main effect of Stimulus [$F(1, 22) = 9.07, p = 0.006$] and significant interactions of Stimulus x Time [$F(5, 110) = 43.54, p = 0.000$], Stimulus x Time x Language Group [$F(5, 110) = 2.93, p = 0.05$], Stimulus x Hemisphere x Time [$F(10, 220) = 4.67, p = 0.002$], Stimulus x Hemisphere x Time x Language Group [$F(10, 220) = 3.17, p = 0.018$], and Stimulus x Site x Hemisphere x Time [$F(10, 220) = 5.27, p = 0.002$]. These significant findings indicate that the deviant was more negative than the standard and that its magnitude was affected by the factors of Language Group, Site, Hemisphere and Time.

Four-way ANOVAs were carried out to see whether there were differences in the MMN related to language group (AE vs. JP) separately that interact with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was no main effect of Stimulus [$F(1, 11) = 0.51, p = 0.50$], but there were significant interactions of Stimulus x Time [$F(5, 55) = 10.40, p = 0.005$] and Stimulus x Site x Hemisphere x Time [$F(10, 110) = 3.28, p = 0.04$]. In the JP group, there was a significant main effect of Stimulus [$F(1, 11) = 15.23, p = 0.002$]; there were also significant interactions of Stimulus x Time [$F(5, 55) = 43.31, p = 0.000$], Stimulus x Hemisphere x Time [$F(10, 110) = 2.63, p = 0.05$] and Stimulus x Site x Hemisphere x Time [$F(10, 110) = 3.06, p = 0.04$]. These findings confirm that the deviant was more negative than the standard for both groups.

To simplify interpretation of the interactions, four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (124-160; 164-200; 204-240; 244-280; 284-320; 324-360 ms) were carried out using the subtraction waveforms (deviant – standard). Significant interactions of Time x Language Group interaction [$F(5, 110) = 2.93, p = 0.016$], Hemisphere x Time [$F(10, 220) = 4.67, p = 0.002$] and Hemisphere x Time x Language Group [$F(10, 220) = 3.17, p = 0.018$] (see Figure 44) were observed. Figure 44 illustrates that the JP group showed a larger MMN than the AE group, especially at the left and midline sites. Two-way analyses examining hemisphere and language group for each time separately reveal that there was significance only at time interval 284-320 ms (time 5) [$F(2, 44) = 3.27, p = 0.05$]. The largest peak amplitude was shown in the time interval 244-280 ms (time 4).

In sum, both language groups showed MMNs for the vowel contrast with the long vowel as standard. The JP group showed a larger MMN than the AE group, especially at the left and midline sites.

[tado vs. taado]: Four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (124-160; 164-200; 204-240, 244-280; 284-320; 324-360 ms) were carried out using the subtraction waveforms (deviant – standard). Significant interactions of Order x Time [$F(5, 110) = 28.94, p = 0.000$], Order x Hemisphere x Time: [$F(10, 220) = 4.69, p = 0.0008$] and Order x Site x Hemisphere x Time: [$F(10, 220) = 5.26, p = 0.0006$] (see Figure 45) were found. Two-way analyses examining standard order and hemisphere for each time separately reveal that there was no significant interaction. The midline-frontal (Fz) hemisphere showed the largest MMN for both standard orders and the largest difference between two standard orders, which was supported by post-hoc tests. Examination of Figure 45 shows that the largest negativity was found at midline frontal sites but was later for the [taado] than the [tado]-as-standard condition. In sum, there has a significant difference in the latency of the MMN for the different orders for both groups, but this did not interact strongly with site or hemisphere. The midline-frontal (Fz) hemisphere showed the largest MMN for both standard orders.

Consonant Contrast:

[mi{i]-as-standard condition: The five-way ANOVAs comparing the ERPs for the standard and the deviant did not reveal a significant main effect of Stimulus [$F(1, 22) = 0.06, p = 0.82$]. However, there were significant interactions of Stimulus x Time [$F(6,$

132) = 6.49, $p = 0.000$] and Stimulus x Hemisphere x Time [$F(12, 264) = 3.31$, $p = 0.007$].

Four-way ANOVAs were carried out to see whether there were differences in the MMN related to language group (AE vs. JP) separately that interact with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was no main effect of Stimulus [$F(1, 11) = 0.11$, $p = 0.75$] but a significant interaction of Stimulus x Time [$F(5, 55) = 2.92$, $p = 0.014$]. In the JP group, there was no main effect of Stimulus [$F(1, 11) = 1.57$, $p = 0.24$], but there were significant interactions of Stimulus x Time [$F(6, 66) = 5.95$, $p = 0.000$]. These findings confirm that the deviant was more negative than the standard in some time-intervals for both groups.

To simplify interpretation of the interactions, four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (204-240; 244-280; 284-320; 324-360; 364-400; 404-440; 444-480) were carried out using the subtraction waveforms (deviant – standard). The Hemisphere x Time interaction [$F(12, 264) = 3.31$, $p = 0.000$] indicated that the midline sites showed the largest negativity from 404-440 (time 6) ms. Post-hoc tests indicated that the greatest difference among regions was in the interval 404-440 ms (see Figure 46). There were no language group differences.

In sum, both language groups showed an MMN for the consonant contrast with the short consonant as standard, but only at certain time intervals. There was no overall language group difference.

[mi]{i}-as-standard condition: The five-way ANOVAs comparing the ERPs for the standard and the deviant revealed a significant main effect of Stimulus [$F(1, 22) = 6.13$,

$p = 0.022$] and significant interactions of Stimulus x Site [$F(1, 22) = 4.31, p = 0.05$], Stimulus x Time [$F(6, 132) = 11.41, p = 0.000$] and Stimulus x Hemisphere x Time [$F(12, 264) = 2.43, p = 0.04$]. These significant findings indicate that the deviant was more negative than the standard and that its magnitude was affected by the factors of Site, Hemisphere and Time.

Four-way ANOVAs were carried out to see whether there were differences in the MMN related to language group (AE vs. JP) separately that interacted with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was no main effect of Stimulus [$F(1, 11) = 0.36, p = 0.56$], but there was a significant interaction of Stimulus x Time [$F(6, 66) = 5.33, p = 0.000$]. In the JP group, there was a significant main effect of Stimulus [$F(1, 11) = 9.46, p = 0.011$]; there were also significant interactions of Stimulus x Site [$F(1, 11) = 6.61, p = 0.03$] and Stimulus x Time [$F(6, 66) = 6.92, p = 0.000$]. These findings confirm that the deviant was more negative than the standard for both groups.

To simplify interpretation of the interactions, four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (204-240; 244-280; 284-320; 324-360; 364-400; 404-440; 444-480 ms) were carried out using the subtraction waveforms (deviant – standard). A Hemisphere x Time interaction [$F(12, 264) = 2.43, p = 0.005$] was observed (see Figure 47). The largest peak amplitude was shown in time 364-400 ms (time 5). In addition, it appears that there was an early negativity in the time interval 244-280 ms (time 2) as seen in the Auditory-Attend experiment, but the amplitude was less than $0.5 \mu\text{V}$. The right site was more negative than the midline, but not at the peaks of negativity.

In sum, both language groups produced a significant MMN for the consonant contrast with the long consonant as the standard, but the MMN amplitude was larger for the JP group although this did not reach significance. There was some indication of an earlier negativity than the MMN to the consonant contrast.

[miʃi vs. miʃfi]: Four-way ANOVAs with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (204-240; 244-280; 284-320; 324-360; 364-400; 404-440; 444-480 ms) were carried out using the subtraction waveforms (deviant – standard). There was no main effect of standard order, but significant interactions of Order x Time [$F(6, 132) = 10.26, p = 0.000$] and Order x Hemisphere x Time [$F(12, 264) = 4.20, p = 0.001$] (see Figure 48) were observed. Figure 48 suggests that [miʃfi]-as-standard condition showed a larger negativity than [miʃi]-as-standard condition in time 244-280 (time 2) and 364-400 (time 5) ms for left, right and midline and additionally from 404-400 (time 6) ms for the lateral sites and 204-240 (time 1) ms for the midline sites.

In sum, the main difference in standard order was the latency of the response with the short as a standard having a later latency than long as standard. The [miʃfi]-as-standard condition showed a larger negativity than [miʃi]-as-standard condition at certain times and certain sites.

Vowel vs. Consonant Contrasts:

The time bins were normalized for this analysis. As in the Auditory-Attend experiment, the two time bins which showed the largest MMN for each contrast-standard order condition (vowels: [tado]: 204-240 ms; [taado]: 244-280 ms; consonants: [miʃi]: 324-260

ms; [miʃfi]: 364-400 ms) were compared to see whether there was a difference in magnitude of the MMN across contrast types. Each time bin was averaged over the short and long standard conditions (e.g., time bin 204-240 ms: [tado] + [taado] /2; 244-280 ms: [tado] + [taado]/2). Five-way ANOVAs with language group (AE vs. JP), contrast type (vowel vs. consonant), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (204-240; 244-280 for vowels) and (324-360; 364-400 for consonants) were carried out using the subtraction waveforms (deviant – standard). There was a main effect, indicating that the vowel contrast produced a larger MMN than the consonant contrast [$F(1, 22) = 21.62, p = 0.00012$], but there were no significant interactions with language group, site or hemisphere.

In sum, for both language groups the vowel contrast elicited much larger MMNs than the consonant contrast.

3.3.2.2 Late Negativity

[tado]-as-standard condition: Only the [tado]-as-standard condition showed evidence of a late negativity in the graphs (see Figure 42). To examine whether this negativity was significant and whether it had the same timing and topography as the LN in the Auditory-Attend Experiment, ANOVAs were performed with stimulus, site, hemisphere, time and group as factors. A significant main effect of Stimulus [$F(1, 22) = 7.76, p = 0.01$] and significant interactions of Stimulus x Site [$F(1, 22) = 18.04, p = 0.000$] and Stimulus x Time [$F(6, 132) = 3.12, p = 0.007$] were observed. These significant findings indicate that the deviant was more negative than the standard and that its magnitude is affected by the factors of Site and Time.

Four-way ANOVAs were carried out to see whether there were differences in the LN related to language group (AE vs. JP) separately that interact with site (frontal vs. central), hemisphere (left vs. midline vs. right), or time. In the AE group, there was no main effect of Stimulus and no significant interactions. In the JP group, there was main effect of Stimulus [$F(1, 11) = 8.46, p = 0.01$] and there was a significant interaction of Stimulus x Site [$F(1, 11) = 13.17, p = 0.004$]. These significant findings indicate that LN is robustly present only in JP language group. The LN was larger at frontal compared to central sites and can be seen in Figure 42 for both groups, even though it did not reach significance for the AE group.

In sum, only the JP group showed the LN for the vowel contrast with the short vowel as standard. In this condition, the frontal site showed a larger LN than the central site.

3.3.3 Discussion

AE and JP listeners both showed evidence of discrimination of JP temporally-cued contrasts even when attention was directed away from the auditory channel, as evidenced by electrophysiological measures in Experiment 2. During ERP recording, the JP group showed slightly better performance than the AE group on the implicit counting of visual deviants task, but there was no overall significant group difference. Thus, we can assume that both groups were attending to the visual input. Overall performance was better in the visual task accompanying the vowel stimuli than in the task presented with consonant stimuli. If this performance is a measure of attention to the visual mode, it could be interpreted as showing that more attentional resources were allocated to the auditory channel in the consonant conditions for both groups. However, the task with the

consonant stimuli has expected to be harder, based on piloting before the study. The electrophysiological data indicated that discrimination was still more difficult on the consonant contrast than on the vowel contrast.

The performance on the behavioral button-press task (with attention) was similar to Auditory-Attend Experiment; the AE group performed above chance, but were significantly worse than the JP group. As for the contrast type, the vowel contrast was easier in general than the consonant contrast for both groups. It appears that there was no learning effect for the AE participants in the Auditory-Attend experiment, since both Auditory-Attend and Visual-Attend AE groups showed similar results in the same behavioral task.

With respect to the ERP data, when attention was directed away from the speech stimuli, JP listeners showed larger MMNs to the vowel duration contrast than AE listeners, particularly when [taado] served as the standard and particularly at left and midline sites. Significant negativities were found for the AE and JP groups for the short vowel as standard from 204 to 240 ms. For the long vowel as standard, significant negativities were found at the midline-frontal site (i.e., Fz) for both groups from 244 to 280 ms. No overall language group difference was found for the consonant duration contrast. For the short consonant as standard, significant negativities were found at central sites for both groups from 324 to 440 ms. For the long consonant as standard, significant negativities were found for both groups at the left-central (i.e., C3) and the midline-frontal (i.e., Fz) sites from 364 to 400 ms.

These ERP results support the claim that listeners are less automatic at discriminating non-native speech contrasts, but the data show this strongly only for the

vowel contrast. One of the reasons that there was no overall language group difference for the consonant contrast was that not only AE group, but also the JP group showed smaller MMNs than expected. That is, with attention allocated to the visual oddball task, the JP listeners appeared to have an attenuated MMN to the consonant contrast. These comparisons are discussed further in Chapter 4.

Vowel versus Consonant Contrasts – Behavioral data supported the conclusion that the vowel contrast was easier than the consonant contrast for both groups, but especially for the AE group. This again reflects the possible role of differences in psychoacoustic salience of duration differences in vocalic segments versus in fricative noise segments, at least for the JP listeners, since both contrasts are phonologically functional in their language. As for the AE group, in addition to this difference in psychoacoustic salience, language experience may have contributed further to the differences in indices of pre-attentive discrimination of vowel and consonant contrasts. There were differences in amplitude and latency of elicited MMNs for vowel and consonant contrasts, with the vowel contrast showing larger amplitude MMNs than the consonant contrast for both groups. One prediction of the study was that consonants would be more difficult to discriminate than vowels for the AE group, but not so for the JP group. In fact, the ERP and behavioral results support the conclusion that the consonant contrast was more difficult, regardless of language experience. This finding was similar to that seen in the Auditory-Attend Experiment; however, unlike the previous experiment, this experiment could assess the relationship between the behavioral data *with attention* and the ERP data *without attention*. Because the results showed that the vowel contrast was still easier than the consonant contrast in this experiment for both language groups, the differences in

relative difficulty of vowel and consonant contrasts can be interpreted as reflecting differences in psychoacoustic salience. Explanations for these patterns are presented in the General Discussion below.

An asymmetry in the latencies of the MMN was again seen, dependent on which phonetic category (short or long) served as the standard. For vowels, the MMN was later when the long-duration stimuli served as the standard. It was somewhat less clear for the consonants which category-as-standard led to the earlier MMN because of an additional early negativity in the long as standard condition. However, this early negativity was too early to be considered an MMN to the consonant duration difference; rather, it may have been an MMN to the small vowel duration difference that correlates inversely with consonant duration. The behavioral data appear to show an asymmetry in the opposite direction, with better performance found for the long-as-standard than for the short-as-standard conditions. However, the RTs are consistently later for the long-as-standard conditions by 70 to 100 ms. These findings will be addressed further in the General Discussion.

In summary, this second experiment supports the claim that speech perception of non-native contrasts is less automatic than that of native contrasts. An overall language effect was found for the vowel contrast (both standard orders). The pattern of findings was different for the vowel- and consonant-duration contrasts, with the vowel contrast showing a larger MMN than the consonant duration contrast for both language groups and showing a significant difference between language groups. In the next chapter the results from the Auditory and Visual-Attend Experiments are compared directly to

determine whether attention to the target auditory stimuli led to a larger MMNs for the AE group or for both groups.

Chapter 4. Experiment 1 vs. Experiment 2

Further analyses were conducted to examine the differences in ERP data between the Auditory-Attend Experiment and the Visual-Attend Experiment. ANOVAs were conducted using the same factors: language group, sites, hemispheres, and time (i.e., the same time intervals used in each experiment). The subtraction waveforms (deviant – standard) were used for all comparisons below.

4.1 Vowel Contrast

To examine whether attention affected the amplitude of the MMN for the vowel duration contrast, four-way ANOVAs with experimental group (Auditory vs. Visual), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (124-160; 164-200; 204-240; 244-280; 284-320; 324-360 ms) were carried out for each language group and each standard order separately. For the AE group in the [tado]-as-standard condition, interactions of Site x Experimental Group [$F(1, 22) = 4.47, p = 0.05$]; Site x Time x Experimental Group [$F(5, 110) = 3.13, p = 0.011$] and of Site x Hemisphere x Time x Experimental Group [$F(10, 220) = 2.27, p = 0.05$] were found (see Figure 49). Three-way analyses examining site and hemisphere and experimental group for each time separately revealed that there was a significant difference only at time interval 324-360 ms (time 6): [$F(2, 44) = 4.90, p = 0.02$]. Figure 49 shows that the MMN was generally larger for the AE participants in the Auditory- versus Visual-Attend Experiment at central sites for most times and particularly at the left central sites. In the earlier time intervals (1 and 2) the frontal sites also showed greater negativity for the Auditory- versus Visual-Attend group. For the JP participants, there were no significant differences between the two experimental groups, with only one interaction approaching significance: Site x

Hemisphere x Time x Experimental Group [$F(10, 220) = 2.02, p = 0.08$]. JP participants in the Auditory-Attend experiment showed a slightly larger MMN at the left-central site compared to those in the Visual-Attend study.

For the [taado]-as-standard condition, there was no significant difference for the AE Auditory- and Visual-Attend groups. The comparison of JP participants found a significant interaction of Time x Experimental Group [$F(5, 110) = 3.63, p = 0.0044$]. The JP participants in the Visual-Attend study showed a larger MMN to this duration contrast than those in the Auditory-Attend study, specifically in time intervals 244-280 (time 4), 284-320 (time 5) and 324-360 (time 6) ms.

The normalized time bins from the Auditory- and Visual-Attend experiments were used to examine whether attention affected the MMN for the vowel contrast ([tado + taado] = vowel); five-way ANOVAs with experimental group (Auditory vs. Visual), language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (204-240; 244-280 ms) were carried out. Interactions of Time x Experimental Group x Language Group [$F(1, 44) = 5.20, p = 0.028$] (see Figure 50) and Site x Time x Experimental Group [$F(1, 44) = 5.33, p = 0.03$] were observed (see Figure 51). The AE group in the Auditory-Attend study showed a larger MMN than in the Visual-Attend study, while JP group in the Auditory-Attend study showed a larger MMN than the JP group in the Visual-Attend study in time interval 204-240 ms and the reverse result was for the time interval 244-280 ms. Both language groups in Auditory-Attend study produced larger MMNs than the Visual-Attention group in both frontal and central sites and frontal site showed larger MMNs than the central site in time interval 204-240 ms. It is possible that the latter time interval may have something to do with the N2b

component which is usually elicited largely at central sites and at later time intervals than the MMN.

4.2 Consonant Contrast

Four-way ANOVAs with experimental group (Auditory vs. Visual), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (204-240; 244-280; 284-320; 324-360; 364-400; 404-440; 444-480 ms) were carried out to examine each language group and each standard order separately. For the AE participants in the [miʃi]-as-standard condition, no differences were found. For the JP participants, a marginally significant interaction of Hemisphere x Time x Experimental Group [$F(12, 264) = 2.19$, $p = 0.06$] was observed. Figure 52 illustrates that the Auditory-Attend group showed a larger MMN particularly at the left sites in time intervals 324-360 (time 4), 404-440 (time 6) and 444-480 (time 7) ms. Two-way analyses examining hemisphere and experimental group for each time separately showed no significant differences.

For the comparison of AE Auditory- and Visual-Attend groups for the [miʃi]-as-standard condition, there were no significant findings. Comparing the JP participants in the Auditory- and Visual-Attend groups for the [miʃi]-as-standard conditions revealed Site x Experimental Group [$F(1, 12) = 9.74$, $p = 0.005$] and Hemisphere x Time x Experimental Group [$F(12, 264) = 3.99$, $p = 0.00001$] interactions. Figure 53 shows that the negativity is larger at frontal sites for the Visual-Attend group and at central sites for the Auditory-Attend group. The graphs displaying the three-way interaction show greater negativity for the Visual-Attend compared to Auditory-Attend group at left and midline sites in the later time intervals.

The normalized time bins which should be the largest MMN were averaged for [miʃi] and [miʃfi] orders and used to examine whether attention affected the MMN for the consonant contrast. Five-way ANOVAs with experimental group (Auditory vs. Visual), with language group (AE vs. JP), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (324-360; 364-400 ms) were carried out. An interaction of Site x Experimental Group [$F(1, 44) = 6.32, p = 0.02$] was observed (see Figure 54). The Auditory-Attend groups showed larger MMNs overall than in the Visual-Attend groups and Auditory-Attend groups showed larger MMNs at central sites than at frontal sites while Visual-Attend groups showed the opposite. This finding is supported by post-hoc test ($p = 0.000$). Unlike the vowel contrast, there was no interaction with Time, but the interaction with Site indicated that there was possibly an effect of the N2b component in the Auditory-Attend Experimental groups.

4.3 Vowel vs. Consonant Contrasts

The normalized time bins were used for this analysis as in the Auditory- and Visual-Attend Experiment. The two time bins which showed the largest MMN for each contrast-standard-order condition were compared to see whether there was a difference between contrasts for the two experimental conditions. This consisted of 204-240 and 244-280 ms for vowels and 324-360 and 364-400 ms for consonants. ANOVAs with experimental group (Auditory vs. Visual) and language group (AE vs. JP) as between group variables, contrast type (vowel vs. consonant), site (frontal vs. central), hemisphere (left vs. midline vs. right), and time (Time 1 and Time 2) were carried out. There was a main effect of contrast [$F(1, 44) = 33.35, p = 0.000$], indicating that the vowel contrast showed larger MMNs than the consonant contrast and there were significant interactions

of Contrast x Site x Experimental Group [$F(1, 44) = 5.44, p = 0.024$] (see Figure 55) and Contrast x Site x Time x Experimental Group [$F(1, 44) = 4.22, p = 0.05$] (see Figure 56). The vowel contrast elicited larger MMNs in the Auditory-Attend condition than in the Visual-Attend condition at both frontal and central sites. The consonant contrast also elicited larger MMNs in Auditory-Attend condition than in the Visual-Attend condition, but the central site showed a larger MMN for the Auditory-Attend group while frontal site showed a larger MMN for the Visual-Attend group, supported by post-hoc test ($p = 0.05$). In sum, MMN was larger for the vowel contrast than in the consonant contrast. Both vowel and consonant contrasts elicited larger MMNs in the Auditory-Attend condition than in the Visual-Attend condition.

4.4 Summary

With respect to the vowel duration contrast, the AE group for the [tado]-as standard condition in the Auditory-Attend experiment yielded larger MMNs than the AE group in the Visual-Attend experiment, whereas the JP groups showed only a small enhancement of the MMN at left central sites for the Auditory- compared to Visual-Attend Experiment. The JP group for the [taado]-as standard condition in the Visual-Attend experiment showed a slightly larger MMN than the JP Auditory-Attend experiment, whereas there was no difference in the AE group. When the two standard-order conditions were combined for the vowel contrast, elicited MMNs were larger for the AE group in the Auditory-Attend experiment than in the Visual-Attend experiment, while the JP group showed no consistent effect of attention task. For JP listeners, larger MMNs occurred in the Auditory-Attend experiment at earlier time intervals, while larger MMNs were elicited in the Visual-Attend experiment at later time intervals.

For the consonant contrast, there was no effect of attention manipulations for the AE listeners. In contrast, the JP listeners showed some enhancement of the MMN at left and midline sites for the consonant contrast with the short as standard and at central sites for the consonant contrast with the long as standard. When standard orders were combined the consonant contrast yielded a strong attention effect at central sites where the Auditory-Attend group showed a larger MMN than the Visual-Attention group for JP listeners.

In general, the vowel contrast produced larger MMNs than the consonant contrast in both attention conditions. One interesting finding was that for both language groups, the consonant contrast elicited a much larger MMN at central sites for the Auditory-Attend task than for the Visual-Attend task. These findings will be discussed more fully in the next chapter.

Chapter 5. General Discussion

A major goal of this dissertation was to examine whether native listeners show more automatic processing of phonologically distinctive phonetic contrasts cued by durational differences than do non-native listeners, for whom the duration differences, at most, serve a secondary role in distinguishing phonemic categories. To do this, attentional focus was manipulated in two experiments: Experiment 1) discrimination with focused attention and Experiment 2) pre-attentive, automatic discrimination were compared in JP listeners and naïve AE listeners using Japanese temporally-cued vowel and consonant contrasts. Within-category variability of exemplars of each phonetic category was a property of the stimuli used in the study to force discrimination at a categorial level. The ERP measure MMN was the principal means of examining this question. Less automaticity of processing was expected to be indexed by a smaller MMN. The main findings of these experiments addressing the question of automaticity are the following:

- 1) In the Auditory-Attend Experiment, the ERP data showed only small differences in MMN amplitude between AE and JP listeners, indicating an attenuated effect of language experience on these brain indices of discrimination. Both groups showed robust brain responses indicating discrimination of vowel and consonant duration contrasts. Thus, when listeners were attending to the auditory input in this categorial oddball task, the effects of language experience were minimized, although there were subtle differences in topography and lateralization of brain indices of discrimination across language groups. Subsequent behavioral testing in the same stimulus conditions revealed a significant advantage for native JP listeners, although both AE and JP groups performed above chance on both contrasts.

2) In the Visual-Attend Experiment, the non-native AE listeners showed smaller MMNs to the vowel contrast than did the JP listeners. The effect of language experience on MMNs to the consonant contrast was negligible. Again, behavioral tests of the contrasts showed better discrimination by JP participants than AE participants, especially for the consonant contrast.

3) The non-native AE listeners in the Visual-Attend Experiment produced smaller MMNs than the AE listeners in the Auditory-Attend Experiment for the vowel duration contrast. Thus, the relationship between attention and automaticity was seen in non-native listeners on this contrast. There was no overall effect of the attention manipulation on AE listeners' MMN indices of discrimination of the consonant contrast. Unexpectedly, the native JP groups showed an effect of attention, with the Auditory-Attend JP group showing larger MMNs than the Visual-Attend JP group on the consonant contrast.

These three results, taken together, suggest that the native (JP) listeners were more automatic in processing categorial vowel duration differences than the non-native (AE) listeners. Only subtle differences in the ERPs were found between language groups in processing the consonant contrast because both groups showed attenuated indices of discrimination in the Visual-Attend condition.

Some of the results are less clear in what they tell us about native versus non-native processing of duration contrasts. The AE group showed a slightly larger negativity than the JP group in the Auditory-Attend Experiment at left and midline sites in later time-intervals of the MMN (404-444 ms). In contrast, the behavioral button-press task indicated that the AE groups in both experiments were less accurate at identifying the

consonant duration contrast than the JP groups. It was also surprising to find that only the JP listeners showed group differences between the Visual- and Auditory-Attend experiments for the consonant contrast whereas the AE group did not. The JP group in the Auditory-Attend Experiment exhibited a larger negativity at left and midline sites than the JP group in the Visual-Attend Experiment for the short consonant as the standard. For the long consonant as the standard, the Auditory-Attend JP group showed greater central than frontal negativity and the Visual-Attend Group showed the opposite pattern.

Overall, the role of attention was seen in both experiments from both language groups. The AE group showed a significant difference between two experiments which indicated that attention was necessary to discriminate the non-native contrasts. The JP group showed a difference for the consonant duration contrast between the two experiments which indicated that the consonant contrast was harder to discriminate than the vowel contrast even for native listeners. Thus, some attentional resources may be necessary for harder contrasts even in the native language. The role of attention also interacted with language experience since the JP group in general showed larger MMNs than the AE group, even though some differences were small.

A secondary goal of the study was to explore whether consonant duration contrasts were more difficult than vowel duration contrasts for AE listeners because vowel, but not consonant duration, serves as a secondary phonemic cue in English. A smaller difference between contrast types, if any, was expected for JP listeners. The ERP results suggested that the consonant duration contrast was more difficult for both groups. For both language backgrounds and both attention tasks, the MMN was smaller for the consonant than the vowel duration contrast. The behavioral data also showed poorer

detection of duration differences in consonants than in vowels. The claim that vowel duration contrasts are easier than consonant duration contrasts is therefore supported by these results with the particular stimulus pairs tested. Since there has been no other direct comparisons between vowel and consonant duration contrasts in ERP studies in the previous literature it is hard to make a general claim about these contrast types. This might be clarified if the same task were conducted using different pairs, including those consonant and vowel pairs tested in Hisagi & Strange (under revision) that were generally easier for AE listeners to discriminate behaviorally.

Several unanticipated results were also observed. A late negativity (LN) peaking between 400 and 700 ms at frontal sites was found for both groups in the Auditory-Attend Experiment and for the [tado]-as-standard condition for the Visual-Attend Experiment. The LN was somewhat larger in the vowel duration conditions for the AE than the JP group in the Auditory-Attend Experiment. This LN, since it was affected by attention, will be considered further in the next section which addresses native versus non-native differences in automaticity of phonetic processing.

An asymmetry in the latency of the ERP and behavioral responses was seen between the different testing orders (short as standard versus long as standard). The long-as-standard led to later latency ERPs and longer RTs for both language groups. Similar asymmetries have been observed in other studies (Shafer et al., 2004; Friedrich et al., 2004). It is possible that the later latency of the MMN is due, at least in part, to refractoriness of neurons indexed by N1 (Naatanan, 1990). Specifically, the longer duration of input reduces the firing of these neurons so that there is less negativity at the N1 latency. This finding, although interesting, is not relevant to the dissertation's goals

and will not be addressed further. The findings related to the main goals of this study will be addressed further below.

5.1 Automaticity of Processing

The Automatic Selective Perception (ASP) (Strange & Shafer, 2006) model characterizes the differences between perception of L1 and L2 phonetic segments/contrasts and describes L1 speech processing in adults as reflecting highly overlearned, automatic selective perceptual routines. In this model, processing of L1 speech input to extract the phonetic message is automatic, due to years of L1 experience, while non-native listeners may require extra attention to discriminate phonetic contrasts that do not function to differentiate phonological categories in their L1. Whether extra attentional resources are needed is dependent on the relationship between the L1 and L2 phonology. The ASP model predicts the level of success in perceiving contrast differences by non-native listeners to be dependent on the degree of attention paid (i.e., cognitive effort) and the relationship of the non-native contrasts to the listener's L1 phonology (see Best, 1995, 2003). The finding of a smaller MMN in the AE listeners perception of the vowel contrast without auditory attention, compared to the listeners with auditory attention, and compared to the JP listeners in both tasks, supports the claim of the ASP model that non-native contrasts are processed less automatically than native contrasts. The absence of a difference, or difference in the opposite direction for the JP groups supports the claim that native listeners process these duration cues more automatically. As Näätänen (1988) suggested, there was an inverse relationship between automaticity and attention in AE performance.

The consonant-duration contrast did not show the predicted difference in MMN amplitude across attention conditions that would indicate differences in automaticity as a function of language experience. It appears that this consonant duration contrast was psycho-acoustically less salient than the vowel-duration contrast, which may have resulted in no group effect for one of two reasons. First, more power may be needed to see group differences with an MMN of the small magnitude seen for this consonant contrast. Second, it is possible that the harder consonant contrast requires some attentional resources even for native JP listeners, and the difficult visual attention task precluded these minimal resources being applied. Note that the JP group showed a larger MMN to the consonant duration contrast in the Auditory compared to Visual-Attend experiment. This supports the notion that few, if any, resources were available for automatic processing of the contrast in the Visual-Attend Experiment. As in the study by Szymanski et al. (1999) and Sussman et al., (2004), the position in the word of the phonetically contrasting segments may have affected the MMN magnitude. Both papers found smaller MMNs to changes in the second syllable than in the first syllable. In the present study, it is possible that a change in the first syllable, as in the case of vowel duration contrast, may have had an advantage over a change in the second syllable as in the case of consonant duration contrast. A study in which the vowel duration contrast occurred in the second syllable (e.g. obasan vs. obaasan) is needed to determine if this accounts for the differences found between contrasts here.

Previous studies have shown that difficult-to-discriminate contrasts show larger MMNs with than without attention (e.g., Szymanski et al., 1999). The counting task was a less accurate way of determining discrimination of the targets, so the apparent good

performance of the AE group here is suspect, given their poorer performance than the JP listeners on the button-press task. In this behavioral task, AE listeners missed 1/3 of the targets, on average, for the consonant contrast. Thus, attention may not have substantially improved their detection of consonant duration deviants.

One might expect the behavioral and the ERP results for the Auditory-Attend experiment to be highly related, since they both index discrimination with attention. In some ways they are. In general the MMN was larger for the vowel contrast and likewise the behavioral performance was better for the vowel contrast compared to the consonant contrast. Furthermore the latency differences dependent on which stimulus served as the standard can be seen for both the MMN and the RTs (see Shafer et al., 2004 for a similar result). However, the behavioral data showed group differences in performance and the ERP results for the Auditory-Attend Experiment showed only subtle group differences in topography.

It is possible that the shift of topography to the central sites indicates that attentional processes indexed by N2b are what led to increased negativity for the AE group. Both language groups generally showed increased central negativity in the Auditory-Attend Experiment, suggesting the presence of N2b. N2b indexes application of attentional resources to processing and is largest at central sites (e.g., Sussman et al., 2004). The larger LN for the AE group versus the JP group may also indicate that the AE listeners were using greater attentional resources to do the task. The functional significance of the LN is unclear at this time, but has been observed in several studies (Čeponienė, Cheour, & Näätänen, 1998; Shestakova et al. 2003; Zachau et al. 2005; Shafer, Kessler, Schwartz, Morr, & Kurtzberg, 2005b; Korpilahti, Lang & Aaltonen,

1995; Cheour, Korpilahti, Martynova & Lang, 2001; Hill, McArthur & Bishop, 2004). It may reflect a second MMN-type response, as suggested by Shafer, et al., (2005b) or index complexity of stimulus processing (Zachau et al. 2005). Recent studies (Datta, et al., in prep) reveal that it is modulated by attention. Three different conditions were conducted in this study: 1) ignore the auditory stimuli and watch a movie during the ERP recording; 2) attend to tone stimuli; 3) attend to speech stimuli. In particular, attention to target spectral cues (ba vs. da) in contrast to F0 (i.e., the third task) led to a larger LN to a non-target deviance difference of spectral cues (ι vs. ϵ). Under these interpretations of N2b and LN, the behavioral and ERP data agree in indicating greater difficulty for the AE than JP listeners in differentiating two non-native contrasts.

Another apparent discrepancy is that the largest MMN did not indicate the best behavioral performance. For example, the AE group in the Visual-Attend experiment showed quite good performance on the [taado]-as-standard task, similar to the JP group, but poorer performance on the [tado]-as-standard task. However, the significant difference in the MMN amplitude between the AE and JP groups was found for the [taado]-as-standard and not [tado]-as-standard condition. This may be explained by a factor of noise in the data. In particular, the figures for [tado]-as-standard condition suggest that more power would result in a significant difference between groups for this condition as well. Furthermore, the AE group in the Auditory-Attend experiment performed equally poorly on the button-press tasks for the two vowel-duration orders. These data partially support the claim by Tervaniemi et al. (2001) that MMN in pre-attentive processing reflects the subject's behavioral responses, but our data suggest that

this relationship is indirect. This indirect relationship has been argued for in other reports (e.g., Shafer, et al. 2005b).

5.2 Vowel vs. Consonant Duration Contrasts

We predicted that the vowel contrast would be more difficult than the consonant contrast for AE speakers because vowel duration serves as a secondary phonetic cue to vowel identity and final voicing contrasts in English. In addition, it was predicted that the advantage of vowel over consonant duration contrasts for AE listeners might be observed only at the pre-attentive level, since Hisagi & Strange (in revision) found no significant differences in accuracy between these contrast types as tested by a categorial discrimination task (92% for vowel contrasts and 88% for consonant contrasts summed over 5 pairs of each type, and 83% for [kado vs. kaado] and 78% for [niʃi vs. niʃʃi]). In the current study, discrimination of a single vowel contrast, as indexed by MMN, was superior to discrimination of a single consonant contrast, not only in the Visual-Attend condition, but also in the Auditory-Attend condition. Furthermore, this result applied to JP listeners as well, although AE listeners showed larger differences between the two contrasts than JP listeners did. The difference between vowels and consonants might be explained in one of three ways.

First, perceptual assimilation patterns of the contrasting segments may have differed across vowel and consonants (Best, 1995). That is, the vowels may have been assimilated in a Category Goodness (CG) pattern, while consonants were assimilated in Single Category (SC) pattern. Since perceptual similarity judgments were not obtained from AE participants, we cannot determine if these contrasts differ in relative perceptual similarity. In addition, while differing perceptual assimilation patterns may have led to

differences in difficulty for AE listeners, they could not explain why JP listeners also showed better ERP indices of discrimination of the vowel than consonant contrast, since both types of duration contrasts are phonemic in Japanese (Hirata, 2004a). The ratio of the short to long consonants was slightly less for the consonants than the vowels (1.54 versus 1.61), but not enough to seem likely to produce the large differences in MMN amplitude found here.

A second possible explanation is that the vowel duration contrast was psycho-acoustically more salient than the consonant contrast. Burnham (1986) suggested that more psycho-acoustically salient contrasts, such as those contrasting in a temporal voice cue, were learned earlier in L1 development, easier L2 contrasts to perceive, to improve by L2 training, and to maintain after training, whereas less psycho-acoustically salient contrasts, such as spectral place cues were later in development and harder to learn in L2, even with training. He did not make any predictions regarding consonant versus vowel duration contrasts, but this explanation can be extended to these contrasts. Specifically, the difference in performance between these two contrast types may be based in psycho-acoustic differences with vowel duration contrasts being more salient than consonant duration contrasts. Studies of discrimination of duration differences in periodic and aperiodic non-speech stimuli might also shed light on this hypothesis.

A third possible explanation is that, as in the study by Szymanski et al. (1999), the first syllable position might be easier to discriminate than the second syllable position. The target vowel duration contrast in this study was the first syllable while the target consonant duration contrast was at the second syllable. Because none of the JP words has

the geminate consonant at the first syllable position, a future studies will need to contrast duration located at the second syllable and this third explanation might be clarified.

The pattern of results in this study suggests that consonant-duration contrasts are less salient psycho-acoustically than vowel-duration contrasts. Both experiments showed that the vowel contrast was easier than the consonant contrast regardless of attentional condition or language experience. A study in which the ratio of the consonant and vowel duration differences was varied in steps (including a match) is necessary to have greater confidence in this claim.

5.3 Topography of Language Group Differences

The reason hemisphere was included as a factor in the analyses was that previous studies suggested that the greatest language-related differences might be seen at left hemisphere sites (e.g., Alho et al., 1998; Näätänen et al., 1997; Shestakova et al., 2002; Shtyrov et al., 1998, 2000; Shafer, et al., 2004; Binder et al., 2005). In the present study, we also found that the left sites showed the greatest differences, not only between language groups, but also between contrast types and between the two experimental conditions.

The [tado]-as-standard condition in the Auditory-Attend experiment showed the language group difference, but particularly on the left: the JP group showed greater MMN at the left sites. Both AE and JP groups for the [tado]- and [miʃfi]-as-standard conditions in the Auditory-Attend experiment showed larger MMNs in the central than frontal region but only on the left. For the LN, both [miʃi]- and [miʃfi]-as-standard conditions showed the largest LN at the left sites. In particular, the LN was largest at the frontal sites of the left hemisphere (i.e., F3). However, there was no left hemisphere bias

of the LN for the vowel contrasts in the Auditory-Attend experiment or in the [taado]-as-standard condition of the Visual-Attend experiment. There was a hemisphere effect in this latter case but the left hemisphere was not the important site.

With regard to attentional differences, for vowels, JP participants in the Auditory-Attend experiment showed a slightly larger MMN at the left-central site compared to those in the Visual-Attend experiment. For the consonant contrast, listeners in the Visual-Attend experiment showed larger MMNs compared to the Auditory-Attend group at left and midline sites in the later time intervals. In addition, the AE group for the [tado]-as-standard condition in the Auditory-Attend experiment elicited a larger MMN across left and right sites than the AE group in the Visual-Attend experiment whereas the JP groups showed only a small enhancement of the MMN at left central sites for the Auditory compared to Visual-Attend Experiment.

In summary, attention typically led to an increase in MMN and LN amplitude that was larger at the left than right sites. Thus, it appears that processes of attention interact more in the left hemisphere than at other sites. In addition, because the JP group showed more robust responses in the left hemisphere than the AE group, in general, it indicates that phonological processing is dominant in the left hemisphere.

Chapter 6. Conclusion

This study provided insight into how the participants discriminated sounds covertly and automatically, as indexed by brain measures, and was significant in several ways. First, the study elucidated whether L2 listeners are less automatic in processing non-native speech contrasts than L1 listeners, even if they show better than chance performance at the behavioral level. The results revealed that AE listeners were less automatic in processing a non-native vowel duration contrast than were native JP listeners. Second, it evaluated the relationship between L1 and L2 in predicting relative difficulties in discrimination. Specifically, temporal differences in short and long consonants should be more difficult than those in vowels for AE listeners, since vowel duration serves as a secondary cue in AE. The brain and behavioral responses to the consonant duration contrast suggested that processing of this contrast is more difficult for both groups than the vowel duration contrast. Why the consonant duration contrast is more difficult cannot be determined from this study. It may be due to psycho-acoustic factors or to linguistic factors (e.g., phonotactics). It is possible that the particular stimuli were responsible for the reported differences in difficulty. Future studies will be needed to clarify this question. Third, the study made a methodological contribution by including both inter- and intra-category variability in the stimulus set, that is to say, it measured categorial discrimination. To our knowledge, this is the first study to examine discrimination of temporal cues in non-native versus native listeners (specifically AE and JP) under different conditions of attention, using the ERP measure MMN. In addition, the performance of naïve AE listeners provided a baseline indicating the perceptual difficulty of these JP contrasts for learners.

The experiments showed a number of topographic differences in the ERP responses with the left sites often showing the greatest differences. Many of these topographic findings were not predicted, and thus replication in future studies will be necessary to have confidence in them. They also revealed an asymmetry in the latency of the ERP and behavioral responses to the different stimulus orders (short as standard versus long as standard). This asymmetry may be due to psycho-acoustic factors, but needs to be further investigated to reveal the cause.

TABLES

Table 1. Target Stimuli: real word models and the nonsense words that served as the experimental stimuli.

	contrasts	short	gloss	long	gloss
1	/a/ vs /aa/	kado	corner	kaado	card
2	/ʃ/ vs /ʃʃ/	nifi	west	niffi	diary/journal
3	/a/ vs /aa/	tado	non-word	taado	non-word
4	/ʃ/ vs /ʃʃ/	mifi	non-word	missi	non-word

Table 2. Summary of acoustic analysis of target nonsense stimulus materials: VOT and vowel consonant durations (ms) and pitch of two syllables (Hz) of selected four stimuli of each phonetic category. Standard deviation (SD) is shown for the contrasting target segments.

taado vs. tado

long	VOT	/aa/	closure (/d/)	/o/	Total	Pitch
1	11	148	23	82	264	402, 277
2	11	132	21	83	248	381, 271
3	11	142	19	79	251	390, 271
4	11	128	28	70	237	388, 297
Mean	11	138	23	78	250	390, 279

SD: 9.21

short	VOT	/a/	closure (/d/)	/o/	Total	Pitch
1	11	88	17	89	206	395, 346
2	14	84	17	100	214	382, 327
3	12	83	23	86	204	378, 286
4	12	88	15	87	202	398, 326
Mean	12	86	18	90	206	388, 321

SD: 2.85
L/S Ratios = 1.61

Total word duration
L/S Ratios = 1,21

miʃʃi vs. miʃi

long	Nasal murmur	/i/	/ʃʃ/	/i/	Total	Pitch
1	61	62	142	28	293	270, 294
2	63	74	128	41	305	271, 279
3	52	75	152	47	326	267, 269
4	71	64	128	29	292	262, 250
Mean	62	69	137	36	304	268, 273

SD: 11.90

short	Nasal murmur	/i/	/ʃ/	/i/	Total	Pitch
1	70	50	92	63	276	256, 275
2	73	49	95	70	287	258, 290
3	66	56	86	66	274	255, 295
4	58	61	84	60	262	253, 286
Mean	67	54	89	65	275	256, 287

SD: 5.26
L/S Ratios = 1.54

Total word duration
L/S Ratios = 1.11

Table 3. An example of the test procedures and time for each task. The orders of contrast type (vowel vs. consonant) and standard order (short-as-standard vs. long-as-standard) were counter-balanced.

	tasks	minutes
Step 1	Hearing screening	25
	Consent form	
	Language background questionnaire	
	Net Preparation	
Step 2	impedance check (1)	5
	Task FAM	2.5
	Condition 1A: vowel short as standard	30
Step 3	impedance check (2)	5
	Condition 1B: vowel long as standard	30
	Short break	10
Step 4	impedance check (3)	5
	Task FAM	2.5
	Condition 2A: consonant short as standard	30
Step 5	impedance check (4)	5
	Condition 2B: consonant long as standard	30
Step 6	offline behavioral	10
		190 (3 hours 10 min)

Table 4. Auditory-Attend Experiment: Behavioral data

The results of Counting auditory targets during ERP recording and Button-press auditory targets after the ERP recording showed the mean of the number correct, standard error of mean, median, mode and standard deviation. One of AE participant for *[missi]-as-standard condition* in Reaction time was not included due to that the participant missed the target.

Counting Auditory targets (14 = perfect score)

	JP	AE	JP	AE	JP	AE	JP	AE
	tado	tado	taado	taado	miſi	miſi	miſſi	miſſi
Mean	12.75	11	12.67	11.75	8.33	8.08	9.92	8.83
Std. Error of Mean	0.61	0.98	0.41	0.76	0.82	0.91	0.57	0.58
Median	14	12.5	13	12	8.5	8	9.5	8.5
Mode	14	13	14	14	9	9	9	11
Std. Deviation	2.09	3.38	1.44	2.63	2.84	3.15	1.98	1.99

Button-press Auditory targets (15 = perfect score)

	JP	AE	JP	AE	JP	AE	JP	AE
Behavior	tado	tado	taado	taado	miſi	miſi	miſſi	miſſi
Mean	14.5	11.58	14.5	12.5	13.17	9.75	13.83	8.08
Std. Error of Mean	0.20	0.85	0.15	0.95	0.21	0.86	0.37	1.37
Median	15	13	14.5	14	13	9	14	8.5
Mode	15	13	14(a)	15	13	8	15	3(a)
Std. Deviation	0.67	2.94	0.52	3.29	0.72	2.99	1.27	4.74

	JP	AE	JP	AE	JP	AE	JP	AE
Reaction Time	tado	tado	taado	taado	miſi	miſi	miſſi	miſſi
Mean	232.33	232.42	289.17	279.75	306.83	303.75	382.5	408.09
Std. Error of Mean	24.37	28.92	26.05	24.52	24.76	26.46	20.67	35.86
Median	242	184	283.5	269.5	289	314.5	378.5	437
Mode	72(a)	82(a)	148(a)	164(a)	289	156(a)	484	437
Std. Deviation	84.44	100.18	90.25	84.94	85.78	91.65	71.61	118.93

Table 5. Visual-Attend Experiment: Behavioral data

The results of Counting visual targets during ERP recording and Button-press auditory targets after the ERP recording showed the mean of the number correct, standard error of mean, median, mode and standard deviation.

Counting Visual targets (14 = perfect score)

	JP	AE	JP	AE	JP	AE	JP	AE
	tado	tado	taado	taado	mifi	mifi	missfi	missfi
Mean	13.58	11.92	13.58	12.5	12.5	11	10.83	10.08
Std. Error of Mean	0.15	0.68	0.34	0.70	0.61	0.75	0.82	0.97
Median	14	12.5	14	14	13.5	11	11	11
Mode	14	14	14	14	14	11(a)	10	6(a)
Std. Deviation	0.52	2.35	1.17	2.43	2.11	2.59	2.82	3.34

Button-press Auditory targets (15 = perfect score)

	JP	AE	JP	AE	JP	AE	JP	AE
Behavior	tado	tado	taado	taado	mifi	mifi	missfi	missfi
Mean	14.42	11.58	14.33	14.33	12.42	8.25	12.92	9.5
Std. Error of Mean	0.15	0.86	0.31	0.28	0.29	1.24	0.34	1.13
Median	14	11.5	15	15	12	8.5	13	10
Mode	14	15	15	15	12	1(a)	12	5(a)
Std. Deviation	0.52	2.97	1.07	0.99	1.00	4.29	1.17	3.90

	JP	AE	JP	AE	JP	AE	JP	AE
Reaction Time	tado	tado	taado	taado	mifi	mifi	missfi	missfi
Mean (ms)	229.33	227.83	298.75	284.17	316.5	320.83	415.5	397.58
Std. Error of Mean	27.66	26.38	22.51	25.77	23.68	25.74	31.87	26.69
Median	228.5	198	295	269.5	325.5	296	397.5	365.5
Mode	114(a)	151	196(a)	164(a)	133(a)	176(a)	315	296(a)
Std. Deviation	95.80	91.40	77.96	89.25	82.02	89.18	110.40	92.47

FIGURES

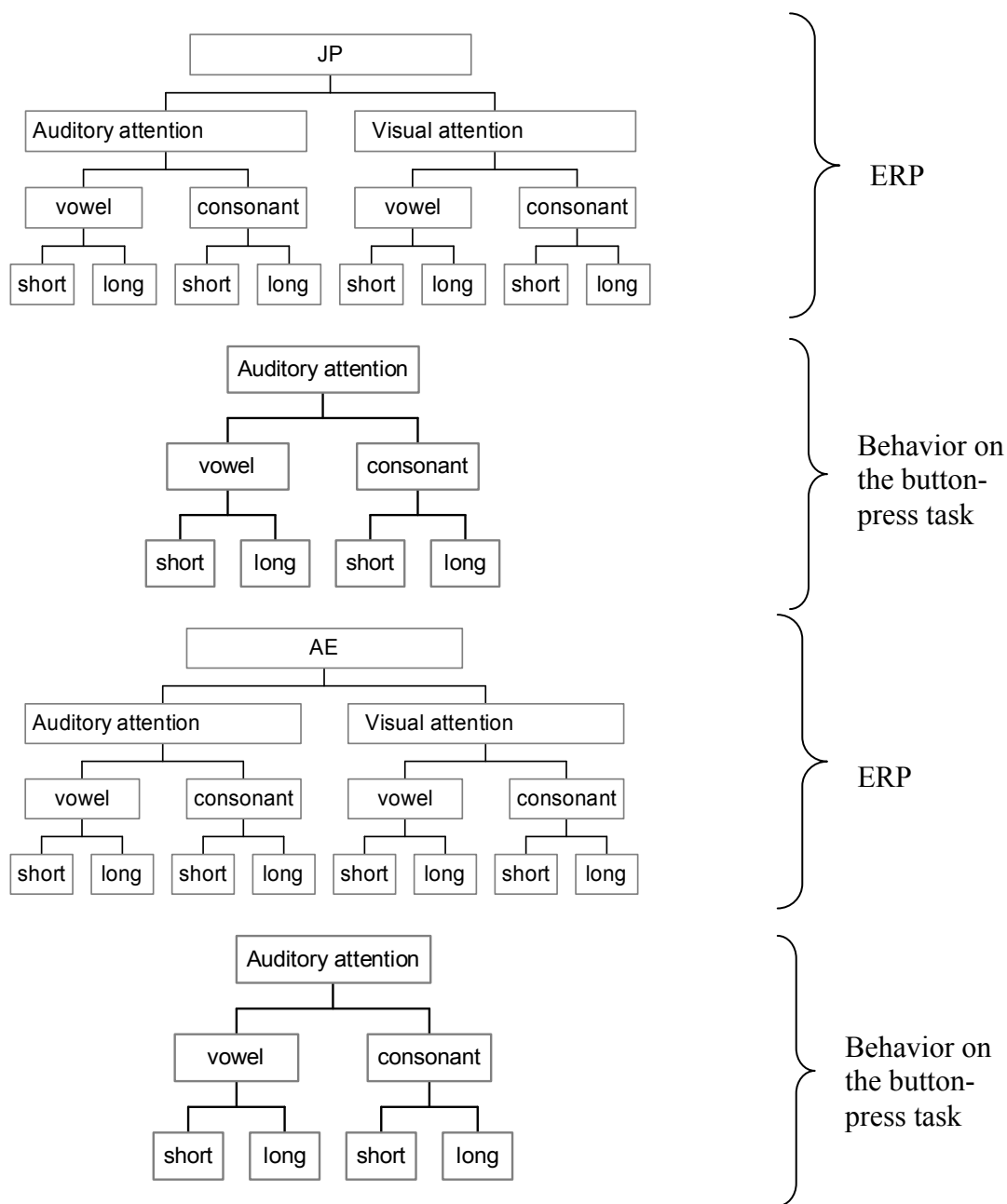


Figure 1. Design of Study. The figure describes the organization of the study. There were two different experiments, auditory- and visual-attention experiments. This is a between-subject group design. Half of the AE and JP groups participated in the Auditory-Attend Experiment and the other half participated in the Visual-Attend Experiment. All participants had the button-press behavioral task after the ERP experiment. Each experiment had the vowel and consonant contrasts; each contrast had two conditions: the short segment as-standard and the long segment as-standard.

Study 1: Auditory-attention

vowel & consonant

One category (standard) occurs on **85%** of the trials
and the other (deviant) occurs on **15%** of the trials

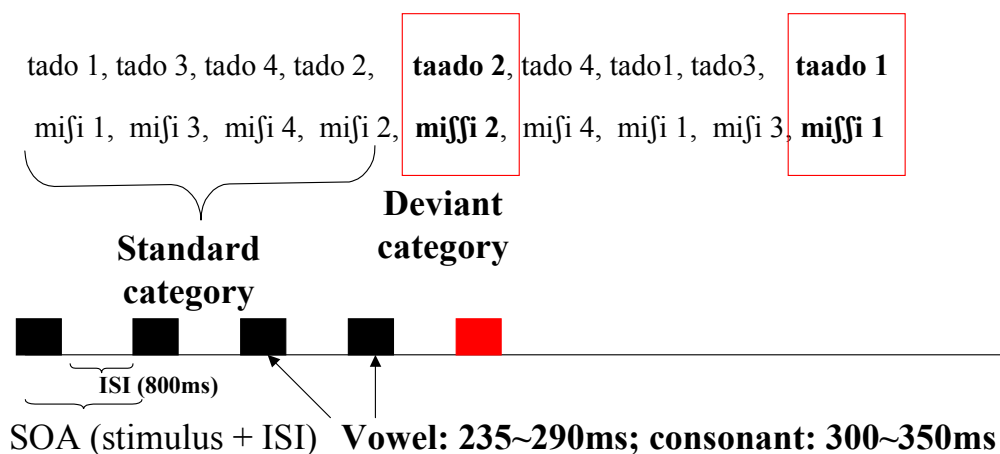
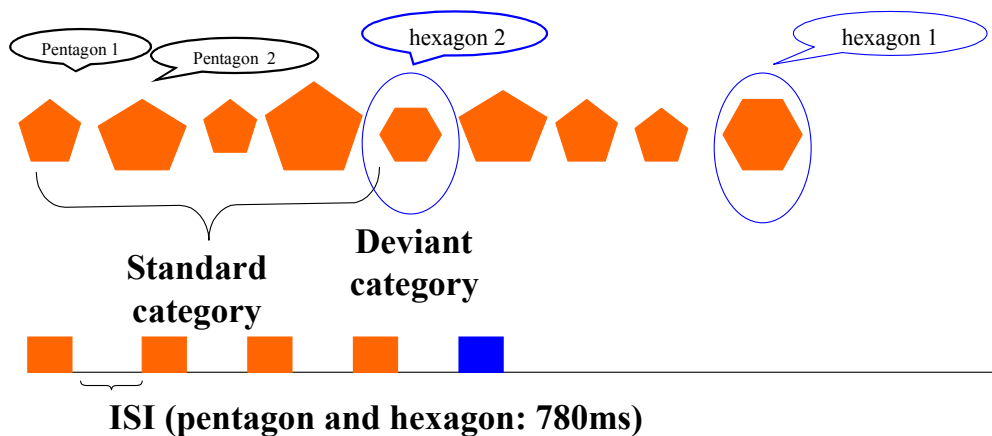


Figure 2. Categorical oddball discrimination paradigm for Auditory-Attend Experiment.

Study 2: Visual-attention

With vowel

One category (standard) occurs on **85%** of the trials
and the other (deviant) occurs on **15%** of the trials



With consonant

One category (standard) occurs on **85%** of the trials
and the other (deviant) occurs on **15%** of the trials

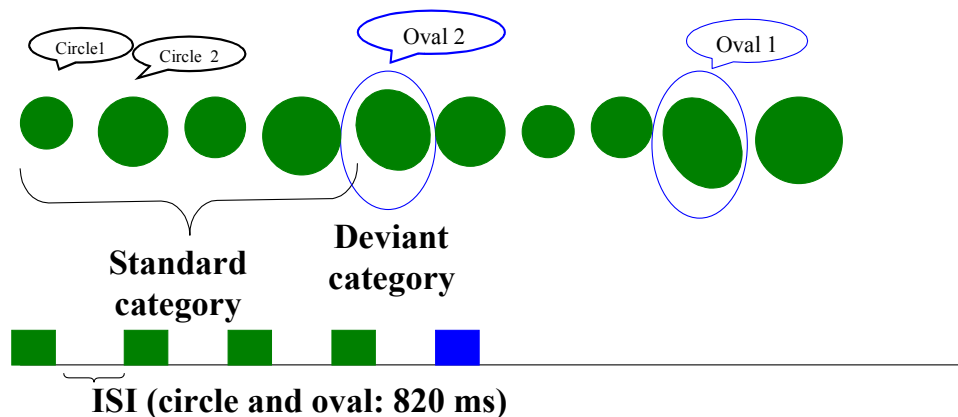


Figure 3. Categorical oddball discrimination paradigm for Visual-Attend Experiment: top describes the paradigm used with vowel stimuli and the bottom those used with consonant stimuli.

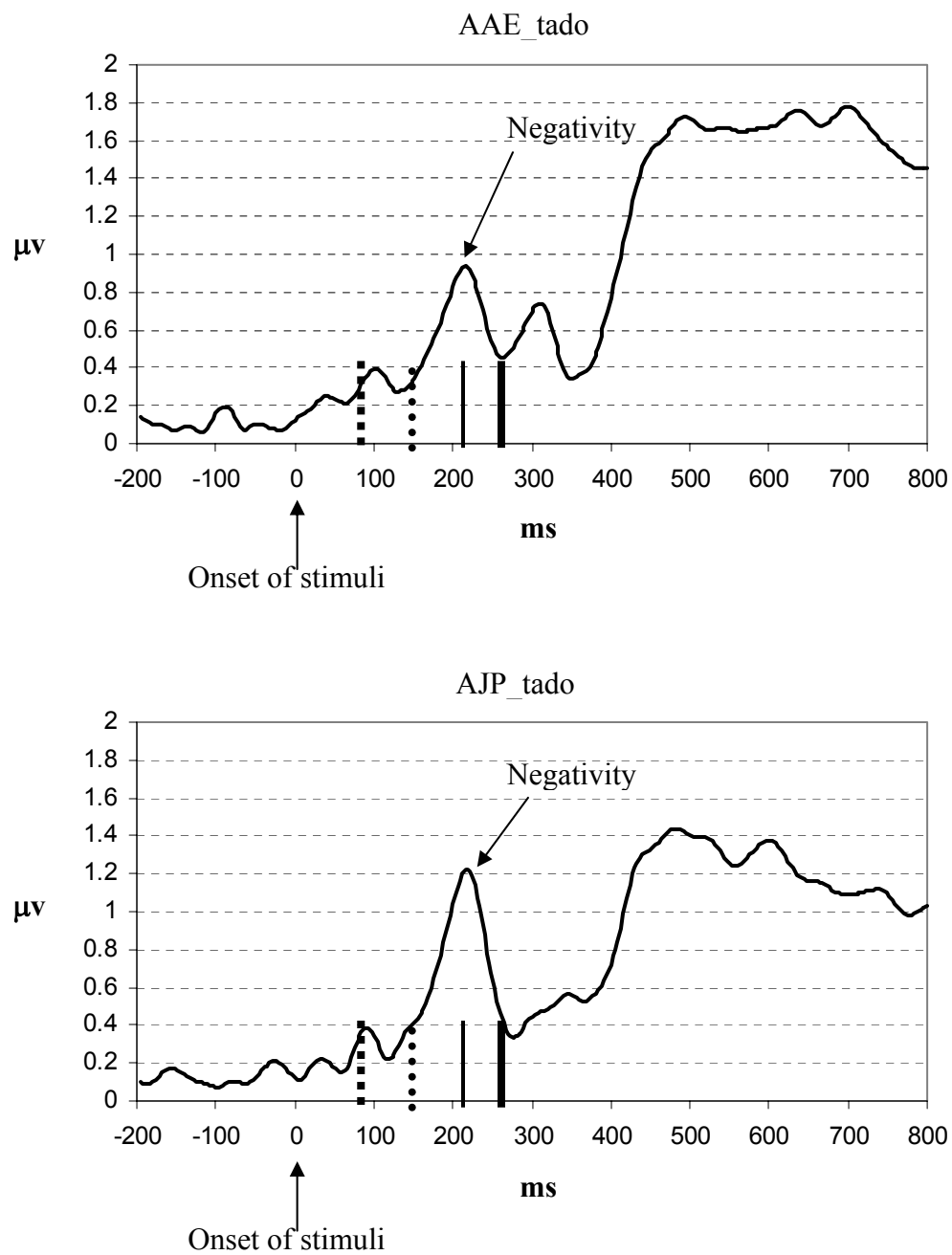


Figure 4a. [tado]-as-standard condition

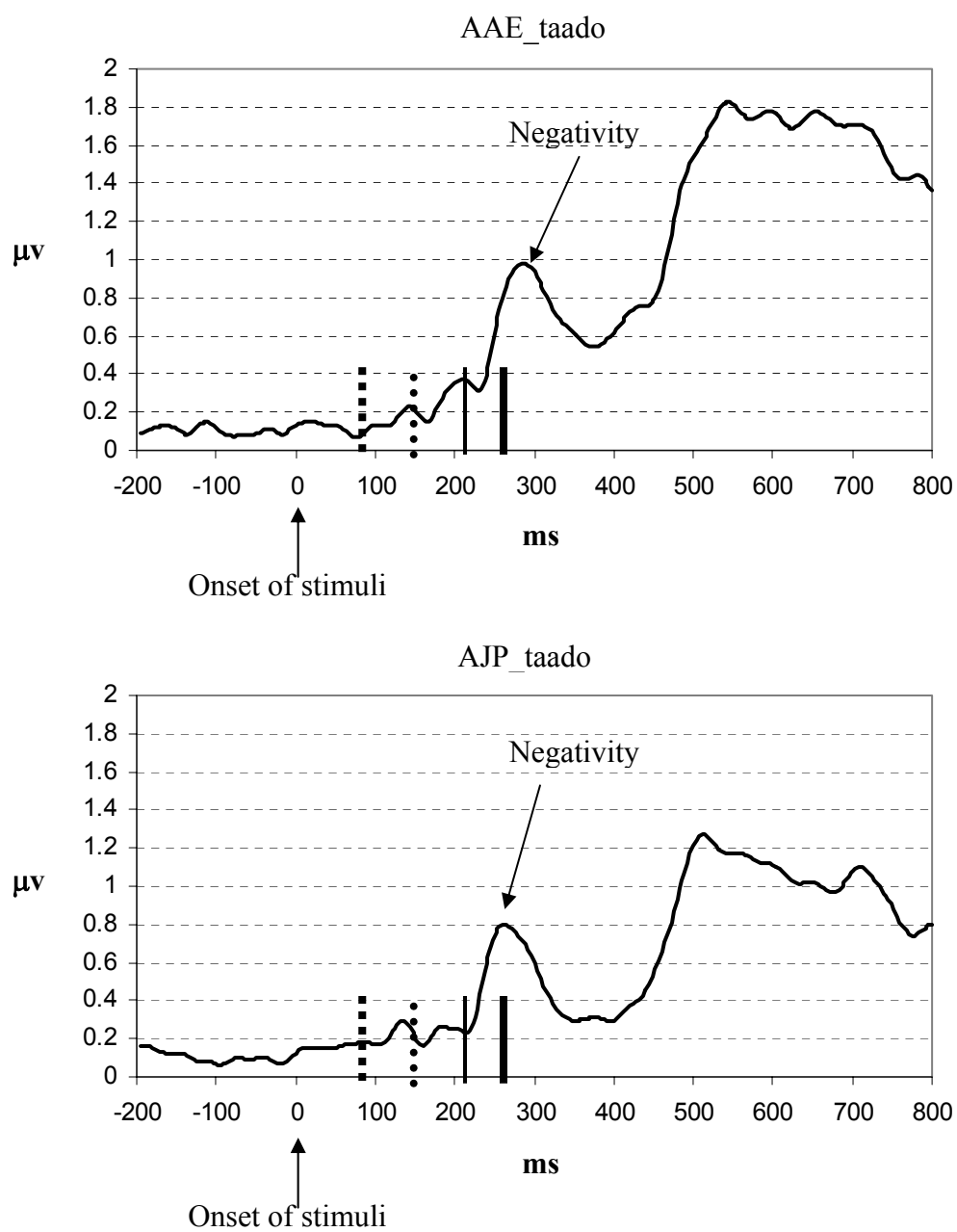
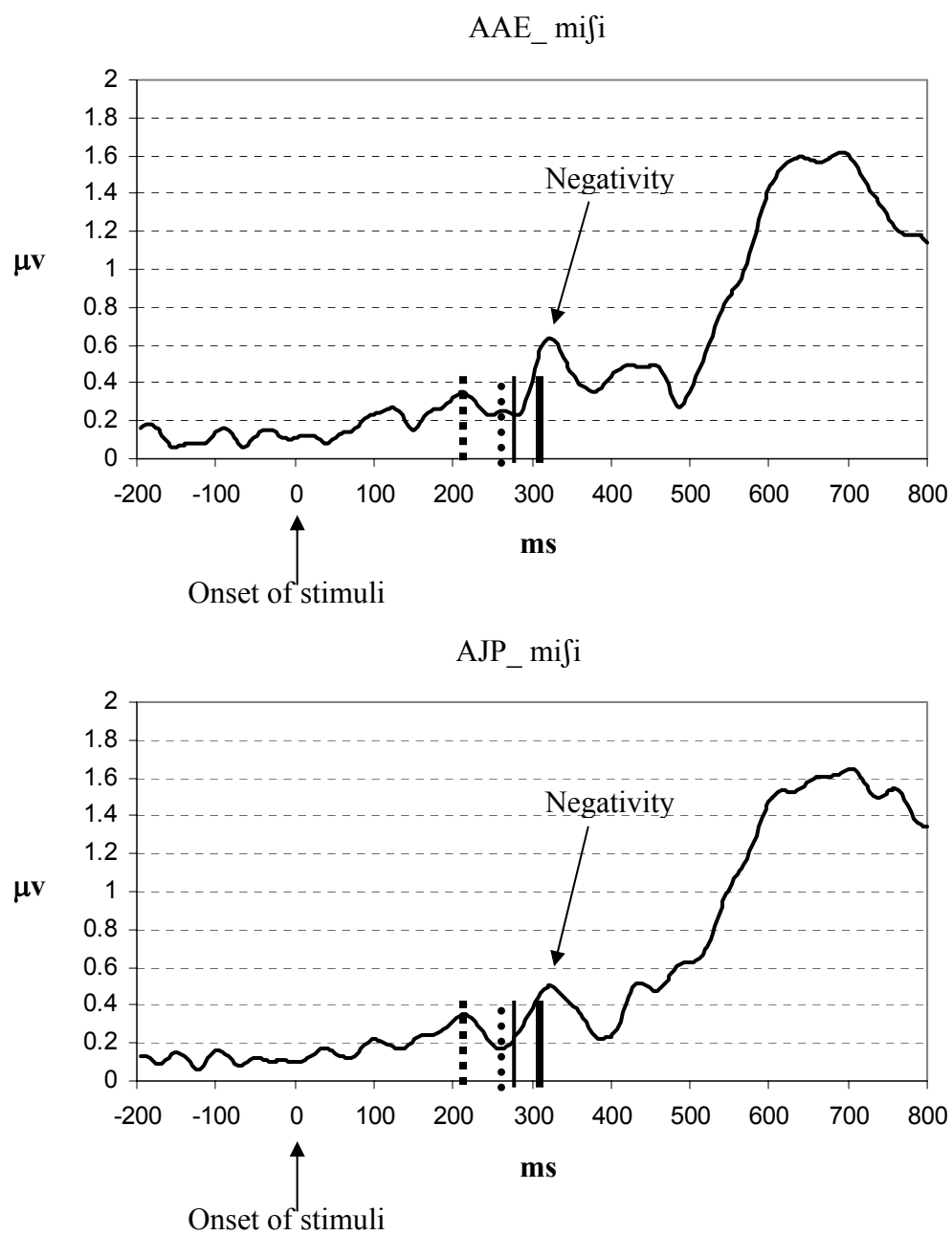


Figure 4b. [taado]-as-standard condition

Figure 4c. [mi*i*]-as-standard condition

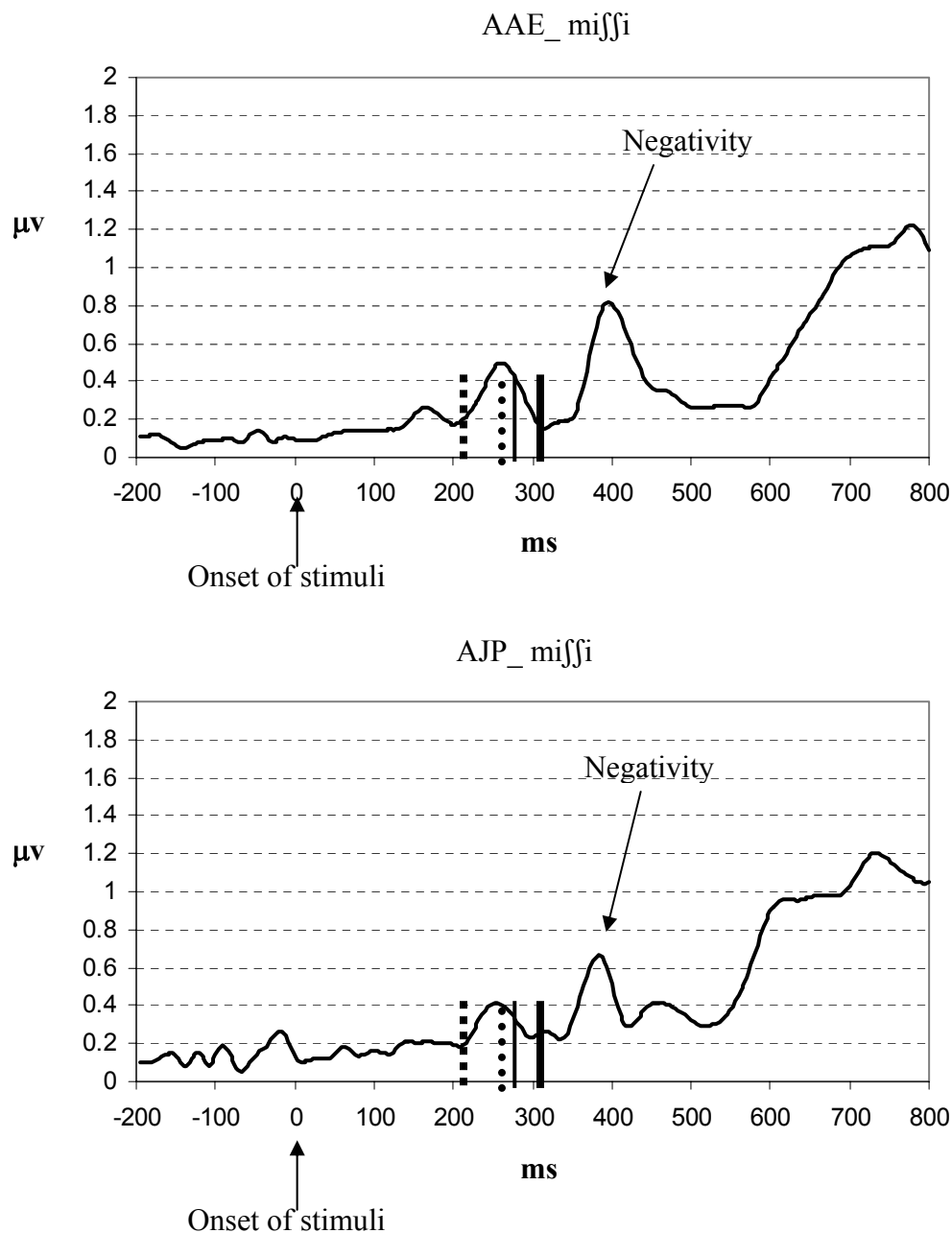


Figure 4d. [miɸfi]-as-standard condition

Figure 4 (a-d). Global Field Power (GFP) for all conditions in the Auditory-Attend Experiment. GFP shows the strongest energy around 200-220 ms for the vowel and 300-400 for the consonant, which is consistent with MMN latencies. The offset of the segments are shown by dotted or solid lines, with Short vowel offset: Long vowel offset: Short word length: Long word length:

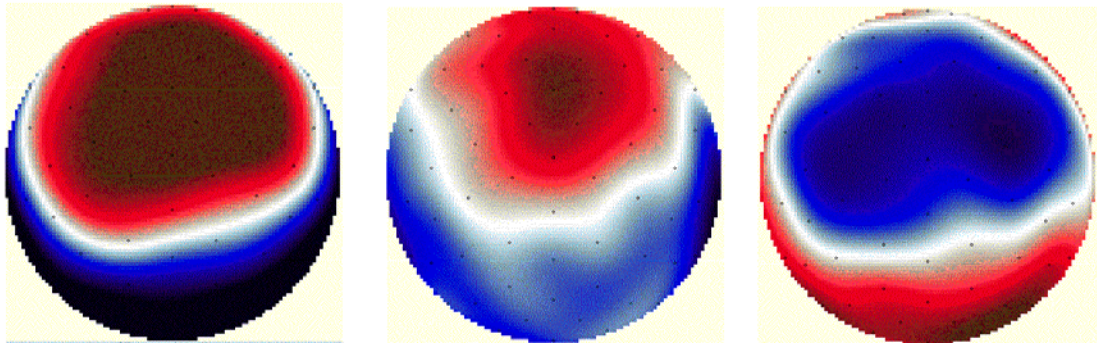


Figure 5. AAE [tado] topographical voltage maps of the standard, deviant, and MMN (from left to right) at 220 ms. Blue shows negativities amplitude and red shows positive amplitudes. The MMN is largest at frontal central sites as anticipated.

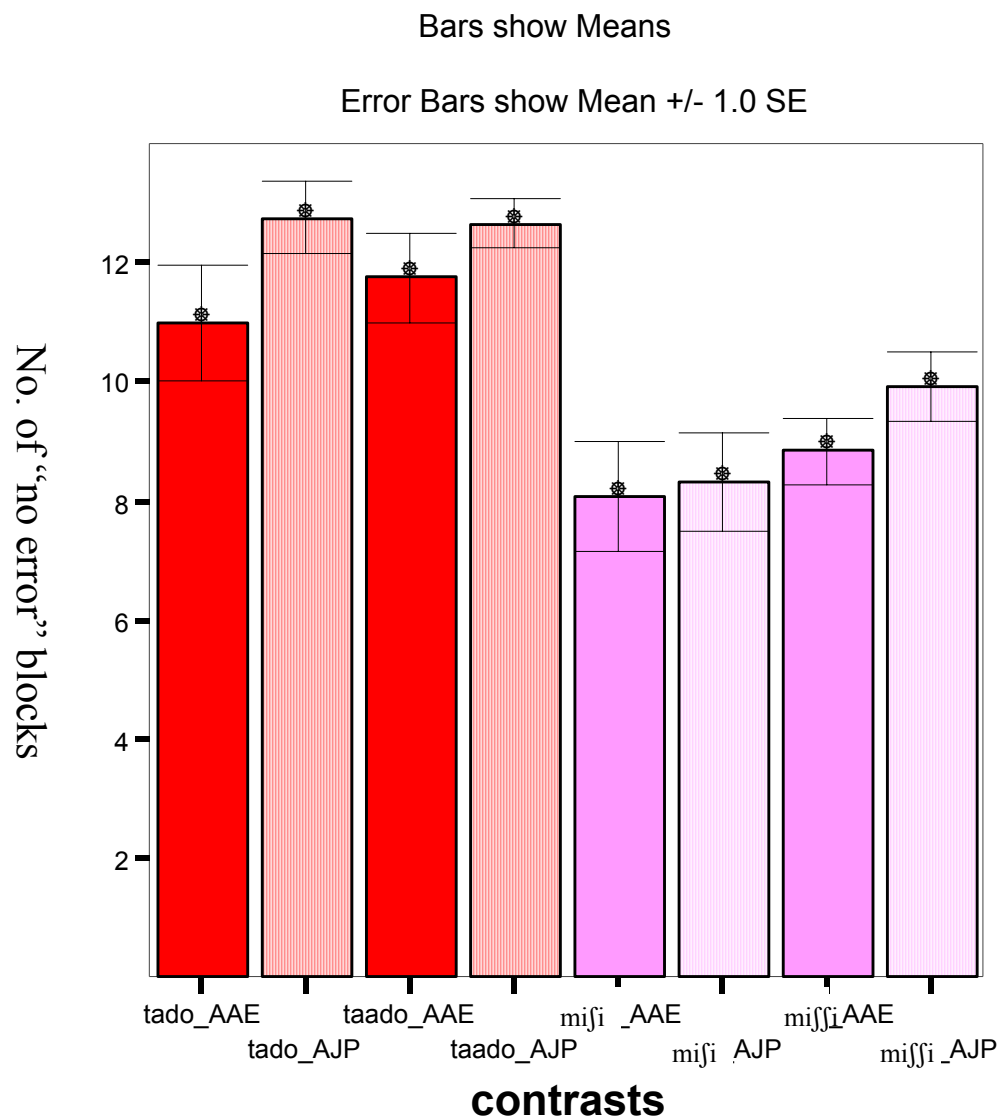


Figure 6. Counting of auditory deviants in the Auditory-Attend Experiment (14 blocks): the numbers of “no error blocks” per condition are described. Solid bars are AE and striped bars are JP listeners. Left two pairs refer to vowel contrasts and right two pairs for consonant contrasts. No language group differences were significant.

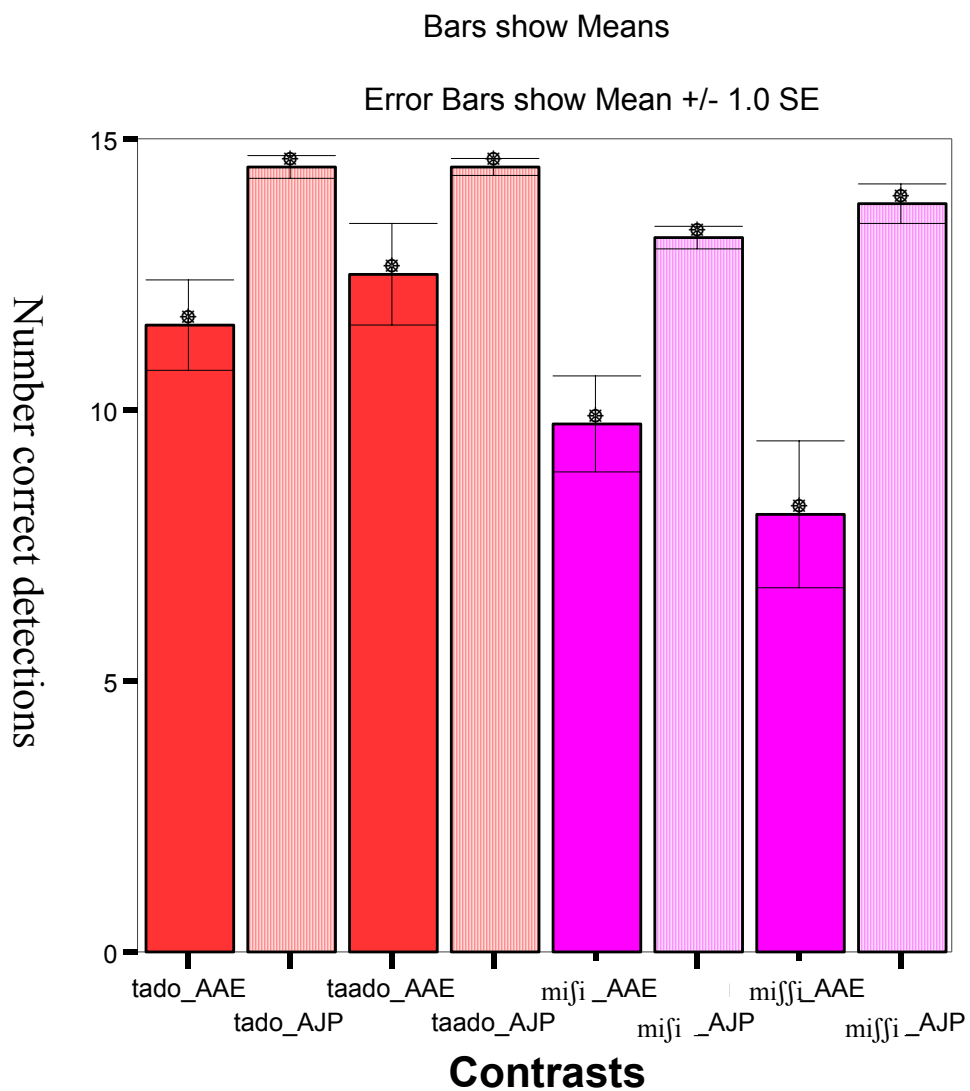


Figure 7. Button-press behavioral data in the auditory task for the Auditory-Attend Experiment (15 possible): solid bars are AE and striped bars are JP listeners. Left two pairs show performance for vowel contrasts and right two pairs show performance for consonants contrasts. There were significant language group differences on all contrast/deviant orders.

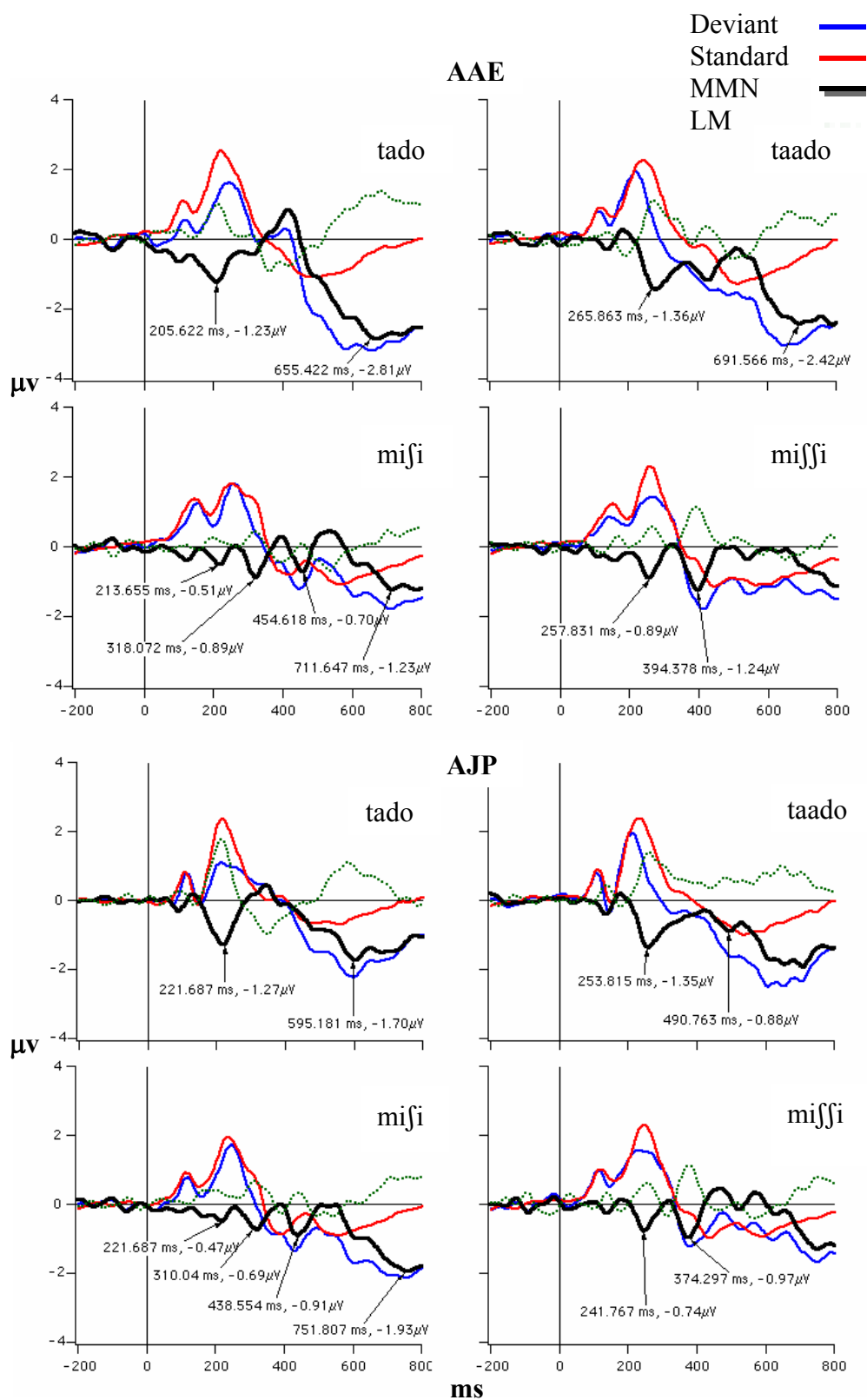


Figure 8. Auditory-Attend Grand Average data (Fz): the grand average at the mid-frontal site (Fz) for the standard, deviant and deviant minus standard and at LM for the deviant minus standard for all four contrasts and for AE (top a) and JP (bottom b) are shown above. Latency and amplitude are shown for peak negativities of the subtraction waves.

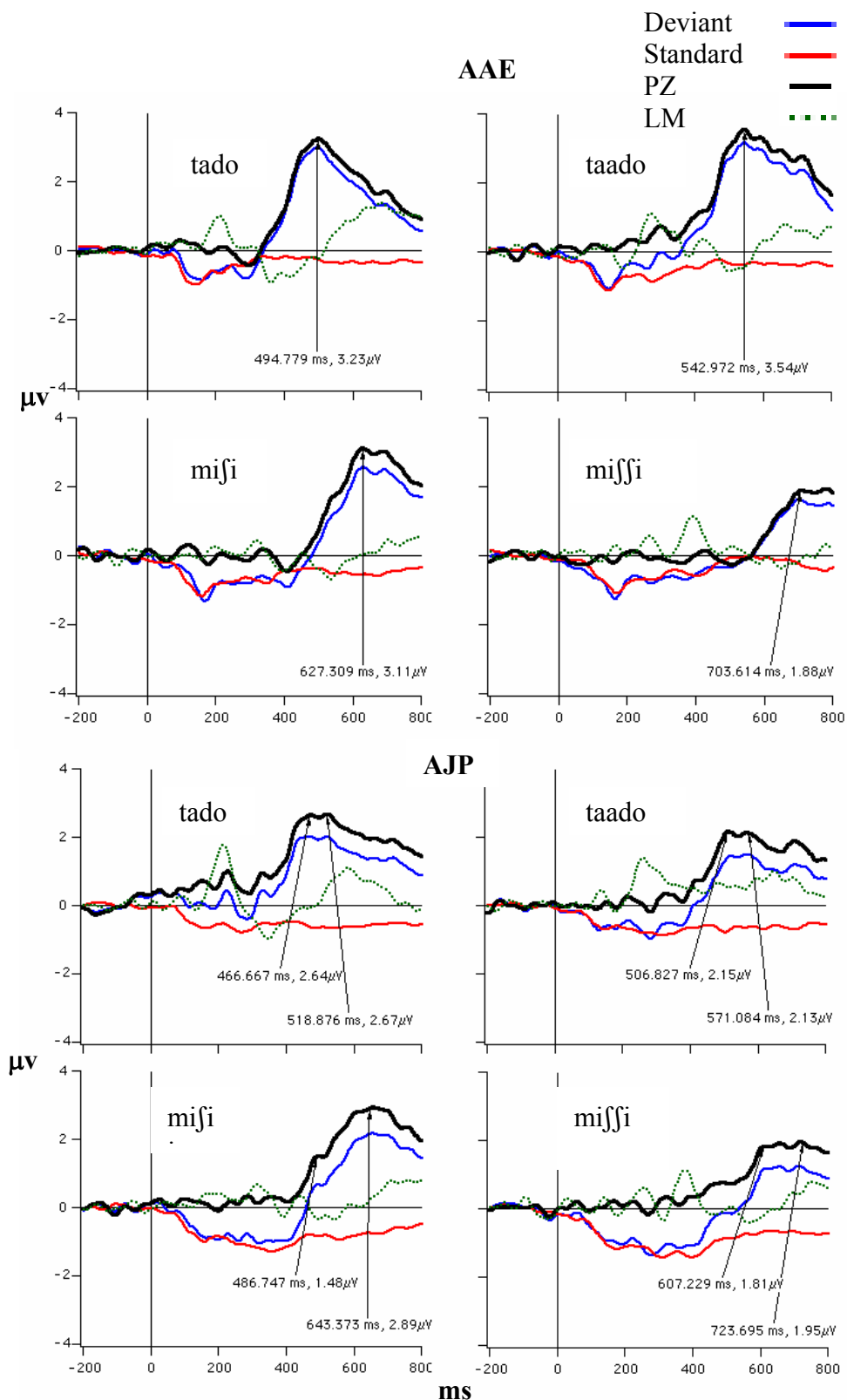


Figure 9. Auditory-Attend Grand Average data (Pz): the grand average at the mid-posterior site (Pz) for the standard, deviant and deviant minus standard and at LM for the deviant minus standard for all four contrasts and for AE (top a) and JP (bottom b) are shown above. Latency and amplitude are shown for peak positivities of the subtraction waves.

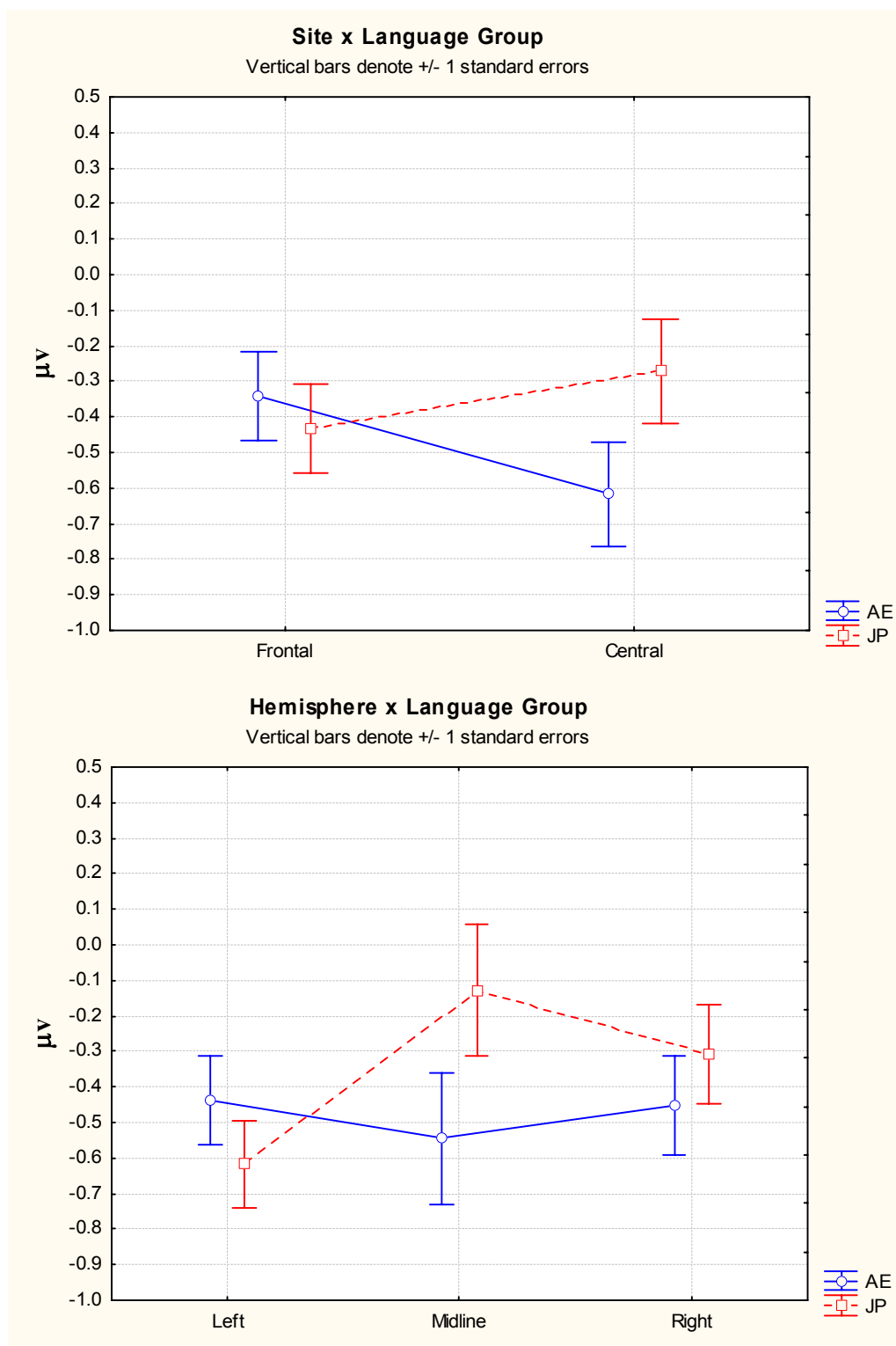


Figure 10. [tado] in the Auditory-Attend Experiment - Language Group effect: the AE group showed the greatest negativity at central sites and the JP group at frontal sites (top). The JP group showed greater negativity at left than midline and right sites, whereas the AE group showed no hemisphere difference (bottom).

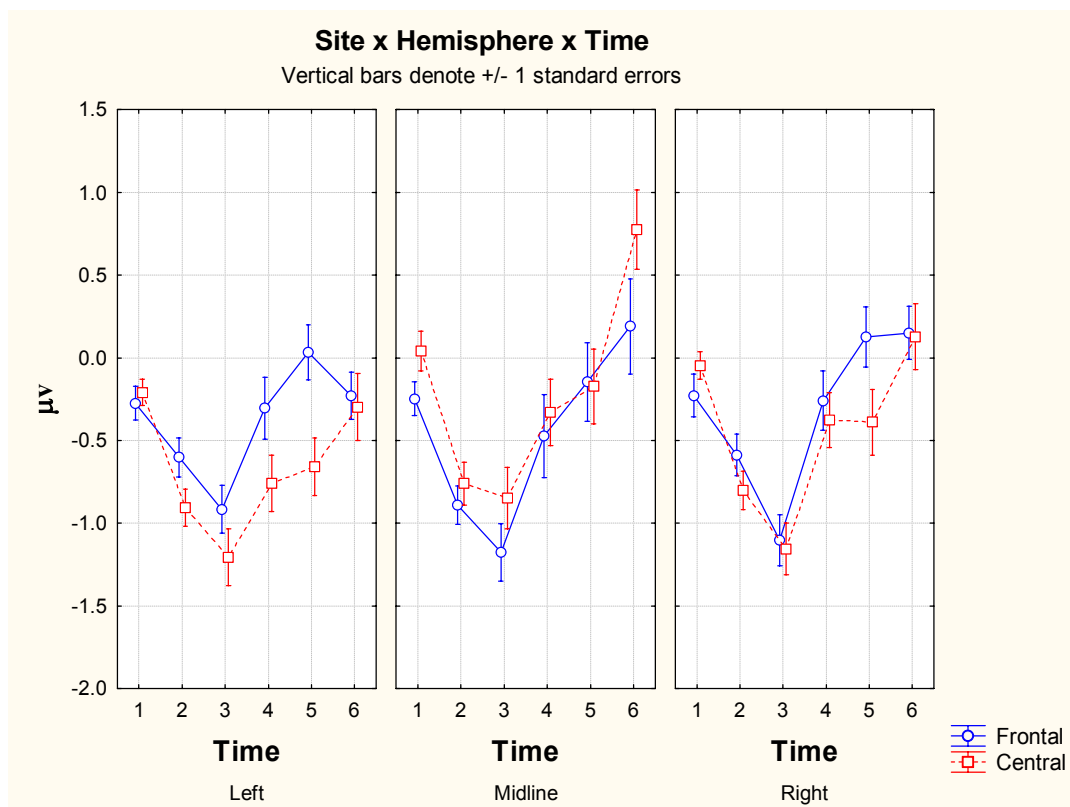


Figure 11. [tado] in the Auditory-Attend Experiment – Site and Hemisphere effect: the negativity is greatest at left-central and midline-frontal sites.

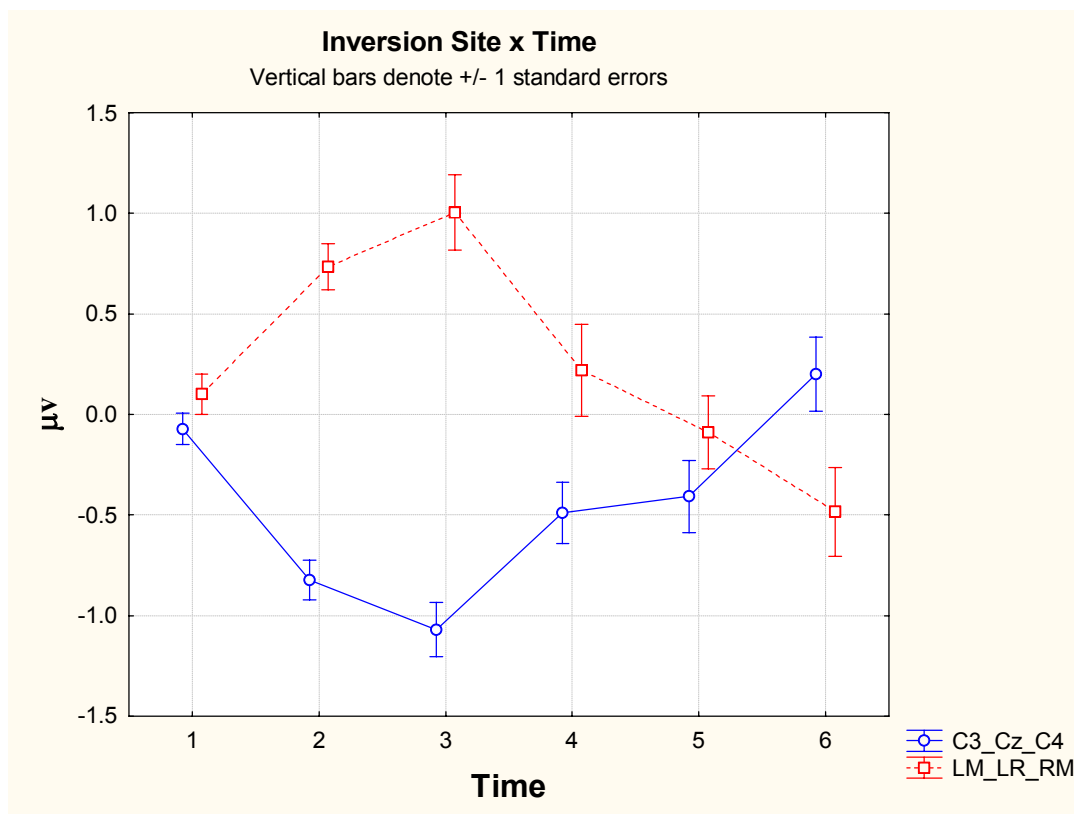


Figure 12. [tado] in the Auditory-Attend Experiment – Inversion effect: the MMN extends from Time interval 3-5 (204-320 ms). The small negativity on Time interval 4 (244-280 ms) and 5 (284-320 ms) may be N2b because it does not differ for central and mastoid sites.

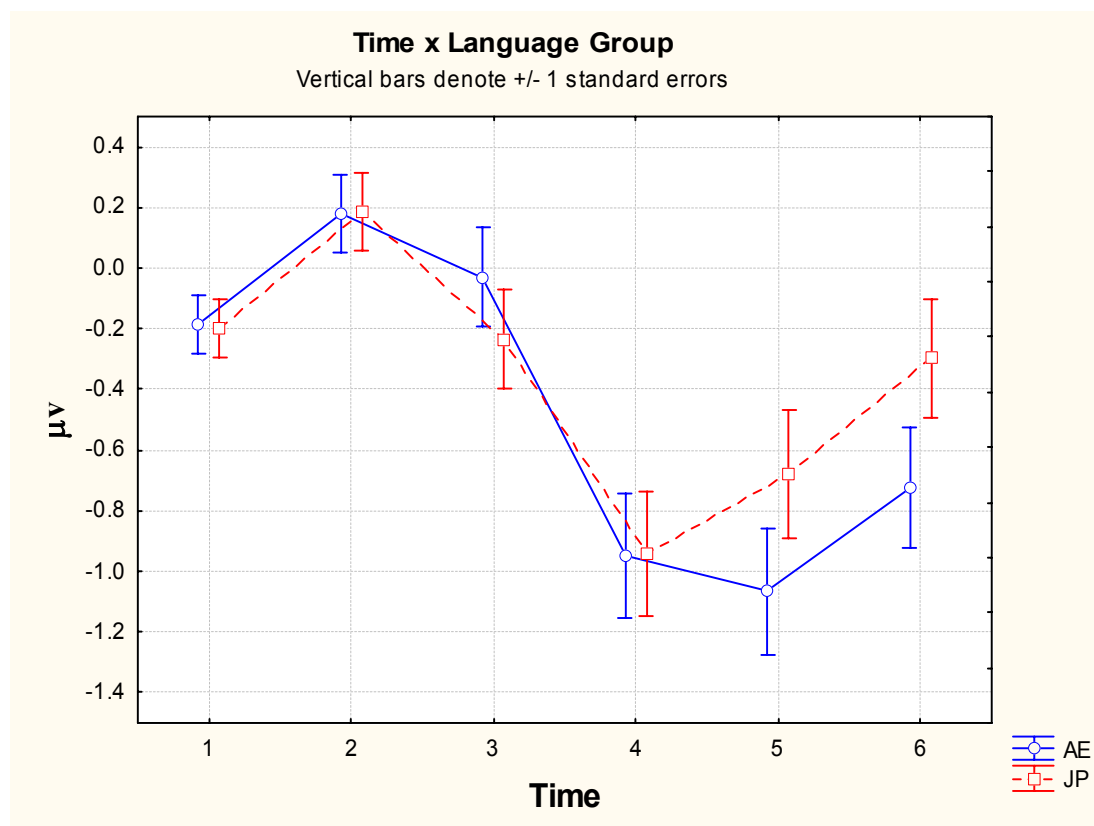


Figure 13. [taado] in the Auditory-Attend Experiment-Language Group effect: the AE group showed a slightly larger MMN than the JP group in Time interval 5 (284-320 ms) and Time interval 6 (324-360 ms). Voltages summed over all six sites.

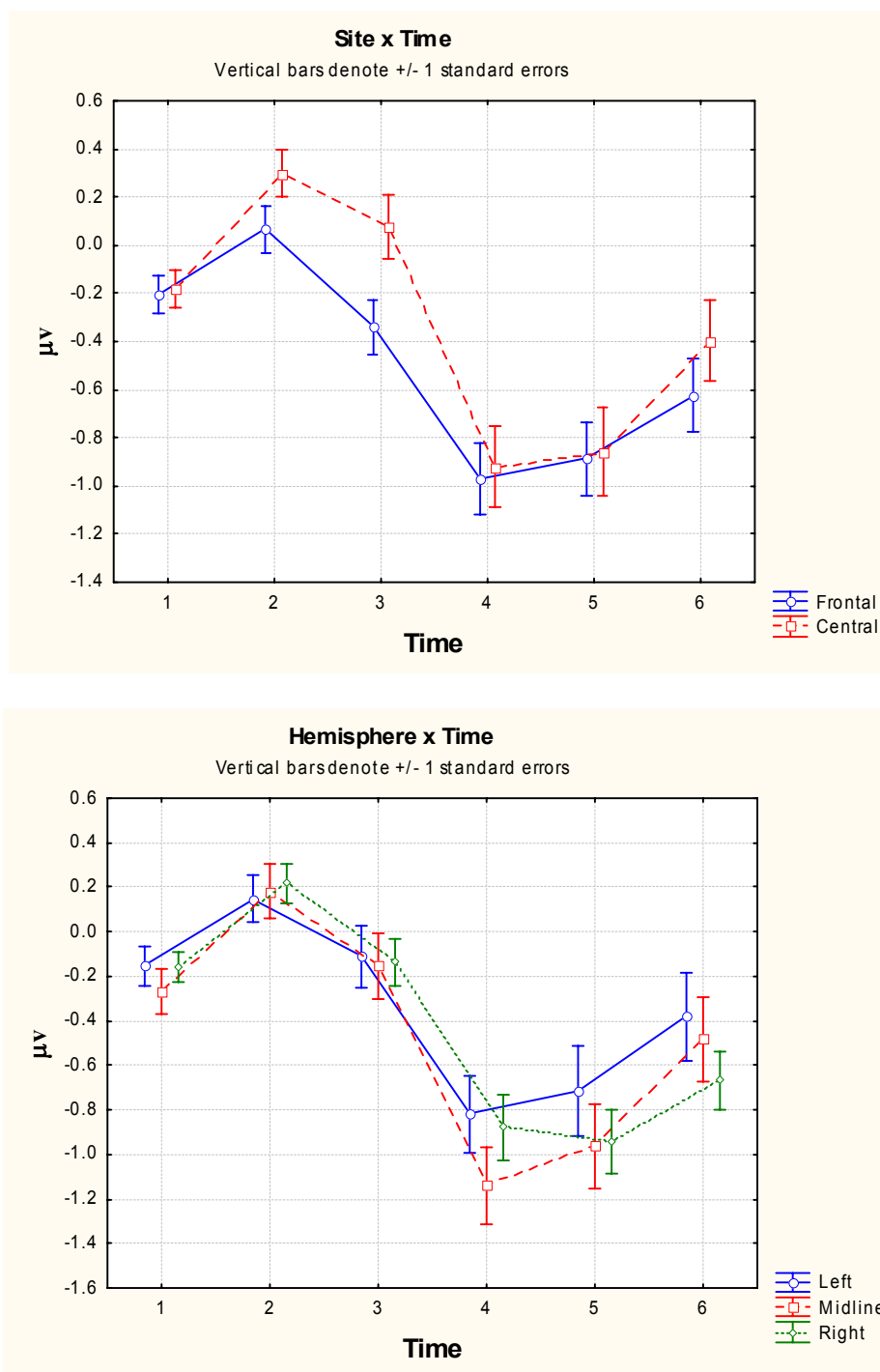


Figure 14. [taado] in the Auditory-Attend Experiment – Site and Hemisphere effect: there was a significantly larger MMN for the frontal compared to the central sites in Time interval 3 (204-240 ms). The midline in Time interval 4 (244-280 ms) showed a larger MMN than the left and right hemispheres.

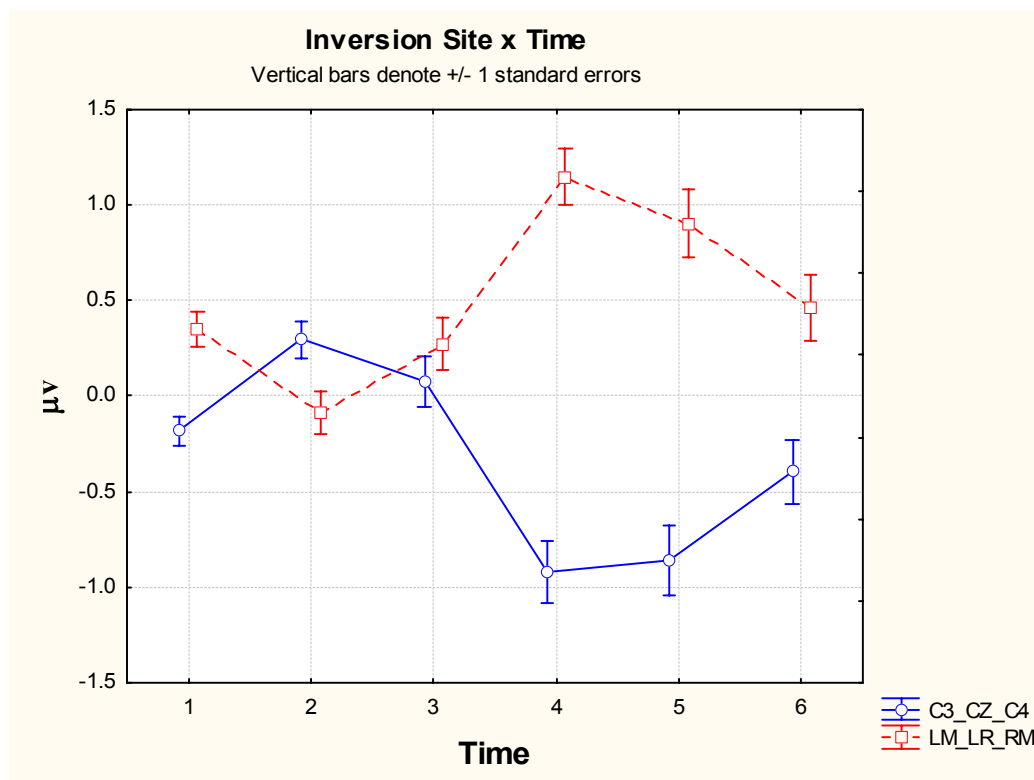


Figure 15. [taado] for the Auditory-Attend Experiment – Inversion effect: post-hoc tests showed that the mastoids were significantly more positive than the central sites in Time intervals 4 (244-280 ms), 5 (284-320 ms) and 6 (324-360 ms). There was no evidence of N2b.

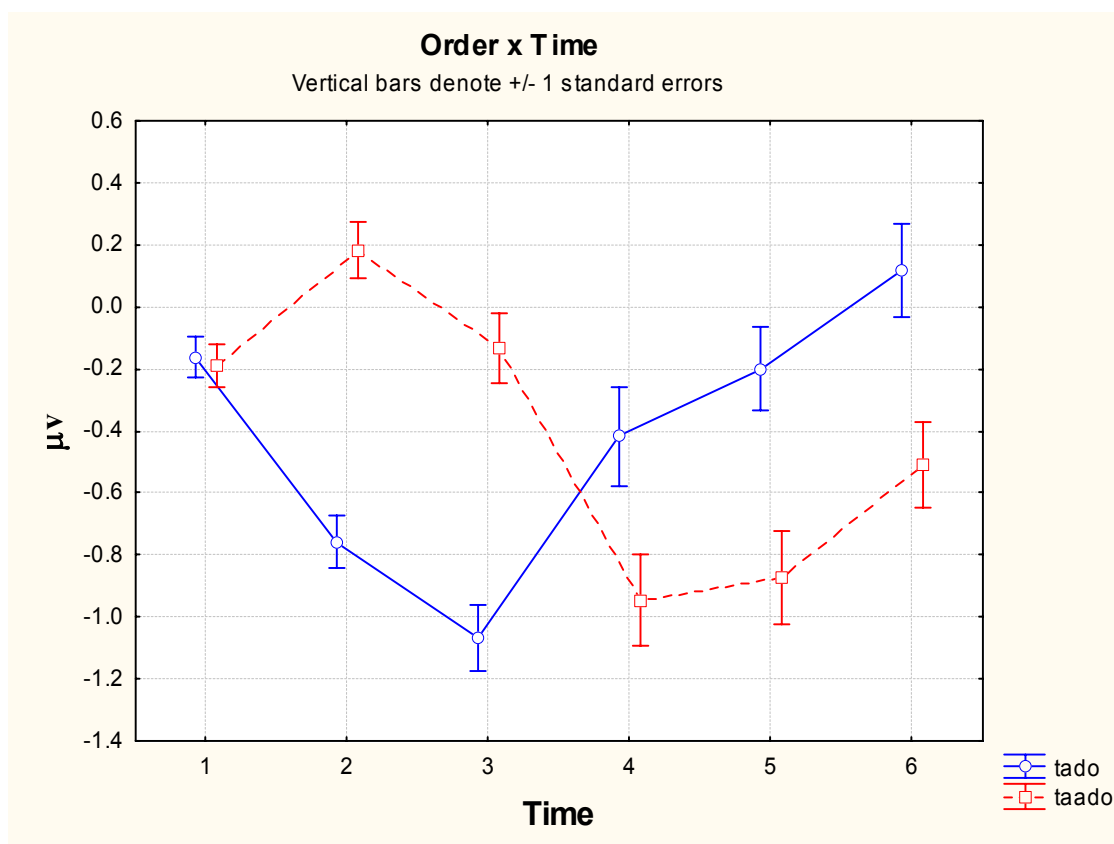


Figure 16. [tado vs. taado] in the Auditory-Attend Experiment – Order effect: the MMN to the [tado]-as-standard condition showed an earlier latency than the MMN to the [taado]-as-standard condition. Voltages summed over all six sites.

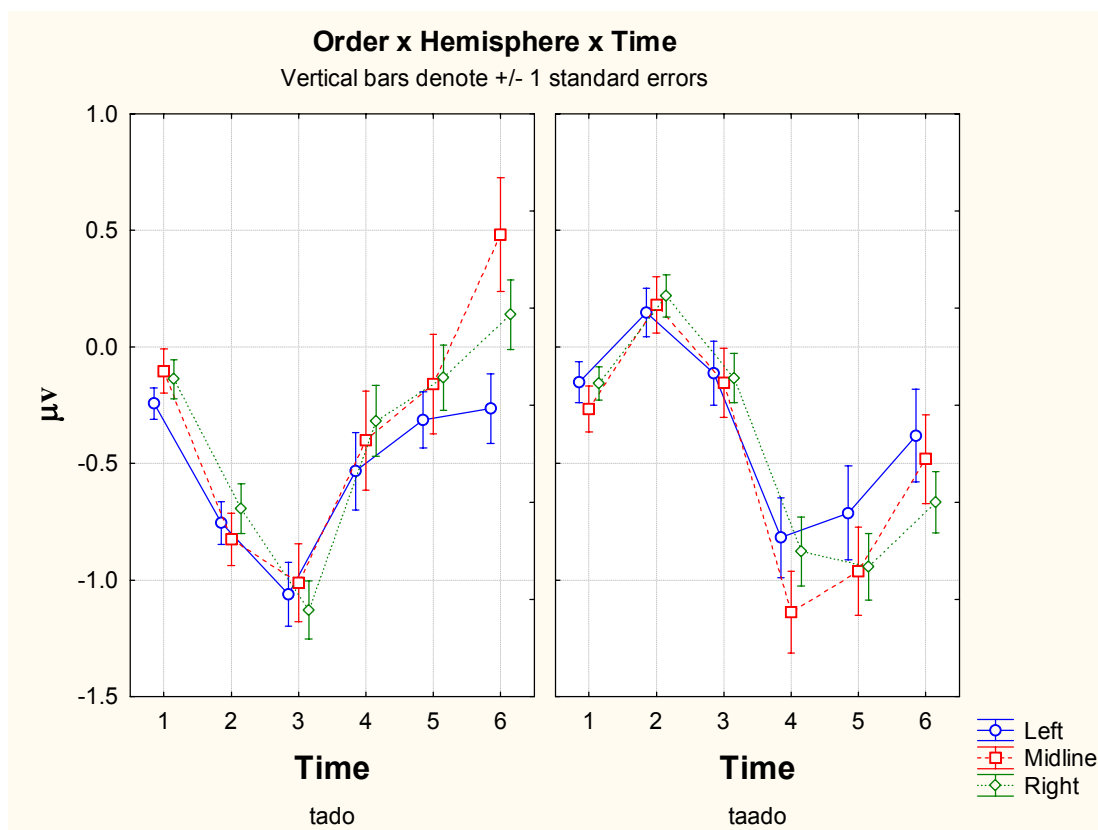


Figure 17. [tado vs. taado] in the Auditory-Attend Experiment – Hemisphere effect: this difference between standard orders is primarily one of time. Specifically, the only clear hemisphere difference is in the final time interval where the negativity ends earlier for [tado]-as standard condition than for [taado]-as standard condition.

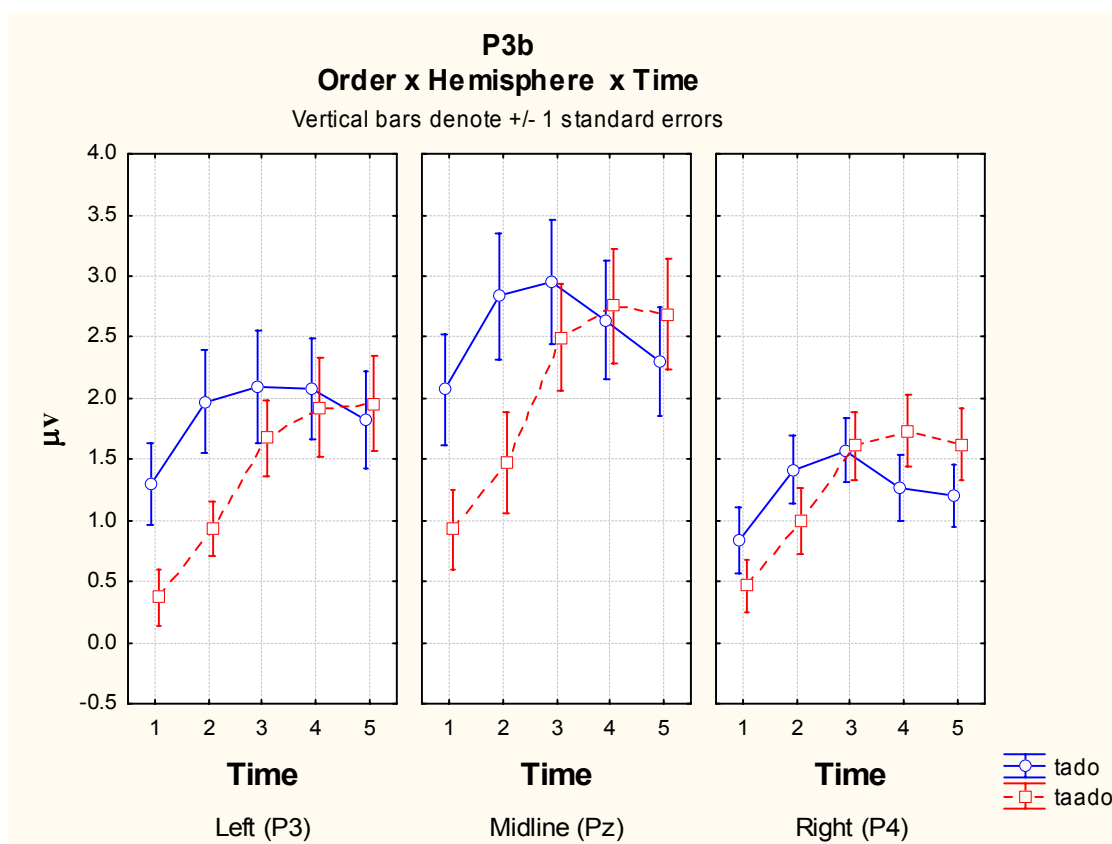


Figure 18. [tado vs. taado] in the Auditory-Attend Experiment – P3b effect: the Pz site showed the largest amplitude P3b for both standard orders, but peaked earlier for [tado] than [taado]-as-standard condition.

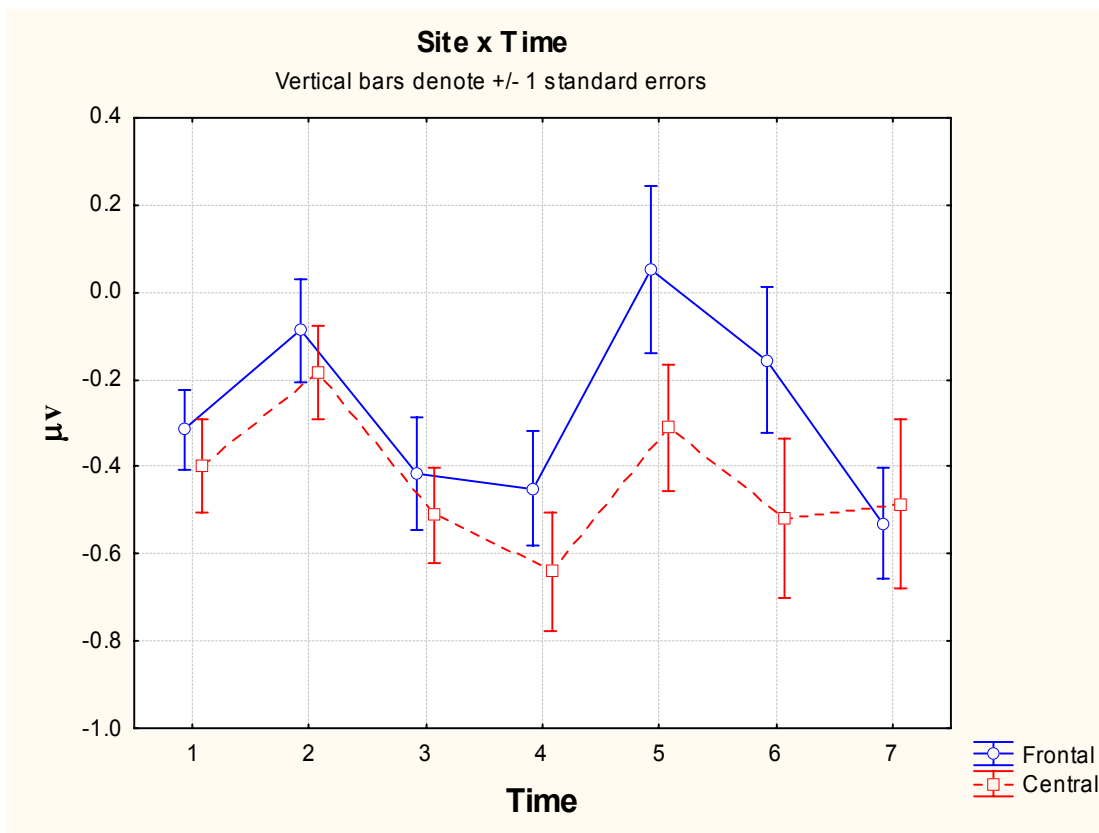


Figure 19. [mif] in the Auditory-Attend Experiment – Site effect: the central sites showed a larger MMN than the frontal sites, and the largest peak amplitude was shown in Time interval 4 (324-360 ms).

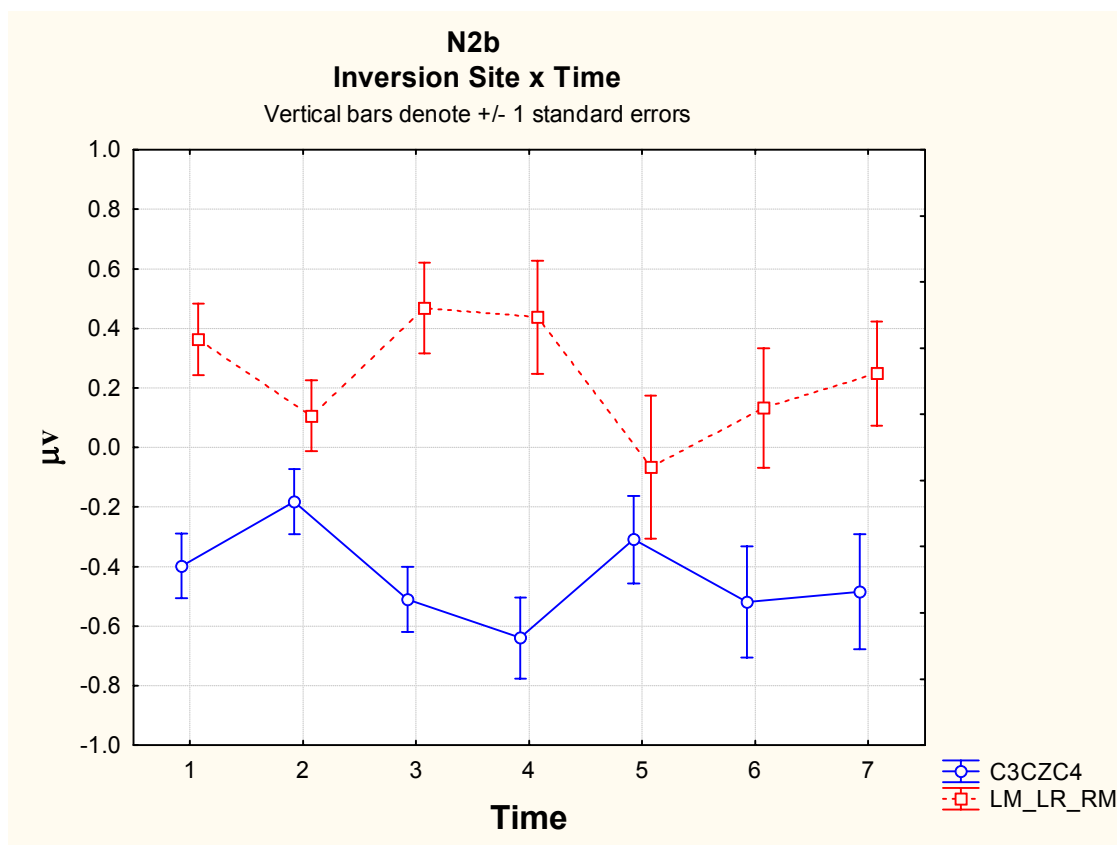


Figure 20. [miʃi] in the Auditory-Attend Experiment - Inversion effect: inversion at the mastoids is found from Time intervals 3 - 4 (284-360 ms) and 6 - 7 (484- 560 ms). There was no clear evidence of N2b.

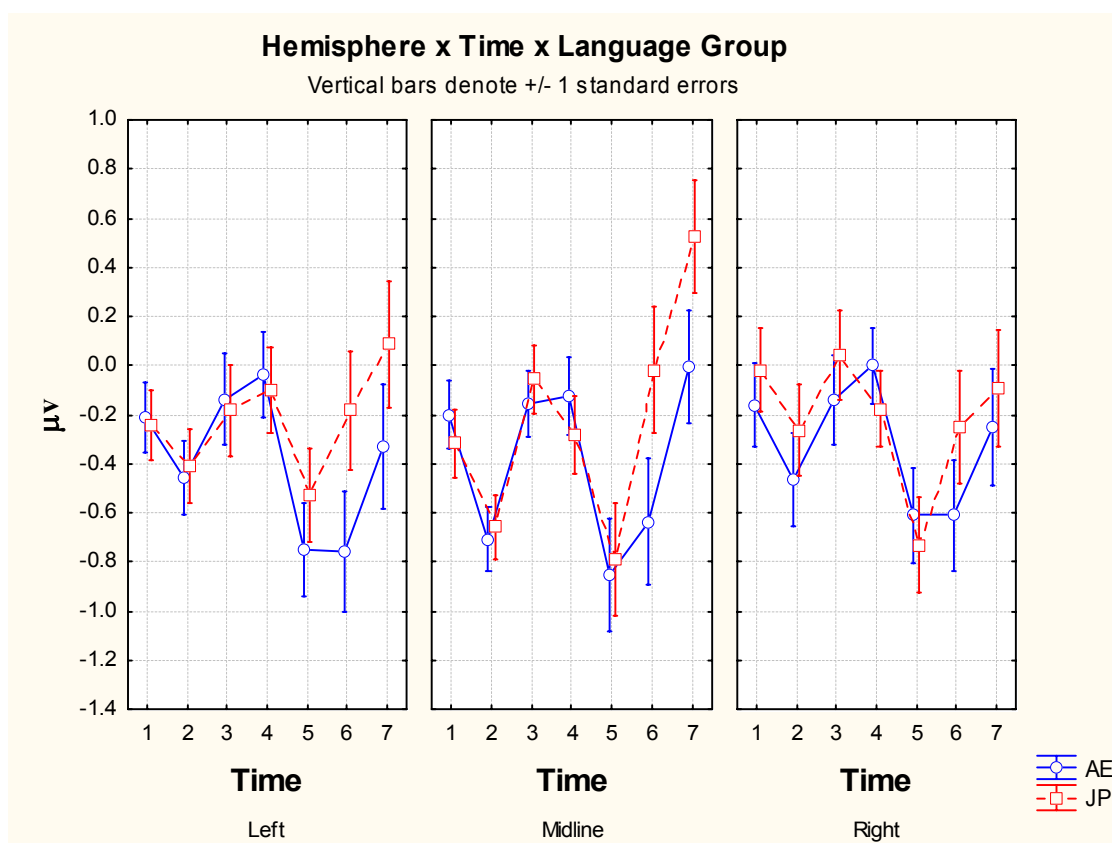


Figure 21. [miʃʃi] in the Auditory-Attend Experiment – Language Group effect: the negativity was larger for the AE compared to JP group particularly at left and midline sites in Time interval 6 (404-444 ms).

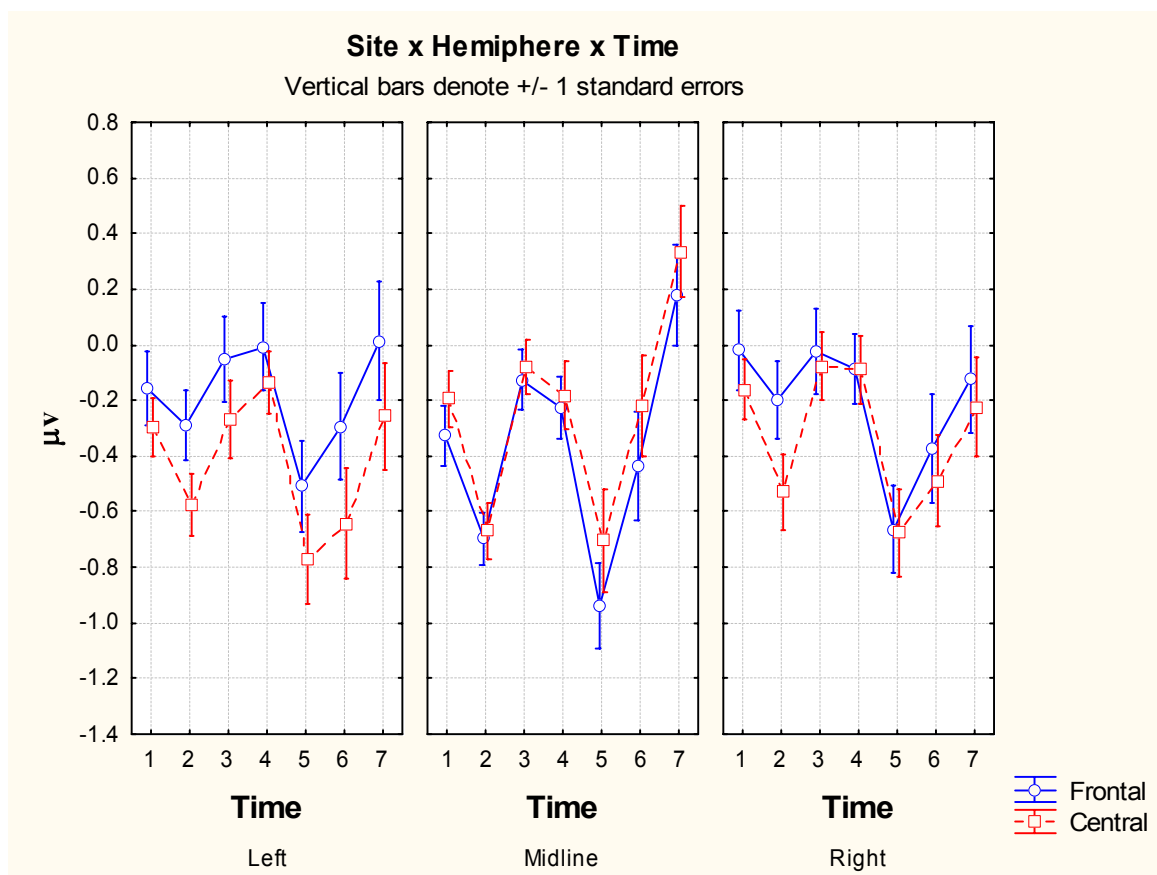


Figure 22. [mi] in the Auditory-Attend Experiment – Site and Hemisphere effect: central sites showed a larger negativity than frontal sites particularly in the left hemisphere.

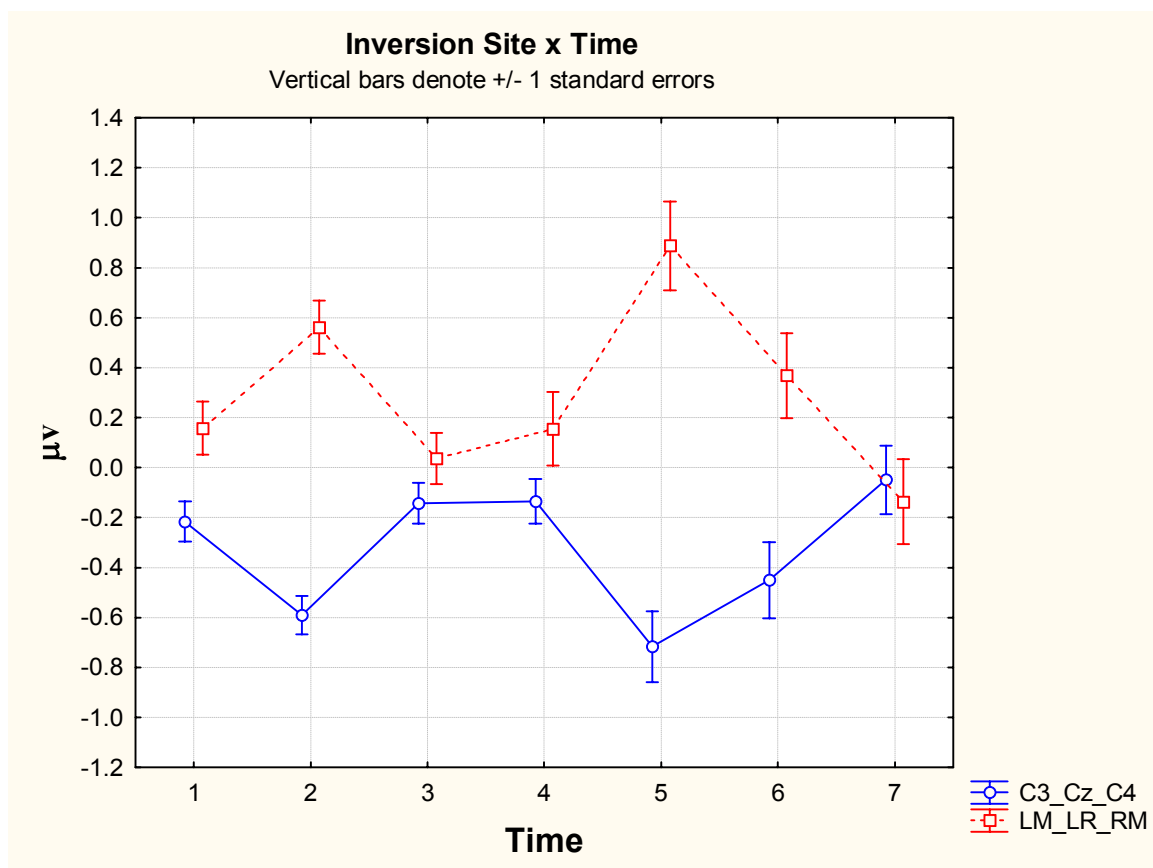


Figure 23. [missi] in the Auditory-Attend Experiment – Inversion effect: post-hoc tests ($p < 0.05$) showed significant differences between the central and mastoid sites for Time interval 2 (244-280 ms); 5 (364-400 ms) and 6 (404-440 ms). There was no clear evidence of N2b.

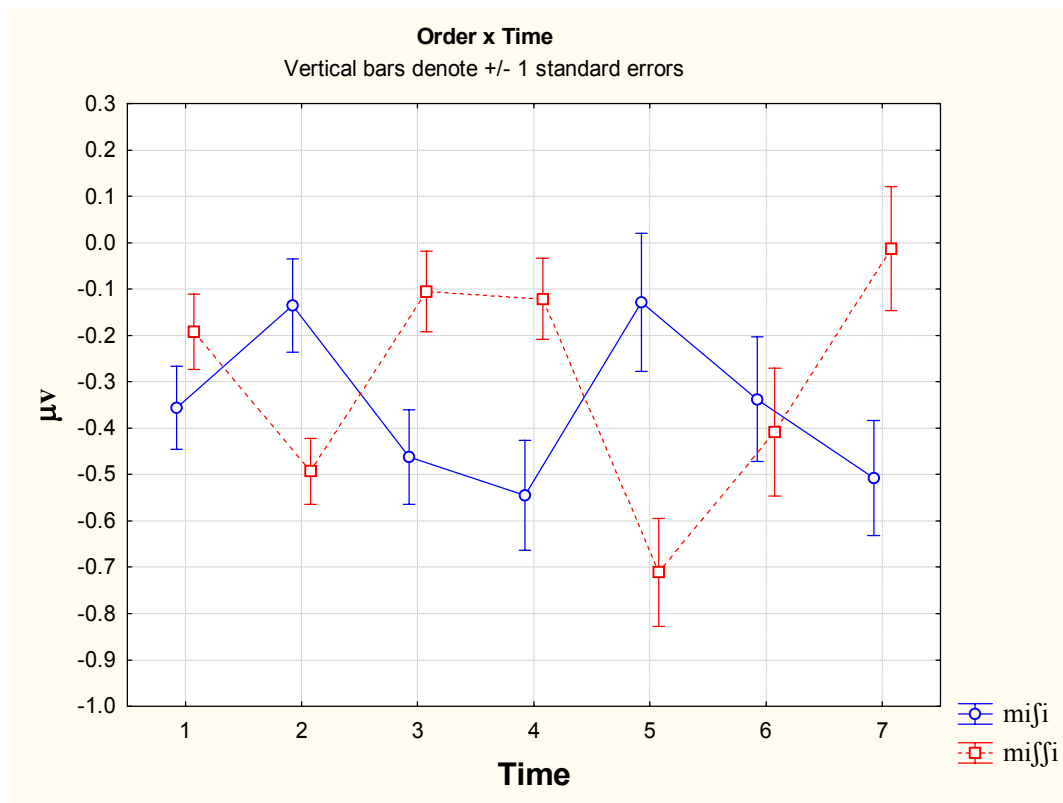


Figure 24. [miʃi vs. miʃʃi] in the Auditory-Attend Experiment – Order effect: the [miʃi]-as-standard condition showed earlier negativities than the [miʃʃi]-as-standard condition, but there was also an earlier negativity in Time interval 2 (244-280 ms) for the [miʃʃi]-as-standard condition. Voltages summed over all six sites.

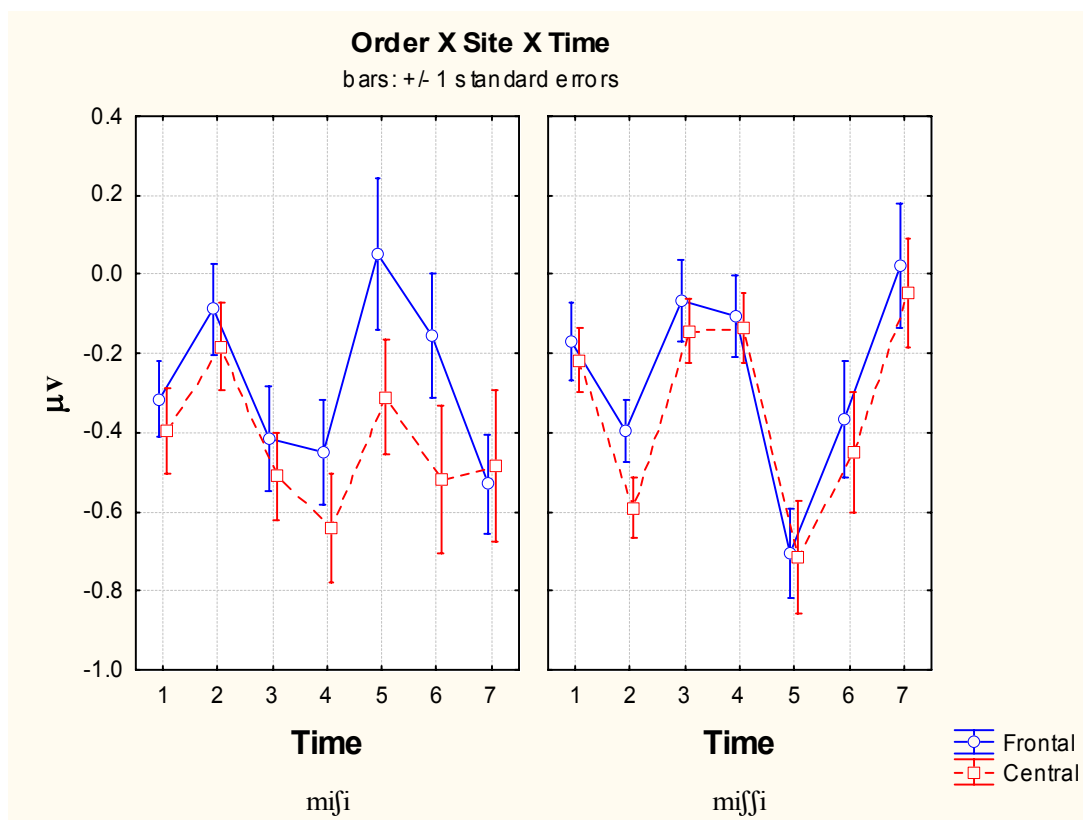


Figure 25. [miʃi vs. miʃfi] in the Auditory-Attend Experiment – Site effect: the [miʃfi]-as-standard condition showed a larger MMN in Time interval 3 (244-280 ms) and 5 (364-400 ms) at both frontal and central sites, while the [miʃi]-as-standard condition showed a larger MMN in Time interval 3 (284-320 ms), 4 (324-360 ms), and 7 (444-480 ms) at both frontal and central sites. There was a larger MMN for the [miʃi]-as-standard condition at central sites, while the [miʃfi]-as-standard condition showed no difference between the two sites.

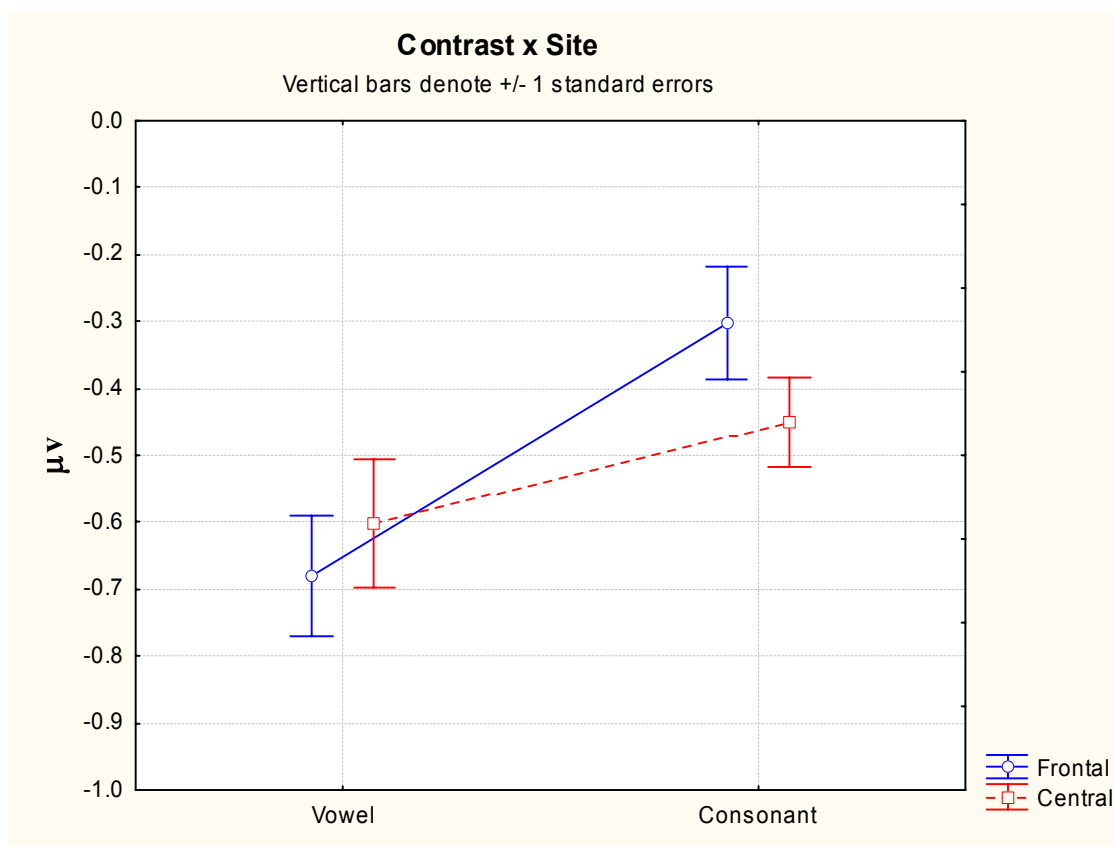


Figure 27. Vowel vs. Consonant Contrasts in the Auditory-Attend Experiment: the vowels show greater frontal negativity than the consonants and the consonants show greater central than frontal negativity.

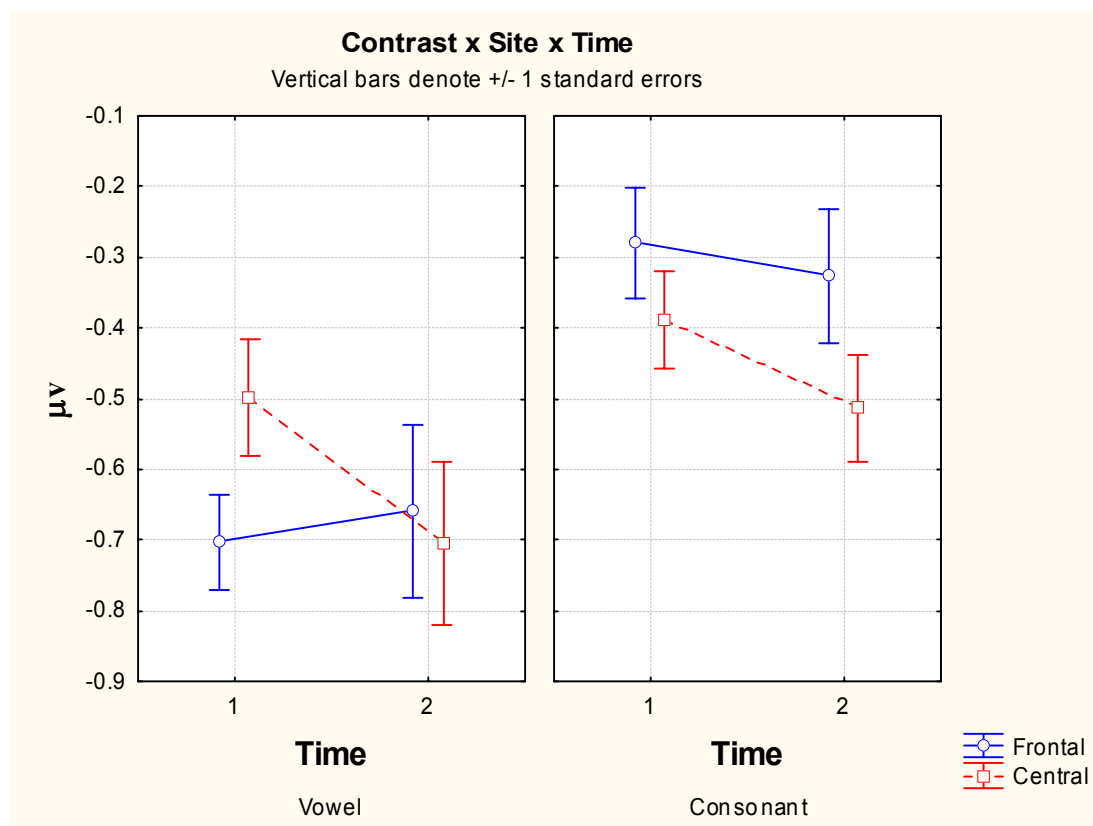


Figure 28. Vowel vs. Consonant Contrasts in the Auditory-Attend Experiment: the frontal sites for both time bins surrounding MMN peaks and the central sites in Time interval 2 (244-280 ms) showed stronger effects for vowels.

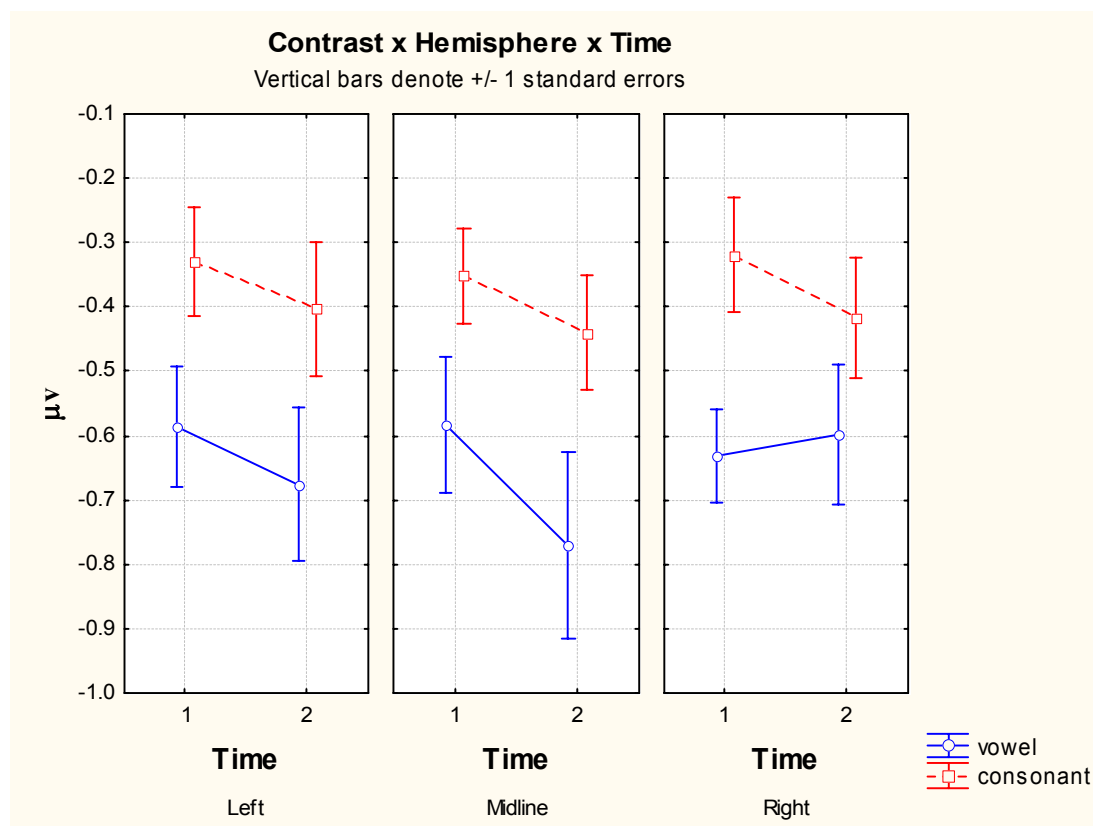


Figure 29. Vowel vs. Consonant Contrasts in the Auditory-Attend Experiment: the negativity to the vowel contrasts was larger than that to the consonant contrasts at all hemispheres and at both time bins surrounding the peak and showed a different pattern across time at the right compared to left and midline sites, but only for vowels.

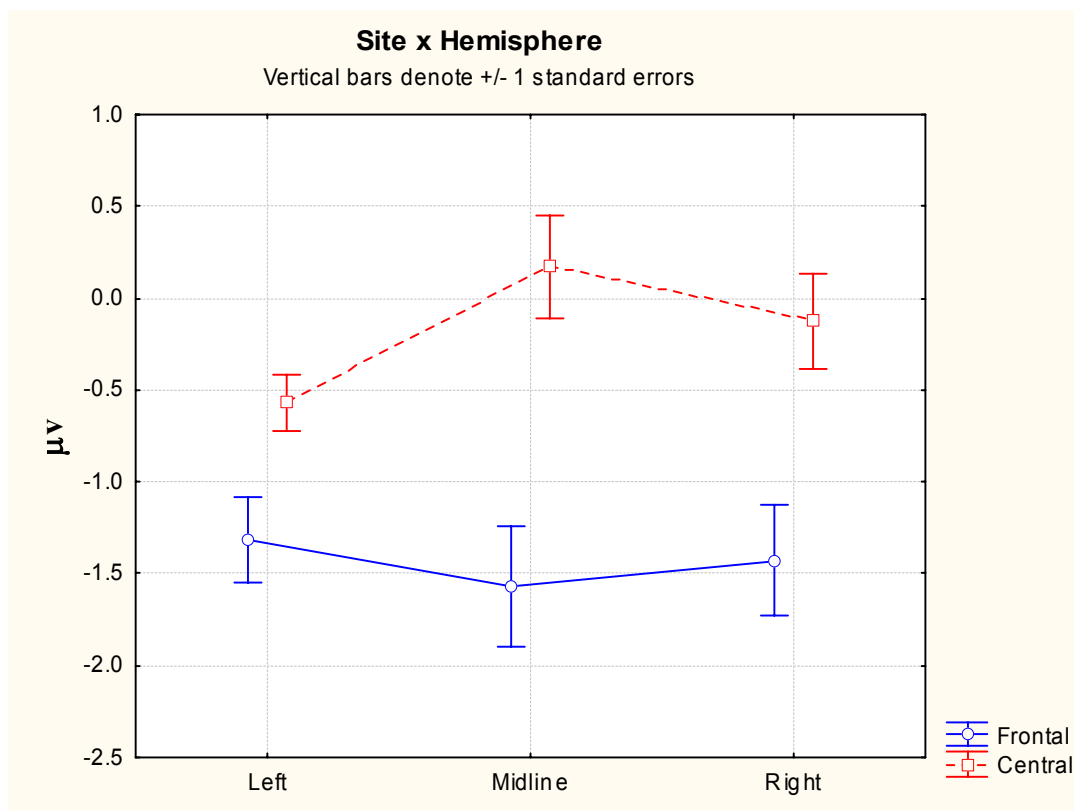


Figure 30. Late Negativity (LN) [tado] in the Auditory-Attend Experiment – Site and Hemisphere effect: the frontal sites were larger than the central sites across left, midline and right sites but more so for the midline and the right sites, and post-hoc tests supported this result.

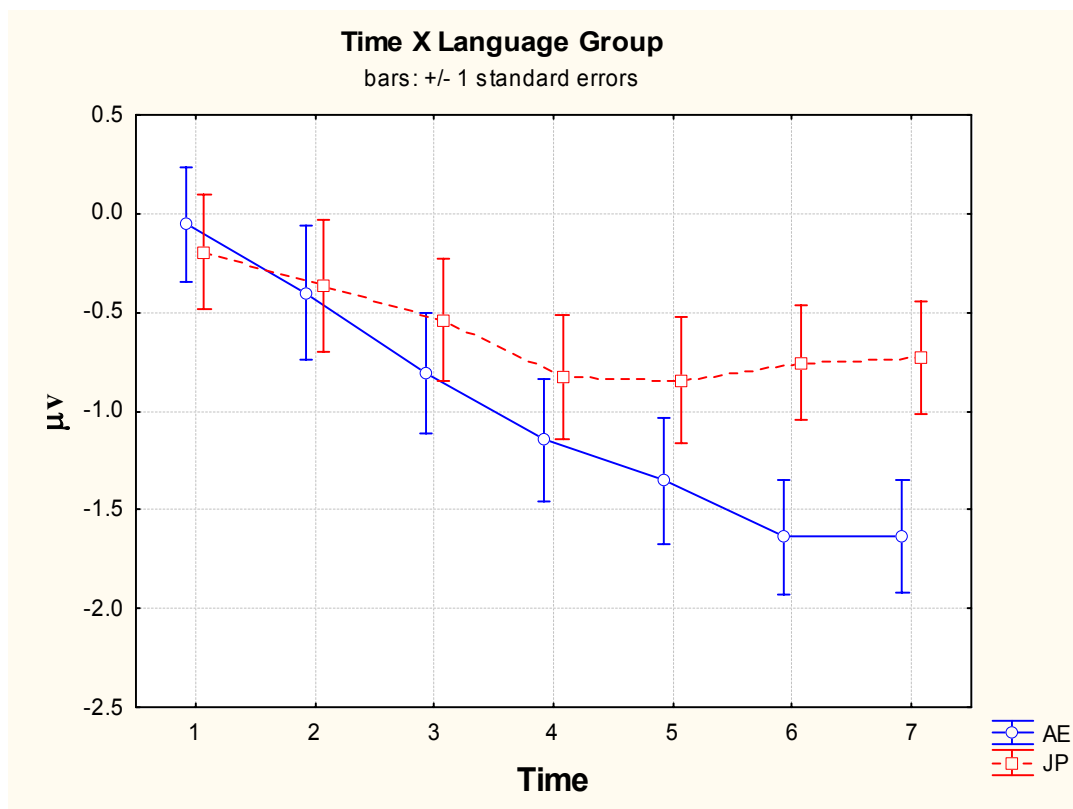


Figure 31. Late Negativity (LN) [tado] in the Auditory-Attend Experiment – Language Group effect: the AE group showed a significantly larger LN than did the JP group. The largest LN was found in Time interval 6 (644-680 ms). Voltages summed over all six sites.

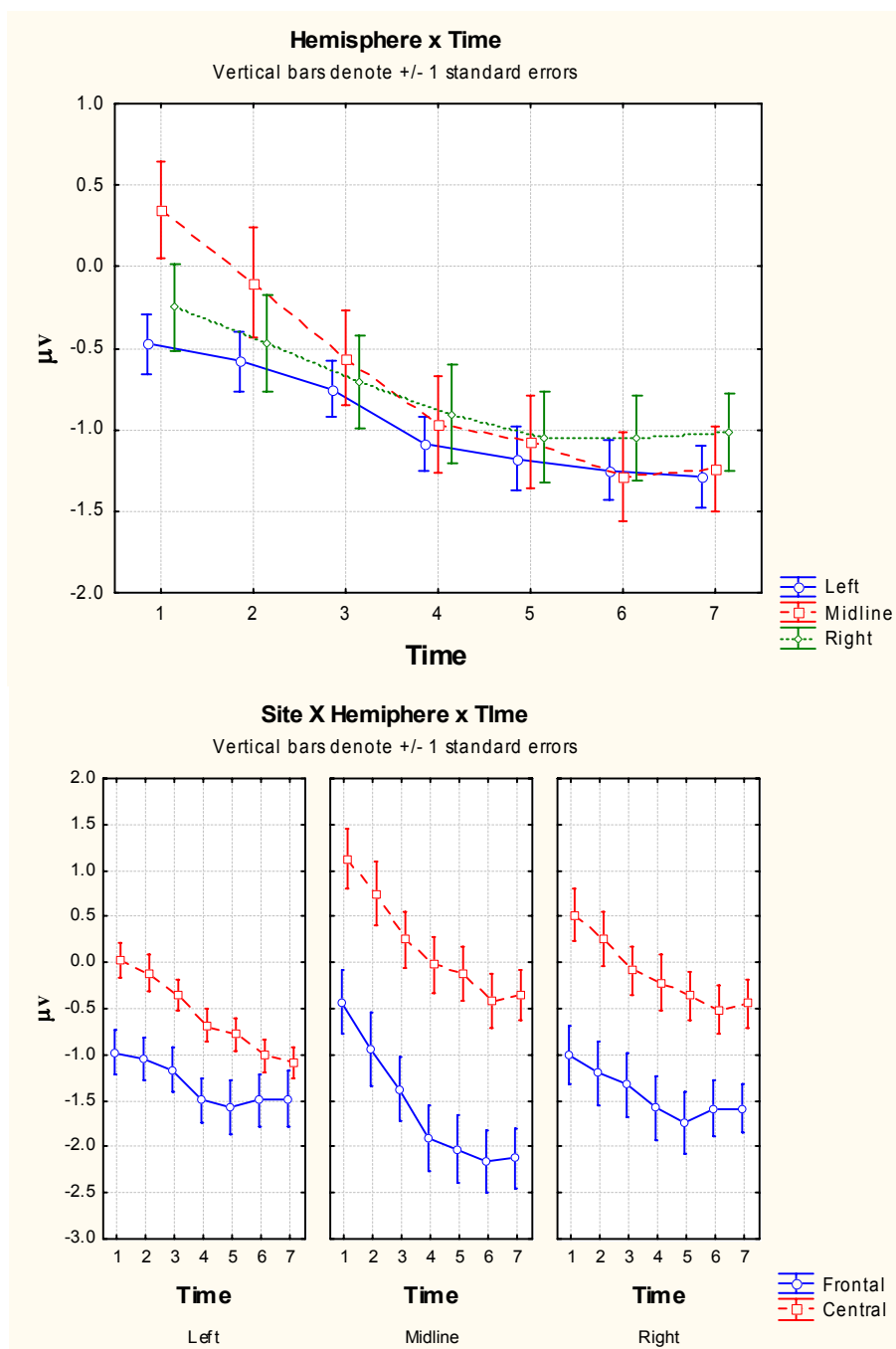


Figure 32. Late Negativity (LN) [tado] in the Auditory-Attend Experiment – Site and Hemisphere effect: two-way analyses examining site and hemisphere for each time separately reveal that the frontal sites elicited larger MMN than the central sites at left-midline and right sites across all the times bins.

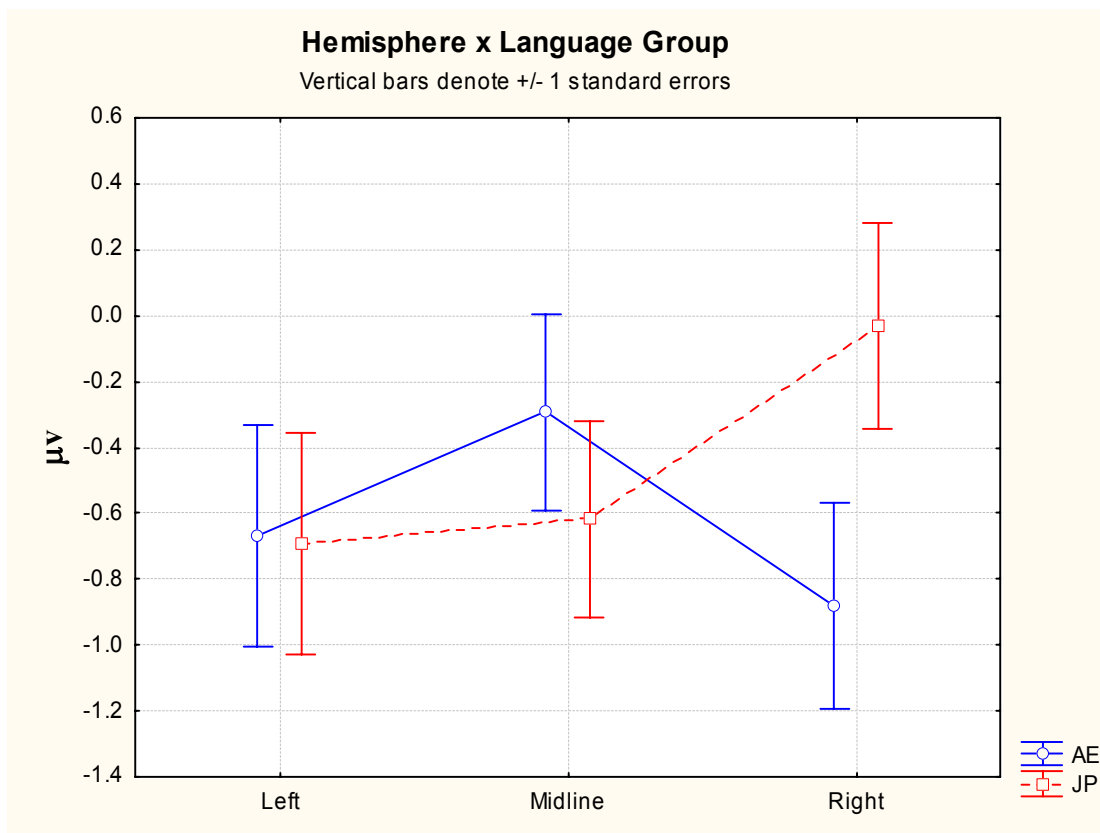


Figure 33. Late Negativity (LN) [taado] in the Auditory-Attend Experiment – Language Group effect: the AE group had the large negativity at the right than midline or left sites, while the JP group showed the smallest LN at the right site.

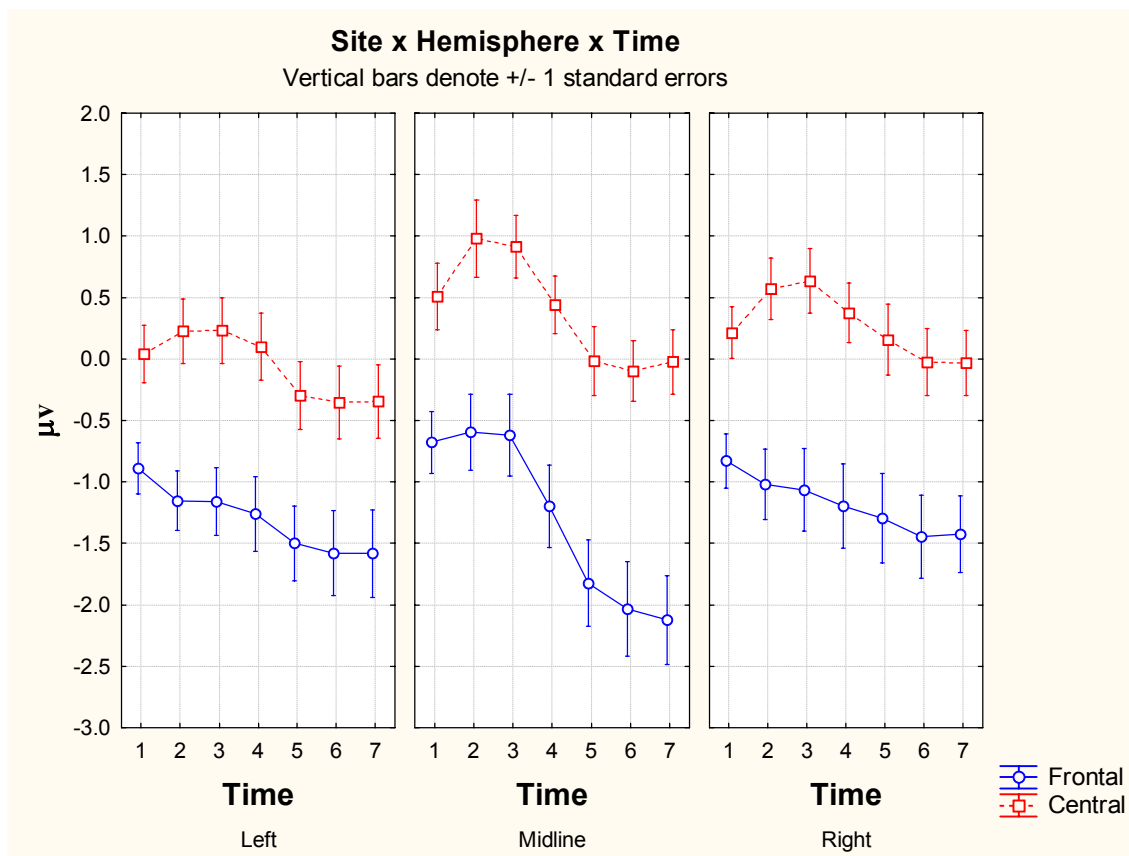


Figure 34. Late Negativity (LN) [taado] in the Auditory-Attend Experiment – Site and Hemisphere effect: the LN was seen more strongly at the frontal sites than at the central sites.

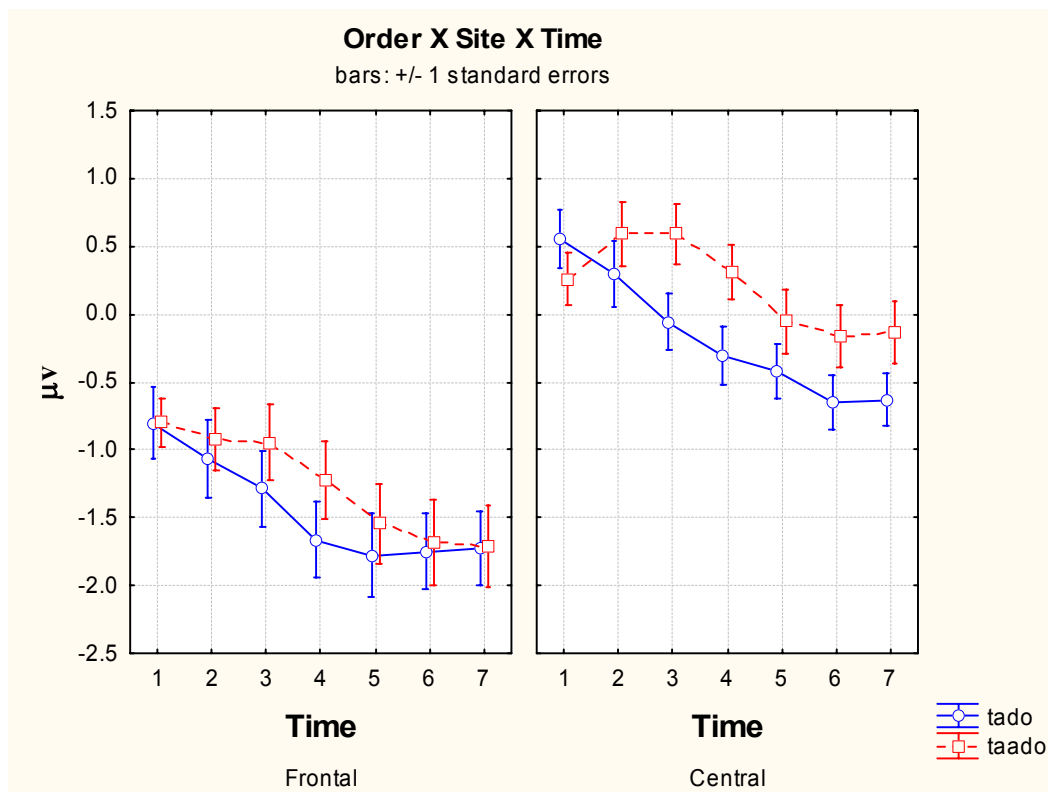


Figure 35. Late Negativity (LN) [tado vs. taado] in the Auditory-Attend Experiment – Site effect: LN was clearly shown in the frontal sites but not in the central sites. The two deviant orders differ across time, but primarily at central sites.

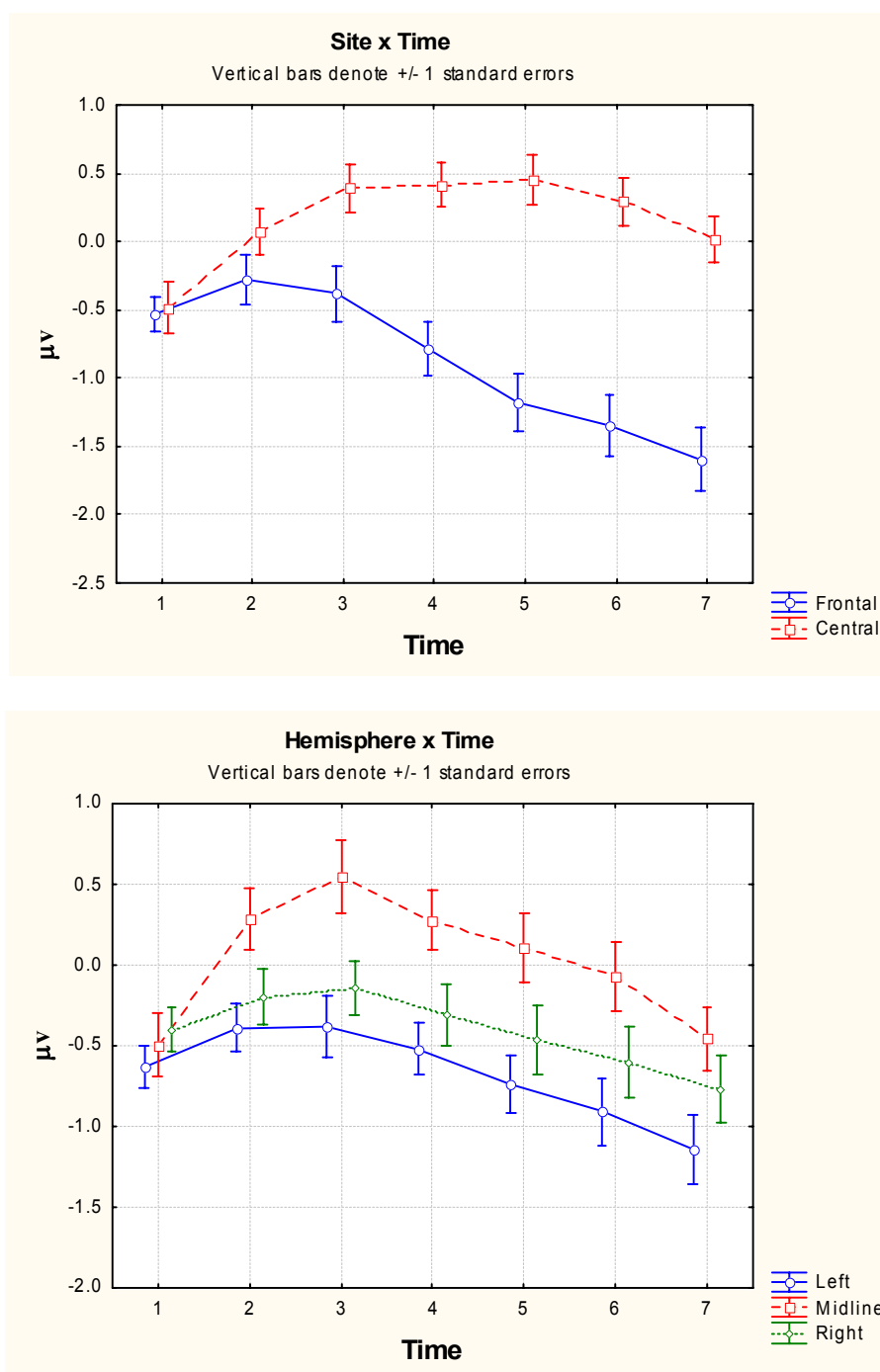


Figure 36. Late Negativity (LN) [mifj] in the Auditory-Attend Experiment – Site and Hemisphere effect: frontal sites showed larger LN than the central sites in all time intervals except at Time 1 (444-480 ms) (top) and the left site showed the largest LN followed by the right and then midline sites except at Time 1 (444-480 ms) (bottom).

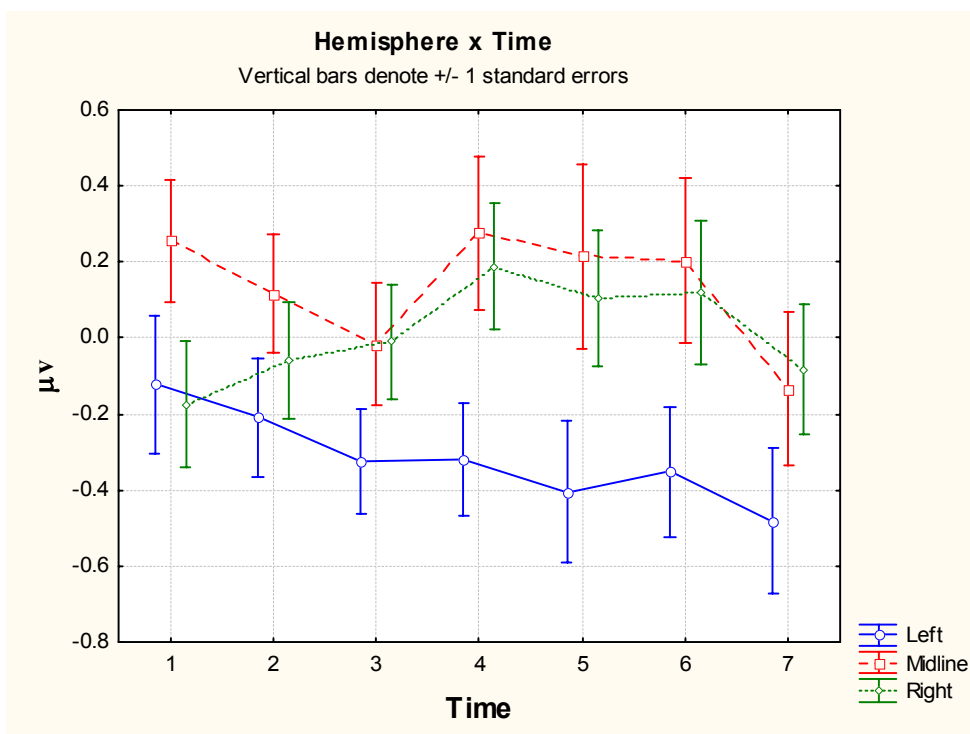
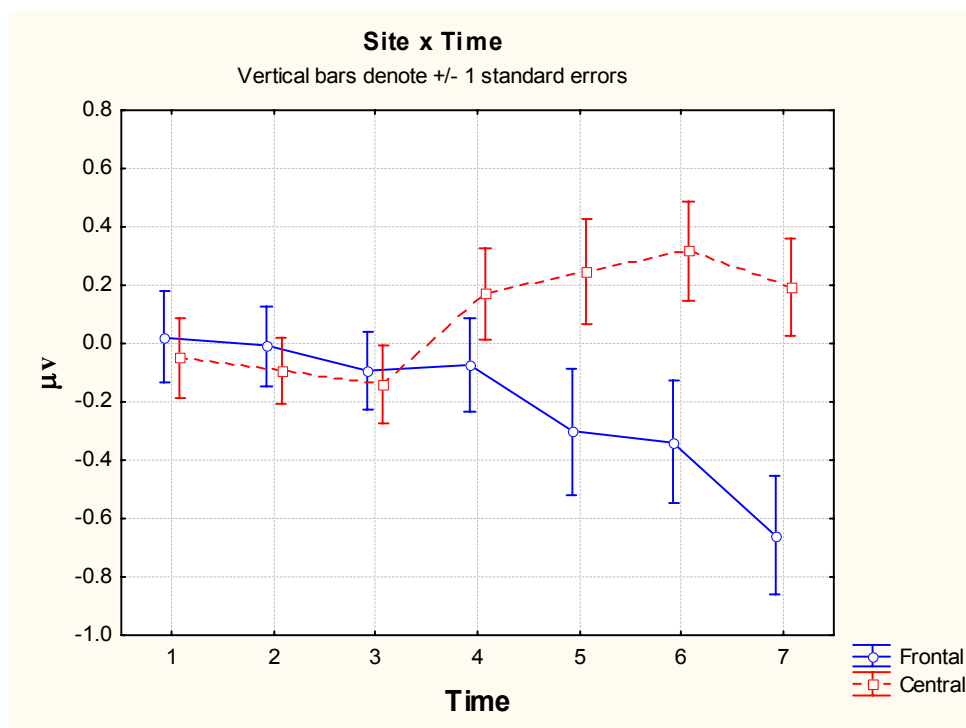


Figure 37. Late Negativity (LN) [miffi] in the Auditory-Attend Experiment – Site and Hemisphere effect: frontal sites showed larger LN than the central sites in Time intervals 4 - 7 (top) and the left site showed the largest LN except at Time interval 1 (444-480 ms) (bottom).

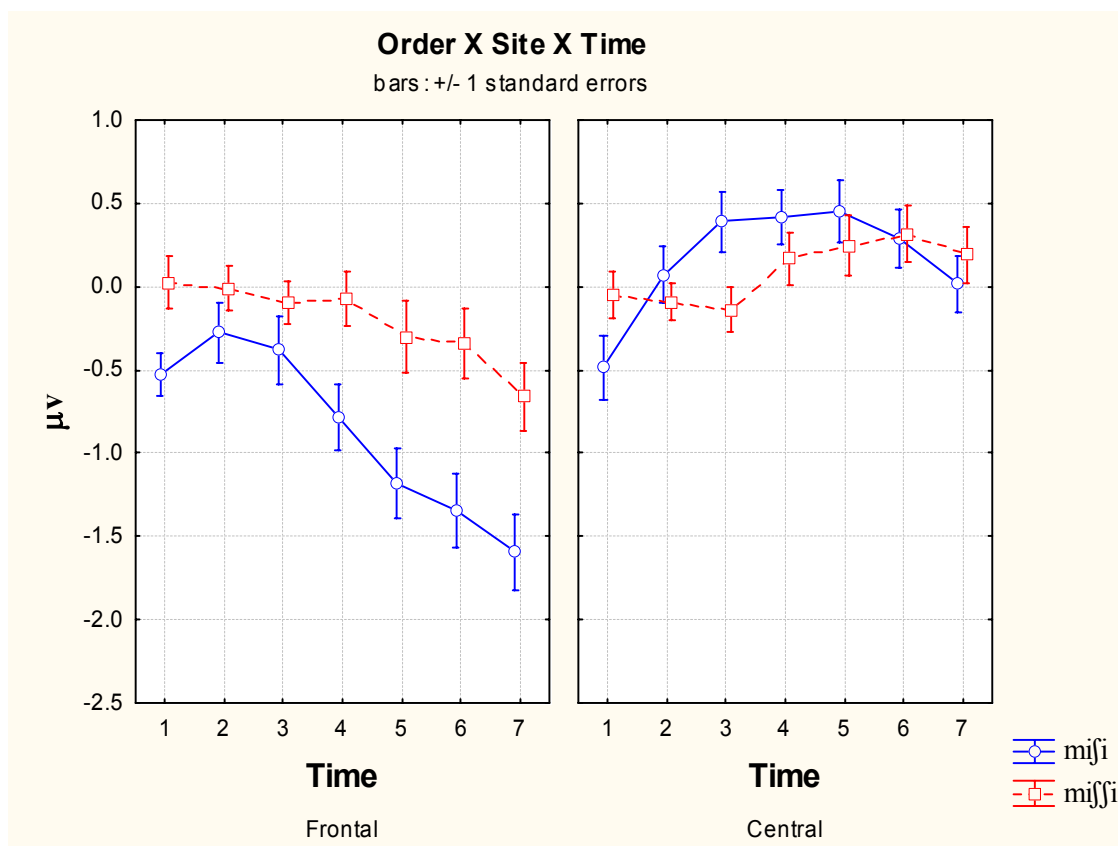


Figure 38. Late Negativity (LN) [miʃi vs. miʃʃi] in the Auditory-Attend Experiment – Site effect: the LN was clearly shown in the frontal sites but not the central sites and is bigger for [miʃi]-as-standard than [miʃʃi]-as-standard conditions.

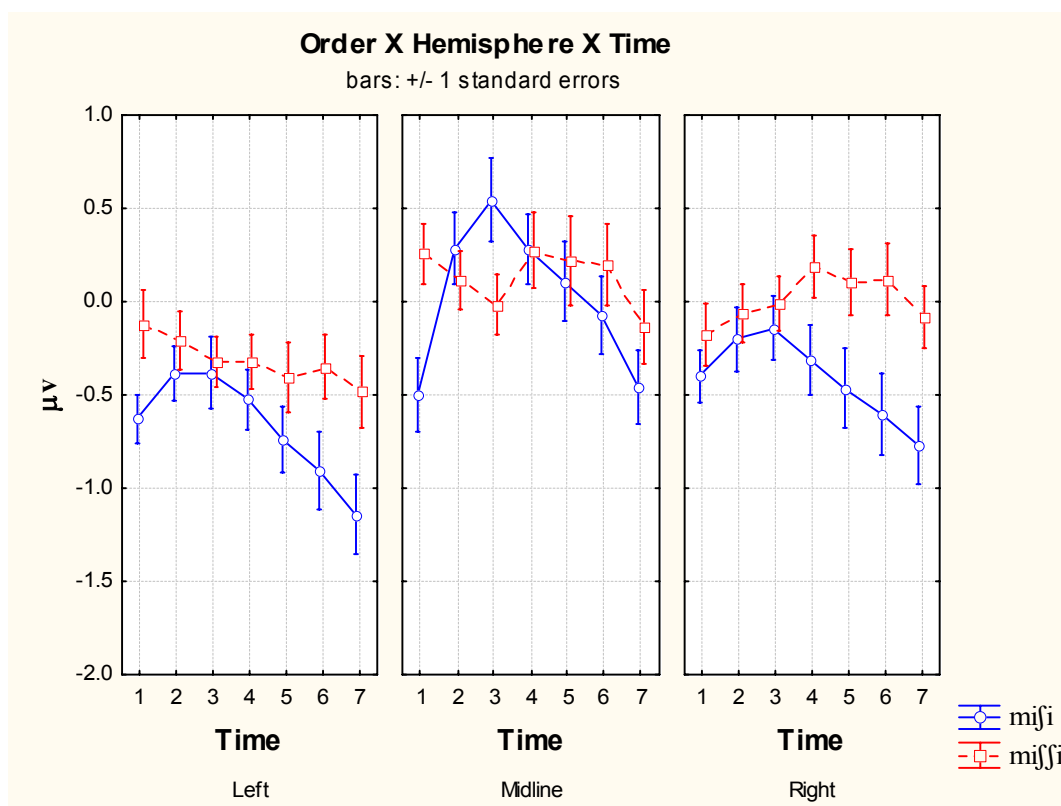


Figure 39. Late Negativity (LN) [$mi|i$ vs. $mi|fi$] in the Auditory-Attend Experiment – Hemisphere effect: the greatest difference between deviant orders is found at lateral sites.

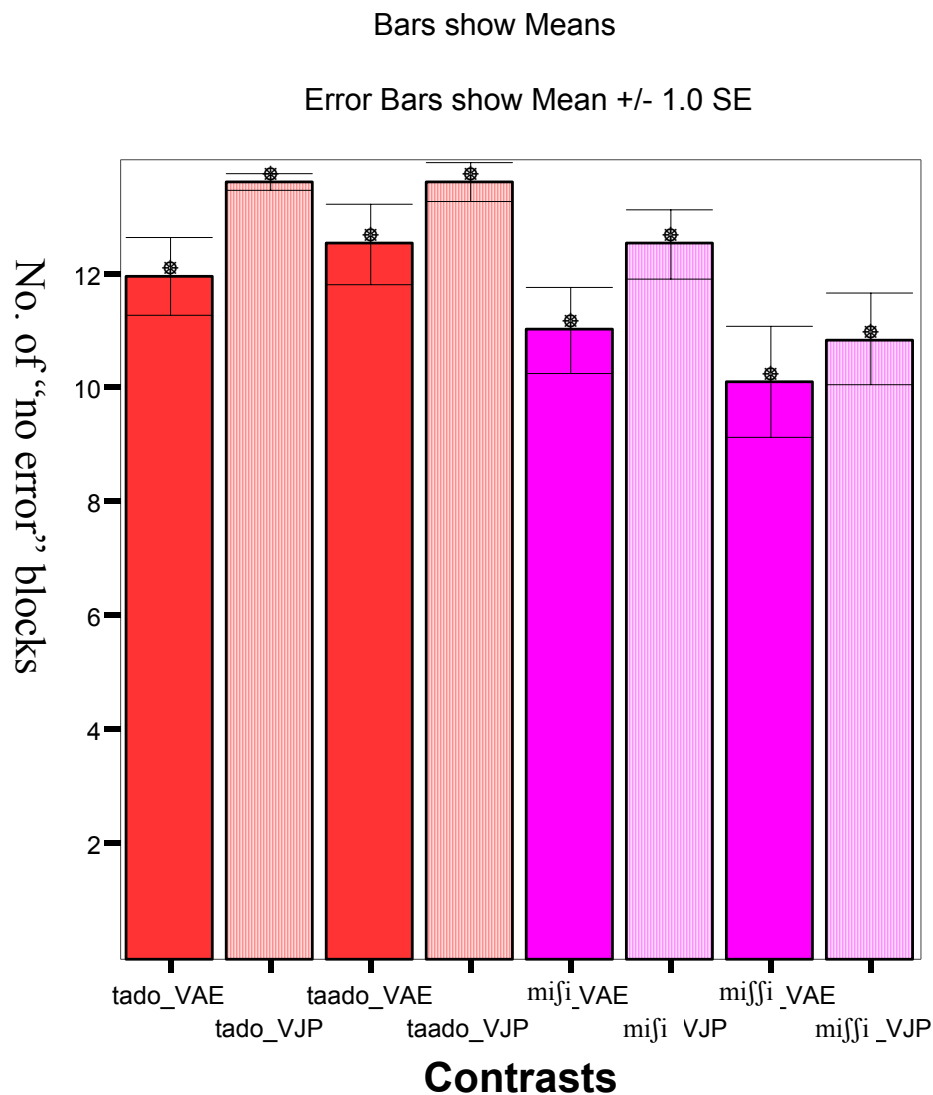


Figure 40. Counting of visual deviants in the Visual-Attend Experiment (14 blocks): the numbers of “no error blocks” per condition are shown. Solid bars are AE and striped bars are JP listeners. Left two pairs refer to vowels and right two pairs for consonants contrasts. No language group differences were observed except [tado]-as-standard condition.

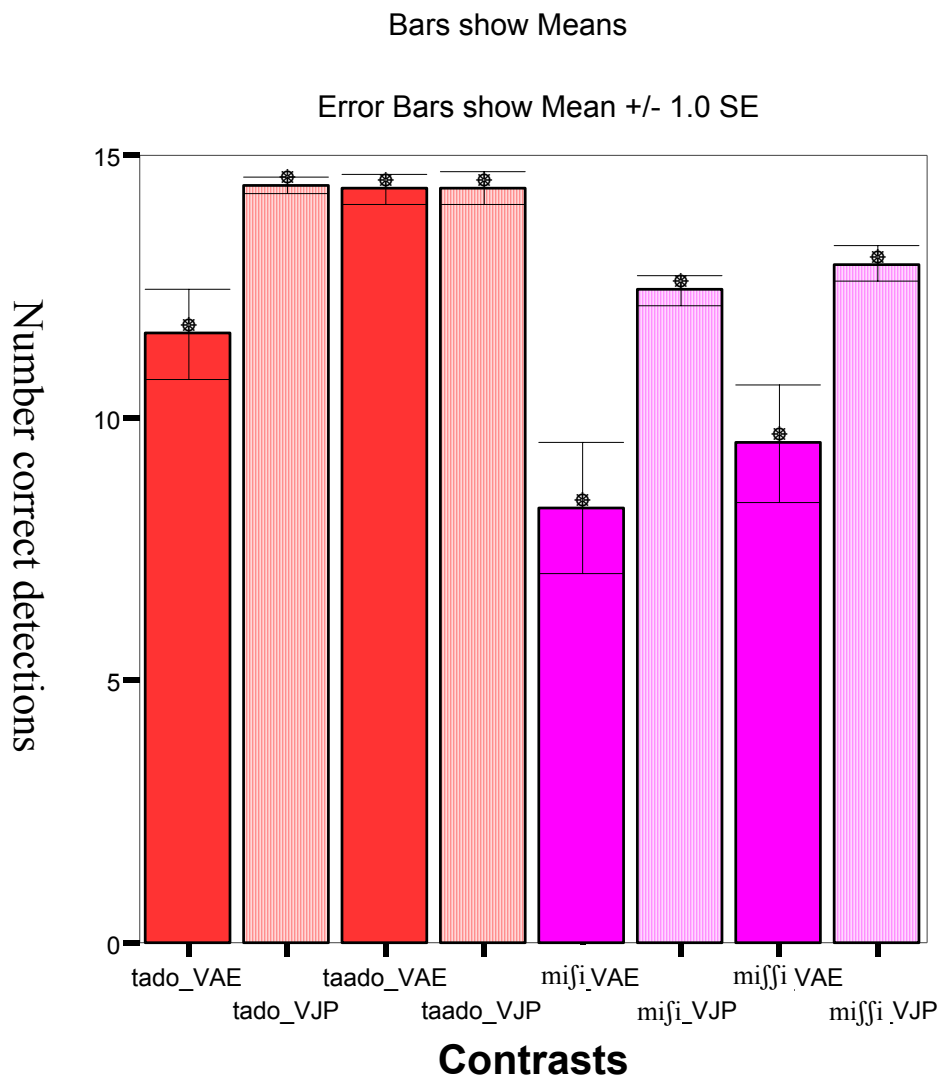


Figure 41. Button-press behavioral data in the auditory task for the Visual-Attend Experiment (15 possible): solid bars are AE and striped bars are JP listeners. Left two pairs shows performance for vowels and right two pairs shows performance for consonants contrasts. There were significant language group differences except for [taado]-as-standard condition.

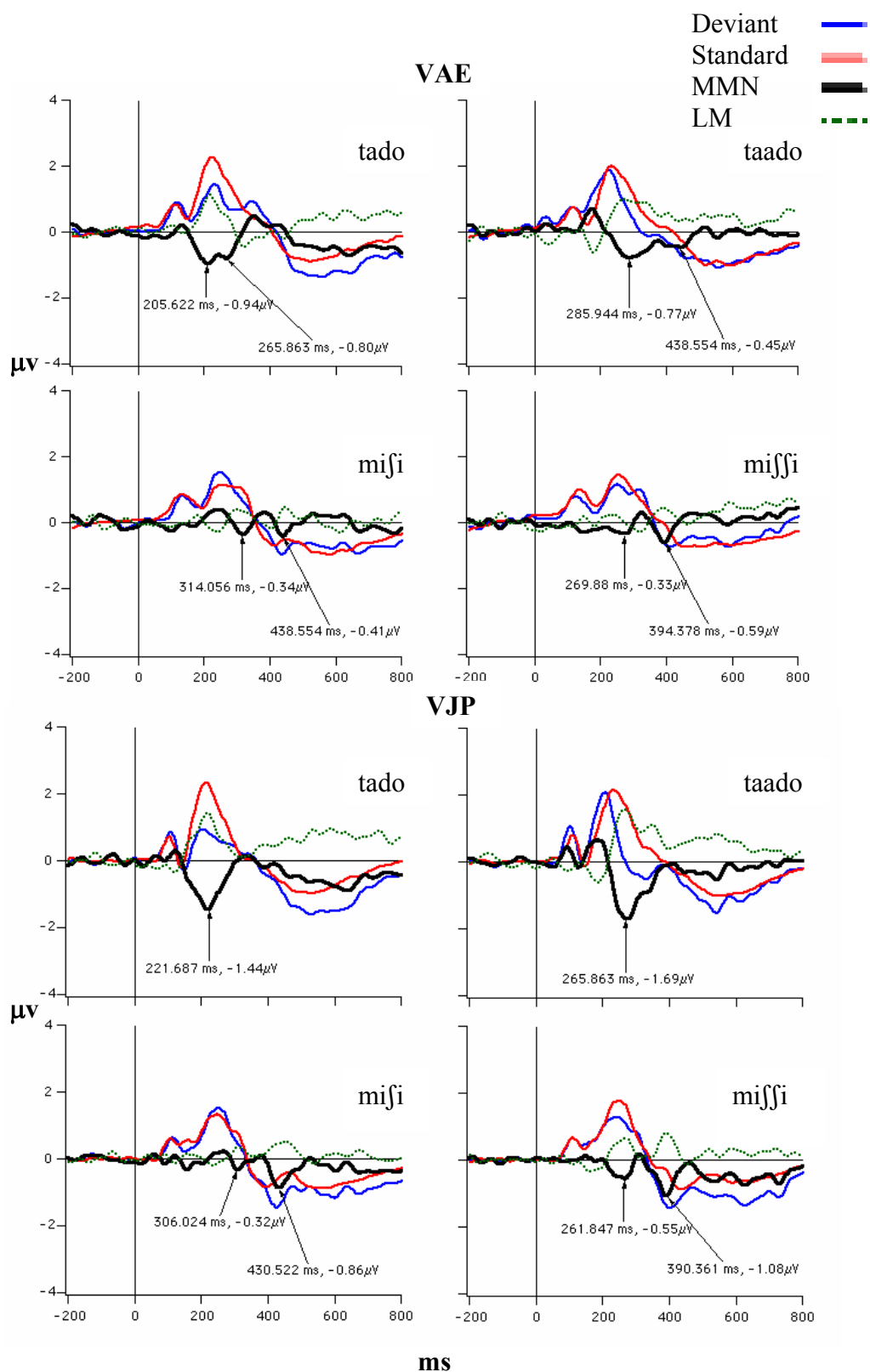


Figure 42. Visual-Attend Grand Average data: the grand average at the mid-frontal site (Fz) for the standard, deviant and deviant-standard and at LM for the deviant-standard for all four contrasts and for AE (top a) and JP (bottom b) are shown above. Latency and amplitude are shown for peak negativities of the subtraction waves.

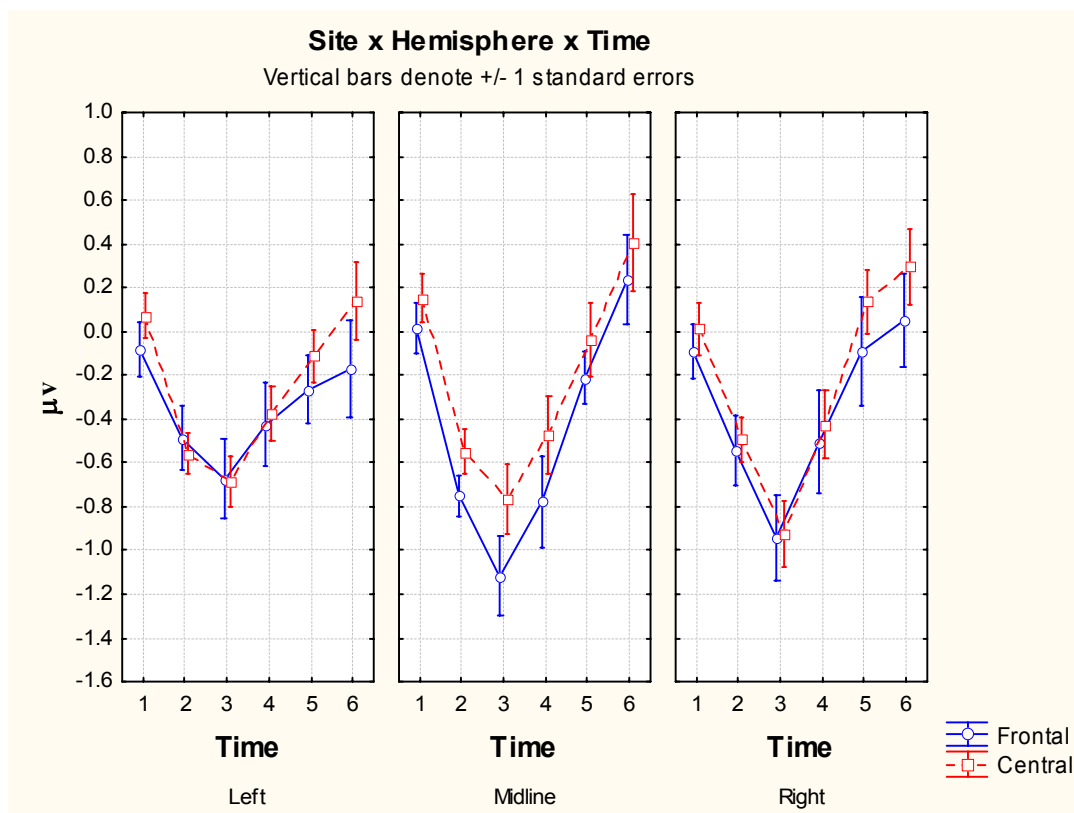


Figure 43. [tado]-as-standard condition in the Visual-Attend Experiment – Site and Hemisphere effect: the midline-frontal site was larger than the central site, but only for Time interval 3 (204-240 ms).

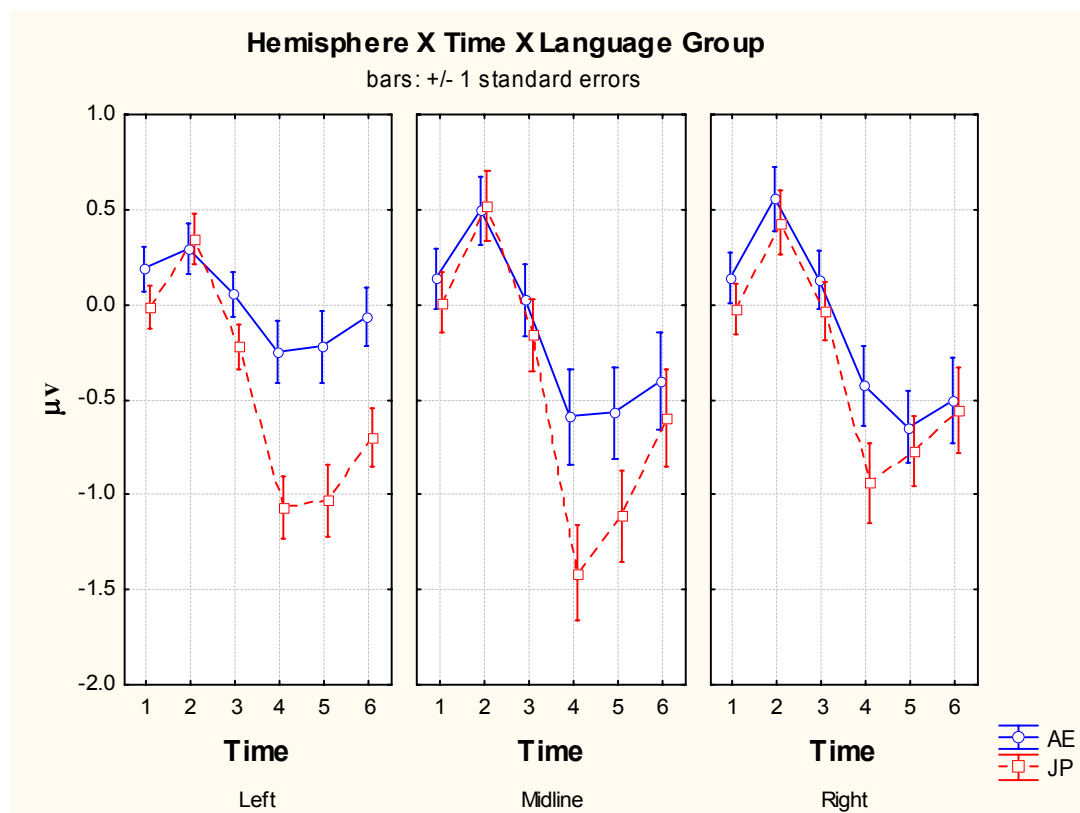


Figure 44. [taado] in the Visual-Attend Experiment – Language Group effect among three hemispheres: the JP group showed a larger MMN for all three regions, but especially at the left and midline sites.

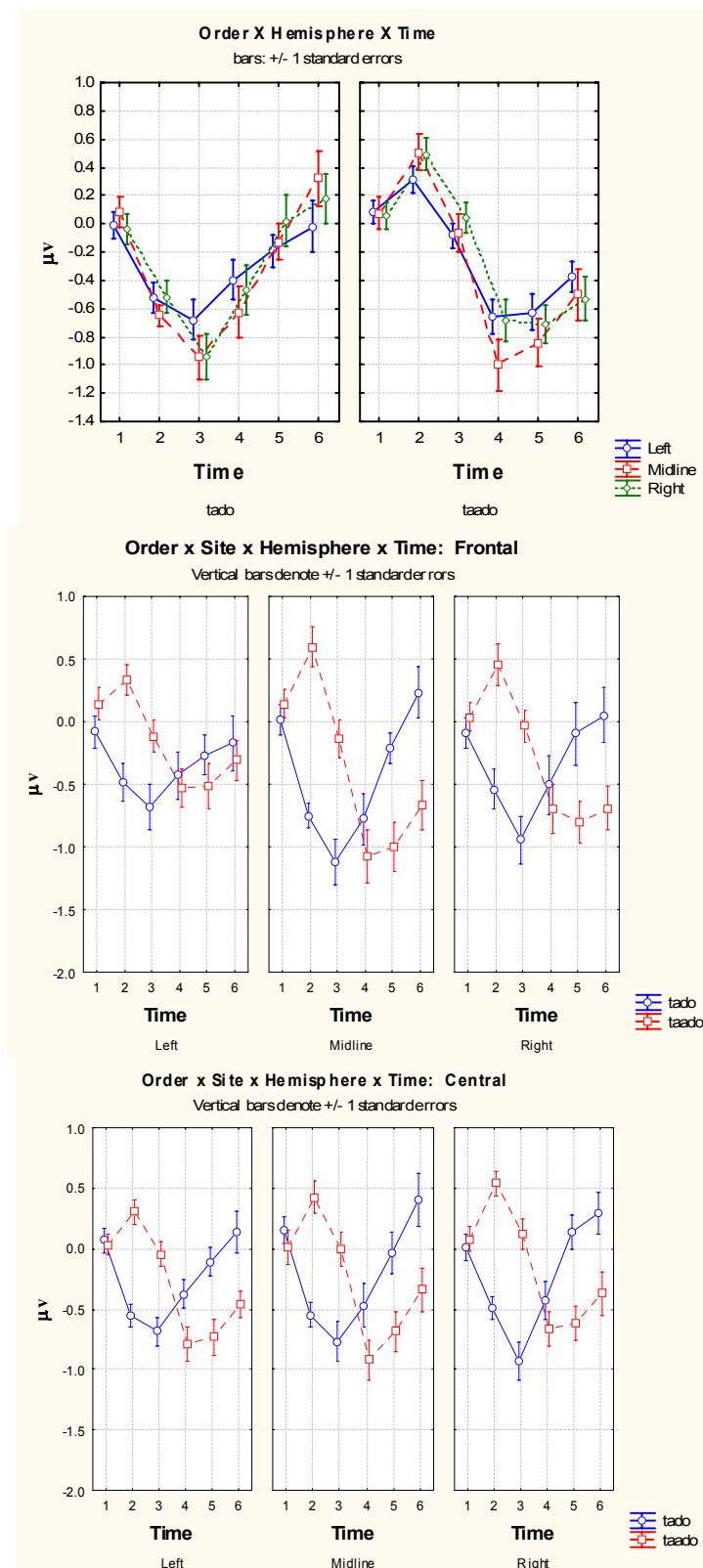


Figure 45. [tado vs. taado] in the Visual-Attend Experiment – Hemisphere effect: the [tado]-as-standard condition had a larger MMN in Time interval 2 (164-200 ms) and 3 (204-240 ms) and for the [taado]-as-standard condition in Time interval 4 (284-320 ms) and 5 (324-360 ms) across all three regions.

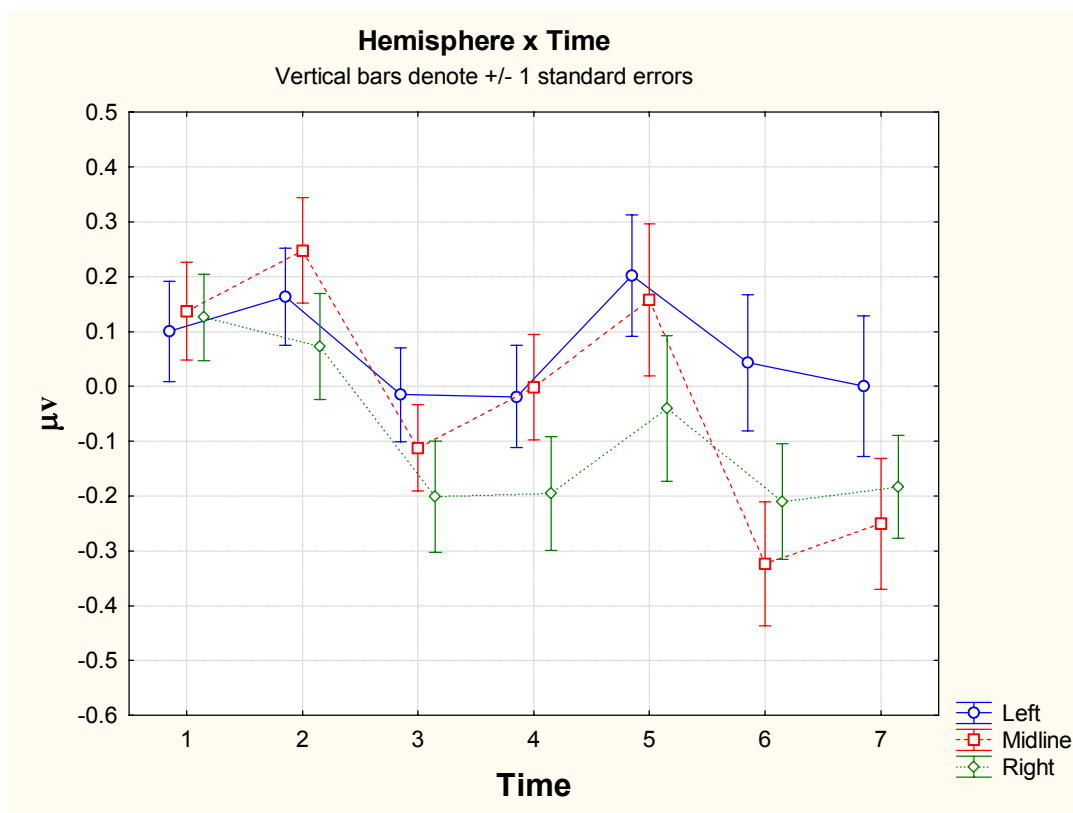


Figure 46. [miʃi] in the Visual-Attend Experiment – Hemisphere effect: the midline sites showed the largest negativity in Time interval 6 (404-440 ms).

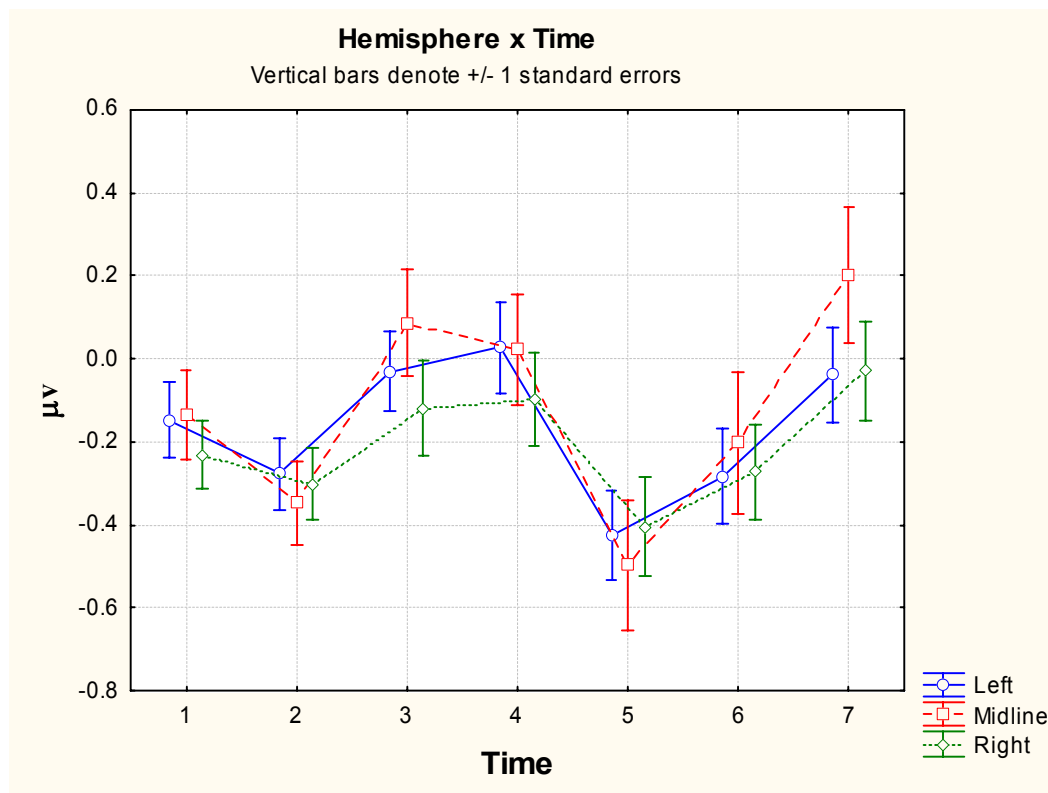


Figure 47. [mi] in the Visual-Attend Experiment – Hemisphere effect: the largest negativity is in Time interval 5 (364-400 ms). There is also a small negativity in Time interval 2 (244-280 ms).

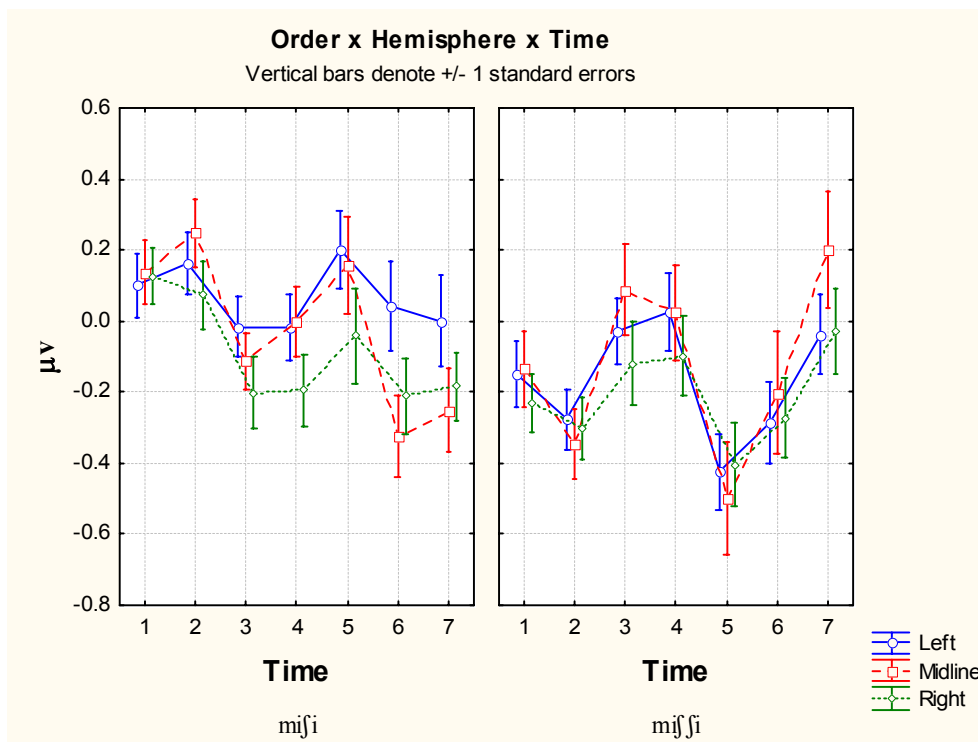


Figure 48. [mi]i vs. [mi]fi in the Visual-Attend Experiment – Hemisphere effect: the peak negativities differed for the two deviant orders, with those to [mi]fi-as-standard condition being more negative.

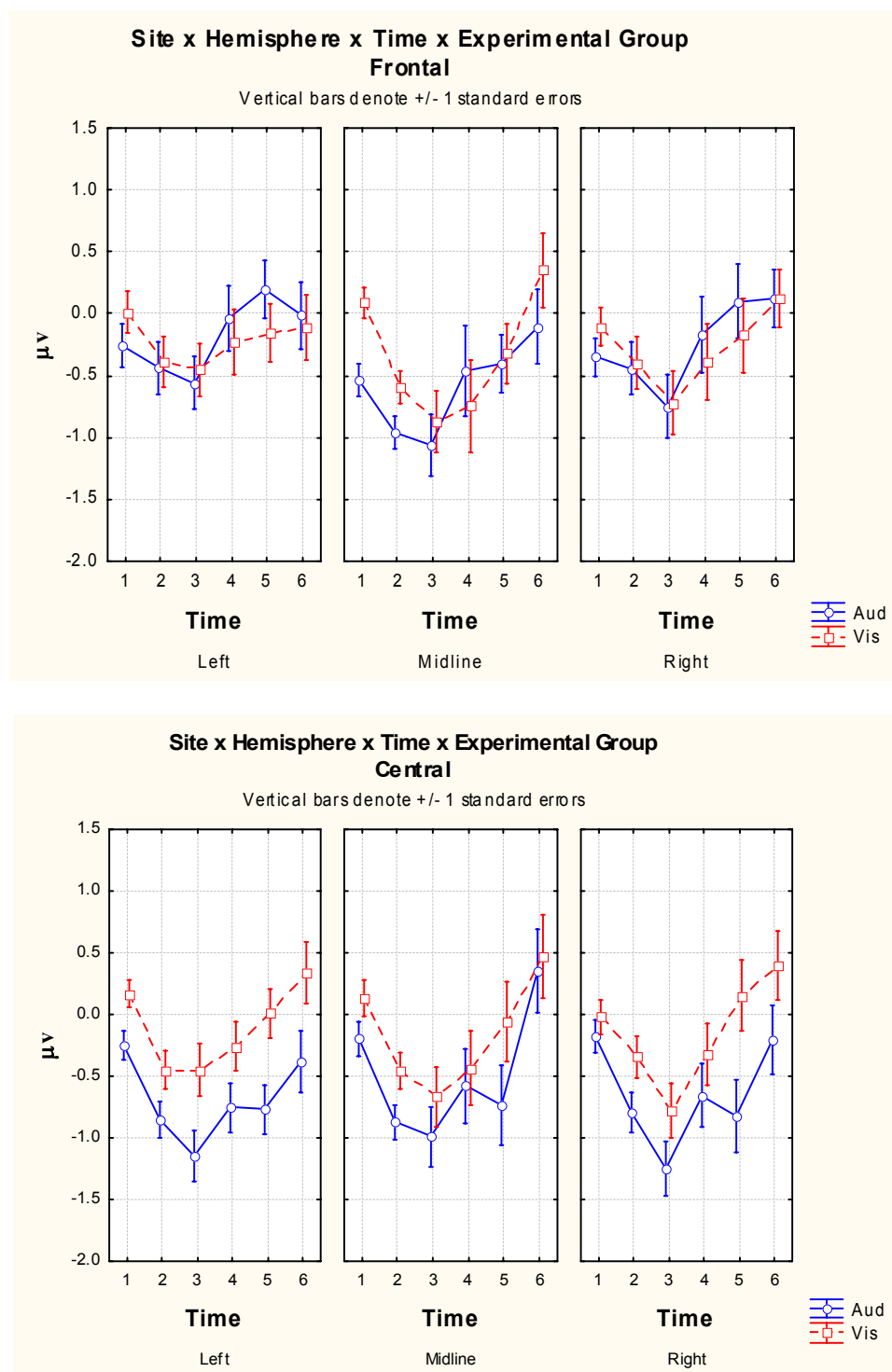


Figure 49. AE group [tado] in the Auditory- and Visual-Attend Experiments: the Auditory-Attend group showed a larger MMN at all three regions at central sites.

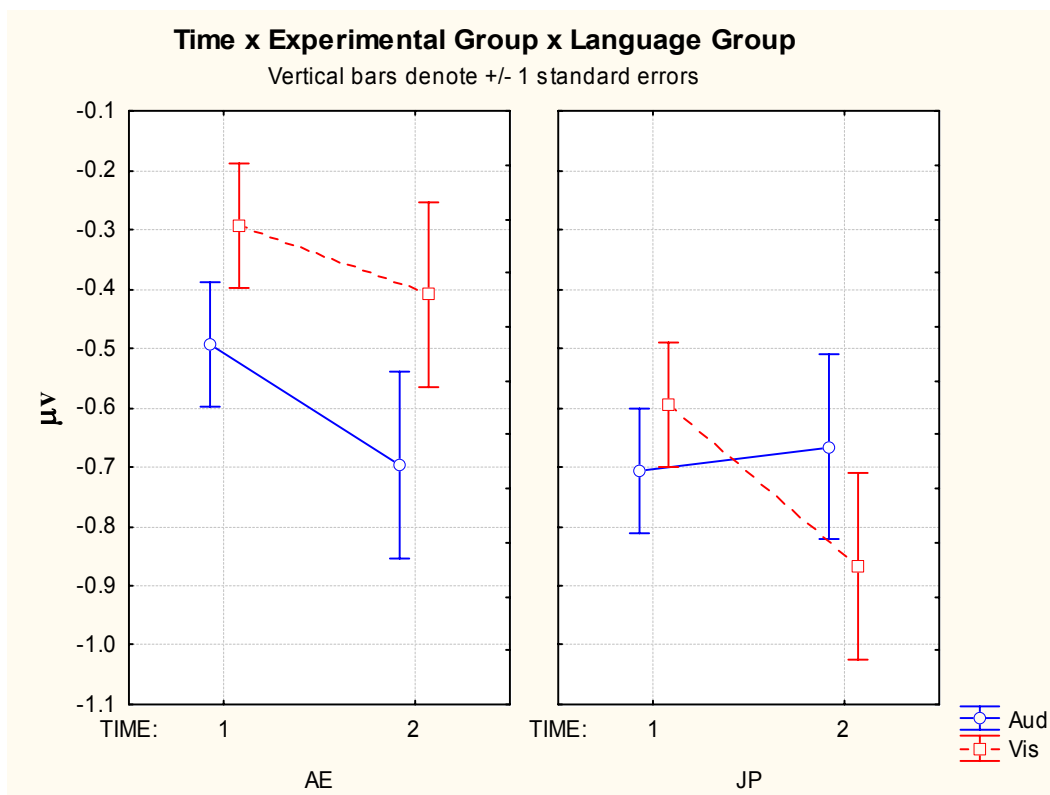


Figure 50. Vowel in the Auditory- and Visual-Attend Experiments: the AE group showed a larger MMN in the Auditory-Attend Experiment than in the Visual-Attend Experiment, while the JP group showed no effect between the two tasks.

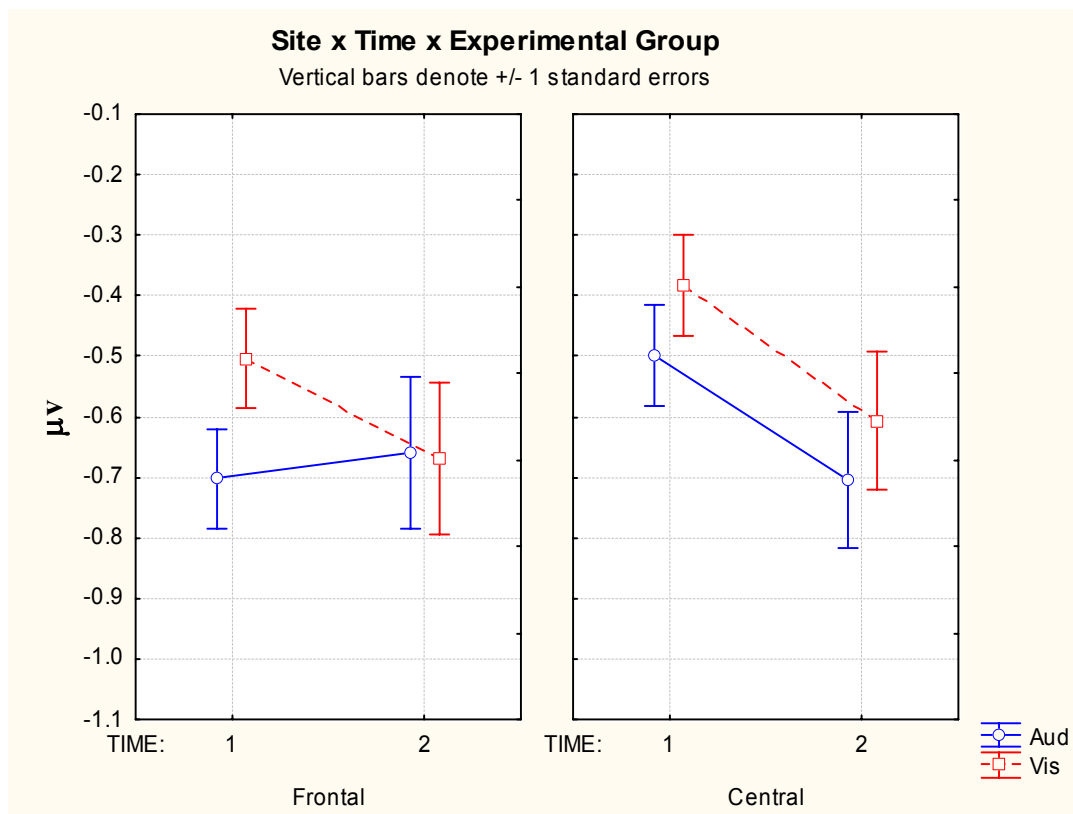


Figure 51. Vowel in the Auditory- and Visual-Attend Experiments: there was larger effect in the Auditory-Attend experiment than in the Visual-Attend experiment, especially in Time 1 for the frontal site.

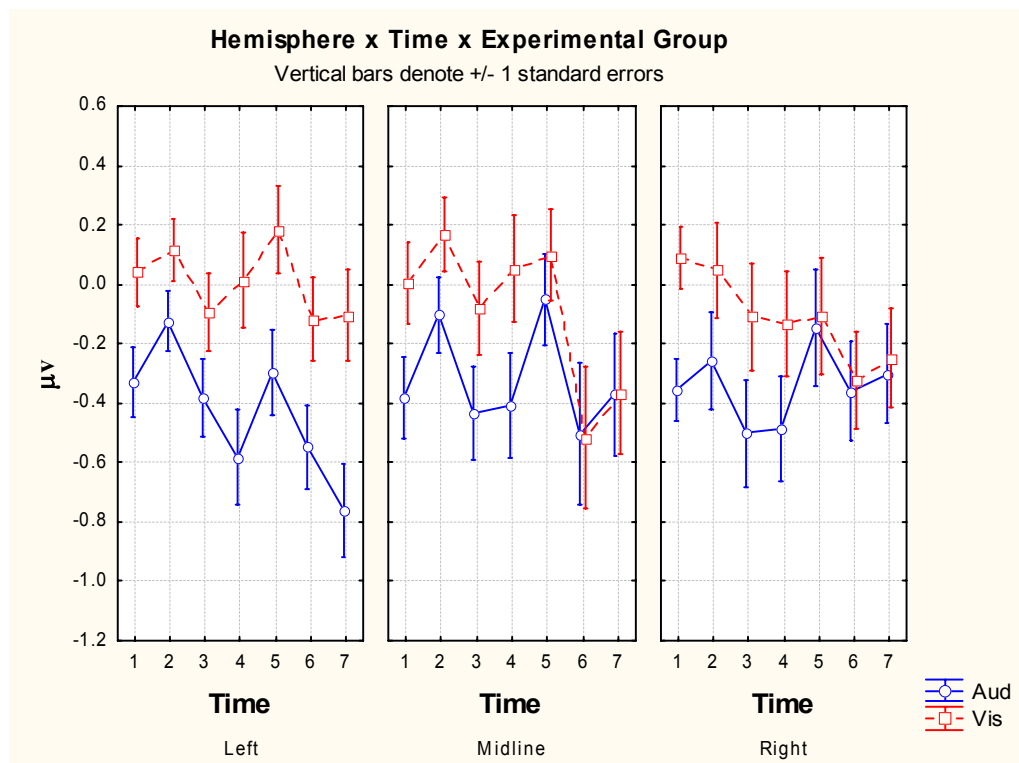


Figure 52. JP group [mifi]-as-standard condition in the Auditory- and Visual-Attend experiments: the Visual-Attend groups showed a larger MMN at all locations.

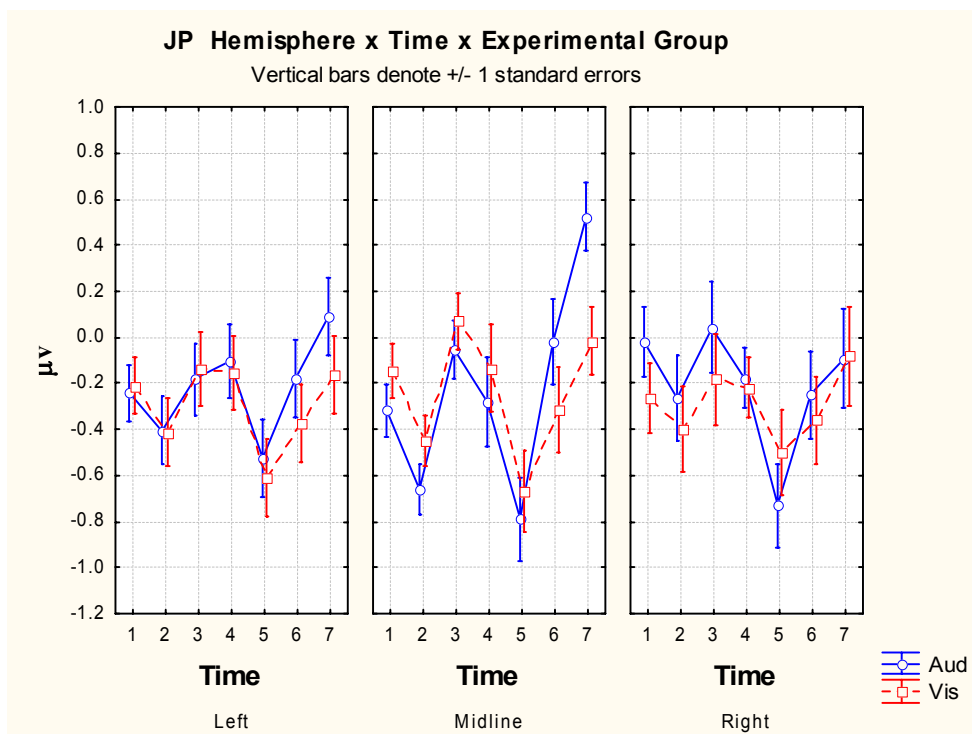
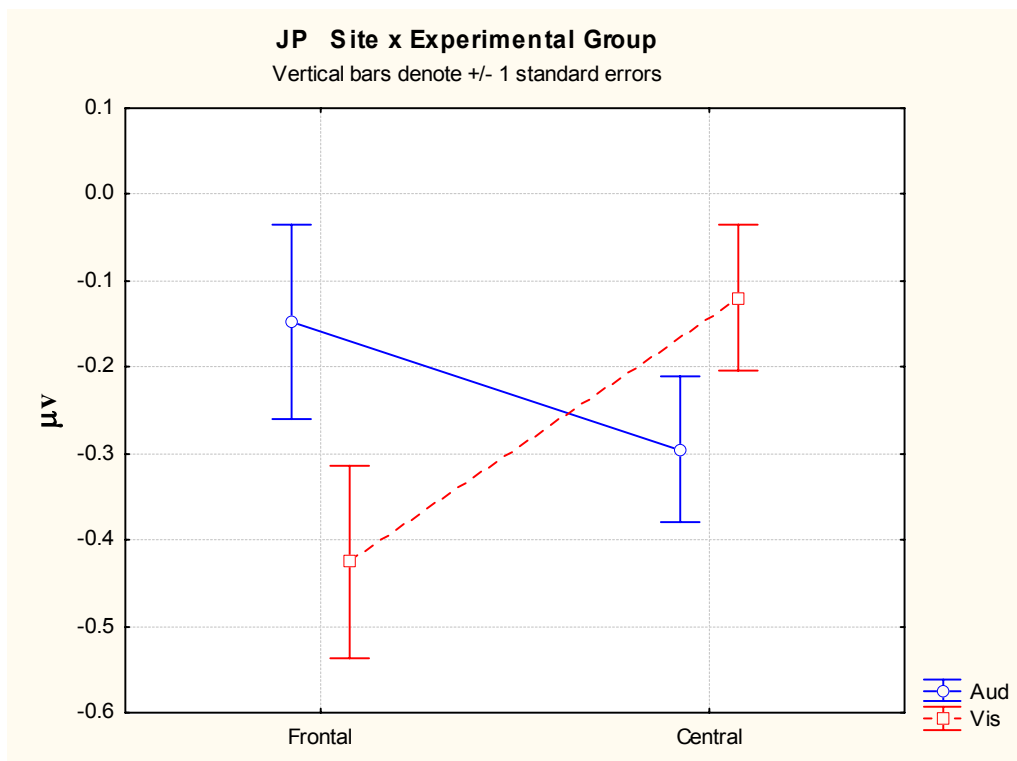


Figure 53. JP group [missi] in the Auditory- and Visual-Attend Experiments: the negativity was larger at midline and right sites for the Auditory-Attend task and at left hemisphere sites for the Visual-Attend task.

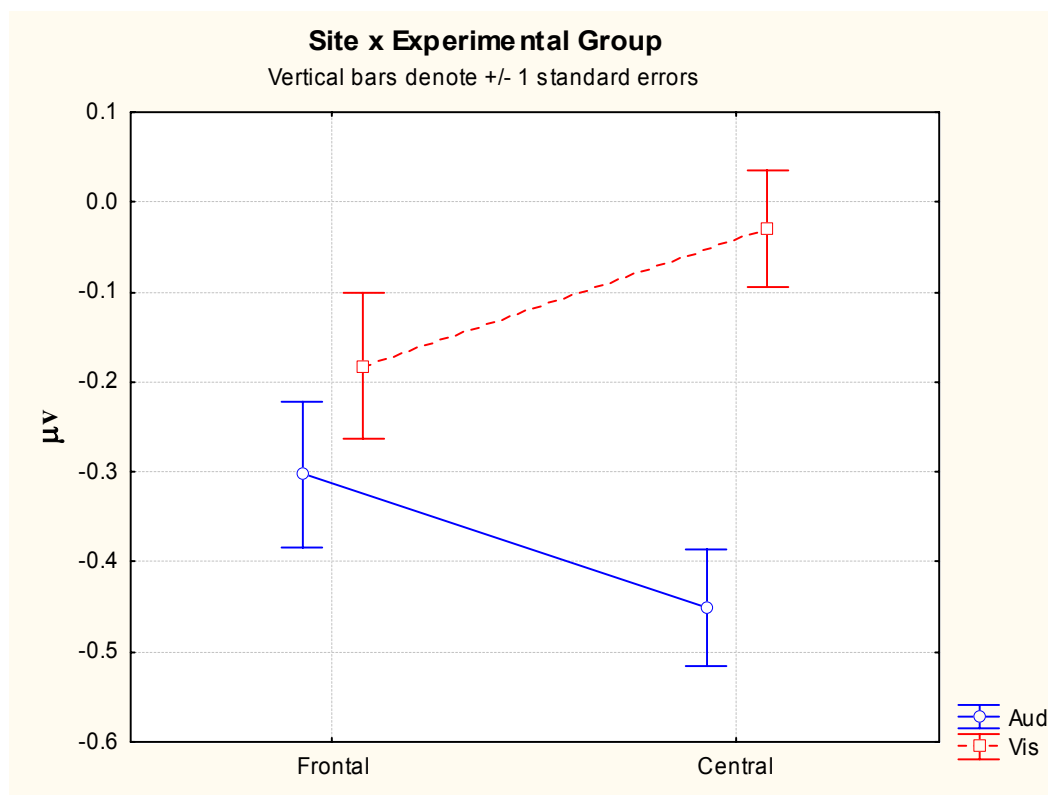


Figure 54. Consonant in the Auditory- and Visual- Attend Experiments: central sites showed larger differences as a function of attention condition for both language groups.

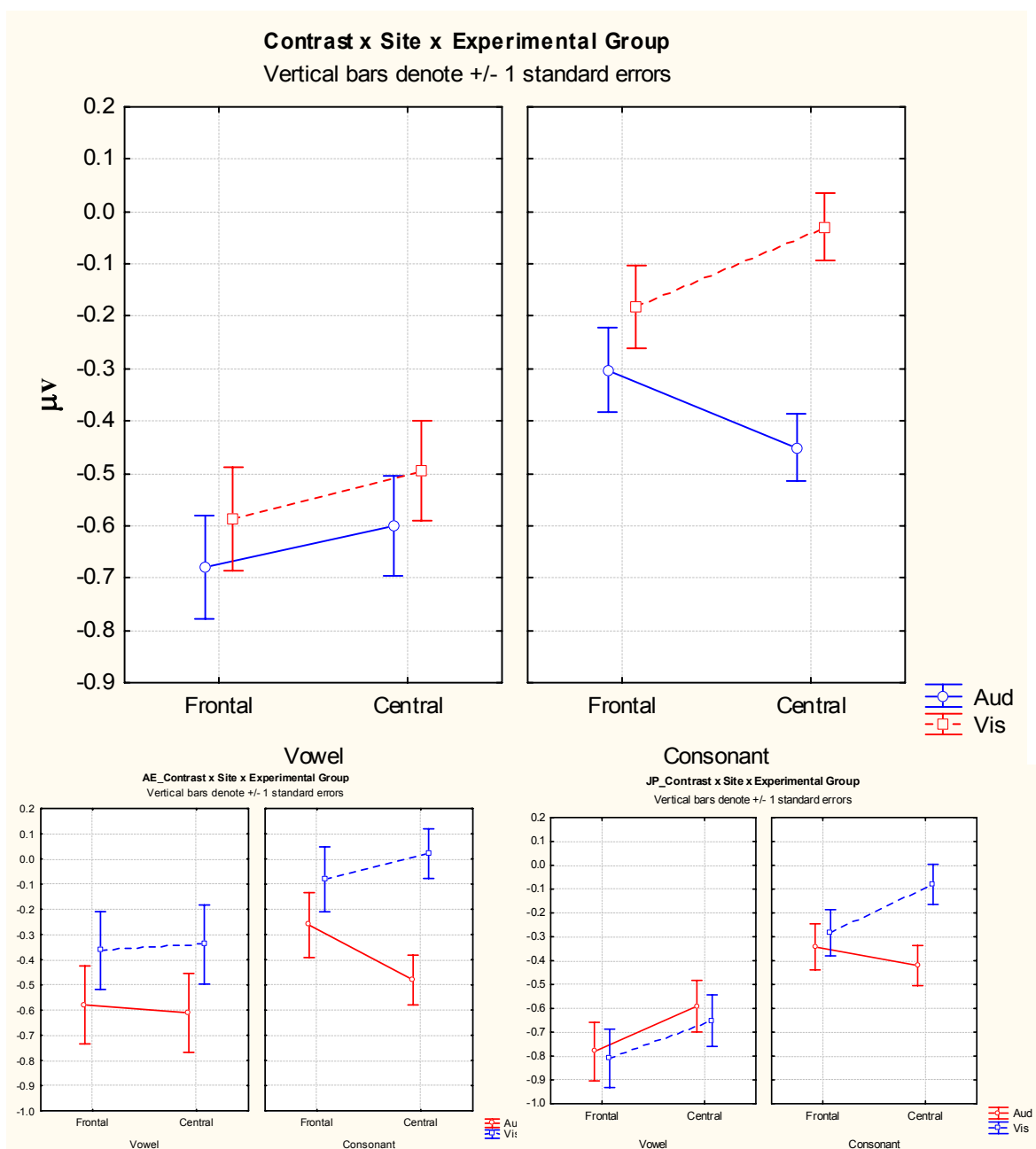


Figure 55. Vowel vs. Consonant in the Auditory- and Visual-Attend Experiments – Site effect: the vowel contrast elicited larger negativities than the consonant contrast for both language groups and there was a larger effect at central sites of attention manipulations for the consonant contrast.

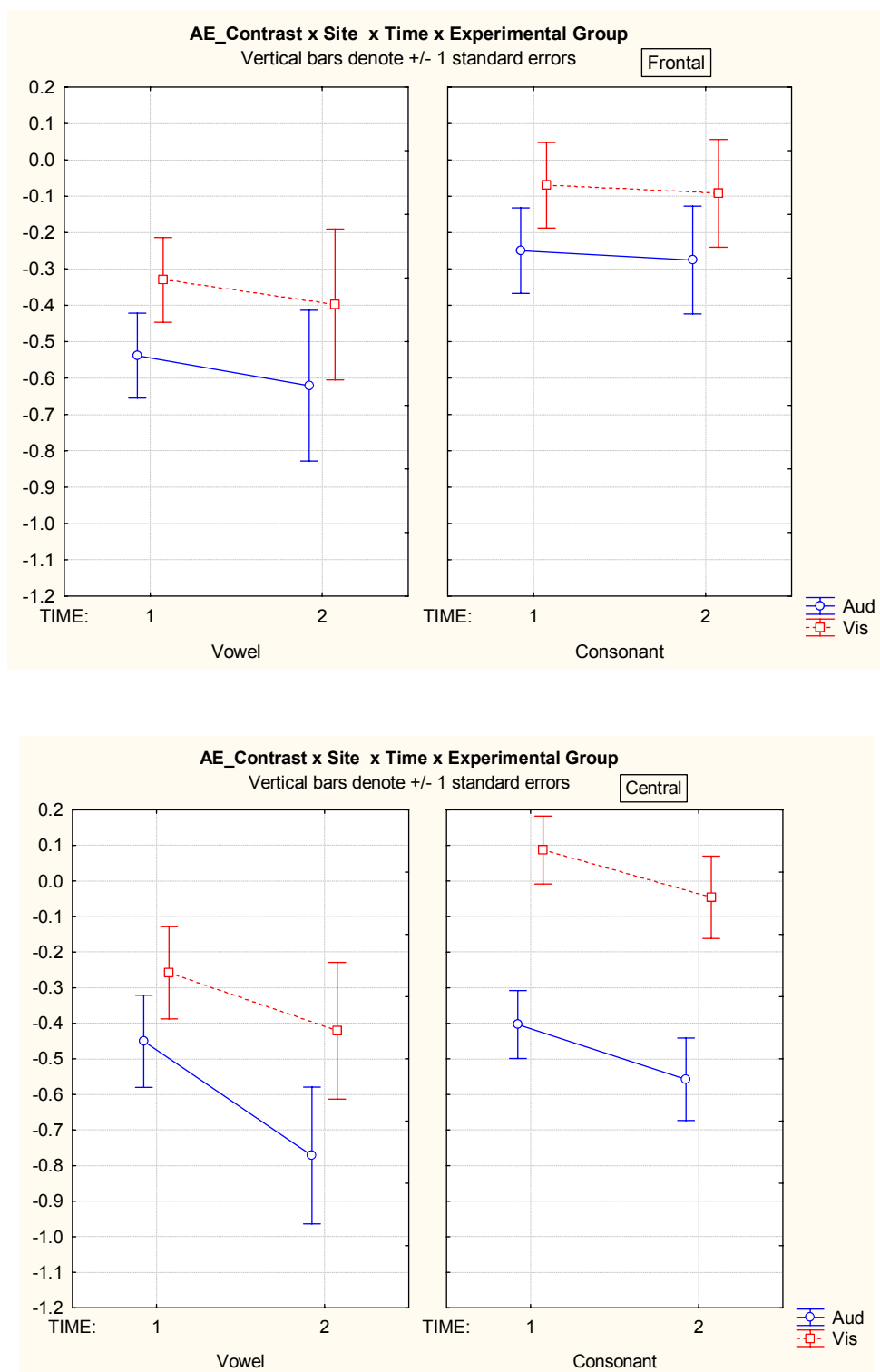


Figure 56a.

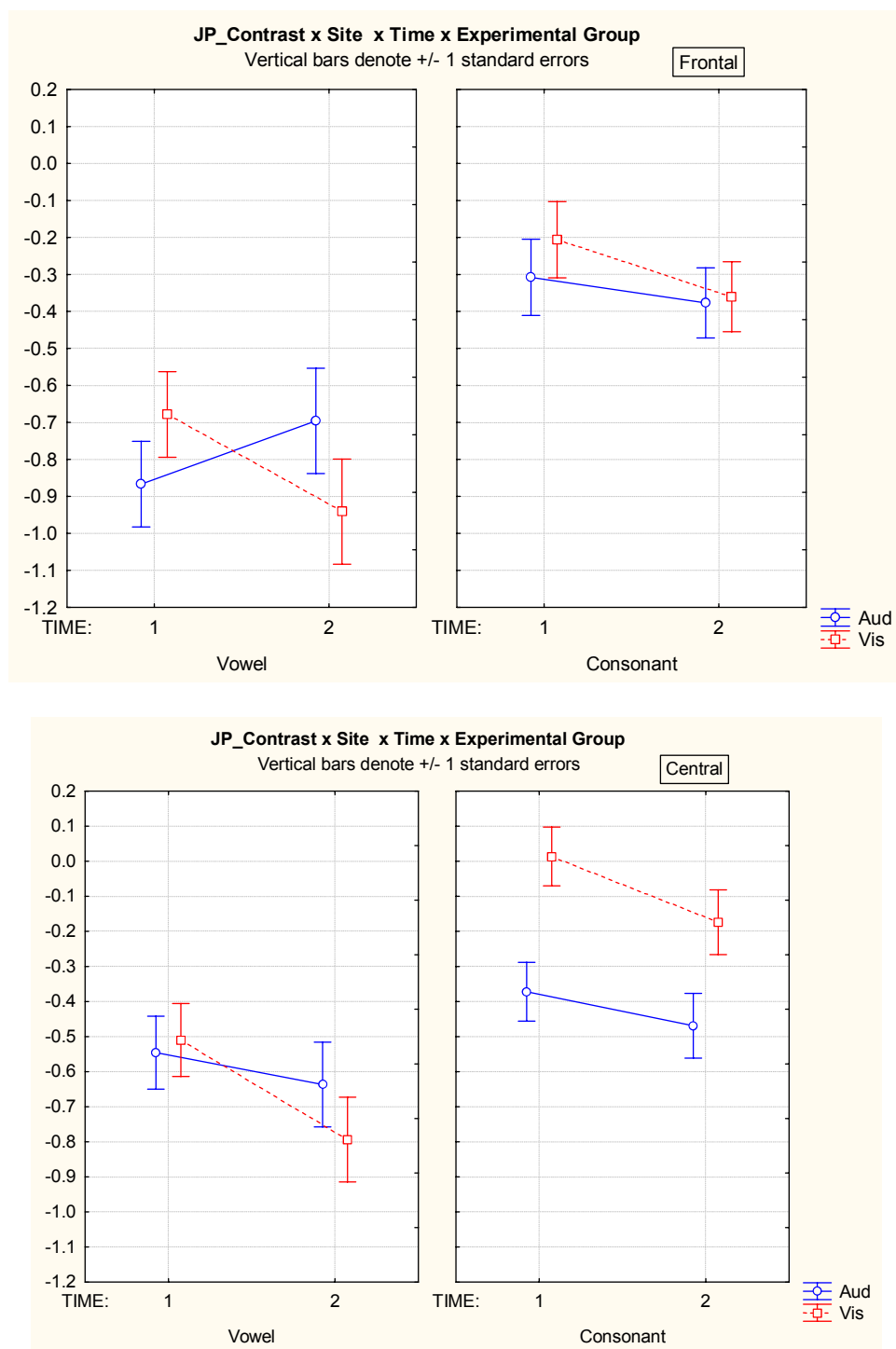


Figure 56b.

Figure 56 (a-b). Vowel vs. Consonant in the Auditory- and Visual-Attend Experiments – Time effect: both AE (a) and JP (b) Auditory-Attend groups showed larger negativities than Visual-Attend groups at frontal and central sites, except for JP group on the vowel contrast at Time 2.

APPENDIX

Appendix A. Protocol for Speakers (V = vowels; G = geminates; P = palatalized)

The sentence is 「これは_____ですね。」 “This is _____.”

Nonsense Stimuli speaker: JP date: / /

Original

- | | | | |
|-----------------------|---|-----------------------|---|
| 1. これは「ピロ (piro)」 | V | 21. これは「ピテ (pite)」 | G |
| ですね。 | | ですね。 | |
| 2. これは「ピーロ (piiro)」 | V | 22. これは「ピッテ (pitte)」 | G |
| ですね。 | | ですね。 | |
| 3. これは「ペキ (peki)」 | V | 23. これは「ポサ (posa)」 | G |
| ですね。 | | ですね。 | |
| 4. これは「ペーキ (peeki)」 | V | 24. これは「ポッサ (possa)」 | G |
| ですね。 | | ですね。 | |
| 5. これは「ポキヨー (pokyoo)」 | V | 25. これは「パコ (pako)」 | G |
| ですね。 | | ですね。 | |
| 6. これは「ポーキヨー | V | 26. これは「パッコ (pakko)」 | G |
| (pookyoo)」 | | ですね。 | |
| 7. これは「クジン (kujin)」 | V | 27. これは「ウケン (uken)」 | G |
| ですね。 | | ですね。 | |
| 8. これは「クージン | V | 28. これは「ウッケン (ukken)」 | G |
| (kuujin)」 | | ですね。 | |
| 9. これは「タド (tado)」 | V | 29. これは「ミシ (mifi)」 | G |
| ですね。 | | ですね。 | |
| 10. これは「タード (taado)」 | V | 30. これは「ミッシ (miffi)」 | G |
| ですね。 | | ですね。 | |
| 11. これは「パサ (pasa)」 | V | 31. これは「ヘコ (heko)」 | G |
| ですね。 | | ですね。 | |
| 12. これは「パーサ (paasa)」 | V | 32. これは「ヘッコ (hekkoo)」 | G |
| ですね。 | | ですね。 | |

- | | | | | | |
|-----|--------------------|-----|-----|-------------------|---|
| 13. | これは「フザン (fuizan)」 | V | 33. | これは「ヘケ (heke)」 | G |
| | ですね。 | | | ですね。 | |
| 14. | これは「フーザン (fuuzan)」 | V | 34. | これは「ヘッケ (hekke)」 | G |
| | ですね。 | | | ですね。 | |
| 15. | これは「トナ (tona)」 | V | 35. | これは「アケン (aken)」 | G |
| | ですね。 | | | ですね。 | |
| 16. | これは「トーナ (toona)」 | V | 36. | これは「アッケン (akken)」 | G |
| | ですね。 | | | ですね。 | |
| 17. | これは「ケキヨ (kekyo)」 | V | 37. | これは「プサ (puusa)」 | G |
| | ですね。 | | | ですね。 | |
| 18. | これは「ケーキヨ (keekyo)」 | V | 38. | これは「プッサ (puussa)」 | G |
| | ですね。 | | | ですね。 | |
| 19. | これは「パド (pado)」 | V | 39. | これは「ケコ (keko)」 | G |
| | ですね。 | | | ですね。 | |
| 20. | これは「パード (paado)」 | V | 40. | これは「ケッコ (kekko)」 | G |
| | ですね。 | | | ですね。 | |
| 61. | これは「パト (pato)」 | Fam | 41. | これは「ピシ (pifi)」 | G |
| | ですね。 | | | ですね。 | |
| 62. | これは「ポト (poto)」 | Fam | 42. | これは「ピッシ (piffi)」 | G |
| | ですね。 | | | ですね。 | |
| 61A | これは「クト (kuuto)」 | Fam | 63. | これは「ケク (keku)」 | G |
| | ですね。 | | | ですね。 | |
| 62A | これは「ケト (keto)」 | Fam | 64. | これは「ケップ (keppu)」 | G |
| | ですね。 | | | ですね。 | |
| 63A | これは「ペプ (pepu)」 | Fam | | | |
| | ですね。 | | | | |
| 64A | これは「ペック (pekku)」 | Fam | | | |
| | ですね。 | | | | |

- | | | | | | |
|-----|-----------------------------|---|-----|-----------------------------|---|
| 43. | これは「ビヨウジ (b'ooji)」
ですね。 | P | 55. | これは「リユク (r'jukku)」
ですね。 | P |
| 44. | これは「ビヨウジ (bijooji)」
ですね。 | P | 56. | これは「リユク (rijuukku)」
ですね。 | P |
| 45. | これは「ピヤク (p'jaku)」
ですね。 | P | 57. | これは「ヒヤム (h'amu)」
ですね。 | P |
| 46. | これは「ピヤク (pijaku)」
ですね。 | P | 58. | これは「ヒヤム (hijamu)」
ですね。 | P |
| 47. | これは「ピヨ一(p'oo)」
ですね。 | P | 59. | これは「ピユ一 (p'uuu)」
ですね。 | P |
| 48. | これは「ピヨ一(pijoo)」
ですね。 | P | 60. | これは「ピユ一 (pijuuu)」
ですね。 | P |
| 49. | これは「チャー (t'jaa)」
ですね。 | P | 65. | これは「バウ (bau)」
ですね。 | P |
| 50. | これは「チャー (t'jiaa)」
ですね。 | P | 66. | これは「ビヤウ (b'auu)」
ですね。 | P |
| 51. | これは「ミユ一 (m'uuu)」
ですね。 | P | | | |
| 52. | これは「ミユウ (mijuuu)」
ですね。 | P | | | |
| 53. | これは「ビヤウジ (b'auji)」
ですね。 | P | | | |
| 54. | これは「ビヤウジ (bijauji)」
ですね。 | P | | | |

Appendix B. Acoustic analysis of target real words: [kado vs. kaado] for the vowel length contrast and [nifi vs. niʃʃi] for the consonant length contrast by the same speaker of the target nonsense words. VOT and vowel consonant durations (ms) and pitch of two syllables (Hz) of selected four stimuli of each phonetic category. Standard deviation (SD) is shown for the contrasting target segments.

kaado vs. kado

long	VOT	/aa/	closure	/o/	Total	Pitch
1	22	155	19	79	275	413, 270
2	19	157	19	93	288	393, 267
3	17	158	24	90	289	365, 273
4	19	153	26	83	281	381, 262
Mean	19	156	22	86	283	388, 268

short	VOT	/a/	closure	/o/	Total	Pitch
1	23	86	33	101	242	380, 288
2	23	88	28	112	251	390, 292
3	24	88	27	110	249	288, 288
4	23	91	20	101	236	374, 301
Mean	23	88	27	106	244	358, 292

Total word duration

L/S Ratios = 1.76

L/S Ratios = 1.16

nifʃi vs. nifi

long	Nasal murmur	/i/	/ʃʃ/	/i/	Total	Pitch
1	57	90	123	55	325	253, 309
2	51	99	136	42	328	258, 295
3	51	100	140	52	343	255, 322
4	51	94	146	65	355	283, 309
Mean	52	96	136	53	337	262, 309

short	Nasal murmur	/i/	/ʃ/	/i/	Total	Pitch
1	40	86	84	50	259	289, 291
2	38	77	93	52	260	277, 311
3	41	75	84	54	254	297, 286
4	45	84	95	47	271	251, 286
Mean	41	81	89	51	261	279, 294

Total word duration

L/S Ratio = 1.54 L/S Ratios = 1.29

Appendix C. Acoustic analysis of Real words of previous study: [kado vs. kaado] for the vowel length contrast and [niʃi vs. niʃʃi] for the consonant length contrast by a female speaker. VOT and vowel consonant durations (ms) and pitch of two syllables (Hz) of selected five stimuli of each phonetic category. Standard deviation (SD) is shown for the contrasting target segments.

kaado vs. kado

long	VOT	/aa/	closure	/o/	Total
1	21	144	40	73	278
2	23	139	19	64	245
3	25	142	25	69	261
4	29	134	25	77	264
5	28	127	38	68	261
	25	137	30	70	262
short	VOT	/a/	Closure	/o/	Total
1	23	79	28	73	202
2	25	90	0	80	195
3	24	87	18	85	214
4	28	85	26	88	227
5	27	74	0	67	168
	25	83	14	78	201
L/S Ratios = 1.65			L/S Ratios = 1.30		

niʃʃi vs. niʃi

long	Nasal murmur	/i/	/ʃʃ/	/i/	Total
1	47	93	129	52	322
2	46	86	130	57	319
3	49	86	124	60	319
4	42	80	129	23	275
5	45	71	121	45	281
	46	83	127	48	303
short	Nasal murmur	/i/	/ʃ/	/i/	Total
1	31	75	83	62	251
2	37	82	72	68	259
3	41	63	85	50	239
4	31	67	84	44	225
5	35	68	90	57	249
	35	71	83	56	245
L/S Ratios = 1.53			L/S Ratios = 1.24		

Appendix D. Consent Forms (For native speaker of English)

INFORMED CONSENT – Adult

The purpose of this information is to help you decide whether you want to volunteer for a research study. Read carefully and ask the investigator about anything you don't understand

Title: Perception of Japanese temporally-cued phonetic contrasts by Japanese and American English listeners: Behavioral and Electrophysiological Measures

Principal Investigator: Miwako Hisagi

Research Assistants: _____

Study Location: CUNY Graduate Center, Developmental Neurolinguistics Laboratory

General Information

- We are asking you to take part in a study on how your ability to hear the difference between certain speech sounds is related to the sounds you produce in your dialect.

Experimental Procedure

- We will use a test with the electroencephalogram (EEG), to see how fast certain speech sounds can be discriminated. The EEG is a safe test that has been used for over thirty years with children and adults.
 - You will wear a net or cap of sensors and your brainwaves are recorded while listening to speech sounds.
 - The sensors pick up brainwaves that occur naturally in everyone's brains. It takes about ten to 20 minutes to put on the net or cap.
 - You may watch a video with the video sound turned off, or may perform a computer task while speech sounds are played over speakers.
 - In the computer task you may press a button when you hear certain sounds.
 - You will also have your hearing tested.
 - The testing will take no more than four hours, including set-up time and breaks.
- We would also like you to fill out a questionnaire, and to produce a sample of your language or languages (if bilingual). The sample will be audio recorded and used to check your dialect, and not for any other purposes. This will take 10-15 minutes. Sign here if you agree to being audio recorded _____.
- With your permission, in the future we may contact you to take part in other studies. Sign here if you agree to be contacted _____.

Benefits and Risks

- Your participation may help to better understand speech and language abilities.
- All of the tests we use have no known risks and have been used for over thirty years. However, you may feel some discomfort with the net of sensors. In this case, we can put a small number of sensors on, which may feel more comfortable.
- We will give referrals for further testing of your hearing, if we identify any possible problem.
- For each visit, you will receive compensation for your time (\$10.00 per hour).

Confidentiality.

- All information is kept in a locked filing cabinet, and will be used only for the purposes of the study.
- Only researchers directly involved with the study will have access to the research information in your records.
- The research staff involved with this study may also have access to your records.
- Results of this study may be published, but will not include information that can identify you.
- You may receive a copy of the published paper on request.

Volunteering

- You should only take part in this research if you want to, and for no other reason.
- YOU MAY STOP PARTICIPATING AT ANY TIME, FOR ANY REASON.
- You may be removed from this study at any time, without your consent

Contact Information

- If you have questions about this research, contact **Miwako Hisagi at (212)817-8842; e-mail: mhisagi@gc.cuny.edu** or **Valerie Shafer, Ph.D. at (212) 817 8833; e-mail: vshafer@gc.cuny.edu**
- To learn about your rights as a research volunteer, contact **Kay Powell, IRB Administrator at (212) 817 7525; e-mail: kpowell@gc.cuny.edu**

Consent – By signing this form you agree that:

- You have fully read this form, or someone has read it to you in your native language.
- You had the opportunity to ask questions about this research, and have received satisfactory answers.
- You understand that you are being asked to participate in research
- You understand that risks and benefits in the procedures described above.
- You may request a signed copy of this form, which is yours to keep.

Print Your Name

Sign Your Name

Date

Investigator Statement

- I have fully explained to the participant the nature of this research. To the best of my knowledge, the participant signing this consent form understands the nature, risks, and benefits of this study.

Investigator's Name

Printed Name

Date

Appendix E. Language Background Questionnaire (for native speaker of English)

Language Background Questionnaire

Name of Experiment: _____

Date: _____

Name: _____ Date of Birth: _____ Gender: _____

Address: _____

Telephone Numbers: (Home) _____ (Work) _____

Handedness: Right Left

Medical history (Optional): _____

Birthplace: _____

Town/City

State/Country

Father's Birthplace: _____

Languages your father speaks fluently: _____

Mother's Birthplace: _____

Languages your mother speaks fluently: _____

Places in which you have lived for more than 1 year:

City/State/Country

Years

_____ from _____ to _____

_____ from _____ to _____

_____ from _____ to _____

_____ from _____ to _____

If you have lived in more places please check here _____ and continue on the back.

As a child, what languages were spoken in your home? (for example, by parents, guardians, grandparents, or relatives) _____

What languages do you speak fluently and understand without effort?

1. _____ 2. _____ 3. _____

What language do you consider your "mother tongue"? _____

What language(s) did you speak/understand as a child (before going to school)?

1. _____ 2. _____ 3. _____

What language(s) were used in your classrooms in elementary school?

1. _____ 2. _____

What language(s) did you study as a foreign language in school?

1. _____ (Circle all applicable: elementary, junior high, senior high, college).
Number of semesters _____. Did you have a native speaker of the language as a teacher or tutor? No ____ Yes ____ (number of semesters with native speaker ____)

Rate your fluency and understanding of this language by checking one of the following which best describes your mastery of this language:

a. speak/understand like a native speaker _____

b. speak with a mild accent and understand native speakers with little or no difficulty _____

- c. speak with an accent and understand, but with some effort _____
- d. speak and understand, but with effort _____
- e. cannot speak or understand this language at all _____

Foreign Language Study (continued)

2. _____ (Circle all applicable: elementary, junior high, junior high, college).
 Number of semesters _____. Did you have a native speaker of the language as a teacher or tutor? No ____ Yes ____ (number of semesters with native speaker ____)

Rate your fluency and understanding of this language by checking one of the following which best describes your mastery of this language:

- a. speak/understand like a native speaker _____
- b. speak with a mild accent and understand native speakers with little or no difficulty _____
- c. speak with an accent and understand, but with some effort _____
- d. speak and understand, but with considerable effort _____
- e. cannot speak or understand this language at all _____

3. _____ (Circle all applicable: elementary, junior high, junior high, college).
 Number of semesters _____. Did you have a native speaker of the language as a teacher or tutor? No ____ Yes ____ (number of semesters with native speaker ____)

Rate your fluency and understanding of this language by checking one of the following which best describes your mastery of this language:

- a. speak/understand like a native speaker _____
- b. speak with a mild accent and understand native speakers with little or no difficulty _____
- c. speak with an accent and understand, but with some effort _____
- d. speak and understand, but with considerable effort _____
- e. cannot speak or understand this language at all _____

Have you ever studied Phonetics (the scientific study of speech sounds) in high school or college level in a linguistics, speech science, or foreign language class? YES / NO

If YES, have you ever done phonetic transcription? YES / NO

If YES, how much? _____

Do you have normal hearing? YES / NO

Did you at any time have therapy for a speech or language problem? _____

 What do you consider your racial/ethnic background to be? Check all that apply.

(Optional: you need not answer)

Caucasian _____

Native American _____

African American _____

Pacific Islander _____

Hispanic _____

Asian American _____

Other- please specify _____

Appendix F. The number of stimuli per block

Condition 1: [tado]-as-standard and [taado]-as-deviant

	Standard	Deviant	Sub-Total
Block 1	96	16	112
Block 2	63	9	72
Block 3	106	20	126
Block 4	82	13	95
Block 5	101	19	120
Block 6	78	14	92
Block 7	79	12	91
Block 8	94	17	111
Block 9	84	16	100
Block 10	79	13	92
Block 11	94	19	113
Block 12	87	15	102
Block 13	72	12	84
Block 14	75	15	90
Total	1190	210	1400

Condition 2: [taado]-as-standard and [tado]-as-deviant

	Standard	Deviant	Sub-Total
Block 1	87	13	100
Block 2	100	20	120
Block 3	86	17	103
Block 4	94	19	113
Block 5	87	15	102
Block 6	67	11	78
Block 7	64	12	76
Block 8	96	16	112
Block 9	63	9	72
Block 10	94	16	110
Block 11	82	13	95
Block 12	110	21	131
Block 13	78	14	92
Block 14	82	14	96
Total	1190	210	1400

Condition 3: [miʃi]-as-standard and [miʃi]-as-deviant

	Standard	Deviant	Sub-Total
Block 1	90	16	106
Block 2	89	17	106
Block 3	86	17	103
Block 4	94	19	113
Block 5	87	15	102
Block 6	72	12	84
Block 7	67	11	78
Block 8	96	16	112
Block 9	71	9	80
Block 10	101	19	120
Block 11	82	13	95
Block 12	95	17	112
Block 13	78	14	92
Block 14	82	15	97
Total	1190	210	1400

Condition 4: [miʃi]-as-standard and [miʃi]-as-deviant

	Standard	Deviant	Sub-Total
Block 1	96	16	112
Block 2	63	9	72
Block 3	106	20	126
Block 4	82	13	95
Block 5	95	17	112
Block 6	91	17	108
Block 7	72	11	83
Block 8	94	17	111
Block 9	92	18	110
Block 10	79	13	92
Block 11	102	19	121
Block 12	87	15	102
Block 13	67	13	80
Block 14	64	12	76
Total	1190	210	1400

Appendix G. Instructions for ERPs (auditory)

Oral instruction for Auditory-Attend Experiment

Audio:

tado vs. taado (along with visual stimuli: pentagon/Hexagon shapes)

mishi vs. mishi (along with visual stimuli: circle/oval shapes)

You will hear the Japanese nonsense words “tado or taado” from the speaker. Your task is to count “taado” (as deviant). Each word may be slightly different from each other, but “tado” is “tado” and “taado” is “taado”. Try to pay attention to the sound “a”. Count the one which has longer “a”.

Let’s practice together. (I will give 15 stimuli (12 standards and 3 deviants)).

There are 14 blocks in this first session and each block is approximately 2-3 minutes. In each block, I’ll open the door and you need to report me how many long “a” (as deviant) you heard.

There are a few things I have to ask.

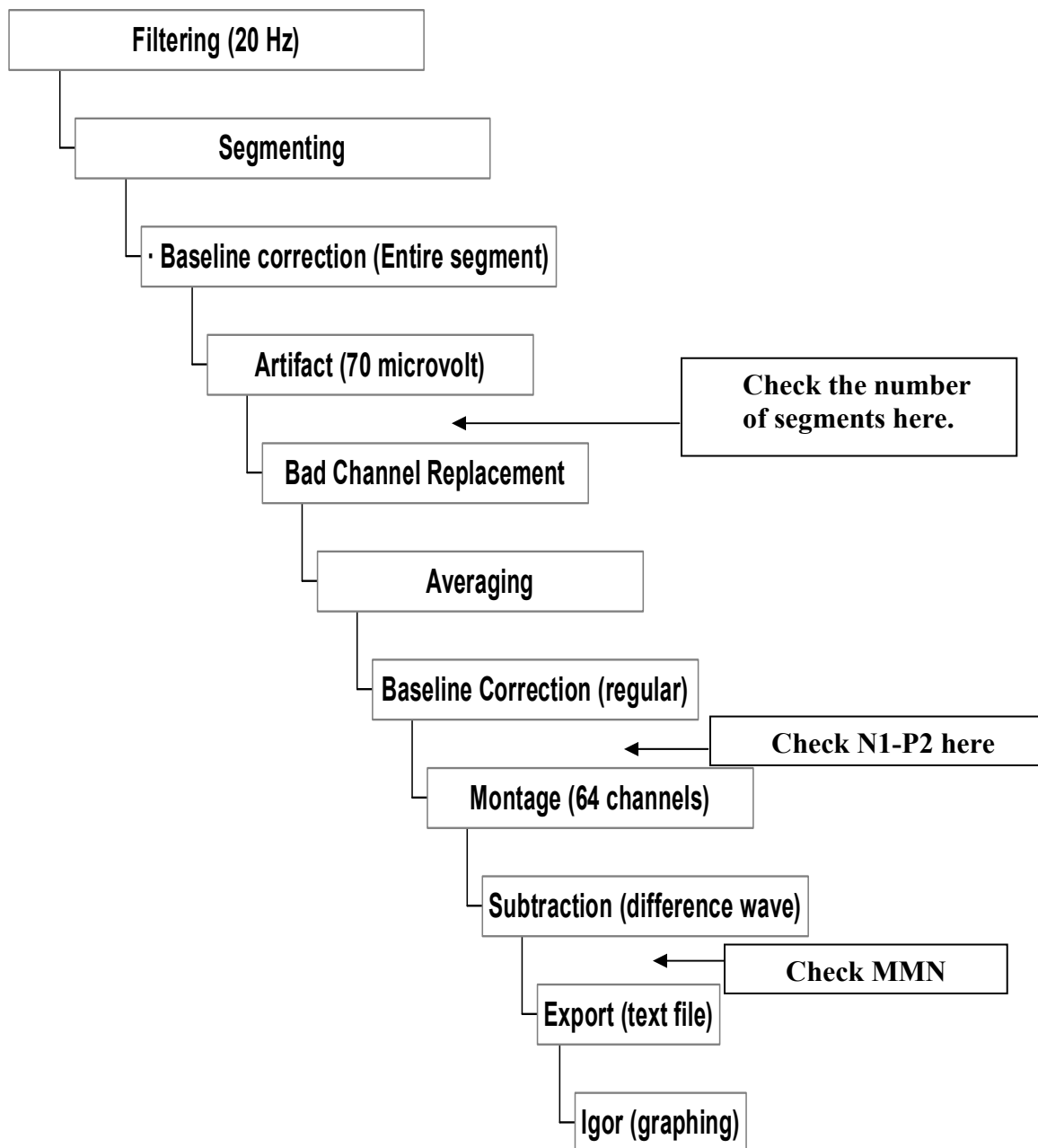
- Please try not to blink too much. If you need to blink, please make a baby blink (like this), but do so infrequently.
- Also please try not to move much. Any motor movement around the face, neck and head is much larger than the brain activity and so we lose information every time you move.
- And please look at the center of the screen.
- Just relax and look at this screen to count the number of oval shape “in your mind”.

In each block, if you count them correctly, I’ll give you “candy”.

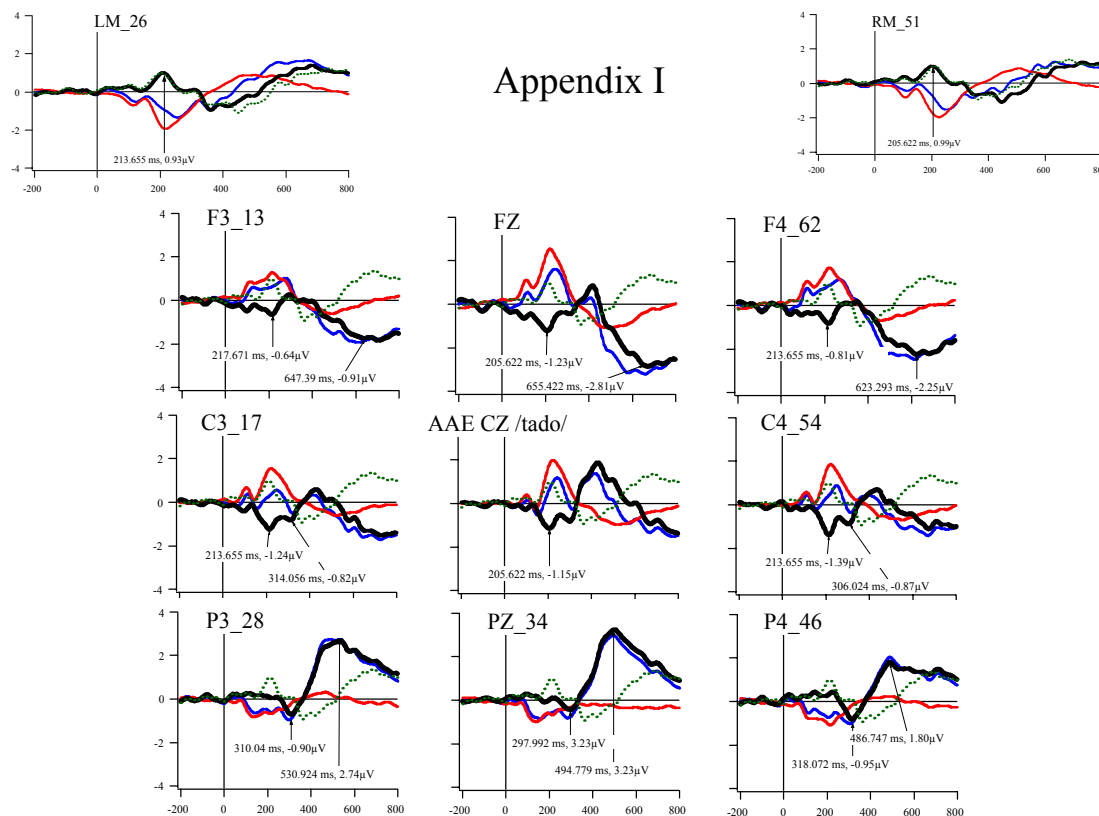
To fix your eyes position, please look at this screen.

Do you have any questions?

Signal Processing



Appendix I. AAE [tado]-as-standard condition: All nine sites with LM and RM: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4.



Appendix J. Instructions for ERPs (visual)

Oral instruction for Visual-Attend Experiment

Visual:

Pentagon vs. Hexagon (along with vowel stimuli)

Circle vs. Oval shape (along with consonant stimuli)

You will see the same shape “pentagon or hexagon” on this screen, but they have different sizes.

Your task is to count the hexagon shape. The size doesn’t matter.

Let’s practice together. (I will give 15 stimuli (12 standards and 3 deviants)).

There are 14 blocks in this first session and each block is approximately 2-3 minutes. In each block, I’ll open the door and you need to report me how many hexagon (as deviant) you saw.

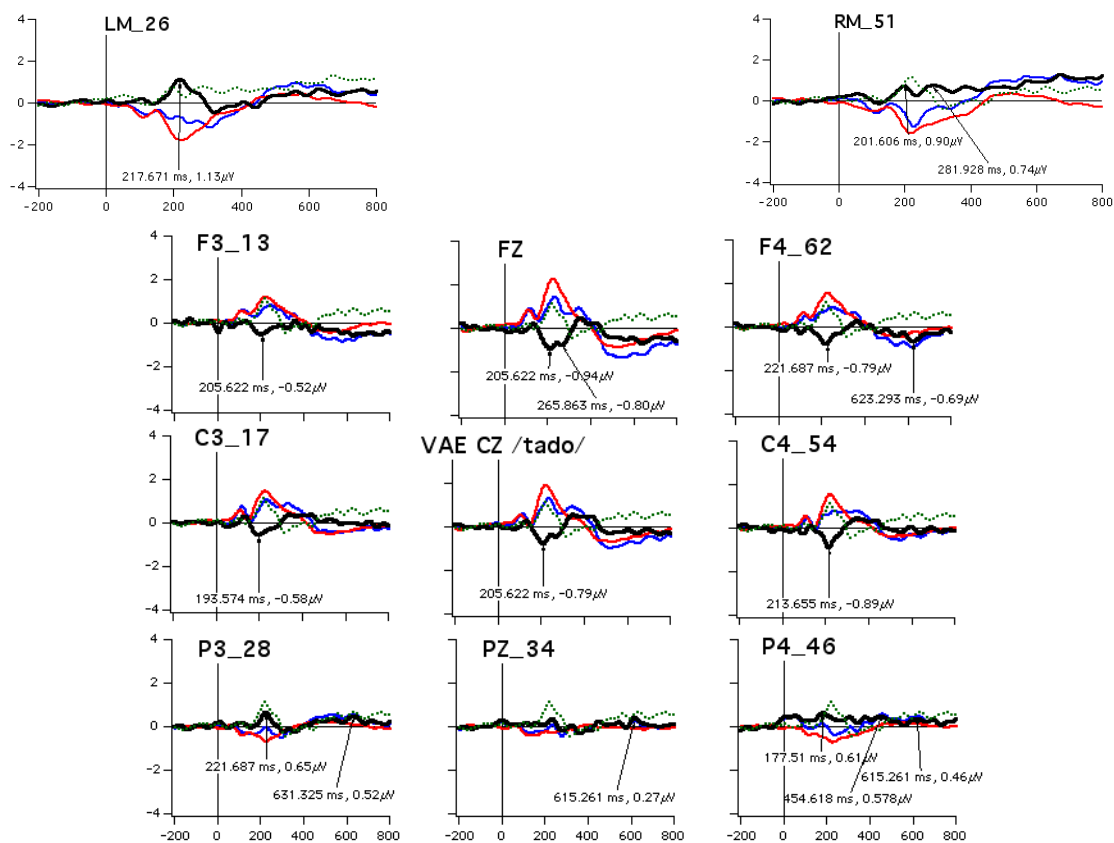
There are a few things I have to ask.

- Please try not to blink too much. If you need to blink, please make a baby blink (like this), but do so infrequently.
- Also please try not to move much. Any motor movement around the face, neck and head is much larger than the brain activity and so we lose information every time you move.
- And please look at the center of the screen.
- Just relax and look at this screen to count the number of hexagon shape “in your mind”.

In each block, if you count them correctly, I’ll give you “candy”.

There will be a background sound, but just pay attention to this screen.
Do you have any questions?

Appendix K. VAE [tado]-as-standard condition: All nine sites with LM and RM: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4



REFERENCES

- Alho, K., Woods, D. L., & Algazi, A. (1994). Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiology* 31(5), 469-479.
- Alho, K., Winkler, I., Escera, C., Huutilainen, M., Virtanen, J., Jaaskelainen, I. P., Pekkonen, E., & Ilmoniemi, R. J. (1998). Processing of novel sounds and frequency changes in the human auditory cortex: Magnetoencephalographic recordings. *Psychophysiology*, 35, 211 -224.
- Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual speech integration falters under high attention demands. *Current Biology*, 15(9), 839-843.
- Arnott, S. E., & Alain, C. (2002). Falling outside the spotlight: Effect of spatial attention on the MMN. *NeuroReport*, 13, 2209–2212.
- Bell, A.J., & Sejnowski, T.J. (1995). An information maximisation approach to blind separation and blind deconvolution, *Neural Computation*, 7, 6, 1129-1159.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press.
- Best, C. T. (2003). Peeling back the layers of time: Integrating developmental, experiential, and stimulus time in speech perception. *Journal of Phonetics*, 31, 613-618.
- Best, C., & McRoberts, G. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language & Speech*, 46, 183-216.
- Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, 17, 905-917.
- Bohn, O.-S., & Flege, J. (1990). Interlingual Identification and the Role of Foreign Language Experience in L2 Vowel Perception. *Applied Psycholinguistics*, 11, 303-328.
- Borden, G. J., Harris, K. S., & Raphael, L. J. (2003). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*, 4th Edition. New York: Lippincott Williams & Wilkins.
- Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics* 7, 207-240.

- Čeponienė, R., Cheour, M., & Näätänen, R. (1998). Interstimulus interval and auditory event-related potentials in children: Evidence for multiple generators. *Electroencephalography and Clinical Neurophysiology*, *108*, 345-354.
- Cheour, M., Korpilahti, P., Martynova, O., & Lang, A. (2001). Mismatch Negativity (MMN) and Late Discriminative Negativity (LDN) in investigating speech perception and learning in children and infants. Invited Review. *Audiology & Neuro-Otology*, *6*, 2-11.
- Datta, H., Wroblowski, M., Hisagi, M., Tessel, C., and Shafer, V. L. (in preparation). Function and Topography of Mismatch vs. Late Negativity and the Role of Attention in Vowel Perception.
- Dehaena-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A Cross-linguistic Study. *Journal of Cognitive Neuroscience*, *12*, 635-647.
- Enomoto, K. (1992). Interlanguage Phonology: The Perceptual Development of Durational Contrasts by English-Speaking Learners of JP. *Edinburgh working papers in applied linguistics* *3*, 25-35.
- Ferree, T. C., Luu, P. L., Russell, J., & Tucker, D. M. (2001). Scalp-electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology* *112*(3), 536-544.
- Friedrich, M., Weber, C., & Friederici, A. D. (2004). Electrophysiological evidence for delayed mismatch response in infants at-risk for specific language impairment. *Psychophysiology*, *41*(5), 772-782.
- Garrido, K., Hisagi, M., & Shafer, V. L. (in preparation). *ERP Indices of Vowel Processing in Bilinguals*. 5th International Symposium on Bilingualism (ISB5) at Barcelona, Spain. March 20-23, 2005.
- Glass, K., Frishkoff, G. A., Frank, R. M., Davery, C., Dien, J., & Maloney, A. D. (2004). A framework for evaluating ICA methods of artifact removal from multichannel EEG. In C. G. Puntonet & A. Prieto (Eds.), *ICA 2004, LNCS 3195* (pp. 1033-1040). Berlin & Heidelberg: Springer-Verlag.
- Gordon, P., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, *25*, 1-42.
- Hill, P. R., McArthur, G. M., & Bishop, D. V. M. (2004). Phonological categorization of vowels: a mismatch negativity study. *Neuroreport*, *15*, 2195-2199.
- Hirata, Y. (1990). Perception of Geminated Stops in Japanese Word and Sentence Levels By English-speaking Learners of Japanese Language. *Nihon Onsei Gakkai*. 4-10.

- Hirata, Y. (2004a). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *Journal of the Acoustical Society of America*, 116(4), 2384-2394.
- Hirata, Y. (2004b). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, 32(4), 565-589.
- Hisagi, M., & Strange, W. (in revision). Perception of Temporally-cued Contrasts in Japanese by American English Listeners. The City University of New York Graduate Center. Speech & Hearing Sciences, Second Project.
- Hisagi, M., Strange, W., Akahane-Yamada, R. and Kubo, R. (2005). Perceptual assimilation and categorial discrimination of American vowels by Japanese listeners. 149th Meeting Acoustic Society of America, Vancouver, Canada, 16-20 May 2005. (poster)
- Inouchi, M., Kubota, M., Ferrari, P., & Roberts, T. (2003). Magnetic mismatch fields elicited by vowel duration and pitch changes in Japanese words in humans: comparison between native and non-speakers of Japanese, *Neurosci. Lett.* 353, 165-168.
- Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Cambridge: MIT Press.
- Kahneman, D., & Treisman, A., (1984). Changing views of attention and automaticity. In R. Parasuraman & R. Davies (Eds.) *Varieties of Attention*. New York: Academic Press, 29- 61.
- Korpilahti, P., Lang, A. H., & Aaltonen, O. (1995). Is there a late-latency mismatch negativity (MMN) component? *Electroencephalography and Clinical Neurophysiology*, 95-96.
- Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences* 9: 75-82.
- Lavie, N. & DeFockert J. W. (2003). Contrasting effects of sensory limits and capacity limits in visual selective attention. *Perception & Psychophysics*, 65, 202-212.
- Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr. Clin. Neurophysiol*, 48, 609-621.
- Menning, H., Imaizumi, S., Zwitserlood, P. & Pantev, C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the JP language. *Learning Memory*, 9, 253-267.

- Muroi, K. (1995). Problems of Perception and Production of JP Morae –The Case of Native English Speakers. (written in Japanese) *Sophia Linguistica*, 38, 41-60.
- Näätänen, R. (1988). Implications of ERP data for psychological theories of attention. *Biological Psychology*, 26 (1-3), 117-163.
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive functions, *The Behavioral and Brain Sciences*, 13, 201-288.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432-434.
- Näätänen, R., & Winkler, I. (1999). The Concept of Auditory Stimulus Representation in Cognitive Neuroscience *Psychological Bulletin*, 125 (6), 826-859.
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, & Winkler, I. (2001). “Primitive intelligence” in the auditory cortex. *Trends in Neuroscience*, 24, 283-288.
- Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cognitive Brain Research* 16, 492-495.
- Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain Language* 92, 26-32.
- Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology and Neuro-Otology*, 5, 111-139.
- Polich, J. (2003). *Detection of Change: Event-Related Potential and fMRI Findings*. Boston: Kluwer Academic Publishers.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, 89, 2961-2977.
- Pugh, K. R., Shaywitz, B.A., Shaywitz, S. E., Fulbright, R. K., Byrd, D., Skudlarski, P., Shankweiler, D. P., Katz, L., Constable, T., Fletcher, J., Lacadie, C., Marchione, K., and Gore, J. C. (1996). Auditory Selective Attention: An fMRI Investigation. *Neuroimage* 4, 159-173.

- Shafer, L. V., Schwartz, R. G., & Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cognitive Brain Research*, *18*, 242-254.
- Shafer, L. V., Morr, M. L., Datta, H., Kurtzberg, D & Schwartz, R. G. (2005a). Neurophysiological Indexes of Speech Processing Deficits in Children with Specific Language Impairment. *Journal of Cognitive Neuroscience*, *17*, 1168-1180.
- Shafer, L. V., Kessler, K., Schwartz, R. G., Morr, M. L., & Kurtzberg, D. (2005b). Electrophysiological indices of brain activity to “the” in discourse. *Brain and Language*, *93*, 277-297.
- Shafer, L. V., Ponton, C., Datta, H., Morr, M. L., & Schwartz, R. G. (in press). Neurophysiological indices of attention to speech in children with specific language impairment. *Clinical Neurophysiology*.
- Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., Ilmoniemi, R. J., & Näätänen, R. (2002). Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. *Cognitive Neuroscience And Neuropsychology Neuroreport*, *13(14)*, 1813-1816.
- Shestakova, A., Huotilainen, M., Čeponiene, R., & Cheour, M. (2003). Event-related potentials associated with second-language learning in children. *Clinical Neurophysiology*, *114*, 1507-1512.
- Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R. J., & Näätänen, R. (1998). Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: Magnetic mismatch negativity study. *Neurosci. Lett.*, *251*, 141–144.
- Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage*, *12*, 657–663.
- Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGRAW-Hill International Editions Statistic Series (second edition).
- Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press.
- Strange, W., & Shafer, V. L. (2006). Speech perception in late second language learners: the re-education of selective perception. Zampini, M., & Hansen, J. (eds). *Phonology and Second Language Acquisition*. Cambridge University Press.

- Sussman, E., Winkler, I., & Wang, W.J. (2003). MMN and attention: Competition for deviance detection. *Psychophysiology*, *40*, 430-435.
- Sussman, E. Kujala, T., Halmetoja, J., Lyytinen, H., Alku, P., & Näätänen, R. (2004). Automatic and controlled processing of acoustic and phonetic contrasts. *Hearing Research* *190*: 128-140.
- Sussman, E. (2004). Chapter 2. Auditory Scene Analysis: Examining the Role of Nonlinguistic Auditory Processing in Speech Perception. In P.L.Divenyi (Ed), *Speech Separation by Man and Machine*. New York: Kluwer Academic Publication.
- Szymanski, M. D., Yund, W. E., & Woods, D. L. (1999). Phonemes, intensity and attention: Differential effects on the mismatch negativity (MMN). *Journal of the Acoustical Society of America*, *106* (6), 3492-3505.
- Tajima, K., Kato, H., Rothwell, A., & Munhall, K. G. (2003). Perception of phonemic length contrasts in Japanese by native and non-native listeners. *15th ICPHS Barcelona*.
- Tees, R. C., & Werker, J. F. (1984). Perceptual Flexibility: Maintenance or Recovery of the Ability to Discriminate Non-Native Speech. *Canadian Journal of Psychology*, *38*, 579-590.
- Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning & Memory*, *8*(5), 295-300.
- Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, *372*(6501), 90-2.
- Toda, T. (1994). Interlanguage Phonology: Acquisition of Timing Control in JP. *ARAL (Australian Review of Applied Linguistics)*, *17*, 51-76.
- Werker, J. F., Gilbert, J., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, *52*, 349-355.
- Werker, J.F., & Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49-63.
- Werker, J. F., & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development*, *1*(2), 197-234.

- Winkler, I., & Czigler, I. (1998). Mismatch negativity: deviance detection or the maintenance of the 'standard' *Neuroreport*, *9*, 3809-3813.
- Winkler, I., Kujala, T., Alku, P., & Näätänen, R. (2003). Language context and phonetic change detection. *Cognitive Brain Research*, *17*, 833-844.
- Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research* *1072*, 175-185.
- Yucel, G., Petty, C., McCarthy, G., & Belger, A. (2005). Graded Visual Attention Modulates Brain Responses Evoked by Task-irrelevant Auditory Pitch Changes. *J. Cogn. Neurosci*, *17*, 1819-1828.
- Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., & Schecker, M. (2005). Extracting rules: early and late mismatch negativity to tone patterns. *Neuroreport* *16 (18)*, 2015-2019.