

ABSTRACT

Title of Document: WINDOWS INTO SENSORY INTEGRATION
AND RATES IN LANGUAGE PROCESSING:
INSIGHTS FROM SIGNED AND SPOKEN
LANGUAGES

So-One K. Hwang

Doctor of Philosophy 2011

Directed By: Professor William Idsardi
Department of Linguistics

This dissertation explores the hypothesis that language processing proceeds in “windows” that correspond to representational units, where sensory signals are integrated according to time-scales that correspond to the rate of the input. To investigate universal mechanisms, a comparison of signed and spoken languages is necessary. Underlying the seemingly effortless process of language comprehension is the perceiver’s knowledge about the rate at which linguistic form and meaning unfold in time and the ability to adapt to variations in the input.

The vast body of work in this area has focused on speech perception, where the goal is to determine how linguistic information is recovered from acoustic signals. Testing some of these theories in the visual processing of American Sign Language (ASL) provides a unique opportunity to better understand how sign languages are

processed and which aspects of speech perception models are in fact about language perception across modalities.

The first part of the dissertation presents three psychophysical experiments investigating temporal integration windows in sign language perception by testing the intelligibility of locally time-reversed sentences. The findings demonstrate the contribution of modality for the time-scales of these windows, where signing is successively integrated over longer durations ($\sim 250\text{-}300$ ms) than in speech ($\sim 50\text{-}60$ ms), while also pointing to modality-independent mechanisms, where integration occurs in durations that correspond to the size of linguistic units. The second part of the dissertation focuses on production rates in sentences taken from natural conversations of English, Korean, and ASL. Data from word, sign, morpheme, and syllable rates suggest that while the rate of words and signs can vary from language to language, the relationship between the rate of syllables and morphemes is relatively consistent among these typologically diverse languages. The results from rates in ASL also complement the findings in perception experiments by confirming that time-scales at which phonological units fluctuate in production match the temporal integration windows in perception.

These results are consistent with the hypothesis that there are modality-independent time pressures for language processing, and discussions provide a synthesis of converging findings from other domains of research and propose ideas for future investigations.

WINDOWS INTO SENSORY INTEGRATION AND RATES IN LANGUAGE
PROCESSING: INSIGHTS FROM SIGNED AND SPOKEN LANGUAGES

By

So-One K. Hwang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor William Idsardi, Chair
Associate Professor Gaurav Mathur
Professor David Poeppel
Assistant Professor Naomi Feldman
Professor Robert DeKeyser

© Copyright by
So-One K. Hwang
2011

Dedication

~ For Shue-Yearn, One, and Tyler ~

Acknowledgements

I am fortunate to have had so many opportunities to learn and be inspired throughout my life, and here, I would like to express gratitude to all those who enriched my graduate school experience.

I would like to thank Bill Idsardi for being my advisor. He has provided me with an excellent foundation for learning about important themes and methodologies in linguistics and speech perception. He always seeks explanations that work with the right level of analysis in the field of cognitive science, with a sharp eye for testable hypotheses. Because of his open-mindedness and encouragements, I have been able to build many collaborations and pursue research projects I can be passionate about. I would like to thank David Poeppel for first introducing me to the importance of temporal processing in language and cognition. He has an extraordinary ability to keep cool while still conveying enthusiasm (or skepticism) for ideas, and I have learned a lot from the way he mentors students and works with colleagues. I would like to thank Gaurav Mathur for first introducing me to sign language research, which has truly been an eye-opening experience for me. He has a passion for cross-linguistic research and a genuine enthusiasm for collaboration, and I am so grateful for his guidance and support. I would also like to thank Robert DeKeyser and Naomi Feldman, whose participation in my committee and valuable feedback also helped shape this work.

This work was made possible by the funding of University of Maryland's NSF IGERT program (#DGE-0801465), the NSF Science of Learning Center on Visual Language and Visual Learning at Gallaudet University (#SBE-0541953), and an NSF Doctoral Dissertation Improvement Grant (#BCS-1025530). In addition to my department, I would like to thank these programs, and in particular IGERT's Colin Phillips and VL2's Tom Allen and Diane Clark for all their support and for providing me with wonderful training opportunities.

I would like to thank Clifton Langdon for his significant contributions to this work and for his friendship. Collaborating with him has been extremely fun and productive, and I am also grateful that he has helped me learn to sign. I would also like to thank Connie Pucci for her work and dedication to these projects, and I really admire her energy, time management, and leadership skills. I am also grateful to Verónica Figueroa, whose own dissertation work had an important influence on me, and who shared her interests in research – and cooking.

There are many people who helped implement this project at various stages. I would like to thank Dave Kleinschmidt, Yakov Kronrod, Vladimir Kronrod, Mirko Santoro, Anika Stephen, and Cecily Whitworth. I would like to give special recognition to Nora Oppenheim, Ji-yun Han, and Lesa Young for their contribution to the analysis of corpus data. I would like to thank Ceil Lucas for sharing with me her ASL corpus (funded by NSF grants #SBR-9310116 and #SBR-9709522). I am also grateful to Karen Emmorey for her interest in this work and our many discussions on

modality and working memory. I would also like to thank all the participants of the experiments for their contribution to this work.

I am fortunate to have worked with many wonderful people in other areas of research. I would like to thank Ariane Rhone for her friendship – she has taught me so much, and our adventures together included co-teaching and two defenses. I appreciated the opportunity to collaborate with Derek Monner, Karen Vatz, Giovanna Morini, and Robert DeKeyser and to learn a lot about bilingualism, working memory, and computational modeling along the way. I would also like to thank Phil Monahan for providing me with my first training in MEG and working with me on speech perception experiments.

I am grateful to everyone at Maryland and at VL2 for their support, encouragement, and friendships. At Maryland, I would especially like to thank Eri Takahashi, Ellen Lau, Mathias Scharinger, Bridget Samuels, Shannon Barrios, Wing-Yee Chow, Sunyoung Lee, Julian Jenkins, Pedro Alcocer, and Diogo Almeida. At VL2, I would especially like to thank Gabrielle Jones, Lynn Hou, Peter Crume, and Shilpa Hanumantha, with whom I served on the Student Leadership Team this past year. It has been a privilege to meet and work with many others who have touched my life during graduate school.

Finally, I am so thankful for my family. Thank you, thank you, thank you.

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	v
List of Tables	vi
List of Figures	vii
1 Introduction	1
1.1 Overview	1
1.2 Why sign language?	4
1.3 Temporal integration windows	11
1.4 Neural correlates of temporal integration windows	16
1.5 Oscillation of sub-lexical units in language	21
1.6 Rates of processing in language	24
1.7 Outline of the dissertation	31
2 Temporal integration windows in sign language	33
2.1 Introduction	33
2.2 Cognitive restoration of locally time-reversed sentences	35
2.3 Flexibility of perceptual parameters to rates	42
2.4 Perspectives from development and bilingualism	47
2.5 Experiment 1 – Effect of modality on temporal integration windows: evidence from local-reversals of ASL sentences	56
2.6 Experiment 2 – Effect of modality-independent mechanisms on temporal integration windows: evidence from compression and local-reversals of ASL sentences	70
2.7 Experiment 3 – Effect of developmental factors on temporal processing: evidence from late-learners of ASL	75
2.8 Conclusion	81
3 Temporal Dynamics in Natural Production	89
3.1 Introduction	89
3.2 Bellugi & Fischer (1972) revisited: Beyond the rate of signs	97
3.3 Perspectives from information theory	113
3.4 Words, signs, morphemes, and syllables	119
3.5 Rates in spoken languages: English and Korean	125
3.6 Rates in sign language: ASL revisited	139
3.7 Conclusion	151
4 Conclusion	163
4.1 Overview	163
4.2 More than meets the eye	164
4.3 Hierarchical coupling in sign language processing?	167
4.4 Innate sensitivity to rhythms in language	170
4.5 Channel capacity for sign language	174
4.6 Availability of two communication channels?	176
4.7 Rates in production and time-course of recognition	182
4.8 General conclusions	185
Bibliography	191

List of Tables

Table 1. Adapted from Krentz & Corina (2008), this table lists some of the qualitative differences between pantomime and ASL.....	6
Table 2. Examples are adapted from Bellugi & Fischer (1972). These pairs of sentences demonstrate differences between English and ASL constructions.	98
Table 3. Adapted from Brentari (2002), who describes the typological distribution of canonical word shapes. These assumptions are reexamined throughout the current discussion in Chapter 3 because they require an examination of syllable and morpheme rates and the ratio of these rates for languages.	120

List of Figures

Figure 1. Reproduced from Petitto, Solowka, Sergio, Levy, & Ostry (2004), this figure shows the distribution of the frequencies (in Hz) of the manual movements among sign-exposed and speech-exposed babies. Sign-exposed babies had movements that were at two different frequencies, where manual babbling in the signing space was marked by a slower rhythm (~1 Hz) than ordinary gestures outside the signing space (~2.5 Hz), whereas speech-exposed babies had movements at a higher frequency (~3 Hz).....	22
Figure 2. Reproduced from Fischer, Delhorne, & Reed (1999), these figures show the intelligibility of stimuli as a function of playback rates for 14 participants. Error bars represent plus or minus one standard deviation of the mean. With sentences, a sharp drop in intelligibility is found at compressions by a factor of 3.	28
Figure 3. Reproduced from Ghitza & Greenberg (2009), this graph shows the percent error in an intelligibility experiment, where sentences were compressed by a factor of 3 and silences were inserted periodically or aperiodically. Error bars represent the standard deviation of the mean.	30
Figure 4. Reproduced from Greenberg & Arai (2001), this figure demonstrates how locally-reversed speech stimuli are created. Here, each 80 ms segment is played backwards, but the original order of the segments is maintained.	36
Figure 5. Reproduced from Saberi & Perrott (1999), this figure shows subjective intelligibility ratings by 7 participants on a single sentence that was repeated for all conditions.	37
Figure 6. Reproduced from Greenberg & Arai (2001), this figure demonstrates 1) the spectrogram of locally reversed sentences, 2) the intelligibility curve as a function of reversal sizes, and 3) the complex modulation spectrum of the sentences. Intelligibility results are from 27 participants tested on 40 sentences. Intelligibility of sentences falls drastically between 40 and 50 ms reversals, falling to 50% at 60 ms reversals, and reaches ~0% by 100 ms reversals.	38
Figure 7. Reproduced from Miller & Licklider (1950), this figure demonstrates the intelligibility of English sentences as a function of frequency of interruption and speech-time fraction (where the duration of interruptions were dependent on the frequency of the interruptions and speech-time fractions and were spaced regularly).....	40
Figure 8. Reproduced from Green & Miller (1985), this figure demonstrates that perceptual boundary, reflected by the percentage of voiced responses for [bi]-[pi] continuum, varies depending on durations.....	43
Figure 9. Reproduced from Figueroa (2009), this figure shows the intelligibility of English sentences as a function of compression and reversal size.	45
Figure 10. Reproduced from Stilp, Kieft, Alexander, & Kluender (2010), this graph shows intelligibility curves of English sentences as a function of the size of local-reversals (segment durations in ms) and speech rates (in syllables per second: slow = 2.5, medium = 5.0, fast = 10).	46

Figure 11. Reproduced from Tweney, Heiman & Hoemann (1977), this figure shows the intelligibility of ASL and English sentences as a function of temporal disruption frequency and signing/speech-time fractions. These results demonstrate that sign language is more resistant to temporal disruptions than speech.....	59
Figure 12. Demonstration of how locally time-reversed stimuli were created for sentences of ASL. This specific example shows reversals 133 ms in duration (reversals by 4 frames).	61
Figure 13. Results from Experiment 1 from 14 participants, demonstrating the intelligibility curve of ASL sentences as a function of reversal size, which implicates ~300 ms temporal integrations windows. 50% intelligibility of even the most degraded stimuli is attributed to spatial-encoding in sign language. Error bars represent plus or minus one standard error of the mean. ...	63
Figure 14. Reproduced from Liddell (2000), this illustration represents the sign for GIVE, where the direction of movement can mean I-GIVE-YOU but the reverse would result in the opposite meaning YOU-GIVE-ME.....	67
Figure 15. Results from Experiment 1 and 2 (14 participants in each experiment), demonstrating the intelligibility curve of ASL sentences as a function of reversal size and compression by a factor of 2, where temporal integration windows are proportional to the input rate (indicated by a sharp drop in intelligibility at ~267 ms reversals at the normal rate and ~133 ms reversals at the 2x rate). These results suggest that temporal integration windows in sign language are determined by the rate and durations of linguistic units. Error bars represent plus or minus one standard error of the mean.	73
Figure 16. Results from Experiment 1 and 3, demonstrating the effects of age-of-acquisition in processing time-distorted stimuli. Note: n=14 in Experiment 1 and n=8 in Experiment 3. Late learners demonstrate greater sensitivity to time distortions in the input, but performance among the early and late learners plateau at similar distortion scales. Error bars represent plus or minus one standard error of the mean.	79
Figure 17. Examples of signs used by Wilson (2001), with images from <i>www.aslpro.com</i> (top) and <i>www.signingsavvy.com</i> (bottom). The top row shows images taken from a video recording of BRIDGE, a two-contact sign that involves hopping motion from the wrist to the elbow. The bottom row shows images from a video recording of CREDIT-CARD, a one-contact sign that involves sliding motion from the palm and outward across the hand.	93
Figure 18. Reproduced from Schroeder, Lakatos, Kajikawa, Partan, & Puce (2008), this figure illustrates the hierarchical coupling of neural oscillations. ...	95
Figure 19. Reproduced from Brentari, Poizner, & Kegl (1995) (and Brentari (1998)), this figure demonstrates sign-internal and sign-external transitions in an ASL sentence. The above sentence is WORD BLOW-BY-EYES MISS SORRY ('The word went by too quickly. I missed it, sorry').....	102
Figure 20. Reproduced from Bosworth, Dobkins, & Wright (2010), this figure demonstrates the 2D movement trace for an elicited sentence containing the sign KNOW.	111

Figure 21. Reproduced from Hale (2001), this figure demonstrates how entropy (or “surprisal”) fluctuates over the course of a sentence.....	117
Figure 22. Adapted from Mathur & Rathmann (2011), this figure demonstrates an example of numeral incorporation in ASL.	121
Figure 23. Reproduced from Mathur & Rathmann (2011), this figure demonstrates the grammatical form for TEN DAY and the ungrammatical form TEN+DAY that would result with numeral incorporation. The latter is believed to be not possible due to phonological constraints against complex movement.	122
Figure 24. Estimated probability density functions for the length in seconds of sentences in two corpora of English: TIMIT (prompted) and CALLFRIEND (conversational).....	129
Figure 25. Estimated probability density functions for words rates (words per second) of sentences in two corpora of English: TIMIT (prompted) and CALLFRIEND (conversational).....	130
Figure 26. Estimated probability density functions for syllable rates (syllables per second) of sentences in two corpora of English: TIMIT (prompted) and CALLFRIEND (conversational).....	131
Figure 27. Estimated probability density functions for morpheme rates (morphemes per second) of sentences in two corpora of English: TIMIT (prompted) and CALLFRIEND (conversational).....	132
Figure 28. Estimated probability density functions for length in seconds of sentences from conversational data in English and Korean.	135
Figure 29. Estimated probability density functions for word rates (words per second) of sentences from conversational data in English (a more analytic language) and Korean (a more synthetic language).....	136
Figure 30. Estimated probability density functions for syllable rate (syllables per second) of sentences from conversational data in English and Korean.	137
Figure 31. Estimated probability density functions for morpheme rates (morphemes per second) of sentences from conversational data in English and Korean.....	138
Figure 32. Estimated probability density functions for length in seconds of sentences from conversational data in English, Korean, and ASL.	144
Figure 33. Estimated probability density functions for word/sign rates (words or signs per second) of sentences from conversational data in English, Korean, and ASL. This comparison word and sign rates replicate the findings from Bellugi & Fischer (1972) for English and ASL. A comparison with Korean demonstrates that word rates depend on grammatical properties of the language.....	145
Figure 34. Estimated probability density functions for syllable rates (syllables per second) of sentences from conversational data in English, Korean, and ASL. Syllables rates in ASL may be the basis for the temporal integration window of ~250-300 ms found in Experiment 1 in Chapter 2.....	146
Figure 35. Estimated probability density functions for morpheme rates (morphemes per second) of sentences from conversational data in English, Korean, and ASL. This figure demonstrates that English and Korean, two	

spoken language with distinct grammars, have the same morpheme rate (~6 per second), in contrast with the morpheme rate in ASL (~3 per second).....	147
Figure 36. The comparison of morpheme:syllable ratios in English, Korean, and ASL suggests that the globally, morphemes and syllables are processed at approximately the same rate. However, the results from ASL are different from spoken languages in that the ratios reveal a trimodal distribution. This may be attributed to properties unique to sign languages, such as productive use of reduplication (resulting in ratios lower than 1:1) and productive use of spatial modulations (resulting in ratios higher than 1:1), in addition to simple signs.....	150
Figure 37. Reproduced from Padden & Perlmutter (1987), where reduplicating circular movement turns the adjective QUIET to mean ‘characteristically quiet’, or taciturn.....	156
Figure 38. Reproduced from Aronoff, Meir, & Sandler (2005), demonstrating a complex ASL classifier construction: ‘A person walks forward, (dragging) a dog squirming behind.’	157
Figure 39. Reproduced from Boyes-Braem (1999), demonstrating the difference between early and learners of Swiss German Sign Language in their lateral torso movements while signing.	173
Figure 40. Reproduced from Jantunen (2010), demonstrating the acceleration peaks in the biomechanics of both hands while signing, annotated for traditional sign boundaries and transitions between signs.	184

1 Introduction

1.1 Overview

The goal of this dissertation is to contribute to a better understanding of the universal temporal processing constraints in the perception and production of language and how they are manifested in particular sensori-motor channels. This endeavor requires a cross-linguistic, and crucially, a cross-modal, study of the temporal dynamics in language processing. Building upon a large body of previous work on spoken languages (Poeppel, 2003; Poeppel, Idsardi, & van Wassenhove, 2008; van Wassenhove, Grant, & Poeppel, 2007; Viemeister & Wakefield, 1991; Yabe, Tervaniemi, Sinkkonen, Huottilainen, Ilmoniemi, & Näätänen, 1998; Ahissar, Nagarajan, Ahissar, Protopapas, Mahncke, & Merzenich, 2001; Saberi & Perrott, 1999; Greenberg & Arai, 2001; Figueroa, 2009; Stilp, Kiefte, Alexander, & Kluender, 2010), this dissertation is the first to investigate temporal integration windows and processing rates in American Sign Language (ASL). A key aim of speech perception research is to determine how an acoustic signal is mapped onto meaningful linguistic representations. However, the existence of sign languages demonstrates that visual signals can also be transformed into rich grammatical meaning. Thus, to more broadly understand how linguistic information is extracted from a sensory signal, common mechanisms in spoken and signed languages must be identified.

This dissertation's focus on temporal aspects of language processing is inspired by two perspectives in language research. Psychophysical investigations of speech reveal an intimate connection between the temporal properties of the acoustic

stimulus and corresponding behavioral and neural responses. The biomechanics of the articulators in the vocal tract create a dynamic acoustic signal with rapidly changing spectro-temporal information, which is transmitted through the air and then through the auditory pathway. Work in theoretical linguistics and experimental neuroscience has suggested that the information in this signal is layered into levels that correspond to units of linguistic representations (phonemes and syllables) and that neural processes underlying speech perception also occur in multiple time scales (Poeppel, Idsardi, & van Wassenhove, 2008).

One possibility is that all the time properties we observe in speech come from the particular properties of the oral articulators, coordination with breathing, and the auditory pathway, but sign language research suggests otherwise. When comparing the rate of production in English and ASL, where the semantic content of the narratives were roughly matched, studies found that the propositional rate in these two languages is the same (Bellugi & Fischer, 1972). Although differences in rates emerged when looking at the level of words and signs, where on average twice as many words are produced per second (~4-5 words per second) as compared to signs (~2 signs per second), the overall result implicates a modality-independent basis for the rates found in language. The fact that words and signs are not always equivalent linguistic units and the emergence of simultaneous morphology and spatial grammar in sign languages (Bellugi & Fischer, 1972; Senghas & Coppola, 2001; Aronoff, Meir, Padden, & Sandler, 2004; Mathur & Rathmann, 2011) have been attributed as the key underlying factors in these findings. Klima and Bellugi (1979: 194) write, “It is possible that the tendency toward compacting linguistic information in signs may

be a response to temporal pressure on language production.” When studying the propositional rate in signed English, which does not employ these strategies and notably is not a natural human language, the propositional rate is twice as slow as ASL (Klima & Bellugi, 1979).

How language design allows for the interaction of core linguistic processes with two completely different sensori-motor systems, as well as other cognitive domains, remains a remarkable puzzle. One of the goals of this dissertation is to show that slower temporal dynamics in sign production results in larger temporal integration windows in perception. I present evidence that these temporal integration windows come from mechanisms that are not just inherent to visual processing but from sensitivity to the time duration of linguistic units in sign language. The methodology employed here is testing the intelligibility of locally time-reversed sentences as a function of reversal size. Although the time-scale of temporal integration windows in sign language differ from the findings in speech, the results point to universal patterns (integration according to durations of representational units). A corpus-based study of rate of production in conversations taken from English, Korean, and ASL also provides greater insights on what are the relationships between time, form, and meaning in natural data. This research fits within the broader aims to disentangle properties that are inherent to core processes underlying language from those that are driven by modality.

1.2 Why sign language?

A valid model for how humans process language requires coverage of typologically diverse languages, crucially including those that use different sensory channels and motor systems for communication. Previous studies on temporal integration windows in language were limited to speech (Poeppel, 2003; Viemeister & Wakefield, 1991; Greenberg & Arai, 2001; Luo & Poeppel, 2007), making it difficult to determine whether ongoing processes were about auditory or more general mechanisms for analyzing linguistic input.

One of the great discoveries of modern linguistic research has been the realization that sign languages are true languages with all of the fundamental properties shown by spoken languages (Stokoe, 1960; Klima & Bellugi, 1978; Emmorey, 2002). A cross-modal approach to language research has been used productively to understand universal grammatical properties, the functional organization of language processing areas in the brain, and the developmental patterns seen in language acquisition. All languages have multiple levels of representation, including phonology, morphology, and syntax, with rules for how units in these domains combine (Sandler & Lillo-Martin, 2006). Beneath the level of signs, sublexical phonological units combine in systematic and rule-constrained ways. Signs can vary in their degree of meaning complexity due to morphological processes. In addition to having structural constituents, sign languages also show sensitivity to island constraints (Padden, 1988; Lillo-Martin, 1991; Ross, 1967).

Lesion and neuroimaging studies show that the same cortical areas support core language functions for both speakers and signers (Hickok, Bellugi & Klima 1998; Emmorey, Mehta & Grabowski 2007; Petitto, Zatorre, Gauna, Nikelski, Dostle, & Evans, 2000). Previously, speculations about the left-hemisphere dominance in spoken languages pointed to a specialization for processing rapidly changing temporal information (Tallal, Miller, & Fitch, 1993). Moreover, the topographic location of Broca's area near speech-production areas of the motor cortex and Wernicke's area near speech-perception areas of the auditory cortex raised the possibility that both areas mainly support the function of spoken languages. However, neuroimaging studies among deaf signers show overlapping activation in these areas (Emmorey, Mehta & Grabowski 2007; Petitto, Zatorre, Gauna, Nikelski, Dostle, & Evans, 2000), and damage to those areas result in similar aphasic profiles (Hickok, Bellugi & Klima 1998).

Although sign languages use manual articulators, the linguistic status of their movements is distinct from gesture (see Table 1 for a comparison of linguistic and non-linguistic gestures). Evidence for dissociations of sign language and non-linguistic gesture has been found in lesion cases, where a patient's production and comprehension of non-linguistic gestures remained intact while performance on sign language was impaired (Corina, Poizner, Bellugi, Feinberg, Dowd, & O'Grady-Batch, 1992). In development, at the age of 6 months, even hearing infants with no previous input to signing treat videos of signing and pantomime movements differently, which has been suggested as evidence that children are born with a

specific interest in linguistic signals, whether they are produced vocally or manually (Krentz & Corina, 2008).

	Pantomime	ASL
Handshapes	Few handshape types Simple handshapes: e.g. broad (open hand) and compact (closed hand) variants	Multiple handshape types Complex handshapes: e.g. broad (open), compact (closed) and intermediate variants ('V' index and middle fingers open, ring and pinky fingers closed; 'F' index finger and thumbing touching, middle, ring and pinky finger open)
Location	Numerous on- and off-body locations (e.g. above head, below waist, behind body)	Adherence to limits of defined signing space (head to waist, directly in front of signer)
Movement	More movement types More undefined movement types Frequent repetitions of movements	Fewer movement types More defined movement types Limited number of repetitions
Eyes and brows	Eyes follow actions that model performs; eyes are not independent of actions	More eye contact with camera; actions are independent of eyes
Facial expression	Expressivity based on actions performed (e.g. frustration at trying to fix hair, satisfaction in finishing a task)	More rapid changes in mouth More varied movement in mouth

Table 1. Adapted from Krentz & Corina (2008), this table lists some of the qualitative differences between pantomime and ASL.

Parallels seen in the developmental course of young children, from babbling to putting words together, has contributed evidence for biological maturation of a modality-independent language faculty. In the same study described above (Krentz & Corina, 2008), at 10 months of age, children no longer treated signing and pantomime movements differently. This is convergent with studies on infant speech perception, where preference for native input is sharpened and sensitivity to distinctions in non-native languages shows significant declines around 10 months (Werker, Gilbert, Humphrey, & Tees, 1981). More broadly, similar linguistic milestones are observed among deaf and hearing children (Newport & Meier, 1985; Lillo-Martin, 1999). Word learning at around the first year of life (Bonvillian & Folven, 1993) is preceded

by babbling, which in of itself has numerous stages (de Boysson-Bardies, 1993; Meier & Willerman, 1995; Masataka, 2003). Early forms of manual babbling are attested among all infants, but only sign-exposed children develop complex handshape and movement patterns (Petitto & Marentette, 1991). In addition, the class of utterances produced in babbling is predictive of the phonological features of first words in both modalities (Oller, Wieman, Dole, & Ross, 1976; Cheek, Cormier, Repp, & Meier, 2001).

Other parallels include patterns in vocabulary growth, increases in grammatical complexity, and even similarities in errors during the acquisition process, which includes phonological errors (Conlin, Mirus, Mauk, Meier, 2000; Meier, 2006; Masataka, 2003), overgeneralization (Meier, 1987), and pronoun errors (Petitto 1987; Jackson, 1989; Meier & Newport, 1990). However, this holds true only in cases where children are receiving signing input since birth. Because >95% of deaf individuals are born to hearing parents (Mitchell & Karchmer, 2004), age of exposure and acquisition of a sign language greatly vary, along with levels of ultimate attainment. Comparison of early and late learners of ASL, as well as comparisons of late learners for whom ASL is either an L1 or L2, provide valuable insights on the impact of critical periods for language development. Similar to the distinctness observations for L1 and L2 acquisition among spoken languages, later acquisition of sign languages is marked by different profiles in perception and production compared to native learners (Kantor 1978; Newport, 1990; Mayberry & Eichen, 1991; Morford & Mayberry, 2000). In the case of spoken languages, differences in L2 performance are confounded with L1 entrenchment, where deficits in L2 may be attributable to

interference effects from L1 (MacWinney, 2006). Signers who are exposed to English as an L1 in early childhood and then later acquire ASL as an L2 outperform signers who receive little to no input in early childhood until exposure to ASL as an L1 in late childhood. This suggests that entrenchment cannot be the main factor behind age-effects in acquisition. Growing research from sign language research (Newport, 1990; Mayberry, 1993; Mayberry, del Giudice, & Lieberman, 2010; Wilbur, 2000; *inter alia*) continues to highlight the importance of early language exposure – whether spoken or signed – for full language development.

Studying sign language users has been relevant to the research of bilingualism more broadly. Signers fit many profiles in bilingualism. As mentioned above, depending on the onset of hearing loss, degree of hearing loss, and type of early education, a spoken language is the first language for many deaf individuals. For hearing individuals who are born to deaf parents, often referred to as “CODAs” (children of deaf adults), sign language is their first language, with acquisition of a spoken language from mainstream society. Finally, deaf individuals who are born to deaf parents and grow up in signing environments both at home and school still have considerable experience using English in the United States through reading and writing.

Growing evidence from bilinguals who use two spoken languages shows that both languages are active even when using only one of those languages (Marian & Spivey, 2003; Kroll, Bobb, & Wodniecka, 2006). One possibility is that these co-activation effects are dependent on shared modality, but recent work on the bilingual activation of English and ASL among deaf signers suggests otherwise (Morford,

Wilkinson, Villwock, Piñar, & Kroll, 2011). Here, the reaction time of deaf signers on making judgments about the semantic relatedness of a given pair of English words was slowed down or speeded up based on the phonological similarity of equivalent ASL signs, which were not presented at any point during the experiment. Such patterns were not found among a group of sign-naïve participants.

Cases of bimodal bilingualism with hearing participants who are fluent in a spoken and signed language provide unique opportunities to study how bilingualism is manifested when two articulatory channels are available. Unlike with two spoken languages, a spoken and signed language can be produced simultaneously, leading to common cases of code-blending (Pyers & Emmorey, 2008; Casey & Emmorey, 2009). Moreover, testing cognitive control on bimodal bilinguals allows for a better understanding of the “bilingual advantage” that is reported for unimodal bilinguals (Bialystok, 2001), specifically whether better cognitive control is caused by switching between any two languages or whether it requires switching within one modality (Emmorey, Luk, Pyers, & Bialystok, 2008). Emmorey et al. (2008) found that bimodal bilinguals did not perform differently from monolinguals. Despite the availability of two different channels, it is important to note that simultaneous production of both English and ASL is extremely difficult because of the large difference in their grammars and other processing constraints, leading performance in both languages to suffer (Wilbur & Petersen, 1998). The degree to which both languages are activated among bimodal bilinguals and the extent to which cognitive control is exercised by bimodal bilinguals while sticking to one modality or code-blending remains unclear. Among deaf adults who are bilingual in ASL and written

English, better performance on higher-order attention tasks are found among those with high proficiency in both languages (Kushalnagar, Hannay, & Hernandez, 2010).

Sign language research has also contributed insights on the interaction of sensory experience and language modality with other aspects of cognition. Cortical reorganization following auditory deprivation results in differences in visual attention, where greater attention is allocated to peripheral areas (Bavelier, Dye, & Hauser, 2006). This is not true for bilingual hearing signers, who show the same profile of devoting greater attentional resources to central fields of vision as non-signing hearing individuals (Bavelier, Dye, & Hauser, 2006). However, experience with a sign language does transfer to differences in visual processing in some cases, in particular with tasks that involve mental imagery and rotation. When compared to non-signers, both deaf and hearing signers were found to have enhanced abilities in mental rotation tasks and in generating complex images (Emmorey, Klima, & Hickok, 1998; Emmorey & Kosslyn, 1996; Emmorey, Kosslyn, & Bellugi, 1993). Early signing exposure has also been associated with enhanced visuo-spatial working memory, as tested by Corsi block experiments (Milner, 1971), where the participant has to identify the sequence of locations that were indicated by the experimenter (Wilson, Bettger, Niculae, & Klima, 1997; Parasnis, Samar, Bettger, & Sathe, 1996). Only deaf children with early exposure to sign language had higher spans than hearing children in this spatial processing task.

Much of past research has shown how spoken and signed languages are similar, but a newer challenge has been to also make sense of their differences. Meier, Cormier, and Quinto-Pozos (2002) provide a useful overview that focuses on these

issues. A critical aspect of determining modality effects in language processing is understanding the physiological properties of the sensori-motor channels. However, it is also important to consider the differences in the diachronic history of these two types of languages, where most sign languages are relatively young and are frequently reinvented by its users, most of whom are not exposed to sign language since birth. While acknowledging the status of sign languages as full-fledged, natural human languages, recognizing and understanding the differences between spoken and sign languages can lead to better targeted strategies to improve learning and education throughout development for each population.

1.3 Temporal integration windows

The experience of many perceptual phenomena, including language processing, feels seamless and continuous, but from a neurophysiological and computational perspective, sensory inputs are analyzed in chunks, or time windows, that lead to discrete units and combinatorial mechanisms. Integration windows exist at more than one time-scale, and the process that occurs at each level may differ (Viemeister & Wakefield, 1991). In perceptual terms, temporal integration windows are considered to be time durations for the summation of the input. Under certain characterizations, the information about dynamics that occur at smaller time-scales may be lost, leading to limits in temporal resolution. Physiologically, the lower limits of temporal encoding may be tied to the duration of action potential spikes and subsequent refractory period, leading to upper limits of sampling rates, and the

properties of individual sensory pathways. At the level of a neuron, the duration of all the processes that contribute to the spiking output is characterized as the integration window (Theunissen & Miller, 1995). Psychophysically, the temporal resolution of the auditory system is ~2 ms and ~20 ms for the visual system. These values refer to the smallest time gaps that can be detected between a sequence of inputs in each respective domain (Green 1971; Kohlrausch, Püschel, & Alpehi, 1992; Chase & Jenner, 1993).

As Viemeister and Wakefield (1991) emphasize, however, it is important to distinguish the phenomenon from the process underlying integration. In their model of auditory processing, integration windows that occur at larger time-scales (~200 ms) arise from different mechanisms than those that set the temporal resolution of sensory processing (2-3 ms). The larger time window is mediated by short-term memory, where samplings of the processed input are stored and remain available for comparisons and further computations. This “multiple looks” model is one way to account for the phenomena of smaller as well as larger temporal integration windows in auditory perception.

A model for integration at multiple time scales is critical to understanding many aspects of cognition, including sensory processing, multi-sensory integration, and sensori-motor coordination (Pöppel, 1997). Representations in language are organized into a hierarchy that includes units in phonology, morphology, and syntax. In on-line processing, these units unfold in different time scales. Determining how acoustic information is mapped onto the building blocks for the representation of words is the focus of speech perception research. One of these challenges includes

identifying the spectro-temporal characteristics of particular features or phonemes (Ladefoged, 2005). More broadly, temporal properties of speech contain correlated information about linguistic features (Rosen, 1992). In particular, fine-structure information in smaller windows (20-40 ms) is critical to the identification of place-of-articulation features of stop segments, and the amplitude-modulated envelopes in larger windows (150-300 ms) for the perception of syllables (Poeppel, 2003).

Temporal integration windows roughly 150 ms in duration are reported in a study that tested facilitation to vowel identification via two types of primes that ranged in durations of 25 – 500 ms (Wallace & Blumstein, 2009). Priming at short durations (25 – 500 ms) was only found when nonspeech tone complexes (which were matched to formant frequencies of the vowels) were used. These facilitation effects were strongest up to prime durations of ~150 ms. Overall, facilitation effects were much more robust with vowel primes, where speeded reaction times peaked with prime durations in the range of 100 – 150 ms. These results are compatible with the multi-time scale model for processing speech proposed by Poeppel (2003).

In the domain of audio-visual processing of speech, temporal integration windows also constrain how information from different sensory sources can be combined (van Wassenhove, Grant, & Poeppel, 2007). The integration of audio-visual information is evident in cases like the “McGurk effect,” where the percept of [ta] is neither the information in the acoustic [pa] nor the visual signal [ka] (McGurk & MacDonald, 1976). Studies have investigated to what degree this phenomenon requires temporal alignment (Dixon & Spitz, 1980; Massaro, Cohen, & Smeele, 1996; McGrath & Summerfield, 1985; Pandey, Kunov, & Abel, 1986; Munhall, Gribble,

Sacco, & Ward, 1996; van Wassenhove, Grant, & Poeppel, 2007). Among those for whom the fused percept is possible at all, approximately ~200 ms of audio lag can be tolerated for the fused percepts to be recognized in these audio-visual speech experiments.

The study of temporal integration windows cover a broad range of processing in cognition, where the temporal resolution of sensory encoding is only one component. Experimental materials also appear to influence the size of these windows in ways that cannot be solely attributable to the sensory process. In contrast to these findings from audio-visual speech, a study that used non-speech stimuli (white noise and LED light) found smaller windows (~100 ms) within which simultaneous percept was possible with asynchronous input (Zampini, Guest, Shore, & Spence, 2005). A possible explanation for the longer duration in the integration of speech-stimuli is associated with the average syllable duration across languages (van Wassenhove, Grant, & Poeppel, 2007; Arai & Greenberg, 1997; Whalen & Liberman, 2000).

The multi-time scale model for speech perception (Poeppel, Idsardi, & van Wassenhove, 2008) is inspired in part by theories on the multiple spatial resolution in vision, where information is analyzed both locally and globally, at various spatial frequencies to recover information (Merigan & Maunsell, 1993). In vision, high and low spatial frequency information is processed by different populations of neurons in different parts of the visual cortex (Singer & Gray, 1995). The fact that a coherent percept can result from more than one anatomical and functional organization is called the binding problem. Singer and Gray propose that binding is achieved by the

time synchronization of neural activity, which is often associated with oscillatory firing patterns. Poeppel et al. (2008:1076) write, “Whereas in the visual case the image can be fractionated into different spatial scales, in the auditory case both frequency and time can be thought of as dimensions along which one could fractionate the signal.” Nevertheless, no such approach has been used to determine temporal integration windows in the visual processing of sign language.

Although this discussion has focused on the role of time windows as durations for integration, their existence may also entail a process of discretization, which remains poorly understood (Van Rullen & Koch, 2003). In vision, the perception of continuous motion is understood to be only “apparent,” meaning that viewers construct the experience by piecing together a sequence of discrete images (Wertheimer 1912; Korte 1915). When reading, the discrete, saccadic movements of the eyes track blocks of linguistic constituents that are not marked by any special spacing or punctuation (Rayner, 1998). Discretization of the input goes well beyond sensory processing to guidance from knowledge about the representational nature of the signal. To distinguish between words like *bear* and *pear*, listeners are categorically tuned to the timing difference between the onset of laryngeal voicing and the onset of the stop release burst. Even when presented with tokens that are varied continuously along this parameter, listeners process these sounds categorically (Lisker & Abramson, 1964; Lisker, 1975; Klatt, 1975). When attending to a continuous speech stream, listeners automatically extract a sequence of segments, syllables, words, and phrases, which is easy to take for granted until one is trying to communicate in a foreign language. Similarly, when non-signers view sentences of

ASL, accurately discerning the boundaries between the signs can be difficult, a process that is automatic for signers (Brentari, 2006). Thus, although many perceptual experiences appear holistic and continuous, research from a wide range of domains reveal underlying mechanisms that are discontinuous.

Although temporal integration windows are not proposed here to account for all of these phenomena, they are listed here to demonstrate some of the challenges in understanding both the continuous and discrete aspects of perception. Van Rullen and Koch (2003) suggest that studies on the temporal dynamics of neural oscillations at different frequency bands may result in a more unified explanation of integration and discreteness in perception. With this idea in mind, the following section summarizes work investigating the neural correlates of temporal integration windows in speech.

A discussion of studies using locally time-reversed stimuli to study temporal integration windows, which is the primary methodology used in the perceptual experiments of this dissertation, will be provided in depth in Chapter 2.

1.4 Neural correlates of temporal integration windows

In speech perception, two particular time-scales, those that correspond to segments (~50 ms durations) and syllables (~200 ms durations), are especially interesting because of the convergence of findings about the time-scales of acoustic fluctuations in speech (Rosen, 1992) and neural oscillations (Poehpel, 2003). Marked by high temporal resolution, electrophysiology is currently the best available

methodology for investigating the potential neural correlates of these temporally dynamic processes in speech. Hypotheses about these neural correlates are based on the acoustic properties of speech, behavioral studies on listeners' response to sounds where the fine-structure and temporal envelopes are manipulated (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Zeng, Nie, Stickney, Kong, Vongphoe, Bhargave, Wei, & Cao, 2001), electrophysiological studies on temporal integration windows auditory processing (Yabe, Tervaniemi, Sinkkonen, Huottilainen, Ilmoniemi, & Näätänen, 1998), and electrophysiological studies on neural oscillations for integration and binding in cognition (Buzsáki & Draguhn, 2004; Engel, Fries, & Singer, 2001; Pöppel, 1997).

Language comprehension requires successful integration of the sensory signal. One study (Ahissar, Nagarajan, Ahissar, Protopapas, Mahncke, & Merzenich, 2001) builds upon previous findings that intelligibility of sentences decreases as a function of the rate of compression (Foulke & Sticht, 1969). Speech compression algorithms that preserve the spectral and pitch content were applied (Portnoff, 1981). In this study, magnetoencephalography (MEG) was used to measure the degree to which cortical signals followed the speech signal modulation. It was found that the frequency of the evoked cortical signals and the temporal envelopes of the stimuli only matched at lower compression ratios (where sentences remained somewhat intelligible). These conditions also showed phase-locking between the speech envelope and MEG signal recorded from the auditory cortex, reflecting entrainment of the neural activity to speech signals and sensitivity to the temporal characteristics of speech. Similar findings are reported by Luo and Poeppel (2007), where

intelligibility of sentences was modulated by creating “auditory chimaeras,” where the envelope of one sound is combined with the fine structure of another sound (Smith, Delgutte, & Oxenham, 2002). Also using MEG, Luo and Poeppel report cortical activity that is correlated with speech intelligibility, more specifically, in the phase patterns of endogenous brain rhythms in the theta band (4-8 Hz), where period durations correspond to the average size of syllables in speech (~ 200 ms). These findings implicate that an important aspect of processing speech is continuous segmentation and integration of the input in ~200 ms temporal windows (Poeppel, Idsardi, & van Wassenhove, 2008).

Temporal integration windows in such time-scales converge with results from other electrophysiological studies on auditory processing (Yabe, Tervaniemi, Sinkkonen, Huotilainen, Ilmoniemi, & Näätänen, 1998). Using MEG, Yabe et al. (1998) varied stimulus-onset asynchronies (SOAs) of pure tones and tested the elicitation of the magnetic counterpart to the mismatch negativity (MMN) in electroencephalography (EEG). The MMN is taken to be an index of change detection, whether caused by a change in stimuli or by an omission, which is thought to be implemented via a comparison to a neural memory trace of a repetitive sound (Cowan 1995; Näätänen, 1992). Shorter SOAs improve the chances that a stimulus is in the same temporal window of integration as the previous item. Results showed that MMNs were only elicited at SOAs shorter than 175 ms, supporting a temporal integration window of 150-175ms in auditory processing. This time window and model for MMN effects are compatible with the “multiple looks” theory of

Viemeister and Wakefield (1991), where integration windows at these larger time-scales are mediated by short-term memory.

Processing speech in ~200 ms time windows for syllables, as well as smaller ~50 ms windows for segments, requires neural substrates that are sensitive to temporal information that fluctuates at those intervals. Such evidence comes from a functional imaging study using fMRI (Boemio, Fromm, Braun, & Poeppel, 2005), as well as a simultaneous EEG-fMRI study (Giraud, Kleinschmidt, Poeppel, Lund, Frackowiak, & Laufs, 2007). In addition to activations that are sensitive to these time windows, both studies report hemispheric asymmetries in these sensitivities. Boemio et al. (2005) used non-speech stimuli with varying spectro-temporal properties, with modulations in segment length and frequency sweep patterns. Among three regions of interest (superior temporal gyrus (STG), superior temporal sulcus (STS), and transverse temporal gyrus (TTG)) in auditory cortical areas, activity in STG was sensitive to the local spectro-temporal structure, where effects were greatest at 45 ms SOAs, a time scale that is important for extracting information about segment boundaries. Activity in STS was not sensitive to segment type but showed hemispheric asymmetry to segment duration (or stimulus rate) and the greatest response was to those with durations >85 ms. Giraud et al. (2007) demonstrate that spontaneous EEG rhythms in the gamma range (30-50 Hz), where periods correspond to the duration of phonemes, and theta range (4-7 Hz), where periods correspond to the duration of syllables, show similar hemispheric asymmetries in the auditory cortex (convergent with the model proposed by Poeppel (2003)), and suggest that

these endogenous oscillations serve as important precursors that support the function of speech processing.

To what degree this multi-time resolution model of speech processing, subserved by gamma and theta band activity, can be applied to visual processing of language remains unknown. Nevertheless, similar temporal integration windows (see Holcombe (2009) for a discussion of multi-time resolution model of vision) and neural activity in these frequency bands have also been implicated in visual cognition. The phase of EEG oscillations in the theta and alpha range is reported to be closely tied to a viewer's ability to detect flashes of light, suggesting that visual detection thresholds fluctuate at these frequencies (Busch, Bubeis, & Van Rullen, 2009). A study using an MMN paradigm with visual stimuli found temporal windows of 150-170 ms in duration (Czigler, Winkler, Pató, Várnagy, Weisz, & Balázs, 2006) like the auditory studies. Drawing upon broader literature on the temporal organization of information (Warren, 1999; Yost & Popper, 1993), Poeppel (2003) points to the prevalence of windows that are ~50 ms and ~200 ms in duration across many sensory systems. The implication of gamma-band oscillation during attentional selection of sensory information and theta (and alpha) range oscillation in top-down effects in processing (Engel, Fries, & Singer, 2001) potentially make periods of these frequencies privileged time-scales that are relevant to all language processing, and even more broadly to core cognitive functions, not just speech.

1.5 Oscillation of sub-lexical units in language

As described above, research on brain rhythms reveals new insights on the basis for oscillatory patterns in speech processing. However, the most prominent theory on the periodic basis of language production focuses on the motor frame of the mandible (MacNeilage, 1998; MacNeilage & Davis, 2001). When its closing and opening are coupled with vocalization, consonant-vowel syllables emerge. In early stages of language acquisition, babbling is marked by its rhythmic qualities. When the oscillatory property of language is attributed to biomechanics that are unique to the jaw, it is predicted that other forms of language production, such as sign language, should not have similar cyclic characteristics.

Those who are familiar with sign language are sensitive to rhythmic qualities in signing, but perhaps the discovery of manual babbling among children exposed to sign language has provided the most powerful counterevidence to claims that rhythms in language come from the biomechanics of the jaw (Petitto & Marentette, 1991; Meier & Willerman, 1995). Manual babbling is marked by repetitive qualities, involvement of possible sign language handshapes, syllabic organization, and production without reference, and is distinct from ordinary gestures. Such movements have also been attested among deaf and hearing babies born to deaf parents and exposed to signing (Petitto, Holowka, Sergio, & Ostry, 2001). When the temporal dynamics of these movements were studied using opto-electronic position-tracking equipment, manual babbling was marked by a slower rhythm (~ 1 Hz) than ordinary gestures (~ 2.5 Hz) and the manual movements of babies who were not

exposed to sign language input (~ 3 Hz) (Petitto, Solowka, Sergio, Levy, & Ostry, 2004) (see Figure 1). Moreover, low-frequency manual babbling was restricted to a smaller linguistic signing space. Because this babbling is quantitatively and qualitatively different from the manual gestures of hearing babies who were not exposed to sign language, it cannot be simply attributed to general motor development.

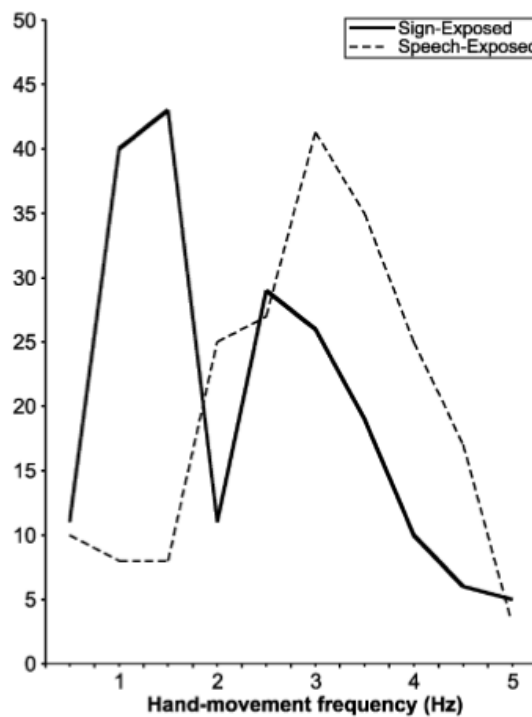


Figure 1. Reproduced from Petitto, Solowka, Sergio, Levy, & Ostry (2004), this figure shows the distribution of the frequencies (in Hz) of the manual movements among sign-exposed and speech-exposed babies. Sign-exposed babies had movements that were at two different frequencies, where manual babbling in the signing space was marked by a slower rhythm (~ 1 Hz) than ordinary gestures outside the signing space (~ 2.5 Hz), whereas speech-exposed babies had movements at a higher frequency (~ 3 Hz).

Although numerous studies have studied the qualitative aspects of vocal babbling (de Boysson-Bardies, 1999; Locke, 1983; Oller & Eilers, 1988; Jusczyk,

1997; Elbers, 1982; Vihman, 1996), few have investigated its frequency characteristics (Dolata, Davis, & MacNeilage, 2008; Levitt & Wang, 1991). In the analysis of infants learning French and English, an average syllable duration of ~300 ms (~ 3.4 Hz) is reported by Levitt and Wang (1991), and similar values were found by Dolata et al. (2008). When compared to the rates found in adults, where syllables are ~200 ms in duration (~ 5 Hz) (Arai & Greenberg, 1997), these frequencies for vocal babbling in infants is notably lower. While these differences suggest developmental constraints in the maturation of rhythm and repetitive production, studies on infant-directed language also reveal the possible role of input. Masataka (1992) found that infant-directed signs (“motherese”) among users of Japanese Sign Language were produced at a mean rate of 1.3 per second, which approximately matches the manual babbling rate reported by Petitto et al. (2004). Nevertheless, since Masataka’s report of 1.5 signs per second in adult-directed production is considerably slower than what is reported in American Sign Language (Bellugi & Fischer, 1972), it is difficult to draw conclusions about the reliability of this connection.

Although sign language production does not have one main oscillator like the mandible in speech, rhythmic patterns underlie many biological systems (Ghez & Krauker, 2002; Fenélon, Casasnovas, Simmers, & Meyrand, 1998), not just speech. Thus, it should be no surprise that these findings from babbling provide evidence for the role of oscillatory patterns in linguistic processing regardless of modality. Although modality seems to have a key effect on the frequency of these oscillations, rhythmic production is a universal precursor to language development (and perhaps

development more broadly), and it may also play an integral role in perceptual processes as well. If frequencies in babbling, where manual babbling is slower than vocal babbling, have any parallels for processing at maturity, it is possible that sign language processing involves larger temporal integration windows than in speech. However, because these windows are posited to occur at multiple time-scales, it is also possible that the size of these windows converge at longer time-scales.

1.6 Rates of processing in language

Bellugi & Fischer (1972) examine the rates of natural production in English and ASL and demonstrate that the rates converge at the overall propositional (or sentential) level although they are different at the word and sign level. This comparison was made by studying three bilingual CODAs who narrate the same story in both languages. To ensure that the rates in ASL were not attributable to the hearing status of the signers, a rate analysis was also conducted on three deaf native signers, which resulted in similar findings (Klima & Bellugi, 1979). Making sense of the differences in rates for words in English and signs in ASL despite the similarities in the global rates requires an understanding of sign language grammar, which employs a signing space, the obligatory use of which cannot be found in any spoken language. Contrary to common myth, ASL is not a manual form of spoken English. Signing systems that are artificially created to help teach deaf individuals learn English, referred to as Manually Coded English, cannot be learned naturally, and when taught, often get reduced to forms that more closely resemble ASL (Supalla,

1991). The fact that production in such signing systems is twice as slow as ASL (Klima & Bellugi, 1979) suggests that natural sign languages follow critical time pressures, and that only the grammar of natural sign languages are compatible with these constraints at the manual-visual interface. Natural language processing seems to require a specific range of rates for informational flow.

As discussed in the previous sections, a prominent theory for the basis of rates in syllables focuses on the motor constraints of the mandible (MacNeilage, 1998), which converge with findings implicating endogenous theta band oscillations in speech perception (Giraud, Kleinschmidt, Poeppel, Lund, Frackowiak, & Laufs, 2007; Luo & Poeppel, 2007). Since syllables provide a frame for speech content, including units of meanings, it is logically possible that syllable rate determines the overall rate of information transfer. However, the evidence for similarity in rates across modalities suggests that temporal constraints in language processing go beyond bottlenecks at particular motor interfaces. Sign languages are phonologically encoded through handshape, location, movement, orientation, and non-manual features, with the dominant hand as the primary articulator. Because the articulators are physically larger in signing than in speech and must overcome greater inertia, it is theoretically possible that signing could be slower than speech. In contrast, the availability of two hands and a potentially richer set of phonological features may also permit signing to result in overall faster rates than speech. Given two very different motor systems, it is puzzling that rates in signing and speech converge the way they do.

Perceptual studies on rate-compressed sentences, where artificially accelerated input remains intelligible (to a certain limit) (Foulke & Sticht, 1969; Foulke, 1971; Ahissar, Nagarajan, Ahissar, Protopapas, Mahncke, & Merzenich, 2001; Fischer, Delhorne, & Reed, 1999) demonstrate that the production system does impose a bottleneck in processing to some degree. Artificially accelerated inputs remain intelligible at rates that are above what the motor system is capable of producing. Although the similarity in rates in English and ASL suggests that this bottleneck is not particular to an individual motor system, parallel bottlenecks are also found in perception across the two modalities. In studies of speech, intelligibility of sentences declines dramatically as a function of the compression rate, where intelligibility is measured as the fraction of correct words in a sentence that the participants are able to produce back. Although compression by a factor of 2 remains almost perfectly intelligible and shows only a slight dip in performance, compression by a factor of 3 results in a steeper decline (to 50% intelligibility) (Ahissar, Nagarajan, Ahissar, Protopapas, Mahncke, & Merzenich, 2001; Ghitza & Greenberg, 2009), and further compressions quickly reach zero levels of intelligibility.

In the first study to investigate the effect of compression in sign language perception, Heiman and Tweney (1981) used compression only by a factor of 2. Using sentences in narratives, intelligibility was measured by performance on a comprehension test rather than by requiring the participants to produce the sentences back and measuring accuracy. They find that comprehension decreased ~ 20% as a result of the compression. To test whether this should be attributed to signal degradation or to an overload of short-term memory, they also tested stimuli where

black, blank films were inserted between “semantically unitary statements” in the compressed versions so that the length of the narratives was equivalent to the noncompressed versions. With these blank film insertions, comprehension scores were similar to (and slight lower than) the compressed condition. This result does not converge with similar experiments in speech that were conducted later by others, where silences in prosodic boundaries improve the intelligibility of compressed speech (Wingfield, Lombardi, & Sokol, 1984; Ghitza & Greenberg, 2009). It is important to recognize that silences in speech are not equivalent to blank films in ASL, where the repetition of a still image of a signer may have been a fairer comparison. Nevertheless, Heiman and Tweney (1981) conclude that memory constraints do not underlie poorer comprehension in compression. When testing the intelligibility of single signs, comparing the control condition to whether they were also compressed by a factor of 2, they find similar decreases in performance as with sentences. When interpreting the results as a whole, Heiman and Tweney (1981:12) conclude that “decrements in comprehension may be due to cumulative decrements in intelligibility.”

Fischer, Delhorne, and Reed (1999) test intelligibility of ASL sentences and single signs as a function of several compression rates in ASL. Intelligibility was measured by taking the percent accuracy of correctly identified signs for each condition. According to their summary, similar patterns in speech and sign, where compression by a factor of 3 results in a steep decline in the intelligibility of sentences (see Figure 2), suggest that even in the perceptual interface, time

constraints for processing are modality-independent at the sentence level. At the level of words and signs, however, different patterns emerged (see Figure 2).

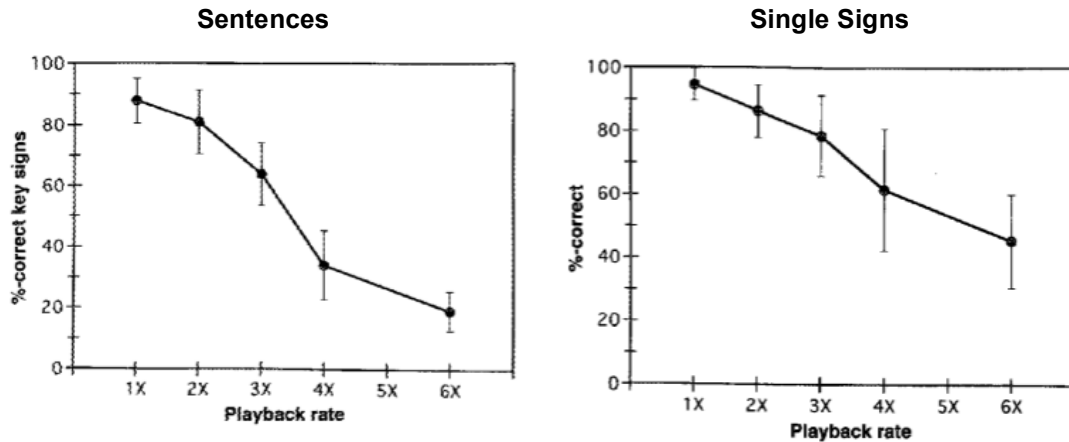


Figure 2. Reproduced from Fischer, Delhorne, & Reed (1999), these figures show the intelligibility of stimuli as a function of playback rates for 14 participants. Error bars represent plus or minus one standard deviation of the mean. With sentences, a sharp drop in intelligibility is found at compressions by a factor of 3.

In the findings from speech, the intelligibility of single (monosyllabic) words is more sensitive to compressions (Beasley, Schwimmer, & Rintelmann, 1972), which is attributed to the benefit of having sentential context (Miller, & Heise, & Lichten, 1951). In contrast, Fischer et al. (1999) find that individual signs are more resistant to compressions than sentences, where same compression factors result in higher intelligibility scores. This result is attributed to the fact that signs in isolation take longer to produce than when produced inside a sentence. An analysis of the number of video frames revealed that signs in isolation took twice the number of frames. The additional frames included those that show a sign in the “final hold” (Liddell, 1984), which involves holding a hand configuration in place or a repetition movement. They report that extra frames contain enough information to reveal the identity of the sign.

Although a follow-up experiment that controls for these factors would be useful, where single signs are extracted from sentences rather than produced in isolation, the overall pattern seems to be that tokens that take longer to produce are more resistant to compression. Similar findings are found in speech, when compression is achieved by time sampling. Vowels are longer in duration than consonants, and they are also more resistant to compressions than consonants (Kurtzrock, 1957). Overall, words that have higher number of phonemes are more resistant to compressions than shorter words (Henry, 1966).

Ghitza and Greenberg (2009) test the intelligibility of sentences through the combination of compression and insertions of silence. Unlike Heiman and Tweney (1981), who selectively inserted blank “silent” films at phrasal boundaries, these silences were inserted either periodically (silences of fixed duration) or aperiodically (silences with randomized duration within a range).

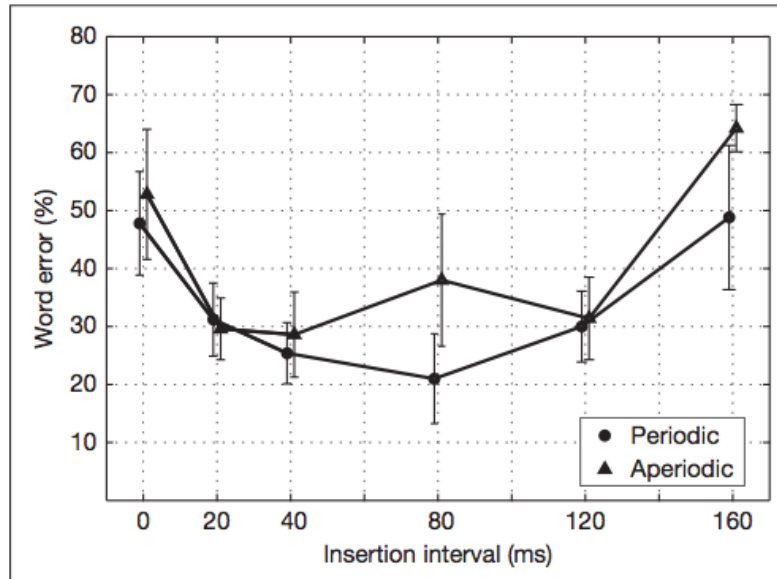


Figure 3. Reproduced from Ghitza & Greenberg (2009), this graph shows the percent error in an intelligibility experiment, where sentences were compressed by a factor of 3 and silences were inserted periodically or aperiodically. Error bars represent the standard deviation of the mean.

As shown in Figure 3, in cases with time compressions by a factor of 3 and no insertions of silence, the intelligibility of the sentences are 50% (error rate ~ 50%). As silences are inserted periodically, intelligibility improves significantly, peaking to 80% (error rate ~ 20%) where 80 ms silences are inserted in between 40 ms of speech material. This alternation creates sentence durations that match the original uncompressed sentences. These results suggest that perhaps only 20% of the 50% drop in intelligibility of sentences compressed by a factor of 3 should be attributed to sensory loss of information. Ghitza and Greenberg use these findings to argue for the importance of endogenous rhythms, specifically in the theta frequency range, for speech decoding.

As Foulke and Sticht (1969:60) emphasize, there is a point at which “a factor in addition to signal degradation begins to determine the loss of comprehension,” in sentences. In their review, there are cases where performance on the identification of words is lower than overall comprehension of sentences, and where it is also higher. The point at which processing full sentences becomes compromised by fast rates in English seems to be 275 words per minute. In addition to identifying words, sentence processing requires temporarily storing and performing operations with those words. Beyond the issues of recovering an input from a degraded signal, there seems to be a channel capacity for processing the flow of linguistic information. The results from ASL in production (Bellugi & Fischer, 1972; Klima & Bellugi, 1979) and perception (Fischer, Delhorne, & Reed, 1999) provide support for a model in which this channel capacity is modality-independent. In addition to better understanding channel capacities, determining the time-windows of integrating information that flows through a channel and determining if these windows are dependent on a sensory modality will lead to an improved model of language processing.

1.7 Outline of the dissertation

Thus far, I have provided background information that motivates the studies that follow. Studying languages that use different sensori-motor systems are essential for understanding core language properties. An important aspect of language processing is its temporal dynamics in on-line perception and production.

Chapter 2 focuses on temporal integration windows in the visual processing of ASL. As outlined in sections 1.3 and 1.4, evidence from temporal integration windows come from a wide range of methodologies and domains. Experiments 1, 2, and 3, tests the intelligibility of ASL sentences as a function of the size of local-reversals, a methodology that is motivated by previous work in speech perception (Saberri & Perrott, 1999; Greenberg & Arai, 2001; Figueroa, 2009; Stilp, Kieft, Alexander, & Kluender, 2010). Through these experiments, it will be shown that time-scales of temporal integration arise from 1) modality, 2) size of linguistic units, and 3) developmental factors.

Chapter 3 focuses on the rates of natural production in English, ASL, and Korean, where rates of words, signs, morphemes, and syllables are reported. Previous studies of English and ASL have attributed differences in word and sign rates to both modality and grammar. By also including an analysis of Korean, which has relevant grammatical properties similar to ASL, a better understanding of the role of modality and grammar in language rates is possible. As will be demonstrated in Chapter 2, an understanding of rates in natural production is a fundamental part of building models for temporal integration windows in language.

Chapter 4 concludes with a synthesis of all the results and a discussion of implications for future research.

2 Temporal integration windows in sign language

2.1 Introduction

Linguistic structures are processed in time, whether listening to acoustic speech or viewing the visual input of sign language. The goal of the three experiments presented in this chapter is to investigate factors that affect the temporal integration windows in language perception. Temporal integration windows are chunks of times during which information is collected and integrated and refer to durations among many phenomena, from the level of the neuron, where output spikes are dependent on the sum of activities (Theunissen & Miller, 1995), to the psychophysical level, where sensory stimuli can be detected and compared to previous inputs (Viemeister & Wakefield, 1991; Näätänen, 1992), to the level of higher cognition, where information is consolidated in memory (Wilson & McNaughton, 1994; Buzsáki & Draguhn, 2004; Furman, Dorfman, Hasson, Davachi, & Dudai, 2007). In studies of speech perception, temporal integration windows refer to durations of time over which the sensory signal are mapped to units of linguistic representations – such as phonemes and syllables (Poeppel, Idsardi, & van Wassenhove, 2008). In audio-visual studies, it has also been used to describe the durations for multi-sensory integration (van Wassenhove, Grant, Poeppel, 2007). Beyond levels of pure sensory processing, some differences between studies that use speech-like and non-speech-like stimuli, where longer lags are tolerated with speech-like stimuli, suggest that the nature of linguistic information in the input also influences the duration of these windows.

The importance of temporal direction in processing can be demonstrated by the simple scenario of playing backwards a spoken sentence that is about 2 seconds in duration – it is utterly unintelligible. In contrast, a sentence that is locally reversed in 20 ms increments is perfectly intelligible (Saberri & Perrott, 1999; Greenberg & Arai, 2001). The mechanisms underlying the mapping of the acoustic signal to meaningful linguistic representations cannot handle distortions over longer time-scales like 2 seconds. Speech is somewhat robust to a variety of adverse conditions, such as noise (Sumby & Pollack, 1954), compression (Foulke & Sticht, 1969), and interruptions (Miller & Licklider, 1950). In these cases, portions of the signal is either masked or deleted, but in the case of backwards speech, all of the input is intact. The unintelligibility of backwards speech is probably a result of several factors, from at the sensory level to higher levels of linguistic processing. For example, simply reversing the order of words in a string (global reversal with local integrity) results in an ungrammatical sentence (*sentence ungrammatical an in results string a in words of order the reversing simply*), and one that can be extremely hard to understand or repeat back. Nevertheless, it would still be possible to pick out a few words from the reversed sentence, which feels like a random list of words. At another level, a sentence can be difficult to understand because no words can be recognized from an acoustic stream (for example, *lacitymargnu ecnetnes* for *ungrammatical sentence*). These examples demonstrate the importance of temporal direction in language processing.

Backwards speech is the most drastic form of temporal order distortion (Saberri & Perrott, 1999). Gradually increasing the degree of this distortion can lead

to a better understanding of the temporal constraints for the construction of linguistic representations from the sensory input. The use of local-reversals has already provided insights to mechanisms for processing the acoustic signal through time and the duration of temporal integration windows in speech. However, without a comparison to languages that use a different modality, it is impossible to determine whether such mechanisms are specific to auditory processing or more generalizable to all sensory processing in language. Experiment 1 is designed to determine temporal integration windows in the visual processing of language through locally time-reversed sentences. It serves as the basis for two follow-up studies, where the duration of temporal integration windows are tested as a function of rate in Experiment 2 and as a function of age-of-acquisition in Experiment 3. The next three sections motivate each of these experiments in turn.

2.2 Cognitive restoration of locally time-reversed sentences

As mentioned above, backwards, or globally reversed, speech is utterly unintelligible, but locally reversed speech demonstrates a different phenomenon (Saber & Perrott, 1999). The creation of locally-reversed stimuli is like rotating the orientation of slats in Venetian blinds. First, a sound waveform is subdivided into intervals of fixed duration, then each interval is reversed, such that only the temporal order of the input within the interval is altered, not the global order of the entire

speech stream (see Figure 4). In these studies, the intelligibility of the sentences drops as a function of the size of the reversals.

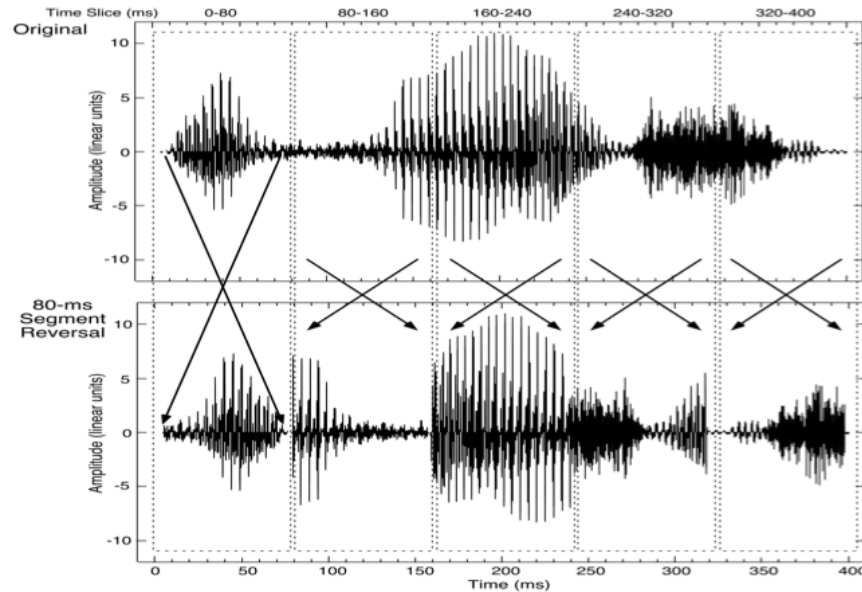


Figure 4. Reproduced from Greenberg & Arai (2001), this figure demonstrates how locally-reversed speech stimuli are created. Here, each 80 ms segment is played backwards, but the original order of the segments is maintained.

In the first study of this kind, Saberi and Perrott (1999) find that intelligibility of a single sentence, as measured by subjective reports, drops to 50% at 130 ms reversals, falling to 0% around 200 ms (Figure 5). Thus, the ability to cognitively restore locally-reversed sentences up to ~100 ms was presented as further evidence for temporal integration windows of these durations. They note that reversals of short durations are less likely to disrupt the temporal envelopes of speech, which have been proposed to be important cues to intelligibility (Greenberg & Arai, 1998).

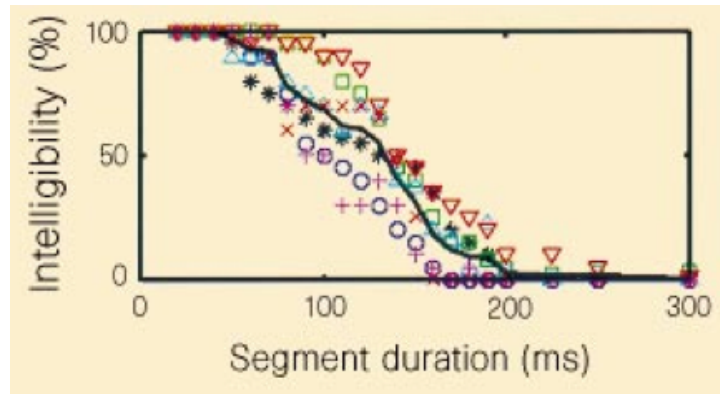


Figure 5. Reproduced from Saberi & Perrott (1999), this figure shows subjective intelligibility ratings by 7 participants on a single sentence that was repeated for all conditions.

However, Saberi and Perrot's (1999) conclusion that "a detailed auditory analysis of the short-term acoustic spectrum is not essential to the speech code" is oversimplified. It is clear from examples like *wolf* and *flow* that the direction of information within a temporal envelope provide critical cues for correct word recognition. As emphasized in the multi-time resolution model of speech perception (Poeppel, 2003; Poeppel, Idsardi, van Wassenhove, 2008), windows that are short and long in duration are important for information processing at different hierarchies. Whether or not Saberi and Perrot's (1999) result support integration at shorter or longer time scales is hard to determine without having data on the stimulus that they used, such as information on the average size of segments and syllables in the sentence. Notably, they tested only a single sentence, which was repeated for all the conditions. As they report, repetition of the sentence, even with larger reversal sizes, improves its intelligibility, reflecting "cognitive recalibration" and learning.

Greenberg and Arai's (2001) replication of the study produced different results, where intelligibility fell quite sharply at smaller time scales, reaching 50% at

60 ms (Figure 6). They argue that information about syllable segmentation as well as fine structure with phonetic details are important for speech perception. In their study, intelligibility was measured quantitatively by scoring the number of words in a sentence that were identified correctly.

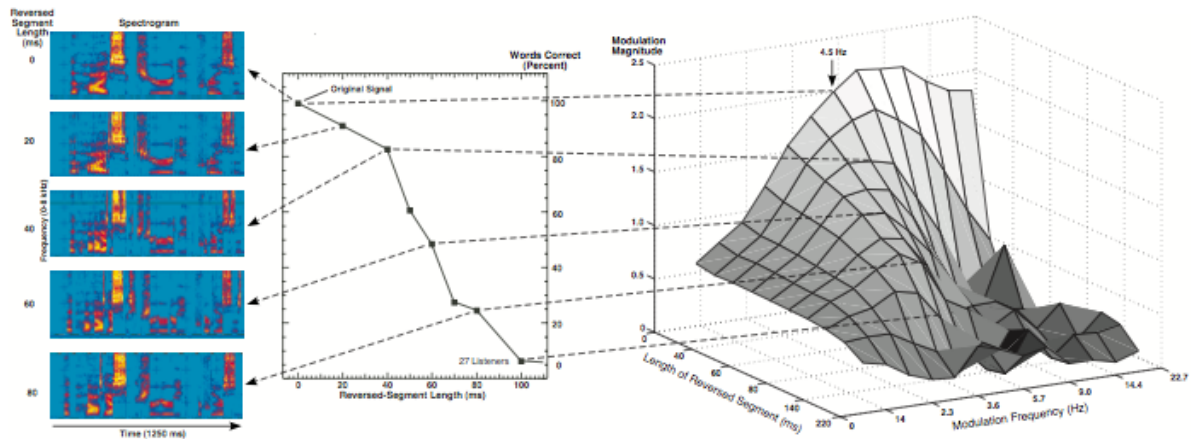


Figure 6. Reproduced from Greenberg & Arai (2001), this figure demonstrates 1) the spectrogram of locally reversed sentences, 2) the intelligibility curve as a function of reversal sizes, and 3) the complex modulation spectrum of the sentences. Intelligibility results are from 27 participants tested on 40 sentences. Intelligibility of sentences falls drastically between 40 and 50 ms reversals, falling to 50% at 60 ms reversals, and reaches ~0% by 100 ms reversals.

40 sentences from the TIMIT corpus were chosen for their low semantic predictability and diversity of speakers. Because this minimizes the influence of context effects in guessing words or adjusting to the acoustics of a single speaker, the study targets sensory processing in sentence comprehension. Moreover, an analysis was conducted on all the stimuli to better understand spectro-temporal consequences of local reversals. Greenberg and Arai (2001) tested the relationship between the size of the reversals and the amplitude component of the modulation spectrum alone, as well as the complex modulation spectrum, which is calculated by taking information about amplitude and phase components of the modulation spectrum at various

frequency bands. The inclusion of phase information, which was referenced with respect to the phase of the control condition, was critical to finding a correlation between the modulation spectrum and intelligibility. This should not be too surprising given that with reversals in duration of syllables, the temporal envelope of the speech stream is almost entirely preserved, yet sentences are utterly unintelligible at reversals of 100 ms, which is smaller than the duration of average syllables, and stay unintelligible with larger reversals. As the size of the reversals are increased, the phases between the original and reversed stimuli become increasingly dispersed. Greenberg and Arai attribute the difficulty in processing sentences with local reversals to the distortion of information that is critical for identifying phonetic information.

Greenberg, Hollenback, and Ellis (1996) find that the median duration for most segments is 60-100 ms (shorter for stops and longer for diphthongs) in natural speech (Switchboard corpus). In a different study using sentences from the TIMIT corpus, Arai and Greenberg (1998) report that the mean duration of a phonetic segment is 72 ms. The sharp decline of speech intelligibility at local reversals >50 ms and falling to 50% at 60 ms, suggests that acoustic signals must be integrated in short time windows to recover phonetic information (Greenberg & Arai, 2001). Reversals in shorter durations preserve not only the relative order of the phonetic segments within words but may also capture the fine-structures within a phonetic segment for recognition in word contexts.

Experiments on the perception of interrupted speech have shown that speech can be almost entirely intelligible under certain conditions where 50% of the speech

material is deleted (Miller & Licklider, 1950) (see Figure 7). Intelligibility of sentences varies as a non-monotonic function of the frequency of the interruptions/deletions. Interruptions at low frequencies occur at longer durations per interruptions, causing certain words to be either entirely captured or missed. At higher frequencies, parts of words get interrupted, and intelligibility is relatively high in cases where some information about every phoneme in the word is preserved. They write, “It appears that one glimpse per phoneme is sufficient [for intelligibility]” (Miller & Licklider, 1950: 168). Taken together with these findings, results from the local-reversal of speech suggests that “looks” to each phoneme and the order of these looks are important for intelligibility, but that the temporal direction within each look is more flexible.

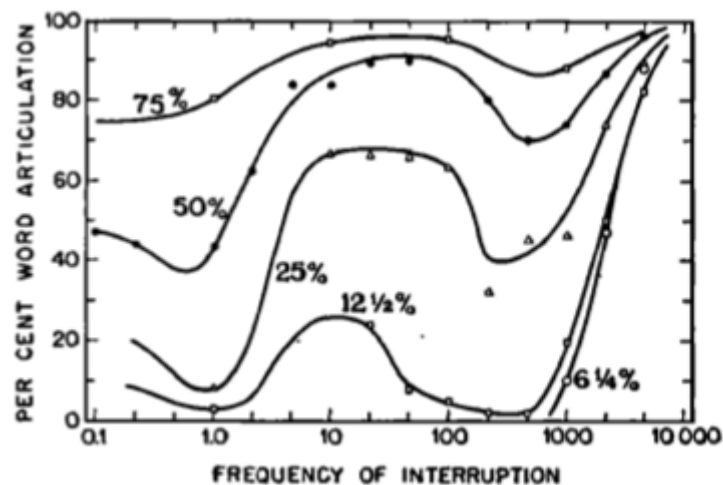


Figure 7. Reproduced from Miller & Licklider (1950), this figure demonstrates the intelligibility of English sentences as a function of frequency of interruption and speech-time fraction (where the duration of interruptions were dependent on the frequency of the interruptions and speech-time fractions and were spaced regularly).

Beyond sensory integration, it is likely that additional factors play an important role in findings from the local reversal of speech. The intelligibility curve is likely to shift rightward (more intelligible with more severe distortions) with the semantic predictability of the sentences (Miller & Isard, 1963). Moreover, the curve is likely to shift leftward (less intelligible with less severe distortions) if a word-list was used rather than sentences (Miller, 1951). As Ghitza and Greenberg (2009) conclude based on evidence where the periodic insertion of silences in time-compressed sentences significantly improved intelligibility, “Intelligibility is not simply a matter of decoding the spectro-temporal pattern.” Listening comprehension and word intelligibility can be dissociated in compression studies (Foulke & Sticht, 1969; French & Steinberg, 1947), and similar patterns may emerge in the case of local-reversals, although this has yet to be tested.

In summary, the ability to detect, integrate, and decode rapid acoustic signals in the speech stream is essential for auditory speech perception. One of the implications from the intelligibility of locally time-reversed speech is that the acoustic signals are not integrated continuously but in a discrete manner. Although the ability to cognitively restore time-reversed stimuli is limited, listeners’ tolerance for local reversals is still remarkable. By manipulating stimuli so that it is chunked in larger sizes, Saberi and Perrott (1999) and Greenberg and Arai (2001) determine perceptual and cognitive limitations for integrating the signal. Findings from Greenberg and Arai (2001) suggests that one important time-window for integrating the speech signal lies somewhere below ~60 ms. Reversals that go beyond these

perceptual integration windows cannot be cognitively restored and no linguistic representations can be recovered.

2.3 Flexibility of perceptual parameters to rates

Speech perception requires flexible mechanisms that can accommodate a wide variety of conditions, created by individual speakers (age, gender, accents/dialect, emotional state, speaking rate, etc.) and environments (noise). Within the speech stream of one individual, one can find phonetic segments that vary in duration (stop consonants are shorter than fricatives, for example), and a given segment may have different phonetic realizations depending on context (whether occurring in word initial or final positions, or in stressed or unstressed syllables, for example). On average, phonemes and syllables are produced at relatively consistent rates (Greenberg, Hollenback, & Ellis, 1996; Arai & Greenberg, 1998), but each unit has its own range of variability. A pattern that underlies rate uniformity is that shorter syllables are flanked by longer syllables, and vice versa (Greenberg, 1999). Since speech is not perfectly periodic, the temporal integration process in speech must be flexible.

The ability to adjust perceptual parameters to a variety of contexts is referred to as perceptual normalization. One area where the effect of speaking rate has been tested (by varying the duration of syllables) is in categorical perception, where perceptual boundaries are found within a continuously varying parameter. For example, a key acoustic cue to distinguish between [b] and [w] is in the duration of

the formant transition at stimulus onset, which is longer for [w]. The transition durations at which the perceptual response changes from [b] to [w] has been shown to shift depending on whether the onset was produced within a long or short syllable (Miller & Liberman, 1979). Another example comes from voiced and voiceless stop onsets, where the timing difference between the onset of laryngeal voicing and the onset of the stop release burst (Voice Onset Time, VOT) is a prominent cue for distinguishing between them (Lisker & Abramson, 1964; Lisker, 1975; Klatt, 1975). Moreover, there is an interaction between VOT and place of articulation, where the boundary between voiced and voiceless stops shifts towards longer VOTs as the closure for the stop is made further back in the mouth. The categorical boundary between a particular set of voiced and voiceless stop onsets is not absolute, however. The boundaries shift towards longer VOTs when the syllable is lengthened either by acoustic or visual cues (Summerfield, 1981; Green & Miller, 1985) (Figure 8).

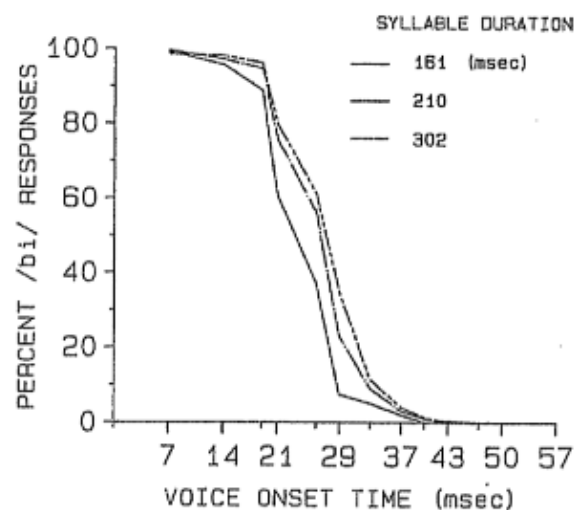


Figure 8. Reproduced from Green & Miller (1985), this figure demonstrates that perceptual boundary, reflected by the percentage of voiced responses for [bi]-[pi] continuum, varies depending on durations.

Greenberg and Arai's (2001) finding that intelligibility falls sharply at reversals >50 ms, reaching 50% at ~ 60 ms reversals, and the measurement of phonetic segment durations with similar materials in a separate study (where phonemes were 72 ms long on average) (Arai & Greenberg, 1998) suggests a link between temporal integration windows and the duration of linguistic units in speech, as has been proposed by Poeppel (2003). Under this hypothesis, temporal integration windows, as revealed through local reversals, should be variable depending on the distribution of phonetic segments and overall rates of speech. Moreover, it has the potential to be generalized as a fundamental language processing mechanism, applying also to sign languages. With the same technique, the intelligibility of signed sentences may also be dependent on the size of locally reversed video segments and rates, albeit at different time-scales. However, another possibility is that temporal integration windows of such time-scales are important to general auditory processing and somewhat independent of the linguistic nature of the acoustic signal. Yet a third possibility is that ~ 60 ms integration windows are important for all language processing, regardless of modality, and not linked to the auditory channel. The discrepancy between the findings in Saberi and Perrot's (1999) and Greenberg and Arai's (2001) studies already suggests that ~ 60 ms is not a perceptual primitive but a result of a combination of factors.

To investigate the relationship between temporal integration windows and duration of linguistic units, Figueroa (2009) (also Figueroa, Howard, Idsardi, & Poeppel, 2009) used a novel combination of compression and local-reversals (see Figure 9 for results). Stimuli consisted of TIMIT sentences, which were presented in

10 conditions, where the reversal sizes ranged from 0 to 100 ms, lengthened in 10 ms increments. At the normal rate of speech, a sharp drop in intelligibility to 50% was found around 60-70 ms reversals, replicating the results of Greenberg and Arai (2001). In addition, intelligibility was tested on conditions where the sentences were either compressed by a factor of 2 or dilated by a factor of 1.5. With the faster rate, intelligibility fell to 50% around 30-40 ms reversals, revealing time windows that are half the durations found in normal speech. In the case of the dilated condition, performance on the intelligibility task was close to ceiling even at 70 ms reversals, falling to 50% around 80-90 ms.

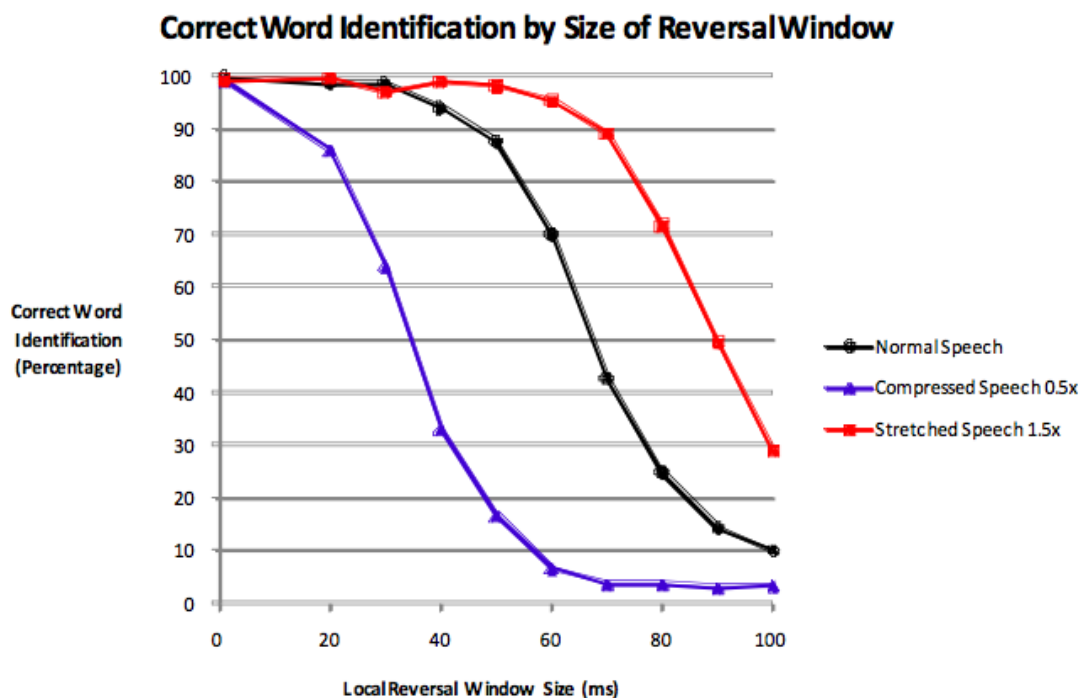


Figure 9. Reproduced from Figueroa (2009), this figure shows the intelligibility of English sentences as a function of compression and reversal size.

Similar findings are reported in a study using synthetic speech, even though the sentences selected from the HINT corpus (Hearing In Noise Test, Nilsson, Soli, &

Sullivan, 1994) were reported to be semantically more predictable and easier than TIMIT sentences (Stilp, Kiefte, Alexander, & Kluender, 2010). The sentences were synthesized to produce 2.5, 5.0 or 10.0 syllables per second, where the average duration of sentences was 2.6, 1.4, or 0.8 s, respectively. Five reversal conditions were used (0, 20, 40, 80, and 160 ms reversals). Across the three rate conditions (slow, medium, and fast), intelligibility reached relative minimum levels when the reversals were roughly the durations of one syllable, suggesting that tolerance for temporal distortions do not arise through absolute perceptual limits for durations but are proportionally relative to the amount of distortion. A visual inspection of the graph shown in Figure 10 indicates that intelligibility falls to 50% at ~30 ms at the slow rate, ~60 ms at the medium rate, and ~120 ms at the fast rate.

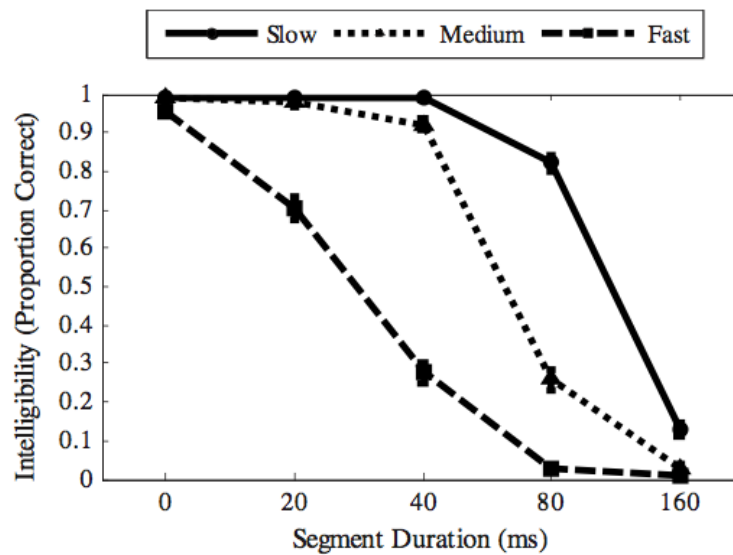


Figure 10. Reproduced from Stilp, Kiefte, Alexander, & Kluender (2010), this graph shows intelligibility curves of English sentences as a function of the size of local-reversals (segment durations in ms) and speech rates (in syllables per second: slow = 2.5, medium = 5.0, fast = 10).

Although these findings demonstrate the flexibility of perceptual processes, to adjust to rate changes as well as directional distortions, they may also reveal limitations. For example, Figueroa (2009), who used 10 conditions of reversal sizes, notes that when doing equivalent time window comparisons, the intelligibility curve does not shift directly in proportion to the rate in the dilated sentences, where performances fall slightly sooner (that is, at smaller reversals than predicted). This is attributed to the possibility that the perceptual system has inherent properties that constrain its ability to adjust to distortions beyond a certain range. This limit (80-90 ms) for temporal integration may be reached even before the stimuli properties are predicted to induce processing difficulty. In the same vein, if speech processing relies upon temporal integration windows that adjust to all speech rates, it is predicted that a wider range of temporal integration windows can be found in this specific processing task. Testing the intelligibility of locally reversed sentences that are further compressed and dilated in gradual steps may reveal that the curves do not shift beyond certain points.

2.4 Perspectives from development and bilingualism

An important part of development is learning the same perceptual tuning process that is present in adults. Some of the abilities underlying perceptual normalization are present at early ages. Using the continuum between [b] and [w], Eimas and Miller (1980) found that infants as early as at the age of 2 – 4 months perceive these sounds categorically. Moreover, they report that perceptual boundaries

shift in relation to the duration of the syllable, as found among adults (Miller & Liberman, 1979).

Although infants are born with high sensitivity to potentially phonemic contrasts found among the world's languages, their capacity to discriminate non-native contrasts declines by the age of 10 months (Werker & Tees, 1984). However, this sensitivity is not lost forever, as demonstrated by the native acquisition of a second language through exposure during sensitive periods (Werker & Tees, 2005). An examination of the conditions under which sensitivity to non-native phonemic contrasts can be recovered reveals the important role interpersonal interaction during exposure (Kuhl, Tsao, & Liu, 2003). It is reported that 9-month old American infants do not show the ability to discriminate the alveolo-palatal affricate and fricative of Mandarin Chinese. These infants who were exposed to Mandarin Chinese either through audio-only or audio-visual recordings through experimental sessions did not show sensitivity to differences among these sounds, whereas those who were exposed to the language input through interpersonal interaction showed a recovery of these phonemic contrasts.

Although children develop sensitivity to native contrasts at early stages, the maturation of perceptual abilities through a variety of speech conditions takes longer periods of time through physiological changes. Studies have shown that children have poorer temporal resolution than adults (Abel, 1972; Wrightman, Allen, Dolan Kistler, Jamieson, 1989), reaching adult performance levels around age 8 (Davis & McCroskey, 1980). The ability to discriminate small changes in frequency may not reach adult-level acuity until about 6 years of age (Olsho, Schoon, Sakai, Turpin, &

Sperduto, 1982; Jensen, Neff, & callaghan, 1987). Children have more difficulty than adults in perceiving speech in noise (Mills, 1975; Elliott, 1979; Nitttrouer & Boothroyd, 1990; Fallon, Trehub, & Schneider, 2000). This is attributed to age-related differences in auditory sensitivity, where children have higher auditory thresholds than adults in especially low frequency ranges, and these thresholds may not reach adult levels until around the age of 10 (Elliott & Katz, 1980). Fallon et al. (2000) report that children even at 11 years of age require higher signal-to-noise ratios than young adults to perform comparably in word identification tasks where sentences are embedded in multitalker babble.

For both children and adults, the amount of exposure to a language plays an important role in speech processing. Familiar words generally require less acoustic information for identification (Rosenwieg & Postman, 1957). Part of children's poorer performance with word detection in noise is associated with their limited language experience (Elliott, 1979; Nitttrouer & Boothroyd, 1990). Similarly, non-native listeners are more adversely affected by noise than native listeners (Buus, Florentine, Scharf, & Canévet, 1986; Mayo, Florentine, & Buus, 1997). In tasks where natives and non-natives may perform similarly in quiet conditions, differences emerge with the introduction of noise. Mayo et al. (1997) investigate the effect of age-of-acquisition on English L2 performance in different degrees of noise. The task for participants, whose first language was Spanish, was to identify the target word at the end of an English sentence, where sentences were either high or low in predictability. Those who learned English before the age of 6 were more resilient to noise (in other words, had higher accuracy with similar noise-levels) than those who

learned English after the age of 14. In addition, late learners did not show a difference in performance between sentences with low and high predictability, unlike the early learners. Because the subjects were all similar in age, late learners overall had shorter duration of exposure than the early learners. Nevertheless, statistical analysis taking into account exposure duration demonstrated that it is not as strong of a predictor as age-of-onset in these results. When exposure duration was matched with the early learners, late learners still showed a significantly poorer performance.

All together, these findings suggest that factors underlying difficulty in speech processing among children and non-native adults differ. While children undergo developmental changes in their auditory processing abilities and can benefit from increasing exposure to their language, late learners seem limited by constraints that are more permanent. Another example of the difference between the developmental constraints underlying children and adult late-learners comes from the differences in performances to conversational and clear speech. Clear speech, which is marked by enunciation and corresponding acoustic-phonetic markers, benefits adults with normal hearing, impaired hearing, as well as children (Picheny, Durlach, & Braidá, 1985; Bradlow, Kraus, & Hayes, 2003). However, a comparison of performance while processing conversational and clear speech among non-native adult listeners did not show the same degree of benefit from clear speech (Bradlow & Bent, 2002). This suggests that benefits from clear speech are derived from rich experience with a language, and it appears that this experience must be gained in early ages of development.

The difference between early and late bilinguals in the study conducted by Mayo et al. (1997) demonstrates the role of input during sensitive periods for developing the flexibility to adjust to a wider range of perceptual environments. However, the difference between monolinguals and early bilinguals in the same study also point to the effect of bilingualism itself on perceptual capacity. Although the performance among monolinguals and early bilinguals (as well as late bilinguals) were very similar in quiet conditions, monolinguals performed better than both groups in noisy conditions. Similar patterns are replicated in another study that tested bilinguals whose first language is Italian (Meador, Flege, & Mackay, 2000). Because accent and proficiency of English was not assessed in Mayo et al.'s (1997) study, and because speakers even in the early bilingual group were reported to have a noticeable foreign accent in Meador et al.'s (2000) study, Rogers et al. (2006) focus their comparison on monolinguals and early bilinguals, whose abilities in English and Spanish were assessed through questionnaires, interviews, and recordings (Rogers, Lister, Febo, Besing, & Abrams, 2006). Only monolingual and bilingual participants that were rated by monolingual speech-language pathology trainees as having little or no regional or foreign accent in English were included for the full study. Among the bilinguals, language assessments suggested that some degree of language attrition in Spanish was present among more than half the participants. To extend the findings of previous work, Rogers et al. included noise conditions with reverberations. Reverberations refer to the persistence of a sound, which is common in enclosed spaces. Noise and reverberations often occur simultaneously, and their combination is more detrimental for speech perception than the sum of the individual components

(Nabelek, 1988). Comparing the consequences of these distortions, which are present in typical environments, among monolinguals and bilinguals contributes to a better understanding of the factors underlying perceptual adaptability.

Roger et al. (2006) replicate previous pattern of results, where both monolinguals and early bilinguals have perfect performance in a word recognition task in quiet conditions, and moreover, the performance of bilinguals decreases more dramatically to the noisy conditions. The findings suggest that in addition to age-of-acquisition factors, acoustic degradations more adversely affect bilinguals than monolinguals. One potential explanation for these results is that bilinguals have a larger number of target phonemes with two languages, forcing them to have more fine-tuned perceptual abilities that are consequently less robust in noise. Another possibility is that language processing for bilinguals requires more cognitive resources to suppress the other language. Because the baseline performance of bilinguals is thought to be already attentionally demanding, less resource may be available to them in adverse conditions. While constant experience with more than one language has been shown to have beneficial effects in the domain-general aspects cognitive control (Bialystok, 2001), these findings suggest that reduced perceptual adaptability to noise is a cost.

Although this summary has focused on studies where differences among native and non-native participants only emerge in adverse listening conditions, age of acquisition has consequences for many areas of language processing, including phonology (Oyama, 1976; Flege, MacKay, Meador, 1999) and morphosyntax (Klein & Dittmar, 1979; Johnson & Newport, 1989; Beck, 1998; DeKeyser, 2000;

DeKeyser, Ravid, Alfi-Shabtay, 2005). Evidence from sign language research demonstrates that age-effects in language acquisition are not specific to auditory perception or vocal production (Newport, 1990; Mayberry, 1993). Late-learners of sign language show “accents” in their production (Cicourel & Boese, 1972; Kantor, 1978; Mirus, Rathmann, & Meier, 2001; Rosen, 2004; Chen Pichler, 2006; Boyes-Braem, 1999). Some features that reveal non-nativeness include handshapes, facial expressions, rhythm, and movements. Accents also exist for those who are native signers in one sign language and learning another (Budding, Hoopers, Mueller, & Scarcello, 1995).

In perception, native signers and late learners differ in judgments about what phonological aspects of signs are most salient (Corina & Hildebrandt, 2002). An eye-tracking study has also shown that native signers and beginning signers fixate on different locations in the signing space (Emmorey, Thompson, & Colvin, 2009). In the perception of handshapes, which sometimes show categorical perception among native signers (Emmorey, McCullough, & Brentari, 2003; Baker, Idsardi, Golinkoff, & Petitto, 2005), late learners show different profiles (Best, Mathur, Miranda, & Lillo-Martin, 2010). In particular, performance of deaf late-learners reflect more attention to fine-grained phonetic properties of signs than deaf native signers and hearing late signers (Best, Mathur, Miranda, & Lillo-Martin, 2010). This converges with previous studies that suggest that deaf late-learners experience “phonological bottlenecks” that have consequences for many other aspects of processing (Mayberry & Fischer, 1989; Mayberry, 2007). Late-learners take longer to identify ASL signs in gating tasks, requiring more phonetic or phonological information than native signers

(Emmorey & Corina, 1990). When testing sentence recall among signers who first acquired ASL at ages ranging from birth to 13, later ages of acquisition were linked to lower performance, due to increasing phonological errors and inefficient sign recognition (Mayberry & Fischer, 1989), even when length of signing experience was comparable (Mayberry & Eichen, 1991). Furthermore, phonological errors were correlated with poorer comprehension. In a probe recognition task, where the participant has to accurately respond whether or not a target sign was present in a sentence, late signers were slower to reject phonologically similar substitutes (Emmorey, Corina, and Bellugi, 1995). In contrast, native signers were only affected by semantic substitutes.

Later ages of acquisition result in difficulty with grammatical aspects as well. In tasks testing morphological processing, late learners show variable use of morphology (where obligatory morphemes are omitted) as well as inappropriate use of whole-word signs that reflected incorrect representation of the morphological structure (Newport, 1990). Grammatical judgment accuracy of sentences decreases with delays in exposure to a first language, (Mayberry, 2003; Boudreault & Mayberry, 2006; Mayberry & Lock, 2003). In cases where late learners have comparable levels of performance on an off-line grammaticality judgment tasks, differences emerged in on-line tasks (Emmorey, Bellugi, Friederici, & Horn, 1995; Emmorey, 1995).

When sign language skills were tested in a sentence shadowing task, where participants simultaneously watch and produced sign language narratives, better performance among native signers also reflected better comprehension (Mayberry &

Fischer, 1989). In this study, performance in good and poor viewing conditions were also compared, where the poor conditions were created by adding visual noise of randomized black and white dots, which looked like video “snow.” Although this reduced shadowing accuracy overall, the effect was similar for both native and non-native signers. As mentioned previously in spoken language studies, late bilinguals were more adversely affected by auditory noise than monolinguals or early bilinguals (Buus, Florentine, Scharf, & Canévet, 1986; Mayo, Florentine, & Buus, 1997). Although more studies on the effect of visual disruptions in sign language processing is necessary, if these patterns persist, it would suggest that some aspects of perceptual adaptability to adverse conditions are modality dependent.

No studies have investigated the effect of late language acquisition on temporal integration windows in processing. Understanding the effects of late learning is particularly important in sign language processing because >95% of deaf individuals are born to hearing parents and do not receive exposure to language from birth (Mitchell & Karchmer, 2004). Because the experiments presented here examine temporal integration windows through locally time-reversed sentences, understanding the consequences of processing imperfect input among native and non-native users of a language are relevant factors to consider.

2.5 Experiment 1 – Effect of modality on temporal integration windows: evidence from local-reversals of ASL sentences

Building upon studies on the cognitive restoration of locally time-reversed speech (Saberri & Perrott, 1999; Greenberg & Arai, 2001), the aim of Experiment 1 is to better understand the universal constraints for temporal direction and integration by testing the visual processing of language. The ability to detect and integrate rapid acoustic signals in the speech stream is essential for auditory speech perception. The intelligibility of temporally distorted sentences suggests that one important time-window for integrating the speech signal lies somewhere below ~ 50 ms (Greenberg & Arai, 2001). Sentences with local-reversals within this window can be cognitively restored successfully, whereas sentences with reversals in sizes that go beyond these perceptual integration windows cannot. What happens in a language that is processed visually? Anecdotal accounts suggest that globally reversed ASL is somewhat more intelligible than backwards English, but not entirely so. To what degree the spatial encoding and the temporal properties of ASL affect time window of integrating linguistic information in the visual modality remains unknown.

If previous findings from the perception of locally-reversed speech can be extended to the processing of ASL, this would suggest that 50 ms time-windows are modality-independent. Given that <50 ms time windows were attributed to the analysis of fine-structures in speech (Greenberg & Arai, 2001), this prediction is unlikely. For ASL, time-windows at longer time-scales would point to differences in processing due to the sensory channel of communication. Finding a different

psychophysical profile for a sign language would point to unique temporal integration windows for the visual interface for language perception, but it does not necessarily rule out the possibility that these windows are dependent on the linguistic nature of the signal. Communication through a different sensori-motor system affects linguistic properties of the language. For example, although the vast majority of spoken languages employ sequential morphology to increase the complexity of words, all sign languages use simultaneous or non-concatenative forms by using spatial modulation and non-manual markers for grammatical inflection. This means that for spoken languages, the addition of each new unit of meaning often takes longer to produce, but this is not necessarily the case in sign languages. What could be the cause of such differences? Simultaneous strategies may be favored in sign languages not only because spatial encoding allows for layered units of information but also because of the time pressures for natural language processing.

Although using ASL involves manual articulators that are massive compared to the organs of the vocal tract in speech, the global rate of naturally produced utterances, as measured by the number of propositions in a given amount of time, is similar to spoken English (Bellugi & Fischer, 1972; Klima & Bellugi, 1979; Grosjean, 1979). When comparing smaller units such as words and signs, however, the rate of signs is half the rate of words in the equivalent measure of time. Similar to English, ASL constructs linguistic structures incrementally over time. One hypothesis is that each step contains more linguistic information and takes longer to produce than an incremental step in English. Although these larger steps take longer to produce, because information is encoded simultaneously in ASL, overall they may

contain similar amounts of linguistic information that is sequentially presented in the same amount of time in speech. Artificially created systems that force sequential grammar on a signing system, such as various forms of Manually Coded English, may not be learned naturally like ASL because even in highly skilled usage, communication is too slow. These findings reveal an intimate connection between modality and linguistic properties.

Moreover, since phonological features of signs such as handshape and location are spatially encoded, is ASL more resistant to the temporal distortion of time-compression? Findings similar to speech indicate that there is a modality-independent upper limit to language processing (Fischer, Delhorne & Reed 1999; Foulke & Sticht 1969), suggesting that the channel capacity for language is not unique to a particular sensory system.

In addition to providing to way to learn about time scales for processing successive input, local-reversals also test a more general ability to tolerate distorted signals. Intuitively, spatial encoding of phonological features is predicted to make sign languages more robust to temporal distortions, but according to Fischer et al.'s (1999) conclusion about their results seen in time-compressed studies, this is not always the case. However, the greater resistance of ASL to temporal distortions of *disruptions* in the signal created by repetitive temporal interruptions, as compared to English (see Figure 11), has been indeed attributed to properties of the visual modality (Tweney, Heiman & Hoemann, 1977).

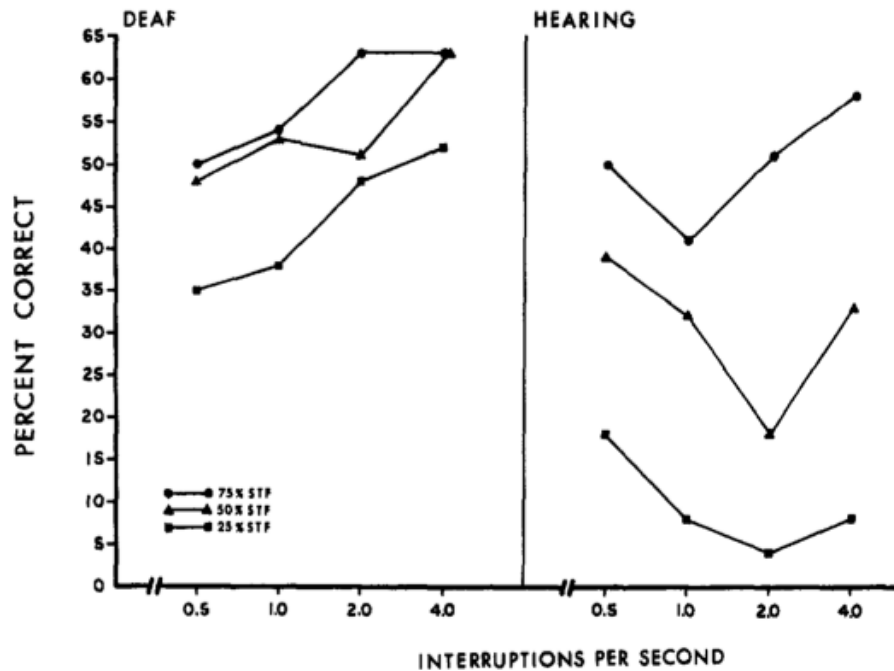


Figure 11. Reproduced from Tweney, Heiman & Hoemann (1977), this figure shows the intelligibility of ASL and English sentences as a function of temporal disruption frequency and signing/speech-time fractions. These results demonstrate that sign language is more resistant to temporal disruptions than speech.

Thus, although spatial encoding of phonological features in ASL may result in higher accuracies overall for local-reversals, its effect on the size of the time-windows is less clear. Based on the hypothesis that there is an intimate connection between modality and the rate at which linguistic units are incrementally presented over time, it is predicted that ASL sentences are integrated over larger-time windows than English.

Materials

40 sentences of ASL were constructed by sign language linguists. These sentences were designed to be natural sentences of ASL, using a diverse range of

phonological parameters, but low in semantic predictability to force the participants to pay close attention to the input rather than guess the sentences from identifying few key signs. These criteria are similar to the ones that are reported for sentences of English in the TIMIT corpus. Although the ideas for semantic content of the sentences were taken from TIMIT sentences, the sentences constructed here were not direct translations but rather followed ASL grammar. Two examples of stimulus sentences translated into English are, “The girl tends to visit the frog on Wednesdays” and “Heavy snow and strong winds make it hard to see the outline of the mountains.” The mean average duration of these sentences was 3.53 s (standard deviation \pm 0.94 s). Since there is currently no large ASL corpus that contains frequency measures for signs (although see Morford and MacFarlane (2003), the first to describe frequency characteristics of ASL), effort was taken to incorporate signs of varying lexical frequency based on intuition. These sentences also incorporated non-manual features that are appropriate to standard use of ASL. Based on the estimate that fingerspelled words appear as frequently as 7%-10% in the overall vocabulary in everyday signing (Padden, 1991), a few fingerspelled words were incorporated into these sentences as well. These words were either lexicalized signs (e.g., B-N-K for ‘bank’) or common proper names (e.g., N-A-N-C-Y for ‘Nancy’).

A female native signer modeled 40 experimental sentences and 8 practice sentences facing the camera. All sentences started and ended with a “neutral” position, which was defined here as a relaxed position with hands crossed below the waist. The best tokens of each sentence produced by the model were then trimmed to 768 x 576 pixel frames (29.97 fps, or 33 ms per frame). Frame sequences in

uncompressed AVI files were locally reversed at increments of 4, 8, 12, 16, 20, 24, and 28 frames (133 – 934 ms) with a control condition without any reversal manipulation (0 ms) (see Figure 12 for an example). This resulted in a total of 320 sentences so that 40 sentences could be randomly assigned to 8 different conditions, with 5 examples per condition, for each participant. All videos were processed with Cinepak codec for stimuli presentation.

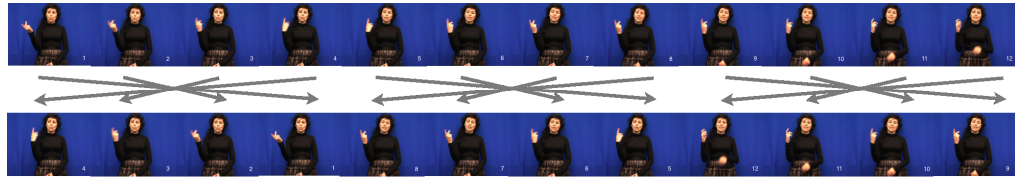


Figure 12. Demonstration of how locally time-reversed stimuli were created for sentences of ASL. This specific example shows reversals 133 ms in duration (reversals by 4 frames).

Procedure

Fourteen deaf participants (11 female, 23 mean age) who all grew up using ASL before one year of age participated in this study at Gallaudet University campus. Participants were compensated \$15/hour for the study. All instructions for the experiment, as well as informed-consent for participation in the experiment and video-recording of responses, were given to the participants both in ASL and written English, following the protocols of the Institutional Review Boards at Gallaudet University and University of Maryland, College Park. Participants were told that they would be viewing videos of sentences of ASL that had been modified by various degrees. They were instructed to sign to the camera whatever they could understand, trying to sign back the original sentence as much as possible. Unlike previous experiments from spoken English, where responses were type-written by the

participants, here the responses were signed and video-recorded. Participants were allowed to view a sentence up to four times before continuing onto the next sentence, moving at their own pace (following Greenberg and Arai (2001)). After the experiment, these responses were coded and analyzed for accuracy in comparison to the original sentences. Accuracy of the responses was calculated by determining how many of the manual signs, facial features for grammatical inflection, and spatial modulation were correctly reproduced by the participants. Signs were scored as being correct only when they matched all target parameters (handshape, location, and movement) and were produced in the original order. Data were analyzed to determine the intelligibility of the sentences as a function of the reversal size.

Results

Results for Experiment 1 are presented in Figure 13, where intelligibility is plotted as a function of reversal size. A one-way, repeated measures ANOVA (using R 2.8.1, R Development Core Team (2005)), revealed that intelligibility varies with the duration of the reversals ($F_{7,104} = 22.924$, $p < 0.001$). A remarkable difference from the patterns seen in the local reversal of speech is that intelligibility in ASL levels off to ~50% for even the most degraded stimuli. Tukey's Honestly Significant Difference (HSD) pair-wise post-hoc tests were conducted to check for differences in accuracy among the reversal conditions. In ASL, reversals 133 ms in duration only resulted in a small decrease in accuracy that was not significantly different from the control condition. A significant decrease in intelligibility is found between reversals 133 ms and 267 ms in duration. Intelligibility continues to fall until reversals of 533 ms, and no differences were found among larger reversals (533 ms – 933 ms).

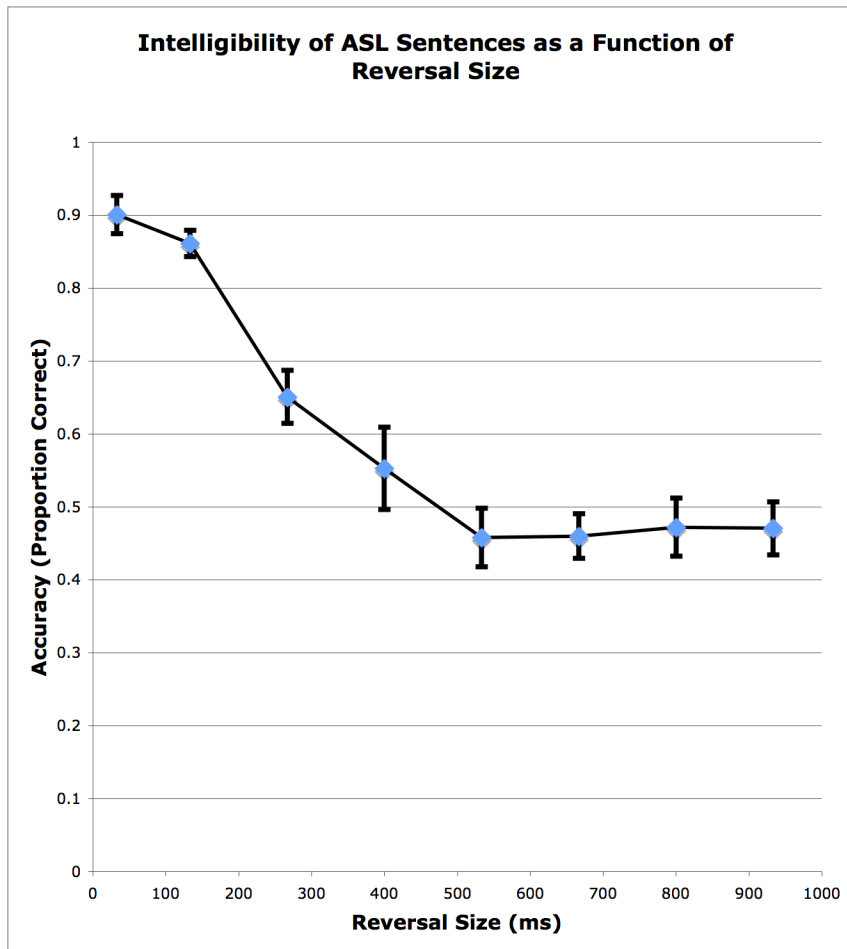


Figure 13. Results from Experiment 1 from 14 participants, demonstrating the intelligibility curve of ASL sentences as a function of reversal size, which implicates ~300 ms temporal integrations windows. 50% intelligibility of even the most degraded stimuli is attributed to spatial-encoding in sign language. Error bars represent plus or minus one standard error of the mean.

Discussion

In speech studies, temporal integration windows have been approximated by looking for the reversal size at which intelligibility falls to 50%, or ~60 ms for normal speech rates, where reversal sizes were increased in 10 ms increments (Greenberg & Arai, 2001). However, it is not possible to use this criterion in these results from a

sign language because even the most degraded sentences remain approximately 50% intelligible. Moreover, in this study, reversal sizes were increased in much larger increments (133 ms). Perhaps one way to compare the results across the modalities is to determine the reversal size for the half-way point between the highest and lowest accuracy scores, approximately 60 ms in speech and 300 ms in ASL. Alternatively, the first experimental condition at which intelligibility falls sharply can be compared across the modalities, 50 ms in speech and 267 ms in ASL. These estimates are not presented as absolute values but as approximations for time scales at which temporal integration occurs.

These results support the hypothesis that ASL sentences are integrated over longer time windows than English. The important differences across the modalities are as follows. In speech, 100+ ms reversal durations leads to intelligibility that is close to 0% (Greenberg & Arai, 2001). In sign language, reversals must be 500+ ms in duration to reach lowest accuracy scores. Even the most degraded sentences remain approximately 50% intelligible. Although 100+ ms reversals result in unintelligibility in speech, 133 ms reversals do not result in performance that is significantly different from the control condition (no reversals) in sign language. This should be contrasted with speech, where reversals do not result in drastic reduction of intelligibility as long as they are 40 ms or less. At 50 ms reversals of speech, intelligibility falls sharply, reaching 50% by 60 ms. In sign language, 267 ms reversals result in a significant decline in intelligibility.

In speech, accuracy in performance ranges from 100% at the control condition to 0% at the most degraded stimuli. Noticeably in this experiment with ASL, the

percent accuracy without local-reversals was at 90%. This is similar to the findings by Fischer et al. (1999), who tested the intelligibility of time-compressed ASL sentences, where percent accuracy was also around 90% for sentences played at the control/normal rate. Also in a sentence-shadowing task (Mayberry & Fischer, 1989), the upper boundary in percentage of ASL sentences that were signed back without any error in good viewing conditions for native signers was 76% in one experiment and 88% in another. The percentage of native signers who made no mistakes shadowing ASL sentences was 91%.

One reason why accuracy scores were ~90% for even temporally intact sentences could be attributed to task differences between this study and speech experiments (Greenberg & Arai, 2001). In speech, participants were asked to type their responses, which suggests that they had a workspace where they could reference and revise their answers while listening to the sentences. If hearing participants are asked to speak back the sentences in the same way that sentences were signed back in this experiment, it is possible that accuracy could be less than 100%. If, however, better controlling how responses are given and collected still results in significant differences at ceiling levels of performance, it may suggest that modality contributes to differences in the working memory for sentence processing.

Fischer et al. provide some speculations on why accuracy was overall higher for speech even with normal sentences in their compression study. They suggest the possibility that the sentences they used were not completely natural in ASL since they were translations of English sentences. Moreover, they attribute the unnaturalness of some of their sentences to the fact that isolated sentences with verb agreement and

topicalization are awkward in ASL because they are discourse-dependent (Lillo-Martin, 1991). If discourse-dependence of a language has a significant effect on the ceiling performance in sentence intelligibility tasks, it is predicted that other spoken languages that are described as being more discourse-dependent than English (such as Chinese, Korean, and Japanese (Lillo-Martin, 1991) will show similar patterns as ASL.

A review of the errors that occurred in the control condition in Experiment 1 indicates that participants did not sign certain pronouns. Because ASL is a pro-drop language, such omissions do not often change the meaning of sentences in conversations, but these were scored as errors in the experiment. Other errors included changes in sign order, selection of different variant handshapes, or selection of different signs that did not affect the meaning of the sentences. For example, one participant substituted WITH for HAVE in a control condition.

The mean average intelligibility of the conditions where performance levels off at its lowest point (533 ms, 667 ms, 800 ms, and 933 ms reversals) is 47%, which is notably higher than 0% found in speech. This is most likely attributable to the spatial encoding of phonological features in ASL. Information on handshape, location, and orientation of the signs, along with facial features, can be extracted from a single frame of a sign. Brentari (2002) compares signs without movement information to words without vowels, where consonants are more informative.

Nevertheless, all signs have movement, and signs can exist as minimal pairs just through the difference of movement. Moreover, native signers consider movement to be the most salient component of a sign (Corina & Hildebrandt, 2000).

Reversing the temporal order of the movement sequence can result in complete differences in meaning. Because of the spatial modulation in the verb GIVE, “I-GIVE-YOU” played backwards results in “YOU-GIVE-ME” (see Figure 14).



Figure 14. Reproduced from Liddell (2000), this illustration represents the sign for GIVE, where the direction of movement can mean I-GIVE-YOU but the reverse would result in the opposite meaning YOU-GIVE-ME.

Some movements also involve changes in handshapes. For example, TAKE played backwards results in DROP, and SEND played backwards results in SHUT-UP (Brentari, 1998). Another example of the importance of temporal order is in fingerspelling. However, there are many cases where movements are not as sensitive to temporal direction. Trilled movements have also been called local movement or oscillations (Liddell, 1990). For example, it is possible to accurately identify the sign for *tree*, which involves trilled radial-ulnar movement, whether it is played forwards or backwards. Finally, there are many signs that would not mean anything if they were played backwards, such as CHINA.

When examining the errors that participants make, some signs seem more sensitive or robust to the local time-reversals. In one sentence, REDUCE was

replaced by the opposite sign INCREASE. Even though the verb ASK involves both a change in movement and hand-orientation, one participant substituted ASK-YOU for ASK-ME even though ASK-ME played backwards has the wrong hand-orientation. A similar example is the substitution of the one-handed sign TELL-ME for the two-handed sign ANNOUNCE. These cases suggest that even in cases where the backwards version of a sign does not perfectly match up to a real sign, the saliency of the movement helps the viewer perceive approximates, which is consistent with Corina & Hildebrandts's (2002) results. Another participant interpreted the sign for PULL-APART, as in peeling an orange, as representing the shape of the orange, where the hands come together rather than apart. For many other signs, the decrease in accuracy scores was simply a result of omissions rather than wrong substitutions. Although fingerspelled words were recognized as fingerspellings, accuracy was low for these signs. The trilled signs DIRTY (wriggling movement) and ORANGE (closing movement) were always identified correctly.

Further analysis that is planned for the future includes studying the error types by determining what percent of errors were due to misses (no guesses) or mis-guesses. None of the previous studies on locally-reversed speech (Saber & Perrott, 1999; Greenberg & Arai, 2001; Figueroa, 2009; Stilp, Kieft, Alexander, & Kluender, 2010) report error types, but a comparison across the English and ASL may indicate modality-effects for error types. Because of the kinds of errors listed above, where playing a sign backwards can result in a close approximation to another sign, one prediction is that ASL will have a higher proportion of errors that are due to mis-

guesses than English, which is predicted to have a higher proportion of misses than ASL.

The fact that even the most degraded sentences in ASL are still ~50% intelligible reveals a modality effect in temporal integration, presumably due to the spatial encoding of phonological features in sign language. This result is similar to findings of Fischer et al. (1999), who tested intelligibility of ASL sentences that were compressed at different rates. Although they do not discuss this aspect of their data, figures show that even in the most accelerated condition (compression by a factor of 6), sentences are ~20% intelligible and individual signs are ~40% intelligible. As compared to the findings from English by Miller and Licklider (1950), Tweney et al. (1977) also find that ASL is more resistant to temporal disruptions. Even at the most disruptive condition of frequency of interruptions and speech-time fractions, ASL was found to be ~35% intelligible (versus 5% in speech). Like the findings from speech, intelligibility of a sequence of words is affected by whether they are presented in a random list or grammatical sentences. In both cases, ASL is more resistant to temporal disruptions than English. Tweney et al. (1977: 255) speculate “whether [resistance to disruption] derives from the linguistic structure of the signs or from the redundancy that would be possessed by any dynamic visual display.” Findings from the local-reversal of ASL and an examination of errors suggest that both factors are involved in the results. While ASL sentences are more resistant to local-reversals overall, the phonological/grammatical or prosodic features of some signs make them more robust than others.

Nevertheless, the sharp decline in intelligibility at 267 ms reversals suggests that reversals in such durations impose a significantly greater processing difficulty than conditions with reversals of shorter durations. In speech, acoustic signals in the speech stream fluctuate over smaller time-scales, and 50 ms reversals cause a significant decrease in intelligibility. This value is mostly likely attributable to the duration of fine-structures in speech, particularly consonants. Linguistic units in ASL have been reported to fluctuate over longer time-scales (Bellugi & Fischer, 1972; Wilbur & Nolen, 1986). Temporal integration windows in the time-scale of 250 – 300 ms may correspond to the average duration of syllables in ASL (Wilbur & Nolen, 1986). A comparison of English and ASL in the cognitive restoration of locally-reversed sentences suggest that temporal processing locks to different linguistic units across modalities (segments in speech and syllables in ASL). Nevertheless, although the values at which intelligibility falls drastically may differ between a signed and spoken language, they both implicate temporal integration windows that are dependent on the duration of linguistic units.

2.6 Experiment 2 – Effect of modality-independent mechanisms on temporal integration windows: evidence from compression and local-reversals of ASL sentences

In Experiment 1, intelligibility as a function of reversal size implicate temporal integration windows that are ~ 250 – 300 ms in duration. Experiment 2 is designed to test whether these results are linked to the size of linguistic units in ASL

or should be attributed to more general visual processing mechanisms. One way to test this is to manipulate the rate of the sentences, where compressions by a factor of 2 reduces the average duration of the articulations by half. Two studies in speech have used the combination of compression and local-reversals and demonstrated that the point at which intelligibility falls drastically is dependent on speech rates (Figueroa, 2009; Stilp, Kiefte, Alexander, & Kluender, 2010). Whereas intelligibility falls to ~50% at ~60 ms reversals at normal rates of speech, it falls the same amount at ~30 ms reversals at 2x rates, proportional to the reduction of sentence durations. These findings from speech, as well as the findings from Experiment 1, suggest that although durations of temporal integration may vary within and across languages, the mechanism of integrating the sensory input according to windows that track the fluctuation of linguistic units is universal.

Materials

The same 40 sentences of ASL from Experiment 1 were used again. The sentences were compressed by a factor of 2 by deleting every other frame from the original videos. The resulting videos were then locally reversed at increments of 4 – 28 frames (133 – 934 ms) with a control condition without any reversal manipulation (0 ms). This resulted in a total of 320 sentences so that 40 compressed sentences could be randomly assigned to 8 different conditions, with 5 examples per condition, for each participant. All videos were processed with Cinepak codec for stimuli presentation.

Procedure

Fourteen deaf participants (10 female, 21 mean age) who all started learning ASL before one year of age were recruited for this study on Gallaudet University campus. The procedure for the experiments were identical to Experiment 1 except the videos of compressed sentences were played. Participants were compensated \$15/hour for the study.

Results

Results for Experiment 2 are given below, superimposed with the results from Experiment 1, where intelligibility is plotted as a function of reversal size (Figure 15). A two-way repeated measures ANOVA (using R 2.8.1, R Development Core Team (2005)) indicated that intelligibility varies significantly by rate ($F_{1,208} = 23.33$, $p < 0.001$) and by reversal size ($F_{7,208} = 55.74$, $p < 0.001$), and that there is an interaction between rate and reversal size ($F_{7,208} = 2.41$, $p < 0.05$), most likely due to similar floor-effects of ~50% in both rate conditions. A one-way repeated measures ANOVA for the compressed sentences showed that intelligibility varies with the duration of the reversals ($F_{7,104} = 39.34$, $p < 0.001$). Tukey's Honestly Significant Difference (HSD) pair-wise post-hoc tests were conducted to check for differences in accuracy among all the reversal conditions for the compressed sentences. A sharp decrease in intelligibility is found between the control condition and 133 ms reversals ($p < 0.001$), and additionally between 133 ms and 267 ms reversals ($p < 0.001$). No differences were found among larger reversals (400 ms - 933 ms). Tukey's HSD post-hoc tests comparing normal and 2x rates at each condition indicate that there were no differences in conditions without any reversals, a significant difference at

133 ms reversals ($p < 0.05$), marginally significant difference at 267 ms and 400 ms reversals ($p \sim 0.10$), and no significant difference among larger reversals (533 – 933 ms).

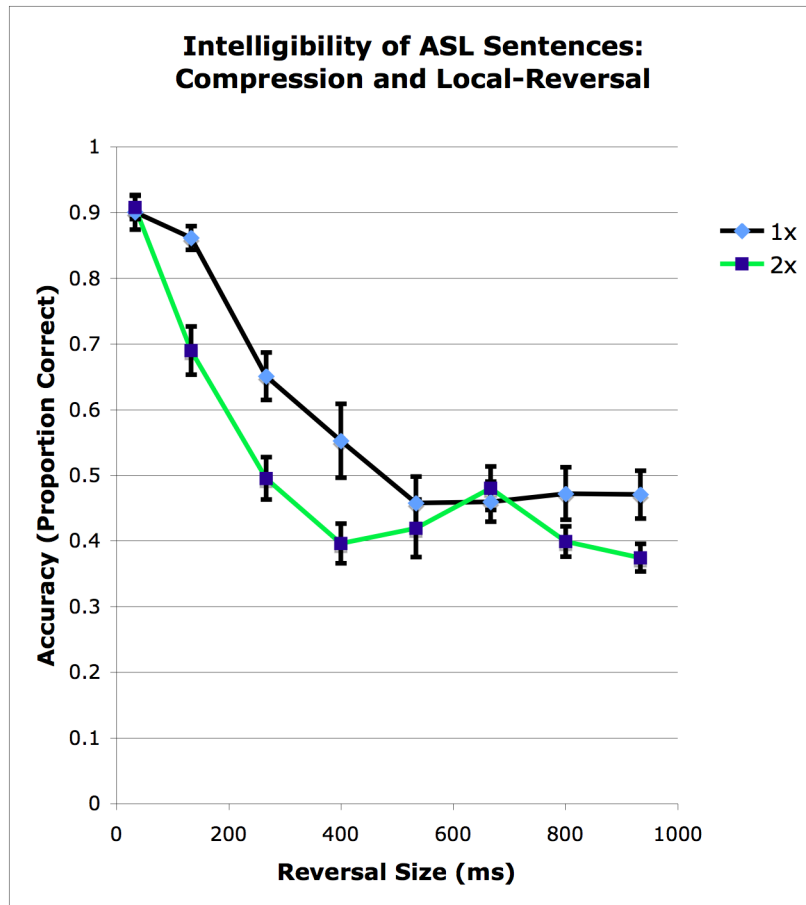


Figure 15. Results from Experiment 1 and 2 (14 participants in each experiment), demonstrating the intelligibility curve of ASL sentences as a function of reversal size and compression by a factor of 2, where temporal integration windows are proportional to the input rate (indicated by a sharp drop in intelligibility at ~267 ms reversals at the normal rate and ~133 ms reversals at the 2x rate). These results suggest that temporal integration windows in sign language are determined by the rate and durations of linguistic units. Error bars represent plus or minus one standard error of the mean.

Discussion

These findings support the assumption that local time-reversals of the sensory input provide insights about temporal integration windows of linguistic units. The results from Experiment 2 suggest that ASL sentences are integrated over time windows that scale to the duration of linguistic units. The key difference between Experiment 2, where sentences were presented at double the normal rate, from Experiment 1, is that intelligibility falls drastically earlier at 133 ms (compared to 267 ms). Moreover, intelligibility of compressed sentences plateaus sooner at 400 ms (compared to 533 ms in normal rate sentences).

These results are similar to the findings from speech (Figueroa 2009, Stilp, Kieft, Alexander, & Kluender, 2010), where the implication of ~60 ms temporal integration windows from local reversals of sentences played at normal rates does not seem to be inherent to auditory processing since windows decrease in duration as the linguistic units also decrease in duration with compression. In the same way, the results from Experiment 1 (~ 250 – 300 ms temporal integration windows) are indicative of both visual and linguistic aspects of integration. Experiment 2 provides evidence that a universal, modality-independent mechanism in sensory processing for language is to integrate the input in time-scales that scale to the fluctuation of representational units.

In speech, findings from local-reversals provide evidence for temporal integration windows that correspond to the sizes of phonemes. The temporal order of phonemes is crucial for word identification. Based on previous reports of sign and syllable rates (Bellugi & Fischer, 1972; Wilbur & Nolen, 1986), it is possible that ~

250 – 300 ms temporal integration windows correspond to the average duration of syllables in ASL. Chapter 3 explores in greater depth the average rate of signs, morphemes, and syllables among native signers of ASL. However, an analysis of signing rate for the 40 sentences used in these experiments indicated an average period of ~500 ms per grammatical component (manual sign, facial features for grammatical inflection, and spatial inflections), and longer for individual signs. In Experiment 1, intelligibility dropped sharply at 267 ms reversals. Wilbur and Nolen (1986) report that syllables are on average 250 ms in duration. Although segment analogies have been made in signs (Liddell, 1984), and reversing the temporal order of these segments can result change in meaning, these results suggest that sign languages can tolerate changes in temporal direction to a greater degree than in spoken languages, suggesting that the nature of how segments are encoded in signing and speech are quite different (Brentari, 1998; Wilbur & Allen, 1991). In addition to simultaneous encodings, repetitions that are possible in sign without producing lexical differences (Channon, 2002) make sign language much more robust to reversals in temporal direction than speech.

2.7 Experiment 3 – Effect of developmental factors on temporal processing: evidence from late-learners of ASL

An important aspect of language acquisition is to recognize – and match in both perception and production – the temporal dynamics of the target language.

Here, preliminary results are presented with deaf late L2-learners of ASL who self-

report English as their first language. Acquiring a language later in life often has consequences for properly learning the temporal parameters that distinguish different representations and processing a rapid sequence of input in on-line processing. The goal of Experiment 3 is to explore the effect of developmental factors on temporal processing by comparing the native signers of Experiment 1 with late-learners of ASL. Studying this group will also lead to a better understanding of whether the mechanism of temporally integrating the sensory signal according to the size of linguistic units in the language is universally present among all users of a language or only the native users. If this aspect of language processing is universally present even among late-signers, it is predicted that intelligibility of sentences would fall drastically at 267 ms reversals and level off at 533 ms. In contrast, if the phonological bottlenecks that late learners experience is due to having temporal integration windows that do not match the time scale at which the sensory information generated, then it expected that intelligibility will fall drastically at 133 ms reversals and level off around 400 ms. Another logical possibility is that intelligibility will fall drastically and level off at larger reversals, but together with previous work on spoken languages, where late learners are more sensitive to distortions in the input (Rogers, Lister, Febo, Besing, & Abrams, 2006), and pilot trials testing the experimental materials on late signers before Experiment 1 was conducted, it is predicted that late-learners are less tolerant to distortions than native signers. This is in contrast to the opposite prediction from the assumption that late learners are more tolerant of variability since they have less robust representations.

Processing locally time-reversed sentences requires perceptual flexibility. Perceptual flexibility may come from more efficient processing, where efficient processing is dependent on the coupling of perceptual processes to the signal that has to be analyzed. Previous work on bilinguals of spoken languages suggests that the ability to adapt to noisy listening conditions is weaker than monolinguals, and that late bilinguals face the greatest degree of difficulty in their second language (Mayo, Florentine, & Buus, 1997). Local-reversals create disruption that is akin to noise. Moreover, since studies on late learners of ASL have shown that they may use finer-grained phonological processing (Best, Mathur, Miranda, & Lillo-Martin, 2010), resulting in “phonological bottlenecks” that affect other aspects of processing (Mayberry & Fischer, 1989; Mayberry, 2007), it is possible that such results can be attributed to temporal integration windows in time-scales that are smaller than those that are characteristic of native signers. The longer temporal integration windows that more precisely match the duration of linguistic units may require language exposure at early stages of development. Experiments on non-linguistic stimuli (using flashes of light) have shown that visual processing can operate in smaller time-scales around ~150 ms (Busch, Buboïs, & Van Rullen, 2009; Perrett, Rolls, & Caan, 1982). However, it is also possible that the nature of signing rates present in the input may still require late-learners to integrate over appropriate durations that map to representational units.

Materials

The same materials from Experiment 1 were used in Experiment 3.

Procedure

8 deaf participants (7 female, 31 mean age) who all started learning ASL as an L2 after age 10 were recruited for this study at Gallaudet University campus. Based on the finding that L2 learners of ASL have distinct profiles from late L1 learners (Mayberry, 1993), it was decided that investigations on late learners should progress in at least two stages. This preliminary experiment tests the group that is assumed to have stronger language skills, so that age of acquisition is not confounded with additional factors associated with late L1 acquisition, although language skill assessment pre-tests were not included in the experiments. Moreover, because visual processing among deaf and hearing are known to be different due to different sensory experiences (Bavelier, Dye, & Hauser, 2006), it was decided that hearing late learners of ASL should not be included at this early stage of the project.

Here, only participants who reported profound hearing loss before the age of two years old and listed English as an L1 were included in the study. Participants were also required to have at least 5 years of ASL signing experience to be eligible. The average number of years of signing experience was 15. The procedure for the experiments was identical to Experiment 1. Participants were compensated \$15/hour for the study.

Results

Results from Experiment 3 are shown in Figure 16, presented together with the results from Experiment 1, where intelligibility is plotted as a function of reversal size. Because of the difference in sample sizes thus far (although plans are made to continue the study to full sample size to match Experiment 1), differences were not

tested statistically. However, the overall patterns indicate that performance among late-signers is lower than early-signers by ~10%. Late-signers also show a drastic decrease in performance at 133 ms reversals. Moreover, their performance levels off at two different time scales, first at 267-400 ms reversals, which is followed by another sharp decrease at 533 ms, after which intelligibility scores do not change significantly.

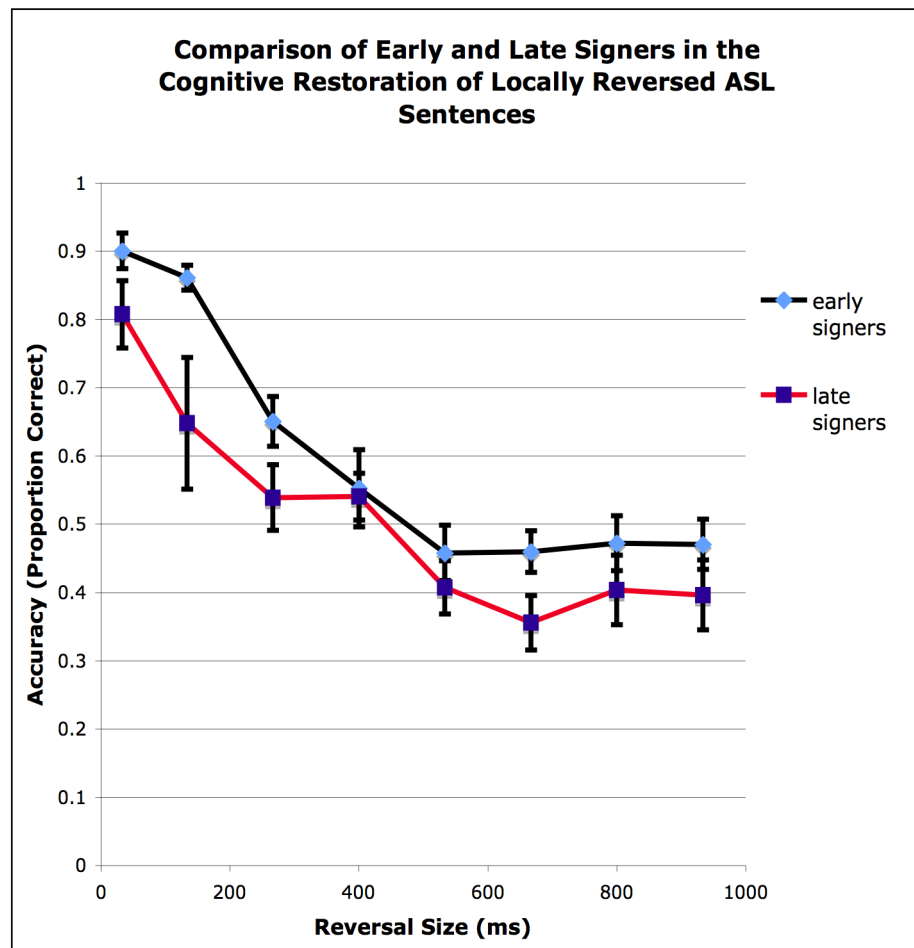


Figure 16. Results from Experiment 1 and 3, demonstrating the effects of age-of-acquisition in processing time-distorted stimuli. Note: $n=14$ in Experiment 1 and $n=8$ in Experiment 3. Late learners demonstrate greater sensitivity to time distortions in the input, but performance among the early and late learners plateau at similar distortion scales. Error bars represent plus or minus one standard error of the mean.

Discussion

These findings point to differences in temporal processing due to acquiring a language later in life. Experiment 3 tested two hypotheses: 1) temporal integration windows for late signers are the same as native signers if the size of linguistic units determines the duration of these windows, and 2) temporal integration windows for late signers are shorter in duration than early signers because the development of longer windows requires early exposure to signing. The current results partially support both hypotheses. The performance of early- and late-signers are similar in that intelligibility plateaus at similar reversal durations (533 – 933 ms). In contrast, Experiment 2 showed that shorter temporal integration windows result in earlier reaches to lowest levels of performance. However, performance among late-learners fell drastically at 133 ms reversals, similar to the findings from Experiment 2. Based on the shape of the intelligibility curve, it is likely that this was not simply due to an overall lower performance due to late acquisition. Although the task is generally more difficult for late-signers, they are also more sensitive to temporal distortions. Processing difficulty at smaller reversals implicate shorter temporal integration windows. Reversals that go beyond these windows become much more difficult to integrate and map to linguistic representations. Nevertheless, knowledge about the fluctuation and duration of linguistic units from signing experience may help late-signers recover information from reversals that exceed these windows.

More data is needed before conclusions can be made about whether late L2-learners processing normal rate sentences pattern more like early-learners processing normal rate sentences or compressed sentences, or whether they have their own

unique profiles as a group, as they seem thus far. Moreover, to better understand what processing of locally reversed stimuli tell us about supporting language skills and other cognitive factors, such as working memory, it would be worthwhile to conduct the experiments with language and cognitive assessments in the future, not only among late learners but all participants in experiments studying the cognitive restoration of time-distorted stimuli.

2.8 Conclusion

The three experiments presented in this chapter are the first to investigate the impact of modality on temporal integration windows in language processing. Locally time-reversing the sensory signals in a sentence provides insights on the mechanisms for recovering linguistic representations. Stimuli with reversal durations that go beyond certain limits cannot be integrated properly and cause problems for comprehension.

Experiment 1 demonstrates two effects that are driven by modality in language processing. Temporal integration windows are much longer in duration in the visual processing of language (approximately 250 – 300 ms) than in speech (approximately 50 – 60 ms). Moreover, spatial encoding in a visual language makes ASL much more resistant to temporal distortions, resulting in ~ 50% intelligibility for even the most degraded sentences (compared to 0% in speech). In speech, distorting the temporal direction of the input reveals smaller temporal integration that scale with the duration of segments. In sign language, locally-reversing sentences tap into the

processing and integration of larger syllabic units (Wilbur & Nolen, 1986). These differences have implications about the impact of modality on the temporal organization of representational units in languages, in addition to structural hierarchies (Brentari, 1998), and when temporal processing converges.

Despite these differences, spoken and signed languages seem to share the characteristic of having temporal integration windows that scale with the size of representational units in the languages. This modality-independent property was confirmed by the results of Experiment 2, where the duration of temporal integration windows were reduced to ~133 ms, in proportion to compression rate.

Results from late-signers in Experiment 3 offer a unique developmental perspective to temporal integration windows that has never been studied before, even in speech research. The findings suggest that longer temporal integration windows found in ASL processing is partially dependent on early exposure to a language where phonological units fluctuate at those time-scales. Early exposure to the target language gives earlier signers the advantage of being less sensitive to temporal distortions, perhaps due to longer integration windows. Moreover, having temporal integration windows that better match the size of linguistic units in the language may lead to more efficient processing. Although late-signers show indications of being sensitive to the duration of linguistic units, they also seem to be integrating the visual input at shorter time-scales. Testing late signers on compressed and locally-reversed sentences in the future will lead to a better understanding of how sensitive they are to time-scales of the input.

Like other studies (Mayberry, 1993; Mayberry & Fischer, 1989), Experiment 3 shows that late L2 signers have overall lower levels of performance on sentence repetition tasks. Many factors seem to contribute to this effect, including difficulty in grammatical processing, phonological processing, and working memory. Experiment 3 also indicates that late L2 signers are more sensitive to distortions in the input than early signers.

A comparison between late L2 deaf signers and late L1 deaf signers has yet to be tested. Performing the intelligibility tasks, where a distorted sentence must be repeated back, is assumed to require strong language skills from early language exposure (Mayberry & Fischer, 1989). Based on previous studies that compare these two types of late learners, where late learners of a first language perform more poorly on language and cognitive assessment compared to their early learning counterparts, (Mayberry, 2003; Boudreault & Mayberry, 2006; Mayberry & Lock, 2003), it is predicted that late L1 deaf signers will overall have lower accuracy with normal and locally-reversed sentences. If resilience to disruptions in the input is dependent on robustness of representations associated with early L1 acquisition, it is also predicted that late L1 signers will show a sharper drop in performance accuracy with local-reversals.

Based on the findings from bilinguals of spoken languages, it is possible that the present results from late L2 signers is due to late-learning or having ASL as another language. Another potential explanation is that having English as an L1 – a spoken language with much shorter temporal integration of segments – impacts temporal integration of ASL. A way to tease apart these explanations is to test

bilingual CODAs, for whom ASL is often the L1 and English an L2. If CODAs also show more sensitivity to noise in the input caused by temporal distortions than deaf native signers, it would suggest that bilingualism reduces perceptual flexibility across modalities. Such a finding would support the hypothesis that greater sensitivity to noise or other distortions in the input among bilinguals is due to the use of greater cognitive resources for language processing associated with suppressing the unused language.

On the other hand, if greater sensitivity to input distortions is due to sharing of phonological space by two languages, such effects should not be found across modalities, where phonological spaces do not overlap to such degrees. However, results from CODAs that show greater sensitivity to shorter reversals would be confounded with experience with spoken English, experience with which might bias shorter integration windows. Thus, it would be valuable to test late-signers for whom ASL is an L1. If this group is equally sensitive to temporal distortions as late-signers for whom ASL is an L2, then late-learning may be the best explanation for the results found in Experiment 3. Finally, it would be valuable to test late-learners of a spoken language (for example, English L2 bilinguals) to better understand the effects of age-of-acquisition and bilingualism on the integration of sensory signals for language processing. Comparing unimodal and bimodal bilinguals would be particularly interesting because time-scales in languages within a modality are much more similar than two languages in different modalities.

The difference in time-scales of temporal integrations windows in sign language ($\sim 250 - 300$ ms) and speech ($\sim 50 - 60$ ms) is attributed to locking to units

of different time-scales. However, another reason that distortions over larger time-scales are tolerated in ASL is that visually represented information may be encoded with greater temporal flexibility than auditory/speech-based representations. In working memory experiments, signers have been found to have spans in ranges that are shorter than hearing individuals (Wilson, Bettger, Nicolae, & Klima, 1997; Boutla, Supalla, Newport, & Bavelier, 2004). Nevertheless, native signers are reported to have an equally easy time recalling a list of items forwards or backwards (Wilson, Bettger, Nicolae, & Klima, 1997). Although Boutla et al. (2004) report that hearing bilingual signers have shorter spans on sign language tasks compared to speech tasks, order flexibility is not mentioned. In a non-linguistic task, Kimura et al. (2010) provide evidence that sequential regularities are automatically encoded in the visual system. Thus, the flexibility for order found among signers may be attributed more specifically to visual working memory or the prosodic features of ASL rather than visual encoding. A promising area of future research is to study the relationship among temporal integration windows, sensory encoding, and working memory, for both linguistic and non-linguistic functions.

Viewing locally-reversed videos and cognitively restoring the movements that encode the original linguistic message requires discrimination between movements actually produced by the signer and apparent motion created by discontinuous video frames. In studies that tested viewer's ability to re-create Chinese pseudocharacters from dynamic point light displays, those who had sign language experience (deaf and hearing signers) were able to determine the underlying discrete stroke movement patterns better than non-signers, although neither group was familiar with Chinese

(Klima, Tzeng, Fok, Bellugi, Corina, & Bettger, 1995; Bettger, 1992). One question that arises is how much experience with signing results in this enhanced ability to analyze movement, and how late signers compare with native signers in these skills.

Results from Experiment 2 suggest that $\sim 250 - 300$ ms windows in Experiment 1 are not absolute values in the visual processing of language but rather are relative to the size of linguistic units in the sentences at different rates of signing. Temporal integration windows in sign language perception may also converge with findings from reading. The average duration of eye fixations in reading is 200-300 ms (Rayner, 1998). One can also consider the possibility that language processing takes advantage of these time-scales that exist also for non-linguistic visual processing. In a study that tested viewer's ability to detect flashes of light, visual detection thresholds corresponded to the phase of EEG oscillations in the theta (4-8 Hz) and alpha (8-12 Hz) range (Busch, Bubeis, & VanRullen, 2009). Experiments on non-human primates have shown that some neurons do not respond to complex visual stimuli until 100-150 ms after stimulus onset (Perrett, Rolls, & Caan, 1982). One way to examine whether time-scales $\sim 250 - 300$ ms in duration are privileged windows for other aspects of visual processing is to test the local reversals of non-linguistic gestural movements. Anecdotal accounts of viewing videos while rewinding suggests that non-linguistic gestures are overall more tolerant of temporal reversals than ASL sentences but testing this more systematically through different degrees of temporal reversals may also reveal a sharp decline in intelligibility at similar time-scales.

In addition to learning how integration windows are linked to the sensory channel and the rate at which information is transmitted in language, a better understanding of the neural mechanisms for language processing may provide insights into its temporal dynamics. Oscillatory neuronal activity are known to be the underlying basis for temporal integration in a wide variety of domains (Buzsaki & Draguhn 2004), including speech perception (Poeppel, 2003). Specifically, activity in the frequency of gamma (~40 Hz) and theta (~5 Hz) bands has been proposed to correspond to integration of segments and syllables of a speech stream, which are on average 50-80 ms and 150-300 ms in duration, respectively (Boemio, Fromm, Braun, & Poeppel, 2005; Luo & Poeppel 2007). How such a model of a multi-time resolution process and neural oscillations, and in particular these frequency bands for integration and comprehension, extends to sign languages has never been investigated.

In addition to entraining to the physical characteristics of the sensory signal, endogenous brain rhythms also subserve other neurocognitive processes. For example, synchronization of neuronal firing with gamma-frequency oscillations is associated with feature binding and attention (Singer & Gray 1995; Fries, Nikolic, & Singer, 2007; Schroeder & Lakatos 2008). Among others, activity in the theta band has been implicated in working memory tasks and also attention (Jensen & Lisman 2005; Deiber, Missonnier, Bertrand, Gold, Fazio-Costa, Ibañez, & Giannakopoulos, 2007). What remains unclear, however, is the nature of the relationship between the neuronal oscillations that subserve language-independent functions and those that entrain to the sensory input in language processing. A better understanding of the

temporal dynamics of sign languages and delineating the similarities and differences with spoken languages will be an asset in these investigations.

3 Temporal Dynamics in Natural Production

3.1 Introduction

Understanding the time properties of perceptual processes in language requires information about the temporal dynamics in natural production. In this chapter, I provide a review of previous studies that have examined the rate at which information unfolds in spoken and signed languages. Then I provide new insights from comparisons of word, sign, morpheme, and syllable rates in English, Korean, and ASL, where data is taken from corpora of natural conversations. In Chapter 2, I demonstrated that perceptual mechanisms for analyzing the sensory signal depend on rate in sign language as well as speech. The goal of this chapter is to replicate and extend the findings from Bellugi and Fischer (1972) by investigating the relationship between linguistic primitives, the time durations over which they are phonologically instantiated, and the grammatical properties of particular languages.

Whether on the perception or production end of communication, one must track how linguistic information unfolds over time. Mismatches between these two interfaces have the potential to overwhelm working memory capacities and create serious information bottlenecks. In this section, I explain how the research program advanced by Poeppel, Idsardi, and van Wassenhove (2008) – with computational, algorithmic, and implementational levels for speech perception – may also be extended to a visual-gestural language. Interestingly, this way of approaching speech perception research with three different levels of analyses is inspired in part by Marr's (1982) model for visual perception, making its application especially relevant

here. In the following sections, I also provide a background on what is currently known about the dynamics of sign language production and representations of linguistic units.

Theories of language processing require models which specify several details: what are the representational units that enter into language-specific computations, how those representational units are recovered from or transformed into sensory signals, and what are the neural bases for these processes. The consideration of lexical and phonological representations are sometimes lacking in other approaches to speech perception that focus on auditory neuroscience. Poeppel et al. (2008: 1017) write:

Speech perception consists of a set of computations that take continuously varying acoustic waveforms as input and generate discrete representations that make contact with the lexical representations stored in long-term memory as output. Because the perceptual objects that are recognized by the speech perception enter into subsequent linguistic computation, the format that is used for lexical representation and processing fundamentally constrains the speech perceptual processes.

In the same way, sign language perception consists of the generation of discrete representations, which compose to form lexical representations, from continuously varying visual light waves. Thus, the assumption adopted here is that processes underlying sign language perception require recognition of perceptual objects that enter into linguistic computations and are constrained by the format that is used for sign representations and processing. Understanding sign language perception undoubtedly also requires principles from visual neuroscience, including knowledge about shape perception, change detection, and the processing of biological, human

motion. In the same way, psycholinguistic studies of speech draw upon auditory theories of spectro-temporal processing, pitch extraction, and object recognition. However, assuming the existence of linguistic representations, there are two possible models in how sensory signals and language-specific representations interrelate. Poeppel et al. (2008: 1074) write:

If one is disinclined to invoke linguistically motivated representations early in the processing stream, then one owes a statement of linking hypotheses that connect the different formats (unless one does not, categorically, believe in any internal abstract representations for language processing). Alternatively, perhaps the representations of speech that are motivated by linguistic considerations are in fact active in the analysis process itself and therefore active throughout the subroutines that make up the speech perception process.

I also assume that the linguistic nature of the information encoded in visual-gestural signals must play a critical role in sensory processing. The effect of linguistic knowledge on visual processing can be found when comparing signers and sign-naïve individuals. Using ASL, Emmorey et al. (2003) report that images of phonemically contrastive handshapes that are varied continuously are perceived categorically by signers, marked by non-linear identification and peak in discrimination around the categorical boundary. No categorical effects were found among hearing non-signers, who showed non-linear identification patterns but not peaks in discrimination around the categorical boundary. Baker et al. (2005) extend these results to other handshapes in ASL, finding again that only ASL signers exhibited linguistic categorical perception and supporting the hypothesis that these effects are based on linguistic categorization rather than purely perceptual categorization (Baker, Idsardi, Golinkoff, & Petitto, 2005). Campbell et al. (1999)

tested whether facial expressions that are used in Yes/No and Wh- questions in British Sign Language (BSL) can be perceived categorically, and how the processing of linguistic facial expressions compare with emotional facial expressions (Campbell, Woll, Benson, & Wallace, 1999). They found that both deaf signers and hearing non-signers showed categorical perception to emotional expressions, but only deaf signers showed categorical perception to the grammatical expressions when identified as a question marker.

Evidence for the role of linguistic knowledge in visual processing can also be found in the perception of apparent motion. Previous studies on the perceptual construction of motion have shown that viewers interpret the shortest possible path in apparent motion (Wertheimer, 1912; Korte, 1915). However, stimuli involving biological motion can cause viewers to interpret the apparent motion as involving biologically plausible motion, even when it is not the shortest path, under specific time-windows that would permit such movements (Shiffrar & Freyd, 1990). Building upon these studies, Wilson (2001) tested viewers' perception of apparent motion using signs that involve movement in ASL. Two-touch signs that involve indirect 'hopping' motion and one-touch signs that involve direct 'sliding' motion were chosen as stimuli (see Figure 17). When presented with two images in rapid sequences, viewers perceived apparent motion. Although hearing non-signers interpreted all signs as involving sliding motion, which involves the shortest biological plausible path of movement, deaf signers of ASL interpreted hopping motion when the motion resulted in a lexical item in ASL.

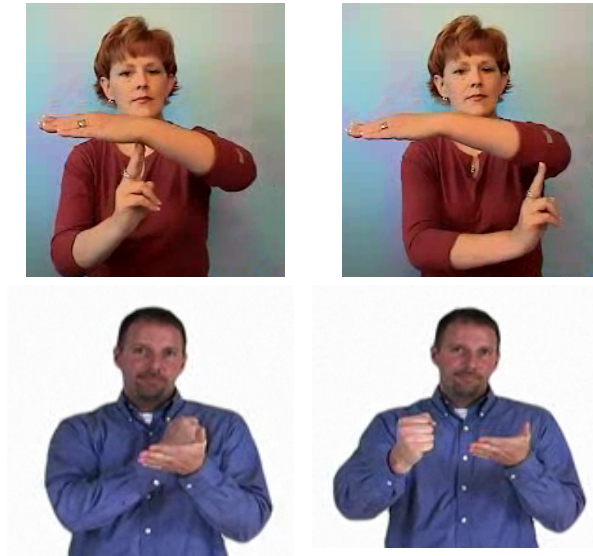


Figure 17. Examples of signs used by Wilson (2001), with images from www.aslpro.com (top) and www.signingsavvy.com (bottom). The top row shows images taken from a video recording of BRIDGE, a two-contact sign that involves hopping motion from the wrist to the elbow. The bottom row shows images from a video recording of CREDIT-CARD, a one-contact sign that involves sliding motion from the palm and outward across the hand.

When adopting a view of sensory processing where linguistic representations play an active role, an explicit theory about the format of these representations is necessary. At the computational level, Poeppel et al. (2008) support a view where words consist of a series of segments, “each of which is a bundle of distinctive features that indicate the articulatory configuration underlying the phonological segment,” as well as syllable-level representations (Stevens, 2002; Halle, 2002; Lahiri & Reetz, 2002; Féry & van de Vijver, 2004; Archangeli & Pulleyblank, 1994; Kabak & Idsardi, 2007). Sign languages also have sublexical units that are organized in a hierarchical way. Parameters of signs include handshapes, locations, movements, orientations, and non-manual features. Bundles of these features combine to form

signs, with internal structure based on features, segments, and syllables (Liddell & Johnson, 1989; Perlmutter 1992; Brentari, 1998), although modality impacts how this set of primitives can be fractionated by time. While these analyses on sign language differ on the structural organization of the sublexical features, they all agree that signs have sublexical structure.

The hierarchical organization of linguistic representations motivates a multi-time resolution implementation for sensory processing, where analyses in short and long time-scales occur in parallel. This model relies on the concept of temporal integration windows, as has been described by previous sections. In speech, this approach is supported by evidence from psychophysics, electrophysiology, and functional imaging. Describing what determines temporal integration windows, Schroeder et al. (2008:109) write: “Because neuronal oscillations cover a wide frequency spectrum, from well below 1 Hz to well over 200 Hz, they enable the integration of inputs on many biologically relevant time scales.”

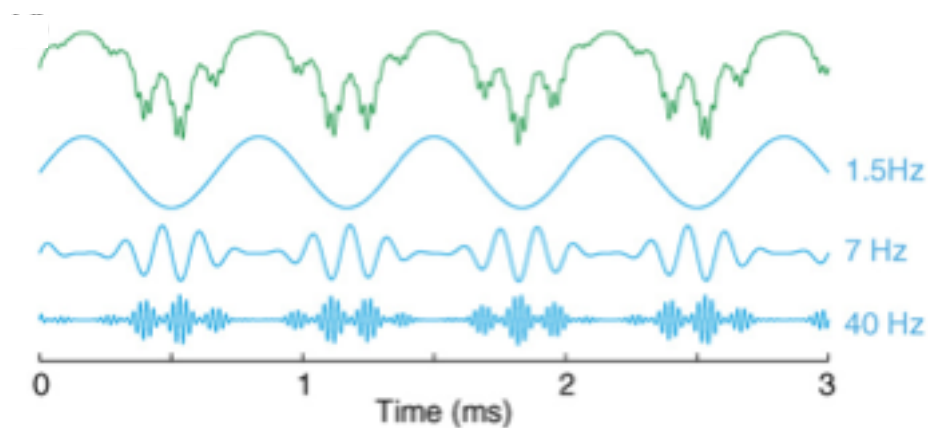


Figure 18. Reproduced from Schroeder, Lakatos, Kajikawa, Partan, & Puce (2008), this figure illustrates the hierarchical coupling of neural oscillations.

In speech, time-scales approximately 20 – 80 ms and 150 – 300 ms correspond to the duration of segments and syllables, which may be reflected by activity in gamma and theta bands. The concurrent analysis of segments and syllables may be possible by the phase-amplitude coupling of oscillations in these two frequencies, which are prominent rhythms in the primary auditory cortex (Lakatos, Shah, Knuth, Ulbert, Karmos, & Schroeder, 2005). The involvement of delta (1 – 3 Hz) oscillations in this hierarchical coupling may be tied to the rhythms of prosodic intonations in speech (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004). Functionally, processing of these temporal signals has been shown to be subserved by the superior temporal gyrus (STG) for high frequency fluctuations and the superior temporal sulcus (STS), with right hemispheric bias, for longer-duration signals (Boemio, Fromm, Braun, & Poeppel, 2005).

At the algorithmic level of description, which specifies the procedure for mapping sensory signals to linguistic representations, Poeppel et al. (2008) adopt an

analysis-by-synthesis model, where perception is driven by internal guesses about the upcoming representations (Halle & Stevens, 1959, 1962; Stevens & Halle, 1967; Yuille, & Kersten, 2006). In this view, the perceptual system does not simply wait for the input to be completed before trying to map the signals to representations. Based on the previous segment or a minimal amount of the current signal, the system may form predictions about the possible inputs that follow, which is a point of comparison to the input signal. In particular, phonological knowledge about how sounds sequence in a given language may be one basis for making such predictions (Hwang, Monahan, & Idsardi, 2010). I will assume that knowledge about the rate at which linguistic representations are generated and the constraints for how those representations are constructed is an important foundation for both speech and sign language perception, and that this is a modality-independent aspect of language processing.

These three levels of analysis (computational, algorithmic, and implementational) for speech perception are thus more broadly applicable to sign language processing. This approach to language research underscores the importance of integrating knowledge about the representations of linguistic units, how they combine, the time-scales at which these representations unfold, and the temporal constraints for supporting perceptual processes at the cognitive and neural levels.

3.2 Bellugi & Fischer (1972) revisited: Beyond the rate of signs

Convergence of the rate of propositions (sentences) in English and ASL despite the discrepancy in the rate of words and signs (Bellugi & Fischer, 1972) presents an interesting puzzle that raises questions about the rate of language-internal computations and what grammatical properties arise due to modality and temporal processing constraints. In the first comparison of rates in sign language and spoken language, Bellugi and Fischer had three hearing bilingual CODAs narrate a story that they knew well. Before starting the rate analysis, the researchers subtracted the times taken for pauses. When the story was told in English, the mean duration of propositions was 1.27 s, and on average 4.7 words were produced per second. When the story was told in ASL, the mean duration of propositions was 1.47 s, and on average 2.36 signs were produced per second. This means that signs are produced on average at 423-ms cycles. Although no statistical tests could be made with only 3 participants, the comparison of these rough numbers suggest no difference in magnitude for propositions but that the ratio of words to signs is 2:1 .

Propositions may be considered simple sentences or clauses, and they were measured in Bellugi and Fischer's study by counting all main verbs or predicates that had overt or covert subjects. From the description of the methodology, it appears that words were counted according to orthographic convention, and contractions like *don't* and *it's* were each counted as one word. It appears that signs were counted according to intuitions about what counts as whole signs for native signers. Signs are regarded as complete bundles of features, including handshape, orientation, location,

and movement. For example, the basic sign for LOOK can be varied to mean ‘YOU-LOOK-AT-ME’, ‘EVERYONE-IS-LOOKING-AT-ME’, ‘THEY-LOOK-AT-EACH-OTHER’, and ‘GAZE-AT-ONE-ANOTHER-LIKE LOVERS’ depending on the number of hands, the orientation of the hands, the movement of the hands, and non-manual features. Even though these constructions convey complex meaning, each was counted as a single sign.

These criteria already suggest that words and signs are not equivalent units in sentential context and that one sign of ASL may be equivalent to several words in English. Bellugi and Fischer discuss three possible reasons for why the propositional rates in these two modalities converge despite the apparent differences in the rates of words and signs: 1) doing without, 2) incorporation, and 3) body movements and facial expression. Sentences in ASL can convey unambiguous meaning with fewer items than English. They note that ASL uses “denser constructions,” which is illustrated by examples below.

English	ASL
and I went back into the kitchen (7)	RETURN TO KITCHEN (3)
So they came in (4)	ENTER (1)
I turned on the gas (5)	ME TURN-ON G-A-S (3)
I pulled open the drawer (5)	I PULL-OUT-DRAWER (2)
And I struck the match (5)	AND STRIKE-MATCH (2)
Until I finally decided to go through the gate. (9)	UNTIL DECIDE GO-THROUGH GATE (4)
OK, so they got off the streetcar (7)	AND ARRIVE GET-OFF TRAIN (4)

Table 2. Examples are adapted from Bellugi & Fischer (1972). These pairs of sentences demonstrate differences between English and ASL constructions.

Bellugi and Fischer remark that ASL lacks redundancy because elements that are not essential to convey the message are deleted. In English, it is possible to reduce redundancy by replacing proper names with pronouns. In ASL because information is preserved by reference to points in signing space, these nouns can often be eliminated. Because signs can incorporate location, number, manner, and shape/size features, a single sign can involve many layers of information. Through non-manual features, it is possible to layer information across a sequence of signs, so that equivalents for 'I understand' and 'I don't understand' take the same amount of time, where a head shake during 'UNDERSTAND' changes the meaning to negation. Thus, the grammatical property of ASL plays a key role in its temporal properties. Bellugi and Fischer (1972:199) write, "It seems to us that (this) condensation [in ASL] may be a response to pressure when the rate of articulation of the language is so different from speech... [ASL] has special ways of compacting and incorporating linguistic information that, because of its nature, are different from spoken language." Bellugi and Fischer entertain the possibility that propositional rate in English and ASL were the same in their study because the participants were hearing bilinguals. However, in a different study with native deaf signers, Klima and Bellugi (1979) found that the rates of signs and propositions were overall similar to the ASL findings from CODAs.

A form of communication that uses the visual-gestural modality while following the grammar of English offers a unique perspective on the interaction of grammar and modality. These systems are referred to as Manually Coded English (MCE), which were developed to make English visible for deaf children (Ramsey,

1989). However, findings that MCE cannot be learned naturally suggest that processing difficulties occur when grammatical structure based on a spoken language is imposed upon sign language (Supalla & McKee, 2002). One form of MCE called Signing Exact English (SEE 2) adapts ASL signs, which serve as roots. Invented and borrowed signs are added as functional morphemes or words, most of which are added linearly as in English. 88% of affixes in SEE 2 have full sign formational structure by having movement and thus are “sign-like.” In ASL, IMPROVING and IMPROVEMENT involve a modification of the sign for IMPROVE, where all three are single signs. In IMPROVING and IMPROVEMENT, the inflections overlap with the root. In SEE 2, IMPROVING and IMPROVEMENT involve the sequence of two signs, IMPROVE plus an affix. In cases where a sequence like KNOW and -ING can be assimilated, so that there is only one movement, the duration of the form is cut in half compared to the unassimilated form. However, the resulting form is not a possible sign because of the relationship of the handshapes in KNOW (with B-handshape) and -ING (with the I-handshape) (Battison, 1978). In many cases, “MCE morphology does not meet the constraints on sign structure” (Supalla & McKee, 2002:156). Giving into time pressures can lead to phonological-ill formedness, which leads to other aspects of processing difficulty or unnaturalness. In terms of rate, Klima and Bellugi (1979) report that the average length of propositions signed in MCE is 2.8 seconds, which is almost double the duration found in ASL and English. Thus, it appears that the temporal properties of MCE also violate more general constraints on language processing than just grammar.

Grosjean (1979) compared the rates of English and ASL, where participants were asked to speak or sign at different subject rates. At the normal rate, signers produced on average 1.94 signs per second and speakers produced 4.57 words per second, similar to the findings from Bellugi and Fischer (1972). Wilbur (2009) also reports that signers produced an average of 1.95 signs per second in normal conditions and 2.43 signs per second in fast conditions. Taking into consideration the amount of pauses in natural production, both studies also report that signers spend more time articulating than speakers. In other words, a higher percentage of the time spent narrating a story was filled with pauses for speaking than signing. In a different experiment, Klima and Bellugi (1979) had signers and speakers produce a list of monomorphemic signs or words at a rate of one per second. They found that twice the amount of time during the one-second intervals is taken for signing than in speaking. Klima and Bellugi (1979:186) point out, “One might imagine, therefore, that signed sentences and their underlying propositions might normally be stretched out in time periods longer than comparable propositions in spoken language.” However, because of the structural and discourse properties of ASL, as outlined by Bellugi and Fischer (1972), large mismatches are avoided.

Although the average duration of signs reported in these studies is ~400 – 500 ms, it can depend greatly on context. As previously described, Fischer et al. (1999) test the intelligibility of rate-compressed ASL sentences and single signs. They compare the duration of five signs (ROOM, MOUNTAIN, APPLE, TELEPHONE, and FATHER) that occurred both in sentential and isolated context. The average duration of these signs were 313 ms in sentences and 553 ms in isolation, 167 ms of

which was attributed to “final hold” (Liddell, 1984). Liddell (1978) analyzed the duration of same signs that appear in different sentence positions and syntactic functions. The duration of signs are shortest in medial position (reaching as low as 233 ms) and highest for topic signs in initial position (reaching as high as 600 ms). Moreover, Friedman (1974) compared the average duration of unstressed signs (367 ms on average) and stressed signs (835 ms on average), where the longer duration of stressed syllables was attributed to longer “holds,” similar to the findings of Fischer et al. (1999). Measuring sign durations by taking the beginning and end of sign boundaries or by taking the length of an utterance and dividing it by the number of signs can also result in considerably different figures. The reason is that there are transition times between the signs (see Figure 19).

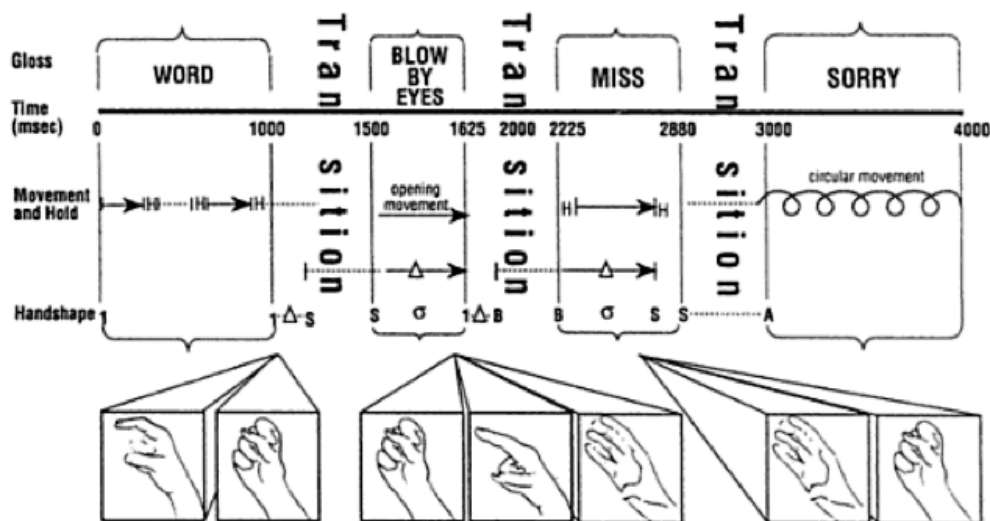


Figure 19. Reproduced from Brentari, Poizner, & Kegl (1995) (and Brentari (1998)), this figure demonstrates sign-internal and sign-external transitions in an ASL sentence. The above sentence is WORD BLOW-BY-EYES MISS SORRY ('The word went by too quickly. I missed it, sorry').

Wilbur and Nolen (1986) analyze the rate of syllables in ASL, which are phonological units that are composed of movements (M) and holds (H) (akin to vowels and consonants, respectively, in spoken languages¹), following Liddell (1984) (see also Perlmutter (1992) for an account with movements (M) and positions (P)). As outlined by Sandler and Lillo-Martin (2006:218), the argument for syllables in ASL is as follows: “1) There is a prosodic unit that organizes the timing of phonetic gestures, 2) there are constraints on the content of this unit, 3) it is referred to by rules, and 4) there is distributional evidence for the following saliency hierarchy: path movement > internal movement > location.”

It has been proposed that most signs in ASL are monosyllabic (Coulter, 1982) and have the following configurations: HMH, MH, HM, or M, where movements are considered to be the nucleus of syllables (Perlmutter, 1992). However, when signs are connected in sentences, transitional movements also occur between the signs. Some of these inter-sign transitional movements occur between lexical signs that have their own internal (intra-sign) movements or provide movements to signs like MOTHER and NOON that do not have their own lexical movements. Taking these factors into account, Wilbur and Nolen (1986) provide a thorough analysis of syllables in ASL taken from natural conversations and prompted sentences.

In conversational data where 889 syllables were measured among 3 signers, the mean duration of syllables was 250 ms. This figure reflects the duration of syllables when inter-sign transitional movements were counted. Surprisingly, the

¹ Consider also the opposite view, where vowels are understood to be steady states (= holds) and consonants as transition states (= movements).

range of syllable durations was 33 – 1300 ms, and the total standard deviation was 162 ms. The shortest syllables were those with only movements (mean duration 195 ms, standard deviation of 128 ms), and these occurred most frequently in the data. Overall, there was a negative correlation between syllable length and frequency, as found in spoken languages (Zipf, 1935). Compared to initial holds, which were 74 ms in duration on average, final holds were 156 ms long.

Wilbur and Nolen (1986) note that the average duration of 250 ms for syllables in their study converges with Liddell's (1978) findings that monosyllabic signs taken from sentential contexts are 233 – 450 ms long. However, Bellugi and Fischer (1972), Klima and Bellugi (1979), and Grosjean (1979) report that approximately 2 signs are produced per second, but they are careful to not make the claim that the average sign is 500 ms in duration, although generally it can be assumed that length (duration) of units correspond to the periods at which they are produced. Perhaps one way to reconcile these findings is to assume some combination of the possibility that many signs are multisyllabic or that many syllables are not part of the signs. Coulter (1982) has argued that most signs in ASL are monosyllabic, which Wilson and Nolen acknowledge. Wilbur (1986) has also argued that multisyllabic signs exist as bidirectional signs, which have two movements. A sequence of movements is highly constrained, however, such that the second movement must be opposite of the first or a 90-degree rotation of the first (Supalla & Newport, 1978). Other cases include reduplicated forms and some compounds. Although many compounds consisting of two signs fuse to become monosyllabic

(Liddell, 1984; Liddell & Johnson, 1986), others retain the syllable from each of the signs.

As mentioned previously, Wilbur and Nolen (1986) take into account transitions between signs into their measurement of syllables. They observe, “Signing differs from speech, where the sound stream may be discontinued while the articulators are in transition. The hands cannot be made invisible while they make transition movement” (1986:273). Among 889 syllables that they measured, 114 syllables consisted of only transitions, where the mean duration of these transitions was 203 ms. When examining syllables with both transitional movement and lexical movement (255 cases), it was found that the ratio of the movements was 1:1. Some transitional movements provide movements to signs that do not have their own lexical movements, and these occurred in 50 cases. Overall, the total number of syllables with inter-sign transitional movements was 419, almost half of all the syllables.

Whether the faster rate of syllables over signs comes from multisyllabic signs or syllables that do not have lexical content, these findings suggest that the ratio of syllables to signs is 2:1. However, it may not be appropriate to draw this conclusion since sign rates were not analyzed by Wilson and Nolen.

In a different experiment, when examining the duration of syllables of elicited sentences with phrasal or compound variants of signs, they found that the syllable rate is slightly slower (where the average syllable duration was 292 ms), which is attributed to the fact that these were not taken from natural conversations. An example of a phrasal sign is FACE CLEAN which can have the literal meaning “clean face” as in “He has a clean face.” The same sequence can be used as a

compound, notated FACE-CLEAN, where the meaning is “handsome” as in “He is handsome.” The ratio of sign to syllable was measured for conditions where these forms were produced in isolation. In other words, inter-sign transitional movements were not relevant for this analysis. The ratios were 3.13 syllables per sign for a combination of simple lexical items and 3.92 syllables per sign for compounds

Wilbur and Nolen remark that a rate of 4 syllables per second is similar to syllable rates found in English (239 ms for unstressed syllables and 301 ms for stressed syllables), taking data from Adams (1979). However, a more recent analysis of English from natural conversations (Switchboard corpus) reveals faster rates where the mean duration of English syllables is 190 ms (Arai & Greenberg, 1998; Greenberg, Hollenback, & Ellis, 1996). Other studies report that monosyllabic English words are approximately half the duration of monosyllabic ASL signs (Emmorey & Corina, 1993; Corina & Knapp, 2006; Capek, Grossi, Newman, McBurney, Corina, Roeder, & Neville, 2009).

These syllables in adult sentence production can also be compared to the rates found for babbling during infancy, which is considered an important stage for phonological development. As discussed in Chapter 1, the average syllable duration was found to be ~300 ms in speech babbling (Levitt & Wang, 1991; Dolata, Davis, & MacNeilage, 2008) and ~1000 ms in sign babbling (Petitto, Solowka, Sergio, Levy, & Ostry, 2004). Petitto et al. (2004) found that non-linguistic gestures of sign-exposed babies move at a frequency of ~2.5 Hz and that the gestures of children who were not exposed to sign language input move at ~ 3 Hz. If frequencies in babbling have any

parallels in adult production, they suggest that the rhythmic properties across modalities are notably different.

Although this provides a good overview of the phonological temporal dynamics of ASL in addition to the rates of signs and propositions provided by earlier studies, no systematic patterns emerge except that the global rate across the modalities is the same. Although the rate of words is double the rate of signs, since signs and words are not equivalent linguistic units, it is difficult to interpret the meaning of these results. One possibility is that signs in ASL contain double the amount of information as words in English on average. Individual signs can incorporate layers of information using nonconcatenative strategies, and additional information can be layered across phrases. However, some of these nonconcatenative strategies, such as reduplication to show aspect on verbs, do lengthen the duration of signs. Examples described earlier in this section demonstrate that certain signs may be much richer in morphology than single words of English. However, without a systematic study of morpheme rates across sentences in both languages, it is difficult to determine to what degree simultaneous strategies make up for differences in word-sign rates quantitatively. Moreover, Bellugi and Fischer (1972) note that ASL sentences can “do with less” and may be less redundant than English. Thus, despite the simultaneous strategies of ASL grammar, the existence of other tactics may suggest that layering information does not sufficiently meet time pressures in language processing.

An analysis of morphemes (units of meaning) as well as an analysis of syllables (units of form) would be helpful in better understanding where and how

rates converge across modalities. In the only study that I am aware of that reports the rate of morphemes in a sign language, Senghas and Coppola (2001) investigate the evolution of Nicaraguan Sign Language (NSL). They describe the emergence of systematic spatial modulations in signing among individuals who were exposed to Nicaraguan Sign Language at different ages and also those who entered signing communities at different stages in the evolution of the language. Overall, the first cohort (the generation that entered the signing deaf community before 1983) showed significantly less spatial modulation per verb in their natural production. The second cohort showed almost double the number of spatial modulations, but only among those who entered the signing deaf community before the age of 10. They also tested whether there was a link between the use of spatial modulations and overall fluency, which they measured as signing rate. Signers from the second cohort who entered the communities before the age of 6 had the highest fluency rate of ~ 350 morphemes per minute (or 5 – 6 morphemes per second). Late learners in both cohorts had the lowest fluency, where morpheme rates were half in comparison.

One conjecture is that spatial modulation emerges as a result of reaching some sort of upper limit of processing without simultaneous layering of information, a limit that is not rapid enough for full-fledged language processing. It is also possible that overall fluency and grammaticalization that leads to aspects of structure like spatial modulation develop together but not in a cause-effect relationship. Insights from Abu-Shara Bedouin Sign Language (ABSL), another new sign language that is still continuing to develop, give indications that verbal agreement through spatial modulations takes time to mature and become grammaticalized (Aronoff, Meir,

Padden, & Sandler, 2004). Aronoff et al. (2004:35) write, “The lesson from ABSL is therefore that even the motivated morphology that we find in all established sign languages requires social interaction over time to crystallize.” It is surprising that with the availability of the visuo-spatial modality, such aspects of sign languages still require time to develop rather than being exploited immediately. Although some aspects of verb agreement are found in rudimentary home sign (the first stage of language creation among deaf children who grow up without exposure to any accessible language input) (Goldin-Meadow, 1993) and even the older generation of NSL users, they lack the systematic and extensive use found among all mature and stable sign languages.

Although the development of simultaneous strategies that do not exist in spoken languages is the focus here, it is important to note that it is not true that early-exposed young generation signers avoid all sequential strategies. In a different study, Senghas et al. (2004) describe how only young-generation native signers discretize manner and path features of movement (e.g., ‘rolling down’) in natural production, whereas others use the more gestural form of expressing these features simultaneously (Senghas, Kita, & Özyürek, 2004). Since this aspect of segmentation and linearization were not present among older generations of signers, it serves as another example of language creation without rich input.

Although the focus of the study by Senghas and Coppola (2001) is on spatial modulation in NSL grammar, the analysis of fluency based on morpheme rates provides a unique insight about their possible interaction. Unfortunately, information about the rate of signs, propositions, and syllables are not reported in the study. What

remains unknown is the average rate of morphemes in a spoken language, and the relationship between morpheme and syllable rates in sign language and in speech.

The underlying assumption about the temporal dynamics of signing has been that they are slow compared to the rapid movements of oral articulators and the fine structures of acoustic signals in speech. Meier (2002:8) summarizes this argument as:

To date, the articulatory factor that has received the most attention in the sign literature involves the relative size of the articulators in sign and speech. In contrast to the oral articulators, the manual articulators are massive. Large muscle groups are required to overcome inertia and to move the hands through space, much larger muscles than those required to move the tongue tip.

However, a look at quantificational measures of velocities in sign and speech production demonstrates that the relationship between speed of the articulators and the grammatical differences is not straightforward. Ostry and Munhall (1984) report that the average maximum velocity of tongue dorsum movements is in the order of 10 cm/s . In contrast, Wilbur (1999) reports that the peak velocity of signs are measured in the order of 300-400 cm/s when measured from diodes placed on the thumb and index finger recorded by cameras. When comparing 2- and 3- dimensional traces of signing motion, Bosworth et al. (2010) find that 2D traces yield slightly slower figures for velocity (see Figure 20) (Bosworth, Dobkins, & Wright, 2010). In their study, the mean velocity of movements was ~ 50 cm/s and the maximum speeds were ~ 150 cm/s in sentence production. The differences in the measurements between Wilbur and Bosworth et al. may be attributed to differences in equipment and distance to recording devices (WATSMART and Virtual Reality InterSense,

respectively) or individual variation. In Bosworth et al. (2010), the maximum velocity of one signer was as high as ~ 300 cm/s.

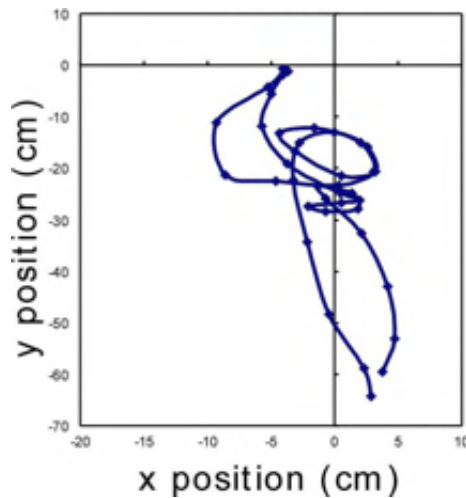


Figure 20. Reproduced from Bosworth, Dobkins, & Wright (2010), this figure demonstrates the 2D movement trace for an elicited sentence containing the sign KNOW.

The maximum amount of displacement in tongue dorsum raising and falling is about 1 cm (Ostry & Munhall, 1984). In signing, Wilbur (2009) reports that the average amount of displacement is about 20-30 cm. However, in the study conducted by Bosworth et al. (2010), a visual inspection of figures reveals displacements ranging 20 – 150 cm, but mean values are not reported. In speech, there is a reliable correlation between the amplitude (maximum distance) of the tongue dorsum movement and its maximum velocity. Bosworth et al. report that since duration and displacement of movements vary linearly, relative speed of movements are kept constant. It is not clear from these figures how to characterize and compare the speeds in production. Although manual movements are executed at higher velocities, they also move greater distances. Given the average peak velocity and distances of

movements in each modality, it would be useful better understand and compare how long an average movement takes in each modality.

Although sign languages have rhythmic properties, they do not have a single predominant oscillator like the mandible in speech. The movements of the mandible are relatively simpler, consisting only of raising and lowering. In signing, multiple joints on the hands and arms can contribute to a wide range of motions, including extensions and rotations. Although the timing of these movements associated with syllable units is rhythmic, they do not have the same cyclic property like the syllables in speech, except for rotations, trilled movement, and repeated movements.

The use of non-manual features, such as mouthing and eye-brow movements, contribute to meaning and also display rhythmic properties (Baker & Padden, 1978; Wilbur, 2009). When mouthing and signs co-occur, oral units entrain to sign syllables (Sandler & Lillo-Martin, 2006). In many sign languages, these oral features are not optional but obligatory (Boyes-Braem & Sutton-Spence, 2001). Lexically specified movements are synchronized with sign movements, but they do not co-occur with transition movements between signs. In ASL, mouthings borrowed from English can co-occur with signs. Meier (2008) report that in cases where the English word and ASL do not match in syllable count, the English word is restructured. One example is the reduction of the mouthing for *finish* to *fish* because the sign FINISH has a single outward twist of the forearm. Woll (2001) also describes the phenomenon of echo phonology in British Sign Language, where movement of the mouth and hands are synchronized and the manner of movements is matched.

Finally, the rate of letters in fingerspelling may provide some insights to the speed of fine-motor changes. Quinto-Pozos et al. (2010) report that approximately 7.5 letters can be produced per second (or 133 ms per letter) by a native signer. Although this may seem rather fast, the degree of coarticulation that takes place in fluent fingerspelling and dropping of letters in signs (for example, B-N-K for *bank* and M-P-H-E for *morpheme*) suggest that these articulation rates are subject to further time pressures. Phonological reduction may be constrained by the average duration of signs and the transition time required between signs.

3.3 Perspectives from information theory

If the rate of signs per second is twice as slow as the rate of words in the equivalent measure of time but the overall propositional rate is the same, this suggests that an incremental sign contains more linguistic information than an incremental word in English. Although these larger units take longer to produce, because information can be encoded simultaneously in ASL, each sign may contain similar amounts of linguistic information that is sequentially presented in the same amount of time (i.e., multiple words) in speech.

In information theory (Shannon, 1951), information is described in terms of entropy, which is a measure of the uncertainty associated with a random variable and can be quantified by taking into account the number of values within a set and the probability of those values. For example, calculating the information bit of a letter in English text takes into account that there are 27 characters (26 letters plus space) and

the probability of each letter. Entropy is used to describe the average uncertainty of an information source, where the maximum entropy is achieved in the scenario where all letters occur with equal probability. In contrast, redundancy quantifies the predictability of the language. Empirically, letters in English are rather predictable because of the differences in frequencies of the letters and the sequence of letters that are possible. The entropy rate of English text is estimated to be 0.6 to 1.3 bits per letter (Shannon, 1951), and similar figures are reported in estimates of phonemes in speech (van de Laar, Kleijn, & Deprettere, 1997). This is well below its maximum entropy, which is estimated to be 3-3.5 bits higher (Chong, Sankar, & Poor, 2009).

Chong et al. (2009) apply a similar approach to sign language by analyzing handshapes of ASL. Their list consisted of 45 different handshapes, 29 that have alphanumeric correspondence and 16 additional ones that are using in signing. Data were collected from video logs (vlogs) found on the Internet and natural conversations that were videorecorded at a deaf school. The frequencies of the 45 handshapes were then computed in order to determine the empirical entropy of the handshapes and to compare them to the maximum entropy. They report that the average entropy of a handshape is approximately 5 bits, which is not very different from a maximum possible entropy of 5.49 bits. They write, “Our findings suggest that a slow rate of sign production in ASL may be compensated for, at least in part, by a low redundancy of handshapes.” Chong et al. speculate that speech requires higher redundancy (it is estimated that approximately half of the text in English can be predicted) because the auditory channel is noisier than the visual channel, but the basis for this assumption is not explained.

This conclusion suggests that ASL should be more sensitive to noise since it is less redundant. However, at least three studies now suggest that ASL is more robust to temporal distortions than spoken languages. Tweney et al. (1977) report that ASL is much more resistant to temporal disruptions compared to speech (Miller & Licklider, 1950). Fischer et al.'s (1999) results show that even with compressions by a factor of 6, 20-40% of signs remain intelligible. In Chapter 2, I demonstrated that ASL is much more resistant to local time-reversals than speech (Greenberg & Arai, 2001).

Chong et al. consider the possibility that although English is more redundant in the sequence of phonemes, ASL achieves redundancy by holding a handshape for longer periods of time. One way to test whether these forms of redundancy are equivalent is to calculate information transfer rates, which is what I describe here. In speech, it has been estimated that 10-15 segments are produced per second (Liberman, 1996). This converges with findings that phonetic segments are on average 72 ms long (Arai & Greenberg, 1998), that there are on average 2.5 segments per syllable in English (Greenberg, Hollenback, & Ellis, 1996), and that syllables in English are on average ~200 ms long. If each phoneme contains 1 bit of information on average, the information transfer rate is approximately 10-15 bits per second. In ASL, each sign has at least one handshape and a maximum of 2 handshapes. Bellugi and Fischer (1972) and Grosjean (1979) report that approximately 2 signs are produced per second. If each handshape contains 5 bits of information on average, the information transfer rate is approximately 10-20 bits per second in regular signing. Quinto-Pozos et al. (2010) find that 7.5 letters are produced per second on

average. Since fingerspelled letters include only a subset of the 45 handshapes analyzed by Chong et al. (2009), the estimates for the information content of handshapes in only fingerspelling contexts would be lower than the estimate of 5 bits. Excepting fingerspelled words, the information transfer rate of English and ASL might be comparable based on a phonetic analysis. Although Reed and Durlach (1998) estimate the information transfer rate differently, they reach the same conclusions about the equivalence of information transfer rates in spoken English and signed ASL.

Chong et al. acknowledge that their analysis of entropy in ASL is incomplete because it does not take into account other phonological features that are essential to the identification of signs, such as location, orientation, movement and non-manual features. Methodologically, it is more difficult to incorporate these features. They explain that orientation has too few variations and that movement has too many. In a study of categorical perception, Emmorey et al. (2003) find that phonemically distinct handshapes are perceived categorically but that phonemically distinct locations are not. The categorical/discrete versus continuous/analogical aspects of signing is still not well understood (Liddell, 2003). Chong et al. also speculate that when combinations of handshapes and motions between the dominant and non-dominant hand are accounted for, greater redundancy would be found in ASL. Depictive gestures in natural signing and the manipulation of classifier handshapes (Liddell, 2003) pose extra challenges for determining what is the set of phonetic features in sign languages. Nevertheless, the development of sign language corpora with annotations for phonological features will be essential to these investigations.

Entropy has also been applied to understand the amount of information contained in whole words in sentences (see Figure 21). Given a sequence of words already encountered in a sentence, the following word is more informative if it is less predictable. Sentence processing is highly sensitive to frequency effects, both at the lexical level and structural level (Hale, 2001).

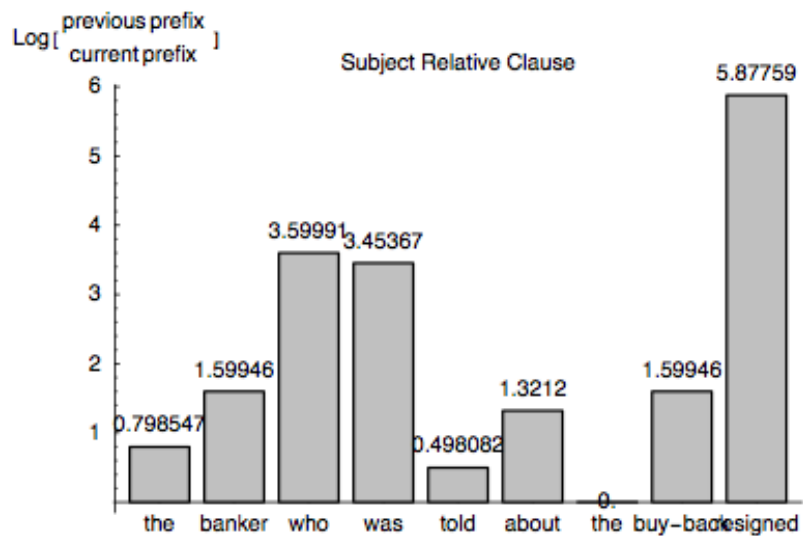


Figure 21. Reproduced from Hale (2001), this figure demonstrates how entropy (or “surprisal”) fluctuates over the course of a sentence.

Words that are the more frequent overall and more predictable in context have shorter phonological forms (Zipf, 1935; Manin, 2006). Given the correlation between length of a form and information content, it is possible that this link applies cross-modally. Since signs on average take twice as long to produce than spoken words, they are expected to carry more information. One proposal for sentence processing is that speakers are sensitive to the amount of information per unit (“information density”) comprising an utterance and try to maintain uniform information density across the utterance (Levy & Jaeger, 2007; Jaeger, 2010). This hypothesis is motivated by a

principle in information theory that sending information at a constant rate is most efficient in noisy channels (Shannon, 1948; Genzel & Charniak, 2002). When the error rate is minimal, it is assumed that information transfer close to the channel's capacity is optimal. Findings by Chong et al.'s (2009) may suggest that sign language processing is more efficient than speech. Across a sentence, some words have more information than others, such that there are "peaks" and "troughs" in information density. These peaks and troughs are modulated to some degree by closed-class words that are highly frequent, are short in length, and make the categories of subsequent words more predictable. In the future, it would be informative to compare information density patterns across different modalities since sign languages involve more simultaneous layering of information.

In summary, understanding rates in natural language processing requires knowledge about the rate at which phonetic units are produced, the rate at which lexical units are produced, and the information content of each unit. The kinematics of oral and manual articulators as well as the sensory pathways in audition and vision reveal considerable differences between the communication systems. The entropy analysis by Chong et al. (2009) suggests that phonetic units in signing have much more information than units in speech, but an extension of their analysis to the duration of the signals suggests that overall information transfer rates may be comparable. Although speech segments are more redundant than sign handshapes, sign handshapes may be as redundant over time. An analysis of spoken sentence production shows that listeners are sensitive to the predictability of upcoming words and that speakers make phonological, lexical, and syntactic decisions based on the

information profile of the utterance (Hale, 2001; Levy & Jaeger, 2007).

Psycholinguistic experiments show that signers are also sensitive to predictability in sentences and show similar neural correlates found in spoken languages, such as the N400 effect in electrophysiology (Neville, Mills, & Lawson, 1992; Capek, Grossi, Newman, McBurney, Corina, Roeder, & Neville, 2009). From an information theoretic point of view, it remains unknown whether information density fluctuations in spoken sentence production (Figure 21) are similar in profile in signed utterances. Highly frequent closed-class words of English do not have direct phonological analogs in ASL. Peaks and troughs seen in sentences of English caused by these shorter words may also emerge in ASL as some parts of signs are more informative than others. Alternatively, the differences between sequential and simultaneous flow of information across the modalities may reveal unique distributions of information density.

3.4 Words, signs, morphemes, and syllables

Since the findings of Bellugi and Fischer (1972), many questions still remain about the convergence of rates across languages and divergence of time properties based on modality and grammatical features. The difference in word and sign rates is difficult to interpret because they may not be equivalent linguistic units. Within spoken languages, the degree of complexity in words is represented by the analytic-synthetic continuum, where analytic languages have little to no morphological inflections on words (e.g., modern Chinese) whereas synthetic languages (e.g., West

Greenlandic) are known for their morphological complexity. Modern English is considered to be closer to the analytic end of the spectrum. Meier (2002) notes that a polysynthetic language like Navajo produces fewer words per minute than English. Thus, the rate of words in English should not be generalized as a property of all spoken languages.

What has never been reported in these studies is the rate of morphemes in languages. Even though fewer words were produced per minute in Navaho than in English, how do they compare in terms of morpheme rates? How do the morpheme rates in these two spoken languages compare with ASL? Brentari (2002) argues that the typological trend among sign languages is that signs are monosyllabic and polymorphemic (Table 3). She also argues that polymorphemic and monomorphemic signs are typically not different in length.

	Monosyllabic	Polysyllabic
Monomorphemic	Chinese	English
Polymorphemic	Sign languages	West Greenlandic

Table 3. Adapted from Brentari (2002), who describes the typological distribution of canonical word shapes. These assumptions are reexamined throughout the current discussion in Chapter 3 because they require an examination of syllable and morpheme rates and the ratio of these rates for languages.

This does not appear to be true in spoken languages, where morphologically complex words tend to have more syllables and thus are longer than morphologically simpler words (for example, *morphologically* can be analyzed as having 4 morphemes and 4 syllables, whereas *simpler* can be analyzed as having 2 morphemes and 2 syllables).

A universal property of all mature sign languages is the use of spatial modulations to mark agreement and the use of classifier constructions. These forms result in great complexity of meaning but can phonologically resemble morphologically simpler signs (Brentari, 1995). Other constructions where semantic information can be layered nonconcatenatively include numeral incorporation, aspectual modulations, nominal and verbal number, and adverbial modifications (Rathmann & Mathur, 2010).



Figure 22. Adapted from Mathur & Rathmann (2011), this figure demonstrates an example of numeral incorporation in ASL.



Figure 23. Reproduced from Mathur & Rathmann (2011), this figure demonstrates the grammatical form for TEN DAY and the ungrammatical form TEN+DAY that would result with numeral incorporation. The latter is believed to be not possible due to phonological constraints against complex movement.

Rathmann and Mathur explain that although these cases are not universal and more flexible to change, they also contribute to increased semantic complexity of constructions. The availability of space in sign language articulation does not blindly allow forms to be combined nonconcatenatively but rather are constrained by phonological and phonetic restrictions. Finally, Napoli and Sutton-Spence (2010) attribute the limitation of 4 propositions that can be articulated simultaneously in sign languages to cognitive limitations, in particular visual short-term memory.

When languages are described as being analytic or synthetic, this usually refers to morpheme:word ratios, where analytic languages are 1:1 and synthetic languages are several:1. Brentari (2002) classifies sign languages as being polysynthetic like West Greenlandic but argues that having the property of both monosyllabicity and polysyntheticism is unique to sign languages (Table 3). A better

understanding of these relationships requires a typological investigation of the ratio of syllables to morphemes.

Bellugi and Fischer (1972) speculate that in addition to incorporation, body movements, and facial expression, which all involve how information is layered without sequential strategies, a possible explanation for the discrepancy in word and sign rates is that ASL can “do without.” For example, a sentence in English like ‘*I ate an apple*’ would be translated in ASL to ‘EAT APPLE’. ASL (and all sign languages) allow pro-drop especially when arguments can be understood from context. Moreover, ASL does not have phonologically expressed function words like ‘an’. Finally, the past-tense information can also usually be understood from the context. In this example, the equivalent of 4 words and 5 morphemes in English can be expressed with 2 signs. This discrepancy cannot be attributed to the fact that ASL has more ‘synthetic’ qualities in this sentence than English.

Taking into account previous work and the theoretical issues that arise, I have chosen to analyze the rates of words/signs, morphemes, and syllables in English, ASL, and Korean. In addition to replicating an analysis of words/signs in English and ASL, an analysis of morpheme rate will lead to a better understanding of to what degree the combination of nonconcatenative morphology and “doing without” leads to true discrepancies in the rate of lexical units in the languages. Given that English and ASL are distinct in more ways than one, it is difficult to assess whether the differences in rates are attributable to modality or grammatical differences. I have chosen to include Korean in this analysis because it is also a pro-drop language and lacks some of the small functional words that exist in English. A perfect comparison

would be between two natural languages that differ in modality but essentially identical in grammar, but this is impossible given that typological distinctions in grammar do seem to be divided by modality. Although Manually Coded English was created to have these features, the fact that it cannot be learned naturally and is globally much slower than ASL suggests that grammatical properties of sign languages are essential for their realization in the visuo-spatial modality.

Another point of interest is the syllable rate in these three languages to compare units of form to units of meaning. The syllable rate has been measured for English by numerous studies but morpheme rates have never been calculated using the same data. A comparison with syllable and morpheme rates in Korean contributes to a better understanding of what trends emerge by looking at typologically distinct spoken languages. Korean is more synthetic than English, it is expected to have a lower word rate than English but to have more morphemes per word than English. Finally, this analysis of ASL builds upon the work of Bellugi and Fischer (1972) and Wilbur and Nolen (1986). Wilbur and Nolen have provided the most thorough report of syllable production in ASL, describing the frequencies of different types of syllables and their lengths, and including sign-external transitions. As a point of comparison, the analysis provided in this present work only counts syllables based intra-sign movements.

3.5 Rates in spoken languages: English and Korean

Studies on the speed of speech have largely focused on the rate of words, syllables, or segments. Here, the goal is to gain a better understanding of the rate of words, morphemes, and syllables for cross-linguistic comparison. The first step in analyzing the rate of linguistic units in speech and sign language production was identifying appropriate materials. With large bodies of data developed for automatic speech recognition, English had the most options, but it was important to also consider if comparable material was accessible for Korean and ASL. Data from each of the three languages was collected from natural conversations. When looking for English materials, one question that arose was whether it would make a difference in the results for rates to use a corpus with prompted sentences (TIMIT) (Garofolo, Lamel, Fisher, Fiscus, Pallet, & Dahlgren, 1993), which were already transcribed and phonetically annotated, or to use a corpus of natural telephone conversations (CALLFRIEND, Canavan & Zipperlen, 1996a) without any annotations. Both corpora were accessed through the Linguistics Data Consortium (LDC) at the University of Pennsylvania.

Whereas speech files in TIMIT comprise of individual sentences that are uttered in isolation, the speech files in CALLFRIEND comprise of full 30-minute telephone conversations between two individuals. Before doing a rate analysis for CALLFRIEND (American English, corpus containing non-Southern dialects only), a set of 363 sentences were extracted from the telephone conversations, 3 sentences from 121 individuals from 60 conversations) for their 1) propositional completeness,

2) lack of long pauses/breaks, and 3) lack of errors and corrections mid-sentence.

The boundaries of the sentences were determined by looking at the acoustic waveforms and spectrograms and measured for overall duration. For TIMIT sentences, rather than blindly taking the duration of the speech files, sentences were also analyzed in a similar way by looking at the onset of the first phoneme and the conclusion of the last phoneme because the sentences were preceded and followed by a short period of silence. 188 unique sentences were chosen from the TIMIT corpus from 188 speakers of a non-Southern dialect (to better closely match the dialects found in CALLFRIEND).

Words have been described as “the free-standing unit that unifies form and meaning” (Sandler & Lillo-Martin, 2006:21), but as discussed previously, languages vary in their definitions of words, which range in complexity. Due to the lack of consistent and linguistically well-motivated definition of words, here words were taken to be units marked by spaces in orthography. As in Bellugi and Fischer (1972), contracted forms (*don't, it's, wanna*) were counted as single words.

Morphemes are considered to be the smallest unit of meaning in language, but making judgments about morphemes is not always straightforward. Debates about the decomposability of words have a long history (see Fiorentino (2006) for an extensive discussion). I adopt the assumption that the lexicon involves structured representations and that morphological parsing is an early process of word recognition (Fiorentino & Poeppel, 2007). Psycholinguistic experiments demonstrate different levels of analysis exist in word processing (Lehtonen, Monahan, & Poeppel, 2011). For example, a word like *corner* contains two morphemes in English – *corn*

and *-er* – but because they do not compose the meaning of the word, the word is considered to be made up of just one morpheme. In on-line processing, evidence suggests that responses to semantically opaque words like *corn* and *corner* show differences to semantically transparent words like *teach* and *teacher* (Lehtonen, Monahan, & Poeppel, 2011). Lehtonen et al. also show that although *-er* in *corner* is not a morpheme, because it is a possible morpheme in words like *teacher*, *corn* and *corner* are processed differently from pairs like *broth* and *brothel*, which only involve an orthographic overlap and no possible morphological decomposition. This three-way separation of the data demonstrates the complexity of morphological processing. Although a word like *corner* may trigger morphological decomposition but which is rejected by subsequent analysis, unlike *brothel*, it is not analyzed in the same way as a word like *teacher*. *Corner* and *teacher* share a decompositional stage of analysis, which succeeds for *teacher* and fails for *corner*.

Although methodologies like priming studies provide a way to probe the psychological reality of a word's subparts, it is not practical to apply them to every single word in a corpus. Understanding the morphological structure of a word may also involve some knowledge about its etymology. For example, *could*, *would*, and *should* are etymologically connected to *can*, *will*, and *shall*, but it is not clear whether native speakers decompose these words as having two parts (where *-ld* was historically linked to a suppletive form for the past tense). Another example is a word like *height*, the noun form of the adjective *high* (where *-t(h)* is linked historically to a Germanic abstract noun suffix). Because judgments about words such as these were not easy, both “conservative” and “liberal” judgments were made about morpheme

counts. Abbreviations like *ESL* were counted with 3 morphemes by the liberal count and 1 morpheme by the conservative count.

Cases of irregular/suppletive forms were judged as being morphologically complex. A word like *didn't* was counted as having 3 morphemes: *do-* + *past* + *negation*. The word *been* was counted as having 2 morphemes: *be* + *-en*. Possessive pronouns like *our* was counted as having two morphemes: *we/us* + *possessive*. This decision was made based on the pattern that *-s* is a productive morpheme that is used with nouns. When *her* was used as a possessive pronoun, it was counted as having 2 morphemes, but when it was used as an object/accusative pronoun, it was counted as having 1 morpheme. This decision was made based on the pattern in English that case-marking is not productive and only used among pronouns.

Syllables in words were also measured with two estimates for similar reasons, although making syllable counts were relatively easier than morpheme counts. Few examples that posed some difficulty include *interesting* – which was perceivable as having 3 or 4 syllables, *actually* – which was perceivable as having 3 or 4 syllables, *several* – which was perceivable as having 2 or 3 syllables, and *you're* – which was perceivable as having 1 or 2 syllables.

Two researchers coded the data from English, where each researcher coded approximately 50% of the sentences from each corpus. These two researchers worked together with frequent discussions to support consistency, but at this current time, inter-rater reliability has not been assessed.

In all of the following figures (from English, Korean, and ASL), results are shown in density plots (using R 2.8.1, R Development Core Team (2005)), which estimates the probability density function of the underlying variable. The kernels in these density plots represent the data (length, syllables per second, etc.) from each sentence.

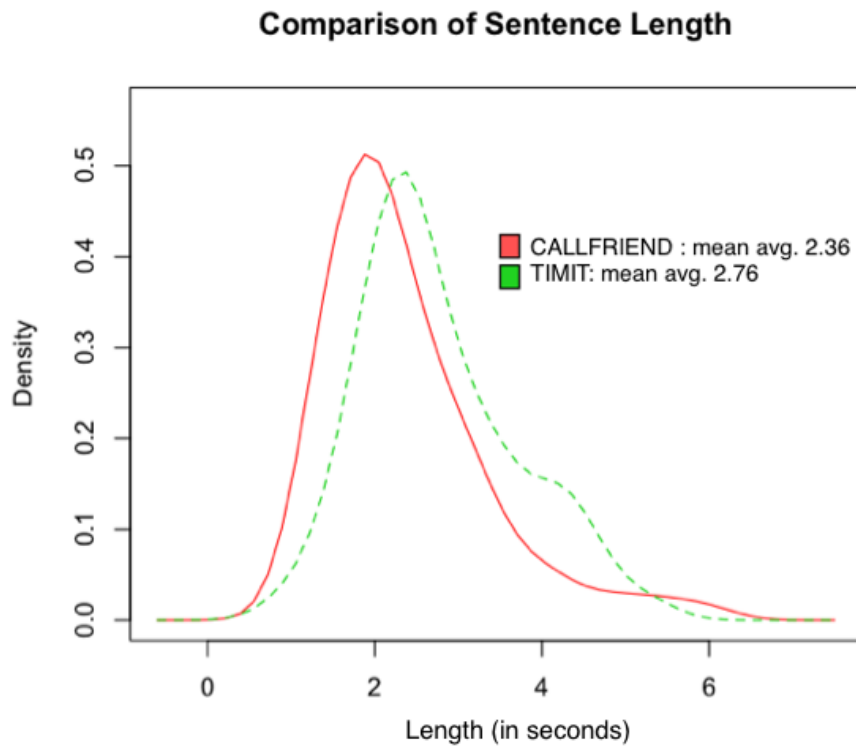


Figure 24. Estimated probability density functions for the length in seconds of sentences in two corpora of English: TIMIT (prompted) and CALLFRIEND (conversational).

A comparison of sentences from TIMIT and CALLFRIEND for English show that rates are significantly different in prompted speech and natural conversational speech. The following figures include “conservative” measures of morphemes and syllables.

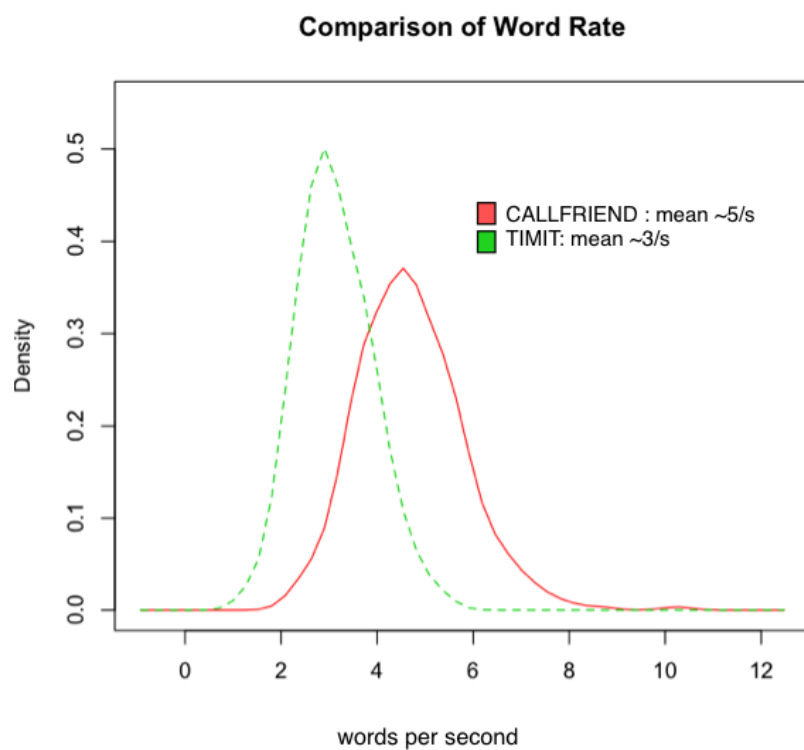


Figure 25. Estimated probability density functions for words rates (words per second) of sentences in two corpora of English: TIMIT (prompted) and CALLFRIEND (conversational).

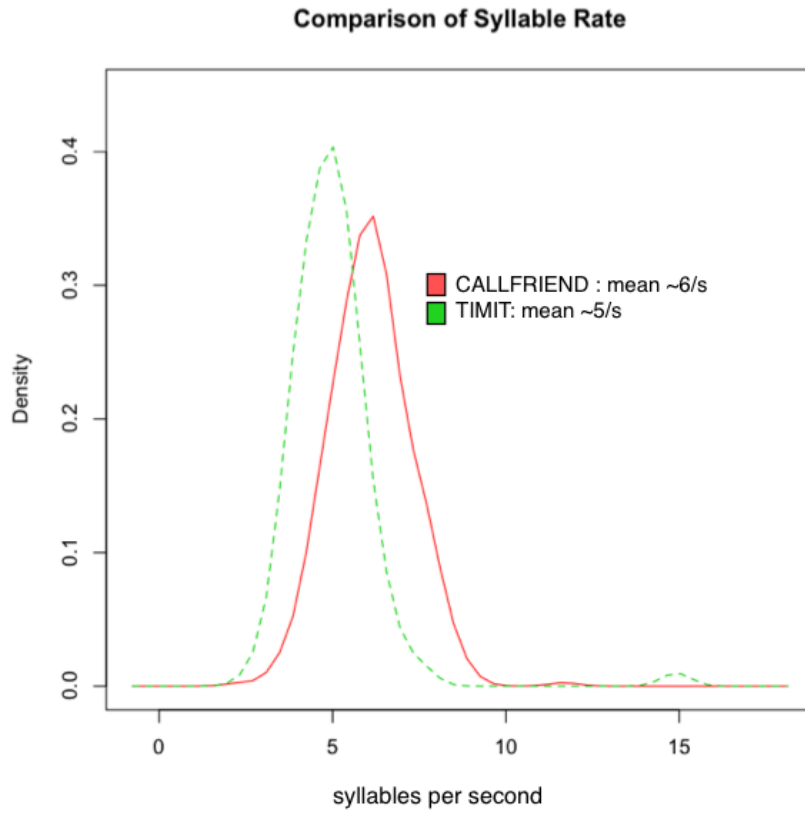


Figure 26. Estimated probability density functions for syllable rates (syllables per second) of sentences in two corpora of English: TIMIT (prompted) and CALLFRIEND (conversational).

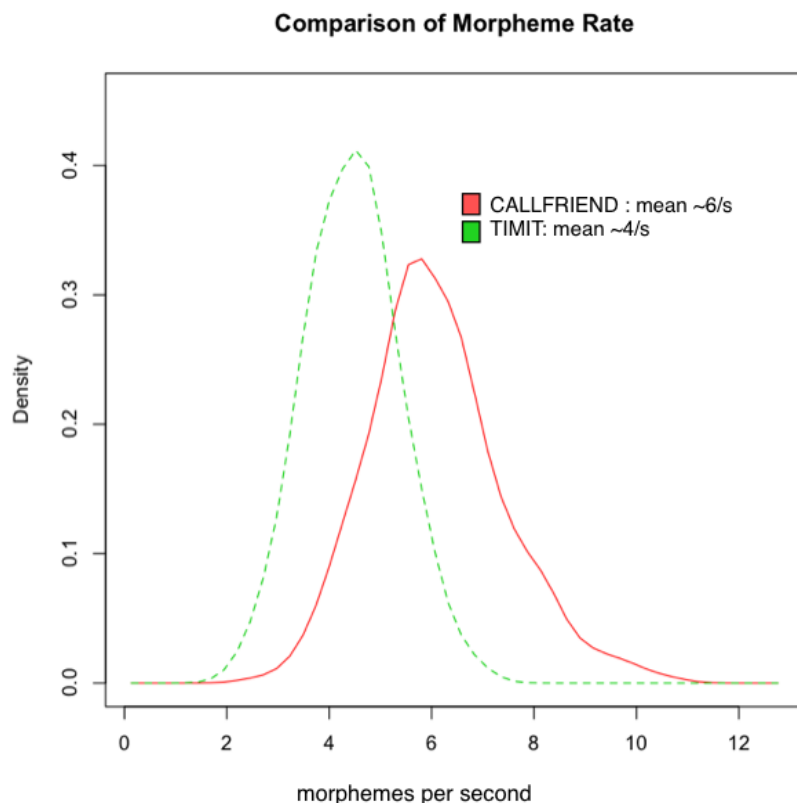


Figure 27. Estimated probability density functions for morpheme rates (morphemes per second) of sentences in two corpora of English: TIMIT (prompted) and CALLFRIEND (conversational).

As may be expected, rates were overall much faster in the natural conversational speech than prompted speech (Figures 25, 26, and 27). A calculation of the mean average syllable rate (on conservative-liberal estimates) reveals a slower articulation in TIMIT (~5.0-5.1 syllables per second) than CALLFRIEND (~6.1-6.2 syllables per second). In addition to the inherent difference between producing self-generated sentences with a communicative partner and reading unfamiliar sentences, other reasons for these differences could be attributed to 1) the oddness of the semantic content of TIMIT sentences and 2) the presence of more low-frequency words in TIMIT. In CALLFRIEND sentences, words were produced approximately

4.7 words per second (mean average), which is similar to the results of Bellugi and Fischer (1972), where stories were narrated by 3 individuals and pauses were excluded from analysis. In contrast, the rate in TIMIT is 3.1 words per second. An examination of the ratio of syllables to words reveals that TIMIT contained words that had longer phonological forms (mean average ~ 1.6 syllables per word in TIMIT compared to ~ 1.3 syllables per word in CALLFRIEND). Although a frequency analysis was not conducted for words in TIMIT and CALLFRIEND, the trend that more frequent words have shorter phonological forms (Zipf, 1935; Manin, 2006) suggests that the words in CALLFRIEND are more highly frequent.

Morpheme rates were also overall slower in TIMIT than in CALLFRIEND. The mean average (on conservative-liberal estimates) was 4.4-4.8 morphemes per second in TIMIT and 6.1-6.4 morphemes per second in CALLFRIEND. An examination of the ratio of morphemes to words reveals that TIMIT (1.4-1.6 morphemes per word) contained words that were more morphologically complex than CALLFRIEND (1.3-1.4 morphemes per word).

Finally, in both corpora there is approximately a ratio of 1:1 between morphemes and syllables. In TIMIT, which presumably contains lower frequency lexemes (a word-like unit that are used to represent all variations of the word in usage) and each morpheme contains more phonological content, the ratio is slightly lower than 1:1, and in CALLFRIEND, the ratio is slightly higher than 1:1.

The mean average duration of the sentences taken from CALLFRIEND was 2.37 s. The mean duration of syllables was calculated by dividing the duration of the sentences by the number of syllables in the sentences. The mean duration of syllables

was approximately 162 ms. This is somewhat faster than 190 ms reported by Greenberg et al. (1996). This could be attributed to at least two reasons: 1) a difference in the corpora used (CALLFRIEND versus Switchboard, where two individuals discuss a specific topic for several minutes) and 2) this study only chose a small subset of the most fluent sentences in CALLFRIEND for analysis, whereas Greenberg et al. used data from full conversations that contained filled pauses and misarticulations.

As explained previously, Korean was chosen for analysis because it is a spoken language that is typologically different from English by being morphologically more complex, being a pro-drop language, and having fewer small functional words (like *a* and *the*) and thereby being grammatically more similar to ASL. A Korean version of the CALLFRIEND corpus (Canavan & Zipperlen, 1996b) with Yale Romanized transcription is also available through the LDC. 378 sentences were extracted from 128 speakers following the same criteria as used for English (fluency, lack of errors and corrections mid-sentence, and lack of long pauses/breaks). Again, words were counted based on orthography. In other words, words were equivalent to *eojeols*, which are the spacing units in Korean orthography. The Romanization used periods (.) to mark syllable boundaries and spaces to mark word boundaries. Similar to English, words in Korean are taken to be free-standing units that can vary in morphological complexity. Conservative and liberal estimates of morpheme and syllable counts were measured for each word. For example, the topic form of the second person pronoun is *ne.nun* ('you-TOPIC') with 2 syllables but it is often reduced to *nen* and perceivable as 1 syllable in fast speech. Case markers were

always counted as morphemes. Examples of words that had different conservative and liberal morpheme counts were *hak.kyo* ('school'), which was counted as consisting of either 1 or 2 morphemes, and *pi.ngwus.ta* ('to mock'), which was counted as consisting of either 2 or 3 morphemes.

The data from Korean was coded by one researcher, and at this current time, inter-rater reliability has not been assessed.

The results from Korean (as compared to conversational data of English) are as follows. The sentences that were extracted from the two corpora were similar in length (Figure 28).

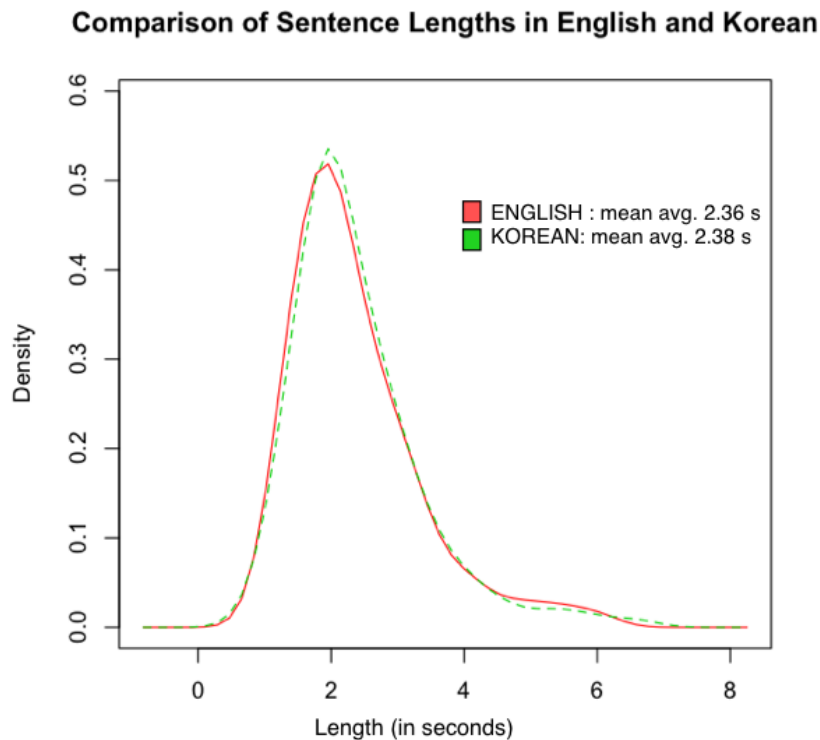


Figure 28. Estimated probability density functions for length in seconds of sentences from conversational data in English and Korean.

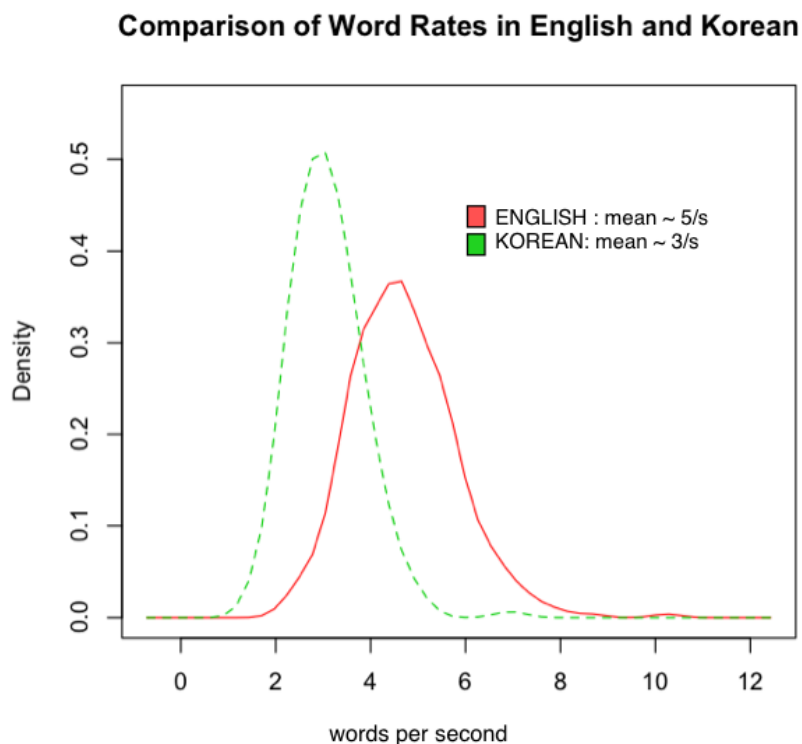


Figure 29. Estimated probability density functions for word rates (words per second) of sentences from conversational data in English (a more analytic language) and Korean (a more synthetic language).

As predicted, Korean had a lower rate of words per second because Korean is more synthetic than English (Figure 29). The results show that the mean average rate is 3.1 words per second (compared to 4.7 words per second in English). However, this does not mean that Korean is slower than English. The mean syllable rate was 7.2-7.3 (conservative-liberal) per second, which is slightly higher than English (6.1-6.2 syllables per second) (see Figure 30). The mean duration of Korean syllables was approximately 138 ms (compared to ~162 ms in English). This may be attributed to the fact that English allows consonant cluster onsets and codas, whereas syllables in Korean are simpler. For example, a long syllable in English like *script* (CCCVCC), would have to be pronounced with 4 syllables ([sukwɪɾɪptu] = CVCVCVCCV) with

epenthesized vowels in Korean. Japanese, like Korean, has simpler syllable phonotactics than English, and Arai and Greenberg (1998) show that the mean average of syllables in Japanese are slightly shorter than in English.

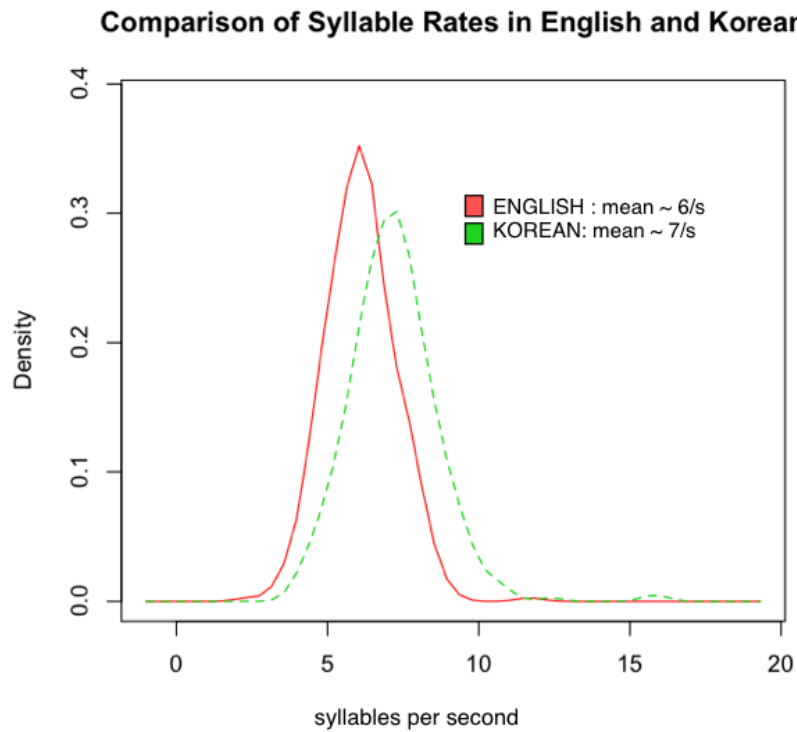


Figure 30. Estimated probability density functions for syllable rate (syllables per second) of sentences from conversational data in English and Korean.

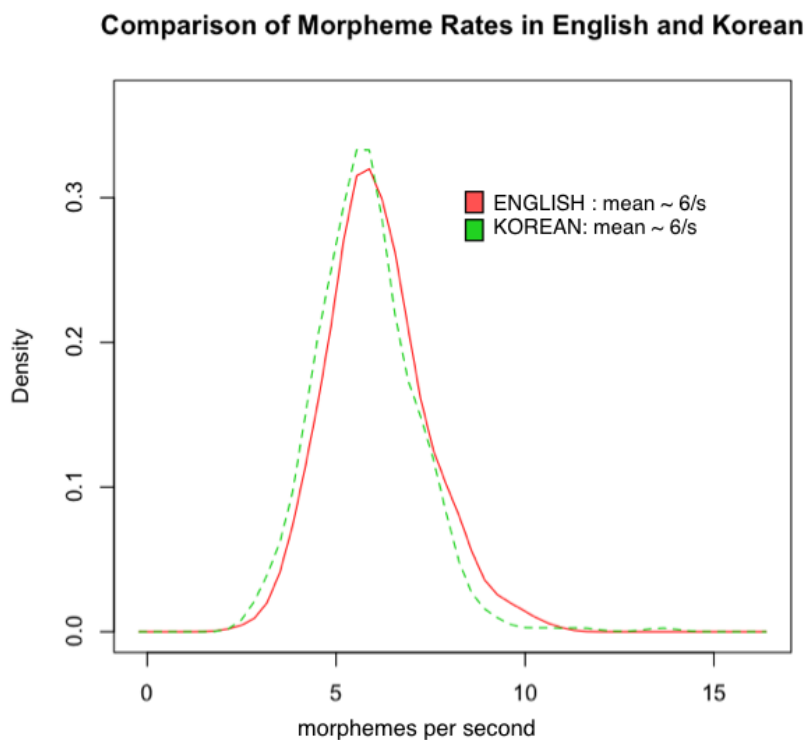


Figure 31. Estimated probability density functions for morpheme rates (morphemes per second) of sentences from conversational data in English and Korean.

The mean morpheme rate was 5.8-6.0 per second in Korean (compared to 6.1-6.4 per second in English) (see Figure 31). As a language that is more synthetic than English, Korean was expected to have a higher ratio of morphemes to words than English. Results show that on average, there are 1.9 morphemes per word in Korean (compared to 1.3-1.4 morphemes per word in English). An examination of the ratio of syllables to words (~2.3 syllables per word) reveals that Korean has words containing more syllables (compared to 1.3 syllables per word in English). Finally, the ratio of morphemes to syllables is 1:1.2, which is slightly lower than the 1:1 ratio found in English.

Although English and Korean are typologically distant languages, similar trends emerge. The main difference between the languages is in the word rate. However, when looking at the smallest unit of meaning, both show rates of approximately 6 morphemes per second. Although the syllable rate is slightly faster in Korean and the morpheme to syllable ratio is slightly lower in Korean, this is most likely due to the simpler syllable structure in Korean. Although Korean does not have small functional words like *a* and *the* in English, it has case markers on nouns and also richer morphology on verbs, resulting in the morpheme rates to closely converge. Overall, the ratio of morphemes to syllables in both languages is approximately 1:1.

3.6 Rates in sign language: ASL revisited

The goal of this study was to replicate previous work that have examined the rate of signs in natural ASL production and extend the analysis to morphemes and syllables within the signs. Sentences that matched the fluency criteria used for English and Korean were taken from natural conversations of ASL collected by Ceil Lucas and colleagues. Lucas's corpus was filmed in the 1990s to study sociolinguistic variations of ASL across the United States. The videos involve free conversations among deaf participants who already know each other and interview sessions with a researcher. The free conversation sessions were recorded without the presence of any researcher. In the interviewed segments, a deaf African-American researcher moderated groups composed of deaf African-American participants. For

the purposes of this study, 179 sentences were taken from 21 participants who are native ASL users.

Sign language linguistics students identified a set of full, fluent sentences within the conversations, which were labeled using ELAN software. These research assistants were instructed to use their intuition about the beginning and end of sentences by doing a frame-by-frame analysis on the first and last signs. Group discussions and viewing of the videos supported consistency in the data, but at this current time, inter-rater reliability has not been assessed.

Each sign in a sentence was first given an English gloss, and sign rates were calculated based on these glosses. For each sign, annotation tiers were then created so that the number of morphemes and syllables could be counted. Morphemes were counted in two ways – with a “conservative” or “liberal” estimate. Before starting the annotation process, it was decided that plain/uninflected verbs like LIKE and HAVE would be counted as having 1 morpheme, agreement/indicating verbs such as SHOW and ASK would be counted as having 2 morphemes (one for the root and one for agreement), and that spatial/locative verbs such as PUT and DRIVE would be counted as having 2 morphemes (one for the root and one for movement). On these verbs, aspectual marking was counted as one morpheme, and aspectual marking that showed number was counted as having an addition morpheme. Depiction verbs were counted as having 2 morphemes, one for the classifier handshape and one for movement. Although these criteria were decided before the annotation process, the vast majority of the verbs found in this set of sentences were plain and uninflected.

Possessive pronouns were counted as having 2 morphemes, one for the palm orientation for indexation and one for the open handshape marking possession. Facial inflections that were used in questions were counted as one morpheme. An expression with noun incorporation such as TWO-MONTHS was counted as having 2 morphemes. The sign for TWO-OF-US was counted as having 2 morphemes. The sign for AGE-THREE was counted as having 2 morphemes. The sign for EVERY-FRIDAY, where the sign for FRIDAY is held in downward movement, was counted as 2 morphemes. There were two cases when the researchers could not identify the sign of short gestures, and these gestures were labeled 'gesture' and counted as one morpheme each.

Liberal versus conservative estimates were used in cases where the etymology of a sign was known to be a compound. For example, the sign for HOME evolved from the combination of the sign for EAT (contact at the chin) and BED (contact at the cheek). HOME was counted as 2 morphemes in the liberal estimate and 1 morpheme in the conservative estimate. The sign for WIFE was counted as 2 morphemes (WOMAN+MARRY) in the liberal estimate and 1 morpheme in the conservative estimate. The sign for TEACHER is traditionally considered to consist of 2 morphemes, one for TEACH and one for an '-er'-like affix that is linked with the sign for PERSON. In natural signing, TEACHER is signed with one fluid motion where separate components for TEACH and PERSON become hard to distinguish. Thus, TEACHER was counted as having 2 morphemes in the liberal estimate and 1 morpheme in the conservative estimate. The sign for PARENT is the combination of the signs for MOTHER and FATHER. PARENT was counted as having 2

morphemes in the liberal estimate and 1 morpheme in the conservative estimate. Fingerspelled words consist of a sequence of letters, each of which represents the letter but as a whole also represents a word. The sign for HIGHSCHOOL, which is a sequence of H and S, was counted as having 2 morphemes in the liberal estimate and 1 in the conservative estimate. For all fingerspelled words, the liberal estimate was the number of letters and the conservative estimate was 1.

Syllables were counted based on the number of movements that occurred within the sign and were based on how they were produced in the video, not citation forms. For example, SCHOOL was sometimes produced with 1 or 2 movements (1 or 2 syllables). Each token was labeled the way it was produced. In another case, the sign for HERE was signed with 1 syllable in one sentence, and when it was emphasized, it was signed with 3 syllables. As has been discussed in the sign language literature, the majority of signs in these sentences were monosyllabic. Examples of disyllabic signs that occurred in this set of sentences included CANCEL and NEVER. In ASL, nominalization of verbs can be achieved through reduplication. As an example, the sign for AIRPORT was the reduplicated version of FLY and was produced with 2 syllables. In cases where a sign involved more than one syllable, it was usually through repetition of a movement, such as in SOMETIMES, VACATION, WORK, FEEL, YOUNG, and TECHNOLOGY. These reduplicated movements are usually produced in a restrained manner. The sign for SIGN was produced with 2 syllables in some cases, and there was one token when it was produced with 4 syllables. When the W handshape was waved three times for WEDNESDAY, it was counted as 3 syllables, and when the M handshape was waved

two times for MONDAY, it was counted as 2 syllables. Syllables in fingerspelled words were generally counted by the number of transitions between the letters but were sometimes lower because of coarticulation of letters. A gesture that was used to indicate 'HEART-POUNDING' was produced with 8 syllables.

The results from ASL are presented together with English and Korean in the following discussion and figures (Figures 32, 33, and 34). The results show that ~2.3 signs are produced per second, replicating the findings from Bellugi and Fischer (1972), Grosjean (1979), and Klima & Bellugi (1979) for ASL (see Figure 33). The main reason word rates are compared here is that previous studies have given much attention to differences between English and ASL at this level. However, languages have different definitions for what a word is as a unit, and here, word rates do not tell us much about modality-based differences. As discussed earlier for English and Korean data, two spoken languages also show significant differences in their word rates. An analysis of a more synthetic spoken language, such as Navajo or West Greenlandic, is predicted to show more similar rates to ASL.

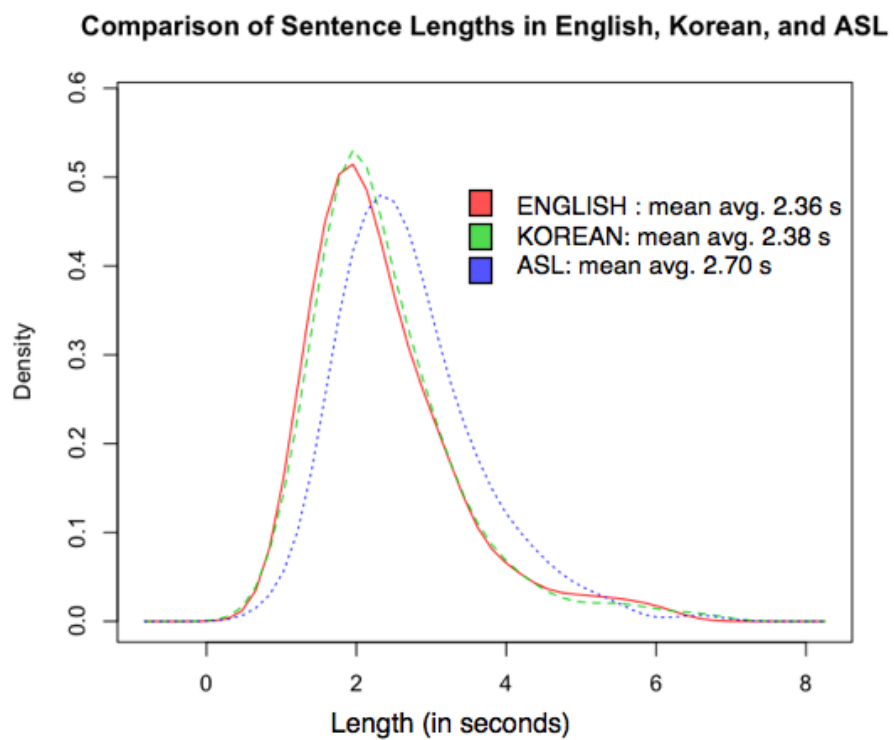


Figure 32. Estimated probability density functions for length in seconds of sentences from conversational data in English, Korean, and ASL.

Comparison of Word/Sign Rates in English, Korean, and ASL

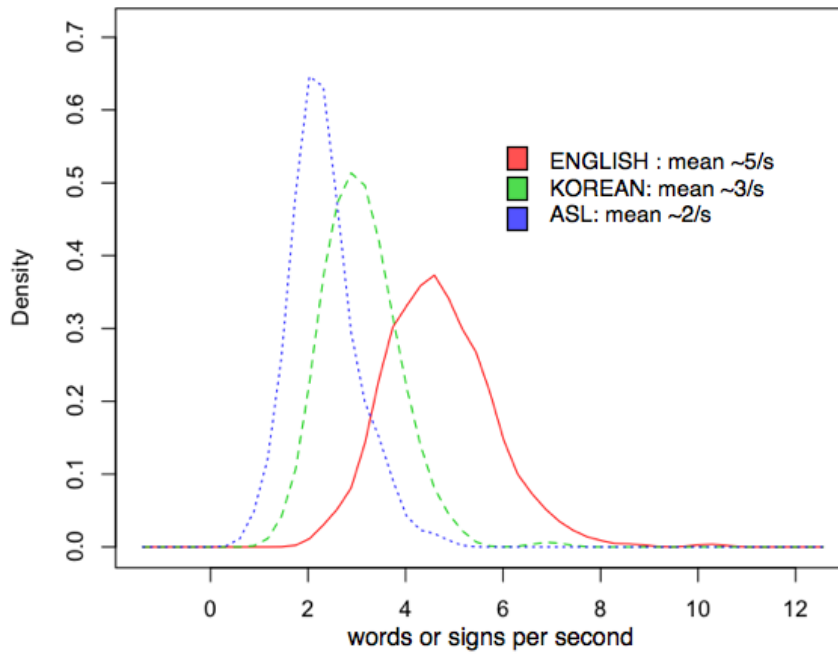


Figure 33. Estimated probability density functions for word/sign rates (words or signs per second) of sentences from conversational data in English, Korean, and ASL. This comparison word and sign rates replicate the findings from Bellugi & Fischer (1972) for English and ASL. A comparison with Korean demonstrates that word rates depend on grammatical properties of the language.

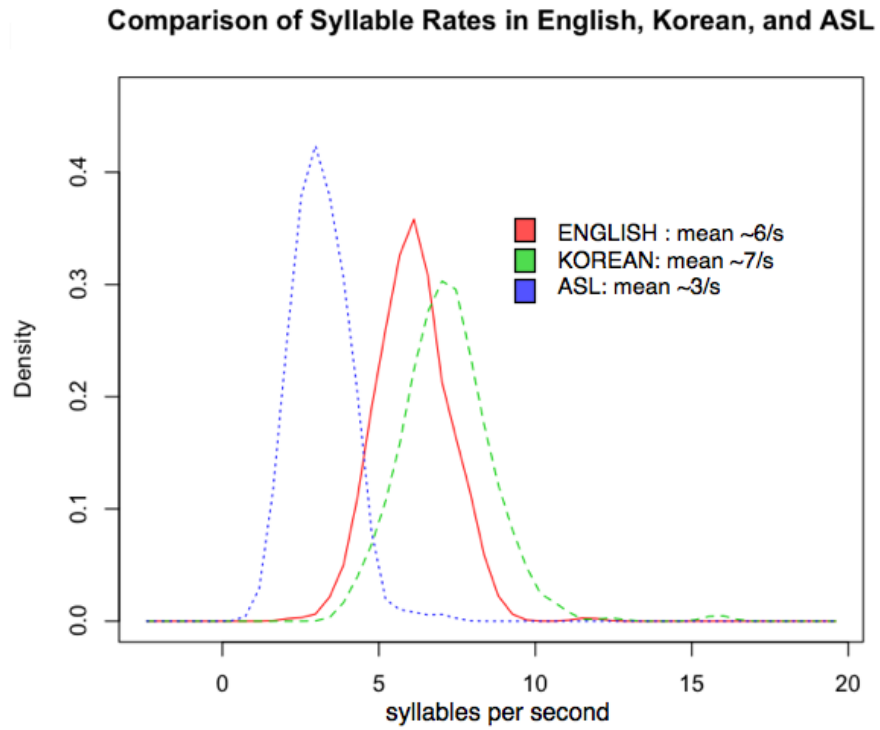


Figure 34. Estimated probability density functions for syllable rates (syllables per second) of sentences from conversational data in English, Korean, and ASL. Syllables rates in ASL may be the basis for the temporal integration window of ~250-300 ms found in Experiment 1 in Chapter 2.

Comparison of Morpheme Rates in English, Korean, and ASL

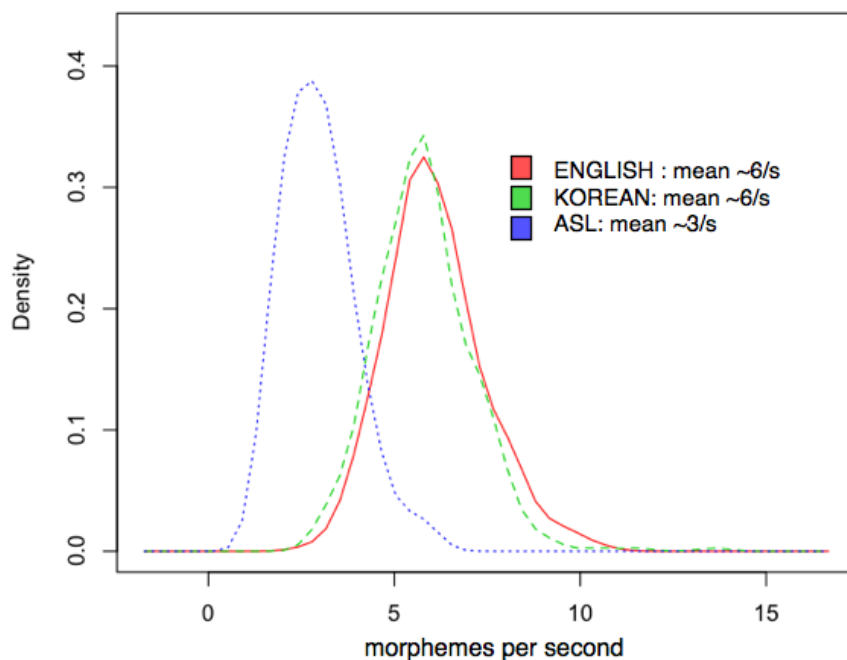


Figure 35. Estimated probability density functions for morpheme rates (morphemes per second) of sentences from conversational data in English, Korean, and ASL. This figure demonstrates that English and Korean, two spoken language with distinct grammars, have the same morpheme rate (~6 per second), in contrast with the morpheme rate in ASL (~3 per second).

The difference between faster English word and slower ASL sign rates have now been discussed widely in the literature, along with speculations on how simultaneous encoding of information (through greater morphological complexity) of ASL signs and more condensed ways of expressing meaning (through “doing without”) may contribute to similar propositional rates. However, in order to test these assumptions, analysis of morpheme rates in these languages is necessary.

By liberal counting methods, morphemes were produced at ~3.0 per second, and by conservative counting methods, morphemes were produced at ~2.5 per

second. These results were surprising given that the rates in English and Korean were both approximately 6 morphemes per second, and Senghas and Coppola's (2001) analysis of rates in Nicaraguan Sign Languages, where 5 – 6 morphemes per second are reported among native signers. A detailed discussion about the theoretic and methodological considerations for why morpheme rate estimates are considerably lower is given in the following conclusion section. However, the present results suggest that strategies for “doing without” may play a bigger role than simultaneous morphology to reach the same propositional rates across modalities.

To test Brentari's (2002) assumptions presented in Table 3, when comparing the ratio of morphemes to words, there are approximately 1.3-1.4 morphemes per sign in ASL. This is the same ratio that was found in English (also 1.3-1.4 morphemes per word) and slightly lower than 1.9 morphemes per word found in Korean.

As explained by Wilbur and Nolen (1986), the articulators in signing cannot be hidden while in transition from one sign to another. Since Wilbur and Nolen already provide a thorough analysis of syllables from an articulatory point of view where all types of movements were measured, here, only intra-sign movements were counted to provide syllable rate estimates. Thus, the mean average number of syllables per second was predicted to be lower than that reported by Wilbur and Nolen (~ 4 syllables per second). In this study, approximately 3.1 syllables were produced per second. This suggests that approximately 25% of the movements during sentence production do not contribute to the articulation of signs.

Similar to the time-scales seen in morpheme rates, among these three languages ASL has a significantly slower syllable rate than English or Korean. These

results are consistent with other studies that report that monosyllabic English words are approximately half the duration of monosyllabic ASL signs (Emmorey & Corina, 1993; Corina & Knapp, 2006; Capek, Grossi, Newman, McBurney, Corina, Roeder, & Neville, 2009).

When examining the ratio of syllables to signs, it was found that there are approximately 1.4 syllables per sign. When examining the mean average of the ratio of morphemes to syllables, a liberal morpheme count resulted in an average estimate of 0.96 morphemes per syllable and a conservative morpheme count resulted in an average estimate of 0.81 morphemes per syllable. In other words, these mean average ratios are very similar in range to the values found for English and Korean.

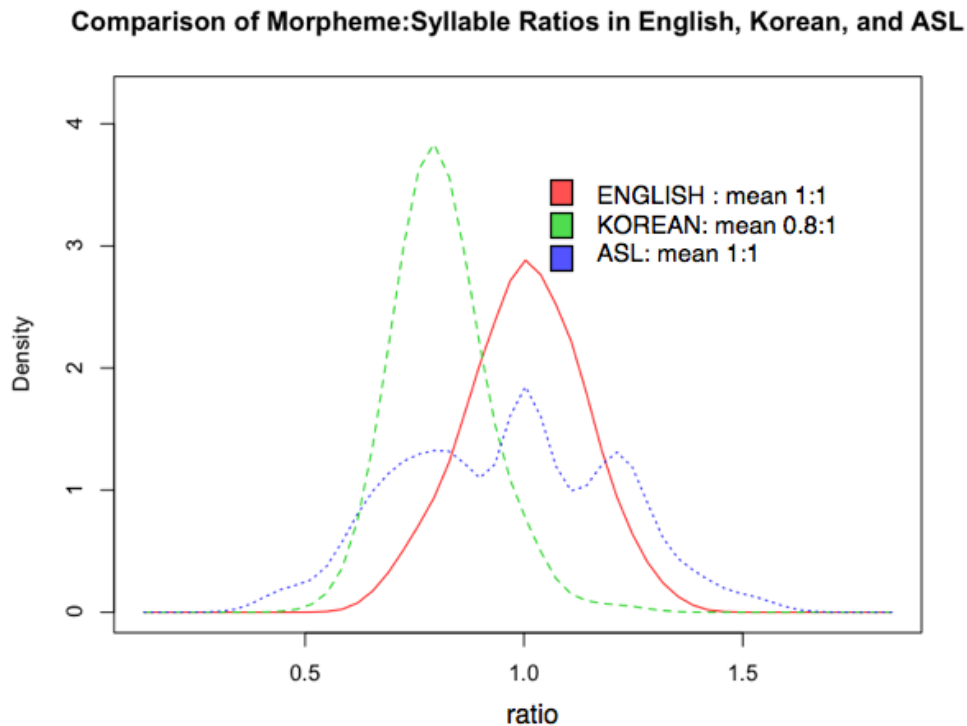


Figure 36. The comparison of morpheme:syllable ratios in English, Korean, and ASL suggests that the globally, morphemes and syllables are processed at approximately the same rate. However, the results from ASL are different from spoken languages in that the ratios reveal a trimodal distribution. This may be attributed to properties unique to sign languages, such as productive use of reduplication (resulting in ratios lower than 1:1) and productive use of spatial modulations (resulting in ratios higher than 1:1), in addition to simple signs.

However, as seen in Figure 36, the ratio in ASL shows a unique tri-modal distribution for sentences where sometimes the rate is lower than 1:1 and other times when the ratios are higher. Sign languages differ from spoken languages by having productive use of reduplication, where a sign can be repeated multiple times, and allowing more compacting of information through simultaneous strategies. Despite

these varied options, the ASL follows the pattern of English and Korean where global rates of morphemes and syllables are approximately the same.

The need for expanding the sample size of ASL sentences and assessing the inter-rater reliability of morpheme and syllable estimates presents some methodological challenges before adopting these findings conclusively. Moreover, this area investigation trying to understand the temporal dynamics of linguistic processes in production and perception requires a better theoretical consensus on how to count all of these units (words/signs, morphemes, and syllables) and compare them. Nevertheless, the emerging trend from this first attempt to study all of these rates together suggests that units of form (syllables) and meaning (morpheme) unfold at approximately the same time scales in all languages.

3.7 Conclusion

By examining the rates of words, signs, morphemes, and syllables, this study provides new insights on the universal time properties of language production and also differences that arise due to grammar and modality. The results from English and ASL converge with previous studies that have examined word, sign, and syllables in these languages (Bellugi & Fischer, 1972; Grosjean, 1979; Wilson & Nolen, 1986; Emmorey & Corina, 1993; Corina & Knapp, 2006). The results from Korean syllables confirm models of speech production based on other spoken languages (Greenberg, Hollenback, & Ellis, 1996; Arai & Greenberg, 1998). The unique contribution of this present work is the demonstration of the relationship between the

physical dynamics of language production and representational units of meaning. Taken all together, these findings reveal consistent patterns in language processing although the particular rates may differ.

Bellugi and Fischer's (1972) original work comparing the rate of a spoken language (English) and signed language (ASL) concluded that at the word/sign level, signed languages are twice as slow as spoken languages but that at the propositional/sentence level, the rates across the modalities are the same. They speculated that the convergence of global rates despite the discrepancy of local rates is due to the differences in the grammatical properties of the two languages. Later work (Klima & Bellugi, 1979) examining a signing system that maintains a similar grammatical structure to English verified that without the special grammatical properties of a true sign language, a manual communication system is significantly slower.

The present results demonstrate that the word-sign comparison is not very meaningful when considering that even among spoken languages, the amount of linguistic information within a word can vary greatly, which is traditionally represented by the analytic-synthetic continuum. A comparison of word rates in English (~ 5 words per second) and Korean (~ 3 words per second) reveals that word rates are not indicative of major differences due to modality but grammar and how word boundaries are determined in languages. Nevertheless, an analysis of morpheme rates in English, Korean, and ASL indicates that Bellugi and Fischer's conclusion about rate differences due to modality, where spoken languages are twice as fast as signed languages, still presents a deep puzzle. Morpheme rates are ~ 6

morphemes per second in English and Korean and ~ 3 morphemes per second in ASL.

Moreover, this work goes beyond Bellugi and Fischer's study by analyzing the rate of syllables among three languages. It also complements Wilbur and Nolen's (1986) study focusing on syllable rates of ASL but differs from their work by focusing on intra-sign movements (or syllable nuclei) that are involved in the articulation of signs, whereas they also included inter-sign transitional movements. Syllable rates reveal the physical dynamics in production and also serve as units for sensory integration in perception. In phonological theory, syllables serve as sublexical units to which constraints and rules apply. Similar to the notion of syllables in spoken languages, syllables in sign languages organize the timing of phonetic segments and arrange them into a sonority/saliency hierarchy. Wilbur and Nolen have speculated that syllable rates are the same across spoken and signed languages. However, the results presented here suggest that time-scales of syllables across the modalities are different – ~6-7 syllables per second in English and Korean, and ~3 syllables per second in ASL (or ~4 syllables per second according to Wilbur and Nolen). These rate differences are consistent with the differences in the frequency of syllables found in babbling, where vocal babbling is faster than manual babbling.

Nevertheless, a consistent pattern that emerges is that the ratio of morphemes to syllables is approximately 1:1 in both modalities. In English, there are certainly polysyllabic and monomorphemic words, such as *apple* and *kitchen*. However, there are also numerous highly frequent monosyllabic and multimorphemic words, such as

went and *men*. The same pattern holds in Korean. As Brentari (2002) has described, in some ways ASL can be described as a language that is monosyllabic and polymorphemic because it has a rich system of simultaneous morphology that exploits the use of space. However, it also has an inventory of bisyllabic signs (like CANCEL and NEVER) and cases in normal usage where monosyllabic signs are reduplicated to polysyllabic forms. Among all these languages, many morphemes are monosyllabic, and monomorphemic-polysyllabic cases are balanced with polymorphemic-monosyllabic cases.

Bellugi and Fischer (1972) listed three reasons for how propositions/sentences in ASL can contain similar amounts of semantic information despite having fewer signs/words than English: 1) doing without, 2) incorporation, and 3) body movements and facial expression. Based on the results of the present study, which took into account the incorporated information in signs by measuring morpheme rates, the factor that seems to play the biggest role in the convergence of rates in spoken and signed languages appears to be the idea of “doing without,” which Bellugi and Fischer characterize as a way of reducing redundancy and increasing information density.

In doing a morpheme rate analysis, this study was not able to replicate the findings from Senghas and Coppola (2001), who measured morpheme rates as an indicator of fluency. Among the group who used the full-fledged version of Nicaraguan Sign Language, the average rate was 350 morphemes per minute, or ~6 morphemes per second. Because the study focused on the use of spatial modulations in the grammar and did not elaborate on the details of the rate analysis, it is not

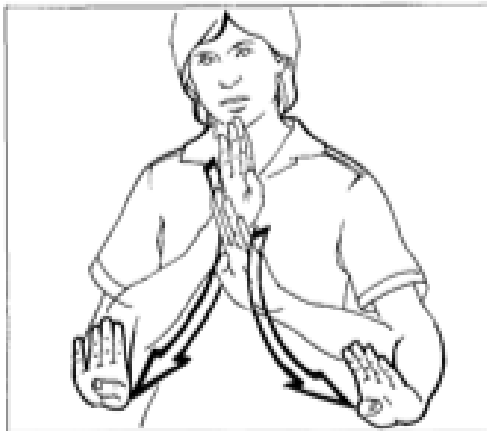
possible to determine whether these differences in results are due to the difference between ASL and NSL or a difference in methodologies on how morphemes were counted.

In addition to assessing inter-rater reliability for all these data, future analyses on the coded data will benefit from considering alternative ways of counting morphemes, especially in ASL. Determining how to count morphemes presents challenges in both spoken and signed languages. Theories of syntax and morphology in generative grammar posit the presence of phonetically null elements that serve functional roles in derivations (Embick & Noyer, 2007; Baker, 1996). Here, only morphemes that were phonetically realized in some way were counted. For example, *men* was counted as having 2 morphemes even though the regular plural suffix is not attached because of a phonetic change to the root. The same approach was taken when analyzing ASL, with most attention given to the manual gestures and where facial features were taken into account in question-marked constructions. It is possible that different criteria could have resulted in a higher estimate of morpheme counts. For example, it was decided that agreeing and spatial verbs would be counted with 2 morphemes, 1 for the root and 1 for the agreement or spatial feature. Another approach would have been to count these as having at least three morphemes, the verb root, and subject and object for agreement verbs, and the source and goal locations in spatial verbs. However, the vast majority of the verbs found in the sentences (that were selected before the annotation process) were plain/uninflecting, and it is predicted that this revision would not significantly change the results. de Beuzeville, Johnston, & Schembri (2009) report similar patterns for plain verbs in

Australian Sign Language. Perhaps one way to increase the number of constructions involving morphologically richer verbs would be to have participants view videos with actions involving many of these verbs and then discuss them with other participants.

Another potential way of increasing morpheme counts in ASL is to take into greater consideration the derivational processes described by Padden and Perlmutter (1987) that change the movement of a sign. Repeated circular movements can change regular adjectives to mean ‘characteristically ____’. Small, quick movements that are reduplicated forms activity nouns from verbs, as in pairs such as SIT-CHAIR. Although it seems relatively simple to systematically count signs like CHAIR as consisting of 2 morphemes (SIT+NOUN), it becomes a tricky issue for nouns like CHURCH and NURSE that phonologically have reduplicated noun forms without corresponding verbs.

QUIET



TACITURN

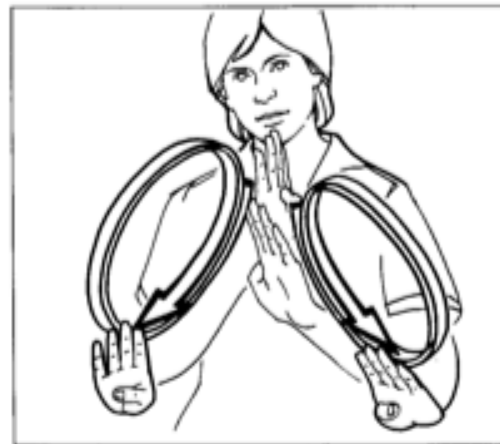


Figure 37. Reproduced from Padden & Perlmutter (1987), where reduplicating circular movement turns the adjective QUIET to mean ‘characteristically quiet’, or taciturn.



Figure 38. Reproduced from Aronoff, Meir, & Sandler (2005), demonstrating a complex ASL classifier construction: ‘A person walks forward, (dragging) a dog squirming behind.’

Perhaps classifier constructions present the greatest challenge in understanding how many units of meaning can be captured in visual imagery. Liddell (2003) provides a useful discussion of these issues. DeMatteo (1977) and others argue that classifier constructions (or “classifier predicates”) are analogical rather than discrete, and that morphemic representations of these constructions are not appropriate. In contrast, Supalla (1982) has proposed that these constructions can be analyzed as being a highly complex, productive, multimorphemic system. Liddell himself has argued for a hybrid of these models where handshapes have lexical status but the use of these handshapes are gradient/analogical. Liddell cautions against attributing morpheme status to metaphorically expressed, depicting movements. For example, he points out that [rl] is not considered a morpheme in English although there are words like *curl*, *swirl*, *whirl*, *twirl*, *furl*, and *gnarl*, which all have meanings related to round, twisted shapes. Dudis (2011) explains that these issues that make depicting verbs hard to analyze morphologically also apply to agreement/indicating verbs since they utilize correspondences in space.

Iconic aspects of signs are highlighted as one of the key modality effects in language and poses interesting challenges in understanding how meaning is composed. For example, mental verbs and nouns tend to involve articulations near the forehead, for example, THINK, KNOW, and DREAM. One possibility is to assume a morpheme for MIND, but it is impossible to distinguish whether such morphemes are computed compositionally or whether the forehead is exploited as an iconic place of articulation in phonology. Signs like BELIEVE and AGREE have been described as originating from compounds: THINK-MARRY and THINK-SAME, respectively. In the case of THINK-MARRY, there is a change from the “1”-handshape for THINK to the “C”-handshape for MARRY. It is now common to see uses of BELIEVE involving handshape assimilation where the “C”-handshape starts near the forehead. Understanding the etymology of this sign motivates a bimorphemic analysis, but there may become a point where this compositional aspect gets lost in on-line processing.

Finally, any future work examining the morpheme rate in a sign language should provide a more careful analysis to non-manual features, which in addition to eye-brow raising/lower includes, eye-gaze, body-shifts, and mouthing. Facial articulations can provide lexical information that adds to the meaning of a sentence. For example, when mouthed at the same time as the verb, the ‘TH’ expression (tongue between the teeth) means ‘carelessly’ and ‘MM’ expression (protrusion of the lips) means ‘with relaxation and enjoyment’ (Corina, Bellugi, & Reilly, 1999). Eye-gaze and body shifts may have provided a phonetic cue of pronouns in cases where

the argument was assumed to be ‘null’. In order to catch these subtleties that may have been lost in these annotations, a corpus with high video quality is necessary.

Taking all these factors into account may show that ASL also displays >6 morphemes per second, suggesting that morphemes rates are universal. If so, the ratio of morphemes to syllables may be >1:1, which may be a unique property of sign languages. However, if transitional movements between signs are also factored in, as proposed by Jantunen (2010), ratios may still remain consistent.

Another way to compare rates of spoken and signed languages in the future may be to compare only open-class/lexical (where lexical is contrasted with functional) morphemes. For example, in a sentence like ‘*I ate an apple*,’ although there are 5 morphemes total (I-eat-past-an-apple), it only contains 2 lexical morphemes (eat-apple), like the ASL sign EAT-APPLE. It is possible that ASL may have more phonetically null morphemes than English or Korean. As discussed by Lillo-Martin (1991) and Fischer et al. (1999), ASL may be more discourse-dependent than English, where the meaning of individual sentences is harder to recover in isolation. Because Korean is a pro-drop language, it may be considered more discourse-dependent than English, but morpheme rates were in the same time-scale as English. Although Lillo-Martin (1991) has suggested that ASL is discourse-dependent like Chinese, Japanese, and Korean, more significant differences in the degree of discourse-dependence may be determined by modality.

The present study did not conduct an analysis of propositional rate for the following reasons: 1) the materials used for the three languages were not matched for semantic content, and 2) the high likelihood that propositional rates across English

and ASL are comparably equal given the task of simultaneous interpreting by professionals. Although discrepancies may exist for particular constructions, the global rates are generally assumed to be the same. Padden (2000:179) summarizes this view by saying, “Languages of different modalities organize timing, prosody and syllable structure differently even if linguistic content is similar. However, over a span of time, the amount of information in any language, signed or spoken, is roughly equivalent.” The slowness of artificially created signing systems adds further support to the idea that natural language processing occurs within a certain range of time constraints.

An analysis of the rate of signing or speech in highly-skilled interpreting and comparison with the rate of the original production may be a useful way to study how global rates become equivalent across the modalities. It is expected that some short constructions in ASL require long English translations, and vice versa. However, to more accurately capture these patterns and fully understand universal time properties, an examination of more sign languages, especially those with different grammatical properties, is needed. For example, Japanese Sign Language is reported to have gender marking on verbs, and Taiwan Sign Language is reported to have auxiliary verbs (Padden, 2000).

Aside from words/signs and propositions, languages also have phrasal units, which are intermediate levels of structure. Due to time constraints, it was not possible to look at phrasal units at the time of the study. Nevertheless, I can speculate about how spoken and sign languages compare at intermediate time scales. Nespor and Sandler (1999) describe how similar principles of dividing sentences into

prosodic and intonational phrases applies to spoken and signed languages. Although there are debates about the degree to which this isomorphism holds, it has been shown that there are phonological constituents that correspond to syntactic constituents (Nespor & Vogel, 1986; Selkirk, 1984). In sign language production, eyeblinks have been recognized as occurring at syntactic boundaries and discourse transitions (Baker & Padden, 1978; Bahan & Supalla, 1995). Nespor and Sandler (1999) provide an analysis of Israeli Sign Language, where cues for prosodic and intonational phrase boundaries are taken from facial features (brows, eyes, cheeks, mouth, tongue, head tilt, mouthing), body shifts, and temporal cues (reduplications, pauses, and speed and size of movements). Although they do not report the time durations of prosodic phrases, which are embedded in intonational phrases, on average their examples show prosodic phrases with 2 signs and intonational phrases with 3 signs. Boyes-Braem (1999) provide an analysis of prosodic rhythms among early and late learners of Swiss German Sign Language. The examples in her work show the time-course of sentences, where signs and prosodic units are labeled. A rough estimate based on the measurements she provides suggests that a prosodic unit is, on average, about 1 second long. Some of these findings may converge with reports of speech, where prosodic information is conveyed at rates of 1–3 Hz (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004).

The intuition behind the finding that global rates in languages are equivalent is that the rate of informational transfer in language is relatively consistent. In this sense, the capacity of the communication channel in language may be amodally determined. The objective in transmitting information is to ensure that the message is

conveyed through a noisy channel, in the shortest amount of time, and with the lowest probability of error. Resistance to error and rate are traditionally considered to be opposing factors, where conveying minimal information with each fragment is the most error resistant strategy but also the slowest. Chong et al. (2009) demonstrate that the ratio of the information bits in units of ASL and English is approximately 5:1. Nevertheless, English is not overall slower than ASL. An extension of their analysis to calculating information transfer rates suggests that they may be quite similar, although the channels that are particular to different modalities may determine how many information bits are transmitted at a time. Based on these characteristics, auditory and visual processing of language may be differentially sensitive to noise in the channel, as demonstrated by the experiments in Chapter 2.

The finding that units of linguistic form and meaning unfold at approximately the same time-scales has broader implications for language processing. This suggests that sensory integration and extraction of meaning proceed in parallel. In addition to research from neuroscience (Hickok & Poeppel, 2007), a typological investigation of patterns in language processing can lead to better models for the architecture of computational and neural networks for language. Although there is still much room for improvement in our understanding of how many meaningful units get phonetically realized, how meaning is constructed, and the discrete nature of meaningful units (Embick & Noyer, 2007; Liddell, 2003), this first attempt to compare units of form and meaning highlights the importance of taking into consideration the time properties of phonological and morphological processing, which are temporally tightly linked.

4 Conclusion

4.1 Overview

In Chapter 2, I described evidence for larger temporal integration windows in sign language perception than in speech. In Chapter 3, I summarized findings that support the claim that units of form and meaning are produced in periods of longer time scales in sign language production than in speech. Despite these differences that are putatively aligned with modality, universal patterns emerge as well. In both English and ASL experiments, intelligibility of the sensory signal falls drastically with severe time distortions created with local reversals. The hypothesis that sensitivity to time distortions is dependent on the size of representational units in the signal has been confirmed in two ways, by comparing the results across English and ASL, and by comparing the results within each language for normal and compressed sentences. Taken together with the findings from Chapter 3, as well as other studies on production rates, the temporal integration windows implicated by these perception experiments corresponds to syllables in ASL.

In this concluding chapter, I discuss temporal patterns in language processing more broadly, providing a synthesis of key findings from speech and sign language research, considering the implications, and outlining future directions.

4.2 More than meets the eye

Sign language perception involves much more than processing visual signals produced by the sign articulators. It is guided by linguistic knowledge about sign language grammar and sensitivity to how signals unfold in time. Temporal integration windows in language processing do not seem to arise just from the properties of a particular sensory system or just from a special property of language but from the interaction of the two. The difference between $\sim 50 - 60$ ms windows in speech and $\sim 250 - 300$ ms windows in sign language clearly demonstrate the effect of modality. The window of $\sim 250 - 300$ ms in sign language perception for sentences played at normal rates in this work is attributed to the perceiver's integration of the visual signal according to syllabic units in ASL (present results from Chapter 3, as well as Wilbur & Nolen (1986)). Studies of compressed and locally-time reversed sentences in both modalities have now shown that the durations over which the signal is integrated must be flexible to a certain extent and adjust to the rate of the incoming linguistic information. Of course, time-compression studies of spoken and sign language (Foulke & Sticht, 1969; Foulke, 1971; Ahissar, Nagarajan, Ahissar, Protopapas, Mahncke, & Merzenich, 2001; Fischer, Delhorne, & Reed, 1999) at increasing rates also demonstrate the limitations of this flexibility, but similar findings from spoken and sign language suggest that perceptual bottlenecks are modality-independent.

The results of Experiment 2, where the duration of temporal integration windows was proportionally reduced by half with sentences compressed by a factor

of 2, parallel to the findings in speech (Figueroa, 2009; Stilp, Kiefte, Alexander, & Kluender, 2010), point towards common mechanisms in the auditory and visual processing of language. Stilp et al. (2010) argue that the findings from locally reversed speech support explanations based on cochlear-scaled spectra. However, the present results from sign language demonstrate the need for more general models, where perceptual processes are more broadly driven by sensitivity to the rate of incoming information.

In studies of low-level visual processing, temporal resolution in vision is ~20 ms (Chase & Jenner, 1993). In an EEG visual MMN paradigm, temporal windows of 150-170 ms in duration are reported (Czigler, Winkler, Pató, Várnagy, Weisz, & Balázs, 2006). The results of Experiment 1 in Chapter 2 implicate longer windows of ~ 250 – 300 ms for sign language processing, suggesting that the linguistic nature of a perceptual task can extend the duration of windows for sensory integration. As described in Chapter 3, studies from categorical perception in signing (Emmorey, McCullough, & Brentari, 2003; Baker, Idsardi, Golinkoff, & Petitto, 2005; Best, Mathur, Miranda, & Lillo-Martin, 2010) and perception of apparent motion (Wilson, 2001) have also shown that sign language knowledge guides visual processing.

At the algorithmic level of language processing, I adopt the assumption that perceptual processes are guided by internal guesses about the upcoming representations (Halle & Stevens, 1959, 1962; Stevens & Halle, 1967; Yuille, & Kersten, 2006; Poeppel, Idsardi, & van Wassenhove, 2008). The analysis of rates in natural production describes what the patterns that influence perception might be. Part of integrating the sensory signal over certain time windows is driven by the

expectation for representations unfolding over those durations. When the sensory signal is manipulated in such a way that those expectations are violated, cognitively restoring the signals becomes much more difficult.

Sentence processing in both spoken and sign languages requires the ability to track rapidly changing sensory signals and integrate them skillfully over long durations. As Foulke and Sticht (1969) note in their review of compression studies, there are cases where performance on the identification of words is lower than overall comprehension of sentences, and where it is also higher. The results in Chapter 2, where intelligibility of locally-reversed input falls sharply at 267 ms reversals and plateaus at ~50% at reversals of ~500 ms and greater, reflect the demands of phonological processing in sentence processing. In a separate pilot study that was designed by Clifton Langdon, we tested the intelligibility of locally reversed single signs and found that accuracy was higher than 50% for most signs. Although many signs are recoverable, when reversals exceed a certain size, it is likely that “un-doing” the motion is not automatic and requires deliberate effort. When all signs in a rapid sequence are distorted in such way, capturing each sign using concerted strategies becomes much more challenging. As Mayberry and Fischer (1989) describe late learners, difficulty in sentence processing can be attributed to phonological bottlenecks (late L1 versus L2 learners of ASL were not distinguished in their study). Late learners are believed to be much less efficient at phonological encoding, which has consequences for many other aspects of language processing. Experiment 1 results suggest that local reversals cause disruptions in the automatic recognition of phonological information that is encoded through time. However, spatially encoded

phonological information provides a buffer that makes signed sentences more robust to time distortions than speech. Late learners are characterized as a group that has difficulty with efficient phonological encoding for even normal sentences.

Experiment 3 results show that late L2 learners of ASL are much more sensitive to distortions in the signal than native signers.

In addition to theories of sensory processing, theories of representations of linguistic units and knowledge about how they combine are integral to complete models of language processing. In particular, assumptions about the status of linguistic primitives motivate psycholinguistic and neurolinguistic investigations about how they fold unfold in real time (Poeppel, Idsardi, & van Wassenhove, 2008). In turn, considerations about the time-scales at which these units are processed may help better inform theories about representations for features, segments, syllables, morphemes, phrases, and sentences.

4.3 Hierarchical coupling in sign language processing?

In speech perception, it has been proposed that endogenous rhythms in the gamma (30-50 Hz) and theta (4-7 Hz) bands serve critical roles for the processing of segments and syllables. More broadly, rhythmic aspects in many biological functions are associated with the frequencies of neural oscillations. Given the findings that the rate of syllables and morphemes in ASL are 3 per second, and that ~ 250 – 300 ms durations are critical temporal integration windows in perception, neural activity in the delta (1-3 Hz) band is implicated for sign language processing.

As is emphasized in the multi-time resolution model of speech perception (Poeppel, 2003), temporal integration windows need not be viewed as serially organized frames for processing. In sign language, different levels of representations are also evident in theories of segments, syllables, prosodic units, and intonational/discourse units. Even in the case of fluent fingerspelling, the letters do not come as a simple sequence of letters but are structured into “chunks” that have been referred to as movement envelopes (Akamatsu, 1982). Thus, while delta oscillations are by no means the only important neural activity, the new psychophysical results presented here strongly suggest that they may have a privileged status in sign language processing.

The analysis of fine-structures in speech that operate at fast rates are attributed to oscillations in the gamma band and bilateral activations in the superior temporal gyrus (STG) of the auditory cortex (Boemio, Fromm, Braun, & Poeppel, 2005). Aside from some trilled movements where the temporal direction is nondistinctive, sign language does not involve fluctuations at such high frequencies. Nevertheless, in experiments that tested the perception of meaningful lexical signs and meaningless (but phonetically plausible) signs, a bilateral activation in STG was found only for deaf signers but not hearing nonsigners (Petitto, Zatorre, Gauna, Nikelski, Dostie, & Evans, 2000). Although the findings from deaf signers may point to explanations where STG might be more generally sensitive to some aspects of visual processing and not just auditory processing, the differences from hearing nonsigners suggest that the activation was driven by the nature of higher order processing of the visual signals, such as phonological processing, lexical access, and subsequent integration

into other computations. Electroencephalographic gamma activity has been used to study neuroanatomy and processing dynamics of speech and sign language production (Crone, Hao, Hart, Boatman, Lesser, Irizarry, & Gordon, 2001), with results that are fairly consistent with other imaging studies that demonstrate overlaps in the functional organization of language-processing areas across modalities. Aside from the special role that it may have for sensory selection in speech perception, gamma activity is more broadly associated with feature binding and attention (Singer & Gray 1995; Fries, Nikolic, & Singer, 2007; Schroeder & Lakatos 2008).

The current findings about the time properties of sign language processing suggest that the brain operates in a rhythmic mode, and more specifically, that neural activity entrains to the low frequency rhythms of signing. Based on the models of oscillatory coupling, especially where gamma synchronies contribute to enhancements in the processing of task-relevant events (Schroeder & Lakatos, 2009) and attention in visual information processing (Müller, Gruber, & Keil, 2001), future work may also find evidence for the critical role of gamma activity in sign language processing for sensory selection as well as higher-order processing as in speech.

Future work investigating the temporal properties of signing and the neural basis for these dynamics requires use of methodologies with high temporal resolution, such as EEG and MEG, complemented by high temporal resolutions measures of sign articulation. It may be predicted that phase patterns of endogenous rhythms in the delta band will be correlated with the sign language intelligibility, where successful processing of the visual signals requires continuous segmentation and integration of the input in ~300 ms temporal windows. At these low frequency rates, the dynamics

of sensory processing in spoken and signed languages may converge. However, these low frequency rates may play a greater role in sign language processing because lexical and prosodic information are processed together at these time-scales. This prediction may be consistent with the prosodic model of sign language phonology (Brentari, 1998:22), who argues that “ASL exploits paradigmatic constraints in a greater range of phenomena than do spoken languages.” Finally, understanding the nature of the relationship between the neuronal oscillations that subserve language-independent functions and those that entrain to the sensory input in language processing should be a broader goal in this research.

4.4 Innate sensitivity to rhythms in language

Sensitivity to rhythms in language is attested in the earliest stages of language acquisition, where newborns are born preferring the voice and language of their mother (DeCasper & Fifer, 1980; Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988). An analysis of newborns’ cry melodies have shown that their productions reflect the prosodic contours of their mother’s language (Mampe, Friederici, Christophe, & Wermke, 2009). After birth, prosodic information in speech continues to shape the language acquisition for young children, for word segmentation (Jusczyk, Houston, & Newsome, 1999) and learning syntactic structure (Gleitman & Wanner, 1982). Infants seem to prefer input where prosodic contours are made salient through infant-directed speech (Cooper & Aslin, 1990; Werker &

McLeod, 1989). Babbling, one of the earliest stages of language production, is marked by its rhythmic qualities.

The importance of rhythm in sign language processing is now evident in a wide variety of cases. Babbling is no longer considered to be a precursor to speech because of the biomechanics of the mandible but to all languages (Petitto & Marentette, 1991). Deaf infants growing up in signing environments also prefer “motherese” versions of the input (Masataka, 2003). The sensitivity to rhythmic aspects of the visual signal does not arise from auditory deprivation. Hearing children who are born to deaf parents and thus exposed to signing also manually babble (Petitto, Holowka, Sergio, & Ostry, 2001). These manual gestures are distinct from other manual movements that might be typical of general motor development because they are produced in the signing space, produced at unique frequencies, and only appear among sign-exposed infants. Evidence that sensitivity and preference for linguistic input is partially innate and not driven by exposure is presented by Krentz and Corina (2008). 6-month old hearing infants that had never been exposed to sign language show a preference to look at videos of signing compared to videos of communicative gestures that are not linguistic.

Even in the domain of fingerspelling, which may be considered a sequence of handshapes to represent letters of the English alphabet, shows hierarchical organization and rhythmic properties. The acquisition of fingerspelling by young deaf children reflect their recognition of movement envelopes, where fingerspelled words are analyzed as whole units rather than individual handshapes (Padden & LeMaster, 1985; Andrews, Leigh, & Weiner, 2004).

Rhythmic characteristics also distinguish native and non-native signers. In subjective ratings, the cues that judges used to determine whether or not a signer was native or non-native were handshape, facial expression, rhythm, and lexical choices (Kantor, 1978). In a quantitative measurement of the production of native and non-native signers of Swiss German Sign Language, Boyes-Braem (1999) found that native signers use side-to-side movement of the torso according to prosodic and discourse units in the signed sentences, and that this was lacking among late learners. Among the three late learners, one who had some limited exposure for one year at an early age, had more of these left-right movements than the other two who had no early exposure (see Figure 39). The results suggest that late learners are following the prosodic patterns of spoken German (their first language) rather than sign language. This production study stands in contrast with perceptual studies where nonsigners had similar sensitivity to sign language prosodic cues (Brentari, González, Seidl, & Wilbur, 2011; Fenlon, Denmark, Campbell, & Woll, 2007). Thus, although some aspects of prosodic rhythms in signing may be perceptually salient and not require sign language knowledge, it is interesting how these characteristics do not become automatic in production for late-learners who have had extensive exposure to signing.

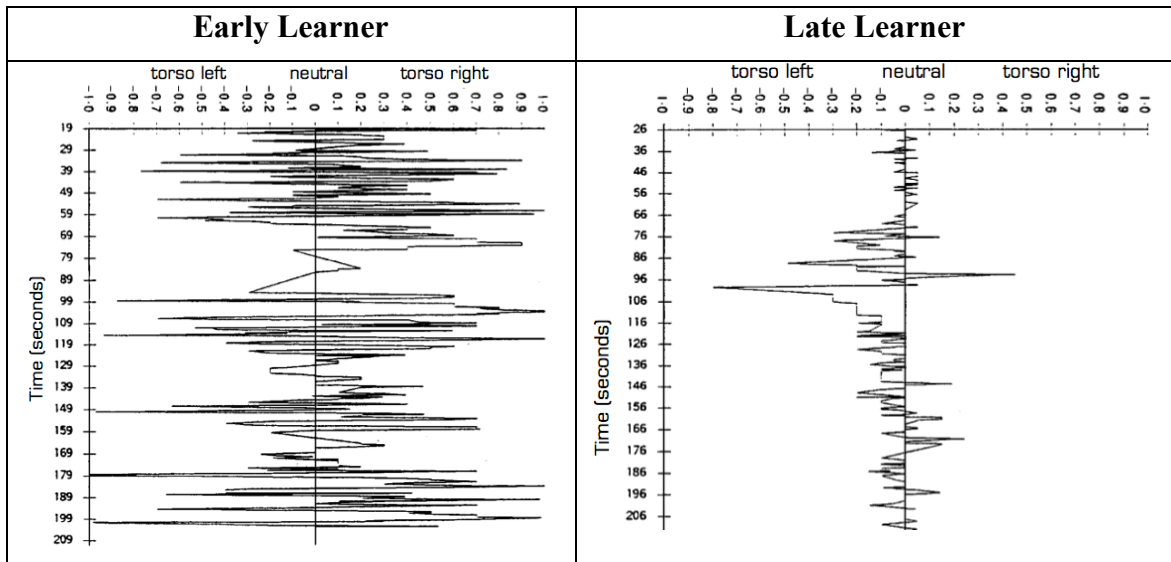


Figure 39. Reproduced from Boyes-Braem (1999), demonstrating the difference between early and learners of Swiss German Sign Language in their lateral torso movements while signing.

By continuing to better understand the rhythmic aspects of sign language production, future research can address what are the temporal characteristics of typical and atypical development. Studying the spectral characteristics of signing (Foulds, 2004) can also lead to better models of what perceptual cues, aside from grammatical organization, distinguish linguistic and nonlinguistic gesture. Finally, such guidelines may help better understand the status of iconic gestures, which seem to straddle these boundaries, in visuo-spatial communication.

4.5 Channel capacity for sign language

Understanding the rate at which linguistic information is transmitted has had practical applications for designing communications devices. The greater bandwidth required for videophones compared to telephones leads to the over-simplistic belief that sign language requires larger channel capacities in natural processing. Chong et al. (2009) demonstrate that a phonetic unit that is realized in some fragment of time in ASL contains 5 times the amount of information compared to a phonetic unit in English. Based on this calculation, estimating the bit rate per second in English and ASL based on production rates of words and signs showed that global information transfer rates are the same.

In an independent information theoretic analysis, Reed and Durlach (1999) also reach the conclusion that auditory processing of English and visual processing of ASL involve the same information transfer rate. Among all the communication systems they analyze (which also included Morse code though different modalities and Braille), the only other system that had comparable information rates with spoken English (auditory form) and signed ASL (visual form) was reading (visual form). Notably, the visual and tactile forms of spoken English and the tactile form of ASL had significantly lower rate measurements.

In a study examining whether it is possible to transmit signs using the bandwidth of one telephone line, Tartter and Knowlton (1981) examined the intelligibility of signs produced with 27 moving spots. This technique has been used to study the gross patterns of biological motion (Johansson, 1973). In signing, 13

retroreflective tapes were anchored to gloves worn by each hand and 1 on the nose to provide a reference for place of articulation. 27 moving spots were sufficient to allow two pairs of deaf subjects to have conversations, although there was some difficulty with understanding fingerspelling. In other studies, spatial image compressions and coding schemes have shown similar results, where videos can be substantially compressed while conveying intelligible messages in sign languages (Sperling, Landy, Cohen, & Pavel, 1985; Abramatic, Letellier, & Nadler, 1982; Pearson, 1981).

In a more recent study examining the compressability of sign language video files, Foulds (2004) approaches the bandwidth requirements from both the perceptual and biomechanical perspectives. Transmission of video with a limited bandwidth involves a trade-off of spatial resolution with frame rates. He explains that most efforts on sign language communication systems have focused on how to achieve lossy spatial compression while preserving temporal information. In perception, high frame rates are necessary to surpass the critical flicker frequency. However, from a kinematic point of view, critical information for sign language perception may be encoded more sparsely. In a separate pilot study, Foulds measures the spectral characteristics of sign language motion by using a sensor that tracks the right index finger of a signer who produced a list of 20 ASL signs. Convergent with the results of the rate analysis presented in Chapter 3, he found that most of the spectral energy is in the lower frequency range of 0-3 Hz.

Based on these findings, Foulds estimated that a frame rate of 6 frames per second may be sufficient to capture the kinematic information necessary for sign intelligibility, the higher standard of 30 frames per second (0-15 Hz bandwidth) is

necessary to avoid flickers in perception. Foulds uses a method that smoothly interpolates the lower bandwidth to the standard 30 frames per second. The results from an intelligibility experiment, where original videos (with 0-15 Hz bandwidth) were compared to stick figure animations (with 0-15 Hz and 0-3 Hz bandwidths), showed that the temporal compression by a factor of 5 (to 6 frames per second) preserved the intelligibility of the stimuli. Foulds concludes that “Earlier reported limitations were imposed by human perception and are not determined by the kinematic bandwidth of human movement associated with sign production.”

Fould’s measurement of the spectral characteristics of signing motion should be extended to articulations of conversational sentences for future studies comparing the dynamics of the sensory signal to the rhythms of neural oscillations, as discussed earlier.

4.6 Availability of two communication channels?

Given the apparently large differences between the vocal-auditory and manual-visual modalities, the convergence of rates in spoken and signed languages is remarkable. What happens when both modalities/channels are available to a language user?

Bimodal bilinguals are individuals who are fluent in a spoken and a signed language, like English and ASL. In natural conversations, bimodal bilinguals have been observed to produce code-blended constructions, even while communicating

with English speaking monolinguals (Pyers & Emmorey, 2008). Based on the finding that bimodal bilinguals used ASL-appropriate facial expressions while speaking English, Pyers and Emmorey propose that “This result provides evidence for a dual-language architecture in which grammatical information can be integrated up to the level of phonological implementation.” In a different study, Casey and Emmorey (2009) found that bimodal bilinguals produce more iconic gestures than nonsigners and that actual signs are used from time to time.

The fact that both channels are available does not necessarily mean that information can be conveyed at a faster rate. In a production task (Emmorey, Petrich, & Gollan, 2009), English-ASL bilinguals’ performance on picture-naming tasks in ASL-only, English-only, and code-blending conditions were compared to English monolinguals and ASL monolinguals. The reaction times among the English monolinguals and ASL monolinguals were the same. The reaction times for bimodal bilinguals in the English-only conditions were similar to the English monolinguals. However, the responses to ASL-only and code-blending conditions were significantly slower, and the reaction times for these two conditions were the same. The results suggest that production in the non-dominant language (ASL) is usually slower, and in code-blending conditions, where reaction times in English and ASL match, the slower response is attributed to time-locking with the slower language. These findings suggest that the vocal and manual articulators are not independent in simultaneous production. However, in a perceptual experiment, where participants had to make semantic judgments to words given in English, ASL, or both languages (code-blended), the fastest reaction times were attested in the code-blended condition. Thus,

in the perceptual channel, the use of both modalities has a facilitating effect, but in the production channel, it has a cost.

In development, a comparison of bilinguals who are acquiring two spoken languages (French and English) and a spoken and signed language (French and Langues des Signes Québécoise (LSQ)) has demonstrated that early linguistic milestones in each language are similar (Petitto, Katerelos, Levy, Guana, Tétreault, Ferraro, 2001). Language mixing occurred to varying degrees in both groups. Even at early stages of acquisition, bimodal bilinguals children are found to produce simultaneous constructions of French and LSQ. 94% of the cases where the languages were mixed involved simultaneous language mixing. Although both channels can be exploited for bimodal bilinguals, simultaneous production seems to only occur in constrained ways. 89% of the simultaneous mixing cases involved lexically congruence. In the other 11% where the signs and words had different meanings, the meanings were cohesive, where, for example “*Ça ressemble*” in French was uttered at the same time MOUCHOIR was signed in LSQ, to result in the sentence, ‘This resembles a [facial tissue]’. The average length of utterances by these children when mixing the languages was around 3 words, the same as for utterances without language mixing. In other words, the availability of two channels did not mean that these children would produce utterances that had double the complexity.

More recent work looking at the development of bimodal bilingualism – among children in the U.S. learning English and ASL, and children in Brazil learning Brazilian Portuguese and Libras – also demonstrate both the constrained and variable aspects of natural, simultaneous use of languages (Quadros, Lillo-Martin, & Chen

Pichler, 2010). Quadros et al. propose that “multiple kinds of blending are possible with multiple articulators,” but that “one proposition is one computation with intermodal expression.” Cases where two separate propositions are uttered in the languages are never attested. However, corresponding words and signs are not always produced together, and mismatches between two languages were common. The results that I have presented suggest that spoken and signed languages operate at different time scales for transmitting lexical items, which may be the cause of these mismatches. These observations from language acquisition should be contrasted with the findings from adults (Emmorey, Petrich, and Gollan, 2009), where code-blending resulted in slower reaction times, which was attributed to the time-locking of the faster, dominant language (English) with the slower language (ASL).

The simultaneous use of a spoken language and a signing system for expressing whole sentences is called simultaneous communication, or SimCom. Given that there are no pairs of a spoken and signed language that have identical grammars, SimCom usually involves a signed version of a spoken language. SimCom does not arise naturally and was originally designed for use in deaf education settings. In Quadros et al.’s work studying natural forms of simultaneous production in early acquisition, it is shown that some mismatches between the modalities can be tolerated. However, the lack of natural, full-fledged simultaneous systems suggests that mismatches created by the differences in grammars of a spoken and signed language cannot be tolerated (Wilbur & Petersen, 1998). Even with the substitution of a natural sign language with an artificial signing system that follows the grammar of the spoken language, errors made in both languages while using

SimCom reflects processing costs. A common observation in SimCom is that signed English can become inaccurate due to omissions of signs. Marmor and Petitto (1979) report that there were errors in 90% of the signed English sentences produced by teachers. One possibility is that these errors in signed English could be a result of the fact that it is not a natural sign language and operates at a time-scale that is globally too slow. However, Hyde and Power (1991) also report that accuracy in Australasian Signed English in SimCom comes at the expense of decreased naturalness in speech production, with much slower prosody. Interestingly, accuracy and rate of speech were correlated, where individuals with faster rates were also more accurate.

In Bellugi and Fischer's (1972) study, production rates were analyzed for ASL-alone, English-alone, and simultaneous signing-speaking conditions. They explain that the participants had much experience and are highly skilled at simultaneous production, but they do not explain whether the signing was closer to ASL or signed English. From their description, it appears that both languages were affected by the other, where translations of ASL signs resulted in unnatural lexical choices in English, and vice versa. In the simultaneous condition, there were more errors in both languages and more time spent for pausing during the narration as compared to the one-language conditions. Somewhat consistent with the results from the picture naming task reported by Emmorey, Petrich, and Gollan (2009), the rate of speaking in the simultaneous condition was slower than in the speaking-alone condition, but the rates in signing were the same.

Wilbur and Petersen (1998) investigate the modality interactions with respect to temporal properties in SimCom. Consistent with previous studies, they find that

sentences produced in English only and ASL only take approximately the same amount of time. Consistent with Klima and Bellugi's (1979) results on signed English, Wilbur and Petersen find that sentences in signed English take considerably longer. The novel finding, however, is that SimCom requires durations that are longer than speech-alone but shorter than sign-alone conditions. In other words, speech production is slowed down in SimCom but signing is sped up. Given that speech and signing occur at different time-scales, SimCom forces the systems to be time-aligned, which incurs costs in time (for speech) and accuracy (for signing, as speeding up results in increased sign omissions).

As suggested by Fischer, Delhorne, and Reed (1999), as well as Foulke and Sticht (1969), bottlenecks in language processing are likely to be rooted in factors beyond motor articulation or sensory processing. In the case of SimCom, even when there is redundancy in meaning, there are minimal overlaps in phonological form, where sublexical units produced in each language contain information bits. Generating this amount greater amount of information has a high cost in production, although processing inputs with redundancy in meaning is advantageous in perception. These bodies of work combined suggest that channel capacities in language processing arise more from cognitive constraints than from the articulatory-perceptual interface.

4.7 Rates in production and time-course of recognition

Evidence now from several studies show that signs are produced at a rate of 2-3 per second in ASL. This implicates that signs are produced at periods of ~400 ms at a time. Although sign durations can be variable, this is not consistent with other ways of measuring the average duration of signs. Signs produced in isolation can be somewhat longer (>500 ms), whereas signs excised from sentences can be shorter (~250 ms), especially depending on the position within the sentence. The discrepancy between periods determined by rate analysis and durations determined by direct measurements is attributed to the fact that there are transitions external to the signs.

Results from perceptual tasks suggest that the time-course of identifying a sign is much shorter than 400 ms. Emmorey and Corina (1990) used a gating task with signs presented in isolation. Participants were asked to identify signs (and report how confident they were about their guesses) after viewing videos of a sign, where one videoframe was added to each presentation. On average, 240 ms of a sign contained enough information for accurate identification. These results are similar to the findings from speech, where although the whole unit is stored (in memory or the lexicon), recognition in perception only needs processing up to point where the unit is distinctive from other units. Grosjean (1981) found that for signs presented in isolation, signs can be recognized from approximately the first half of a sign. Extending these results, Clark and Grosjean (1982) tested signs that were produced within sentences and tested recognition times with or without the sentence context.

When presented in context, signs could be recognized from the first 40% of the sign.

When analyzing what percent way through a sign makes it possible to identify each of the four formational parameters of signs (orientation, location, handshape, and movement), they found that sign recognition was linked to the identification of movement (Grosjean, 1981; Clark & Grosjean, 1982). Although the four parameters could be isolated at around the same percent way into a sign, movement took the longest.

If the average period per sign is about 400 ms, but the true duration of the sign is in fact shorter, and only about 50% of the sign has to be viewed for lexical recognition, this suggests that much of the sign period is not contributing lexical information, which is somewhat puzzling. Clark and Grosjean do not explain how the onset of the sign was measured, especially with respect to how much of the transitional movement from the previous sign was included.

As discussed earlier in Chapter 3, sign-internal movements are distinguished from sign-external movements. Jantunen (2010) explains that “Standard theory treats transitions as nonlinguistic, unintentional, meaningless, automatic, nonsalient, unmodifiable, holistic, etc. (e.g. Wilbur 1990, Perlmutter 1990, Wilcox 1992, van der Hulst 1993).” However, Jantunen argues that this characterization of transition movements and traditional annotations of sign boundaries need to be revised based on two experiments (see Figure 40). In the first experiment, Jantunen studied the biomechanics of signing by measuring the acceleration peaks during sentence production and found that movement dynamics of sign-internal and sign-external transitions were quite similar. In the second experiment, he tested the intelligibility

of video clips that were created by excising the signs and concatenating the remaining transition frames in a sentence. More than 60% of such “signless” video clips were understandable. Given the saliency of these transitions both in terms of phonetic attributes and meaningful content, Jantunen proposes that they should be viewed as being internal parts of signs.

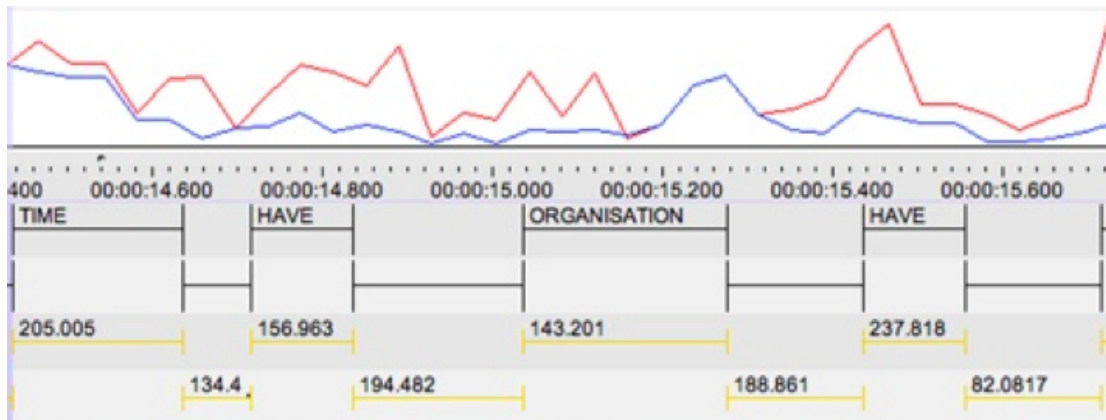


Figure 40. Reproduced from Jantunen (2010), demonstrating the acceleration peaks in the biomechanics of both hands while signing, annotated for traditional sign boundaries and transitions between signs.

If Jantunen’s model were adopted, it would support the idea that duration of signs and the periods of signing rates are in fact the same. Nevertheless, it still presents the question on why signs are produced over long periods of time when the time course of lexical recognition is shorter by as much as 50%. From an informational theoretic perspective, some amount of redundancy is not only expected but desirable. A potentially meaningful connection is that redundancy in printed English is also reported to be approximately 50% (Chong, Sankar, & Poor, 2009).

Jantunen’s model also suggests that the idea that most signs are monosyllabic needs to be revised. Future investigations examining the relationship between the

rate of form and meaning will need to consider these new ways of counting phonological units as well as morpheme units. Moreover, interpreting all movements to be sign-internal would mean that sign languages are not different from spoken language in that all phonetic components are meaningful even though the articulators are not hidden. Combining Jantunen's methods of measuring the patterns of acceleration peaks (Jantunen & Takkinen, 2010) and Fould's (2004) methods of measuring spectral frequency will lead to a better understanding of the rhythmic characteristics of signing.

4.8 General conclusions

The goal of this dissertation was to investigate temporal integration windows and the rates at which form and meaning unfold in language processing from a cross-linguistic and cross-modal perspective. Psychophysical experiments using locally-reversed speech demonstrated that temporal integration windows that capture direction-sensitive input are much longer in sign language ($\sim 250 - 300$ ms) than in speech ($\sim 50 - 60$ ms). Despite the differences in these absolute values, the universal pattern across languages is that temporal integration windows are sensitive to the size of representational units in language. This was demonstrated by the reduction of temporal integration windows in proportion to time-compressed sentences and by the comparison between English and ASL. The analysis of production rates from corpus data of natural conversations also contributed to a better understanding of what is the temporal dynamics in language production that might shape expectations in the

perceptual process. A common mechanism for mapping sensory signals to abstract representations that are used towards higher-order linguistic computations is integration in time-scales that match the rate of the linguistic units. Construction of meaningful representations may also operate at the same time-scale as syllables since the ratio of syllables and morphemes is approximately 1:1 . Although there may be different requirements in the technological implementation, auditory and visual channels for language processing seem to involve similar global rates of information transfer and the same amount of redundancy.

Studying the rates at which linguistic units are produced and the time-windows over which this information is integrated leads to a better understanding of how information is chunked and organized when processed through a particular channel. The degree to which information is encoded sequentially/simultaneously affects the time-scales of integration windows. In speech, the unintelligibility of sentences where fragments were reversed at durations exceeding the average size of segments demonstrates the importance of the temporal direction of the segments. In sign language, although syllables can also be decomposed to segmental units (Liddell, 1984; Perlmutter, 1992), the robustness of signed sentences to reversal sizes up to the length of syllables and signs suggests that the way temporal direction of these segments are encoded is different in nature from the segments in speech. These results are consistent with the findings of Wilbur and Allen (1991), who argue against internal structure in ASL syllables.

The difference in performance between early and late learners of ASL supports the view that part of being a native user of a language is the ability to

efficiently decode the sensory signal and map it to lexical representations. Part of having robust phonological representations may be having tolerance to distortions in the signal. Following the work on bilingualism in spoken languages, future work investigating the effect of developmental factors in sensory integration should examine the role of noisy environments and noisy inputs for processing sign language. In the case of temporal distortions, late learners are more vulnerable than early learners of a sign language. This remains untested among late learners of a spoken language. In the case of snow-like visual noise, Mayberry and Fischer (1989) found that early and late learners were equally vulnerable. In spoken languages, Rogers et al. (2006) report that late bilinguals are more vulnerable to noise and reverberations than early bilinguals, who are more vulnerable than monolinguals (Rogers, Lister, Febo, Besing, & Abrams, 2006). It is possible that modality, type of noise, age of acquisition, and bilingualism all interact to produce different sensitivity to noise. Understanding the effect of noise in communication for late learners is particularly relevant in the sign language community where 95% of deaf individuals are born to hearing parents and thus are not exposed to signing since birth. For example, although a noisy channel of communication may allow successful communication between two native signers, it may not be adequate for late signers. Given the findings that early auditory deprivation leads to a reorganization of visual attention to peripheral fields of vision (Bavelier, Dye, & Hauser, 2006), future research should also investigate the effect of peripheral noise on cognitive processing for both early and late language learners.

Understanding how information is chunked through time also has implications for working memory. Advantages in many cognitive functions are associated with larger capacities for holding information, integrating its contents, and comparing it to other sets of information (Baddeley, 2003; Duncan, Seitz, Kolodny, Bor, Herzog, & Ahmed, 2000). Among the many factors that contribute to working memory function, such as attention (Engle, 2002; Conway, Cowan & Bunting, 2001; Kane & Engle, 2003) and inhibition or filtering mechanisms (Vogel, McCollough & Machizawa, 2005), strategies that can expand working memory capacity include rehearsability (Baddeley, 2003; Gathercole & Baddeley, 1993; Wilson & Emmorey, 1997) and “chunking” strategies (Miller, 1956). Findings that using ASL results in shorter short-term memory spans than when using English implicate a difference between auditory versus visually-based representations in taking up resources within working memory (Boutla, Supalla, Newport, & Bavelier, 2004). These results were consistent among deaf signers and English monolinguals as well as within hearing bilinguals who sign. More specifically, it is possible that the sequential nature of units in speech-based input and processing these units in smaller time-scales results in higher spans that are measured serially.

Among the many consequences of having delayed language exposure, individuals who are late learners of a first language have smaller working memory capacities than their early learning counterparts (Mayberry, 1993). Newport (1990) explains that the acquisition process and errors made by late learners reflect a lack of understanding of the internal structure of signs. By acquiring language through the development of memory, early learners may learn the discrete components in sign

even though they are produced simultaneously. The ability to analyze language in a finer-grained way may contribute to a larger working memory capacity that can also be exploited for other cognitive functions.

Knowledge about how linguistic information is chunked when viewed through the eyes may also be relevant to research in reading, especially among deaf children for whom vision is the primary channel for communication. A key challenge in deaf education is improving the rates and achievement-levels of literacy. For hearing children, correspondences in written print can be found with the language that they already speak. For deaf signers learning to read, the process involves learning the grammar of a second language. Increasing evidence for the importance of early language exposure suggests that reading skills depend on strong language foundations (Mayberry, del Giudice, & Lieberman, 2010; Wilbur, 2000). Most of the focus on literacy efforts has been on phonological coding and awareness skills (Wang, Trezek, Luckner, & Paul, 2008; Allen, Clark, del Giudice, & Koo, 2009). Since spoken language in its natural usage operates over short time scales that are not relevant in sign language processing, one might also imagine that this poses extra challenges. In reading eye movement patterns, although hearing children (inexperienced readers) exhibit frequent small saccades, hearing adult readers (skilled readers) exhibit fixations in relatively long time scales (200-250 ms on average) and mean saccade sizes of 7-9 letter spaces (Rayner, 1998). However, it is possible that different early language experiences, especially using distinct time windows for integrating language input, may shape the process of learning to read differently among hearing and deaf children. Ahissar et al. (2001) note that understanding temporal response patterns to

sensory signals is highly relevant in many cognitive functions. In particular, they mention that individuals with “poor successive-signal processing” in audition and vision tend to be poor readers, and that they are more vulnerable than good readers to time compressions of sentences.

By examining perspectives from speech perception and sign language processing, as well as information theory, grammatical theory, development, and neuroscience, and summarizing new experimental work, I have demonstrated how critical temporal dynamics are for language processing and outlined new challenges for future research. Besides making specific contributions to our understanding of temporal integration windows and rates in language processing through cross-linguistic comparisons, this work supports an approach to language processing that takes into account the representations of linguistic units, the information in those units, and the time course over which they unfold. In addition to comparisons in grammar and functional organization in the brain, temporal relationships in on-line language processing demonstrate universal patterns. Key differences also contribute to the model for the architecture of language, where interaction with the sensori-motor interfaces results in unique properties in each modality. The time properties seen in language processing, which impact the grammar of spoken and signed languages, may be best understood from the perspective of the temporal dynamics of underlying neural processes, a claim that motivates future interdisciplinary work in speech and sign language research.

Bibliography

- Abel, S. M. (1972). Discrimination of temporal gaps. *Journal of the Acoustical Society of America*, 52, 519-524.
- Abramatic, J. F., Letellier, P. H., & Nadler, M. (1982). A narrow- band video communication system for the transmission of sign language over ordinary telephone lines. In T. S. Huang (Ed.), *Image sequence processing and dynamic scene analysis* (pp. 314-316). New York: Springer-Verlag.
- Adams, C. (1979). *English Speech Rhythm and the Foreign Learner*. The Hague: Mouton.
- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M.M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings from the National Academy of Sciences*, 98, 13367–13372.
- Akamatsu, C. T. (1982). The acquisition of fingerspelling in pre-school children. Unpublished doctoral dissertation, University of Rochester, Rochester.
- Allen, T. E., Clark, M. D., Del Giudice, A., Koo, D., Lieberman, A., Mayberry, R., & Miller, P. (2009). Phonology and reading: A response to Wang, Trezek, Luckner, and Paul. *American Annals of the Deaf*, 154(4), 338–345.
- Andrews, J., Leigh, I., & Weiner, M. (2004). *Deaf People: Evolving Perspectives From Psychology, Education, And Sociology*. Boston: Allyn & Bacon.
- Arai, T., & Greenberg, S. (1997). The temporal properties of spoken Japanese are similar to those of English. In *Proceedings of Eurospeech: Vol. 2*, 1011–1014.
- Arai, T., & Greenberg, S. (1998). Speech intelligibility in the presence of cross-channel spectral asynchrony. *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 2*, 933–936.
- Archangeli, D. & Pulleyblank, D. (1994). *Grounded Phonology*. Cambridge, MA: MIT Press.
- Aronoff, M., Meir, I., & Sandler, W. (2005). The paradox of sign language morphology. *Language*, 81(2), 301–344.
- Aronoff, M., Meir, I., Padden, C., & Sandler, W. (2004). Morphological universals and the sign language type. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 2004* (pp. 19–39). Kluwer Academic Publishers.

- Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, 36, 189–208.
- Bahan, B., & Supalla, S. (1995). Line segmentation and narrative structure: A study of eyegaze behavior in American Sign Language. In K. Emmorey & J. Reilly (Eds.), *Language, gesture and space* (pp.171-191). Hillsdale: Lawrence Erlbaum Associates.
- Baker, C., & Padden, C. (1978). Focusing on the nonmanual components of American Sign Language. In P. Siple (Ed.), *Understanding language through sign language research* (pp.27-57). New York: Academic Press.
- Baker, M. C. (1996). *The Polysynthesis Parameter*. Oxford University Press.
- Baker, S. A., Idsardi, W. J., Golinkoff, R. M., & Petitto, L. A. (2005). The perception of handshapes in American Sign Language. *Memory & Cognition*, 33(5), 887–904.
- Battison, R. (1978). *Lexical Borrowing In American Sign Language*. Silver Spring, MD: Linstok.
- Bavelier, D., Dye, M. W. G., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends in Cognitive Sciences*, 10(11), 512–518.
- Beasley, D. S., Forman, B. S., & Rintelmann, W. F. (1972). Perception of time-compressed CNC monosyllables by normal listeners. *Journal of Audiology Research*, 12, 71–75.
- de Beuzeville, L., Johnston, T. & Schembri, A. (2009). The use of space with indicating verbs in Australian Sign Language: A corpus-based investigation. *Sign Language & Linguistics* 12(1), 53-82.
- Beck, M. (1998). *Morphology and its interfaces in second language knowledge*. Amsterdam: Benjamins.
- Bellugi, U., & Fischer, S. (1972). A comparison of sign language and spoken language: Rate and grammatical mechanisms. *Cognition*, 1(3), 173-200.
- Best, C. T., Mathur, G., Miranda, K. A., & Lillo-Martin, D. (2010). Effects of sign language experience on categorical perception of dynamic ASL pseudosigns. *Attention, Perception, & Psychophysics*, 72(3), 747-762.
- Bettger, J. G. (1992). *The effects of experience on spatial cognition: Deafness and knowledge of ASL*. Doctoral dissertation, University of Illinois, Urbana-Champaign.
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.

- Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 8, 389–395.
- Bonvillian, J. D., & Folven, R. J. (1993). Sign language acquisition: Developmental aspects. *Psychological Perspectives on Deafness*, 1, 229.
- Bosworth, R.G., Dobkins, K.R., & Wright, C.E. (2010). Analysis of visual properties in American Sign Language. Presentation given at the *10th Theoretical Issues in Sign Language Research Conference*, Purdue University, West Lafayette, IL.
- Boudreault, P., & Mayberry, R. I. (2006). Grammatical processing in American Sign Language: Age of first-language acquisition effects in relation to syntactic structure. *Language and Cognitive Processes*, 21(5), 608–635.
- Boutla, M., Supalla, T., Newport, E.L., & Bavelier, D. (2004). Short- term memory span: Insights from sign language. *Nature Neuroscience*, 7, 997–1002.
- Boyes-Braem, P., & Sutton-Spence, R. (2001). *The Hands are the Head of the Mouth*. Hamburg, Germany: Signum.
- Boyes-Braem, P. (1999). Rhythmic temporal patterns in the signing of deaf early and late learners of Swiss German Sign Language. *Language and Speech*, 42, 177-208.
- de Boysson-Bardies B. (1993). Ontogeny of language-specific syllabic productions. In B. de Boysson-Bardies, S. de Schonen, P.W. Jusczyk, & P. McNeilage (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp.353-363). Dordrecht, Netherlands: Kluwer.
- de Boysson-Bardies B. (1999). *How Language Comes to Children: From Birth to Two Years*. Cambridge, MA: MIT Press.
- Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, 112, 272-284.
- Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, 46(1), 80-97.
- Brentari, D. (1995). Sign language phonology: ASL. In J. Goldsmith (Ed.) *The Handbook of Phonological Theory* (pp.615–639). Oxford, England: Blackwell.
- Brentari, D. (1998). *A Prosodic Model of Sign Language Phonology*. Cambridge, MA:MIT Press.

- Brentari, D. (2002). Modality differences in sign language phonology and morphophonemics. In R. Meier, K. Cormier, & D. Quinto-Pozos (Eds.) *Modality and Structure in Signed and Spoken Languages* (pp.35–64). Oxford University Press.
- Brentari, D. (2006). Effects of language modality on word segmentation: An experimental study of phonological factors in a sign language. *Papers in laboratory phonology*, 8, 155–164.
- Brentari, D., González, C., Seidl, A., & Wilbur, R. (2011). Sensitivity to visual prosodic cues in signers and nonsigners. *Language and Speech*, 54(1), 49-72.
- Brentari, D., Poizner, H., & Kegl, J. (1995). Aphasic and Parkinsonian signing: differences in phonological disruption. *Brain and Language*, 48(1), 69–105.
- Budding, C., Hoopes, R., Mueller, M., & Scarcello, K. (1995). Identification of foreign sign language accents by the deaf. In L. Byers & M. Rose (Eds.) *Gallaudet University Communication Forum, Vol. 4* (pp.1-16). Washington, DC: Gallaudet University Press.
- Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *Journal of Neuroscience*, 29(24), 7869-7876.
- Buus, S., Florentine, M., Scharf, B., & Canévet, G. (1986). Native French listeners' perception of American-English in noise. *Proceedings of Inter-noise*, 86, 895–898.
- Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679), 1926-1929.
- Campbell, R., Woll, B., Benson, P.J., & Wallace, S.B. (1999). Categorical processing of faces in Sign. *Quarterly Journal of Experimental Psychology* (52A), 62–95.
- Canavan, A., & Zipperlen, G. (1996a). *CALLFRIEND American English-Non-Southern Dialect*. Linguistic Data Consortium, Philadelphia.
- Canavan, A., & Zipperlen, G. (1996b). *CALLFRIEND Korean*. Linguistic Data Consortium, Philadelphia.
- Capek, C. M., Grossi, G., Newman, A. J., McBurney, S. L., Corina, D., Roeder, B., & Neville, H. J. (2009). Brain systems mediating semantic and syntactic processing in deaf native signers: Biological invariance and modality specificity. *Proceedings of the National Academy of Sciences*, 106(21), 8784 -8789.
- Casey, D. S., & Emmorey, K. (2009). Co-speech gesture in bimodal bilinguals. *Language and Cognitive Processes*, 24(2), 290–312.

- Chase, C., & Jenner, A.R. (1993). Magnocellular visual deficits affect temporal processing of dyslexics. *Annals of the New York Academy of Sciences* 682, 326-329.
- Cheek, A., Cormier, K., Repp, A., & Meier, R. P. (2001). Prelinguistic gesture predicts mastery and error in the production of early signs. *Language*, 77(2), 292–323.
- Chen Pichler, D. (2006). The development of sign language. In K. de Bot & R.W. Schrauf (Eds.) *Language Development over the Lifespan* (pp. 217-241). New York, NY: Routledge.
- Chong, A., Sankar, L., & Poor, H. V. (2009). Frequency of Occurrence and Information Entropy of American Sign Language. *arXiv:0912.1768*.
- Cicourel, A., & Boese, R. (1972). Sign language acquisition and the teaching of deaf children. *American Annals of the Deaf*, 177(1), 27-33.
- Clark, L. E., & Grosjean, F. (1982). Sign recognition processes in American Sign Language: The effect of context. *Language and Speech*, 25(4), 325-340.
- Conlin, K. E., Mirus, G. R., Mauk, C., & Meier, R. P. (2000). The acquisition of first signs: Place, handshape, and movement. In C. Chamberlain, J. Morford, & R. Mayberry (Eds.), *Language Acquisition by Eye* (pp. 51–69). Mahwah, NJ: Lawrence Erlbaum.
- Conway, A.R.A., Cowan, N., & Bunting, M.F. (2001). The cocktail party phenomenon revisited: The importance of WM capacity. *Psychonomic Bulletin & Review*, 8, 331-335.
- Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, 61(5), 1584–1595.
- Corina, D. P., Bellugi, U., & Reilly, J. (1999). Neuropsychological studies of linguistic and affective facial expressions in deaf signers. *Language and Speech*, 42, 307.
- Corina, D. P., & Hildebrandt, U. C. (2002). Psycholinguistic investigations of phonological structure in ASL. In R. Meier, K. Cormier, & D. Quinto-Pozos (Eds.), *Modality and Structure in Signed and Spoken Languages* (pp.88–111). Cambridge University Press.
- Corina D.P., & Knapp H.P. (2006). Lexical retrieval in American Sign Language production. In L.M. Goldstin, D.H. Whalen, & C.T. Best (Eds.), *Papers in Laboratory Phonology 8: Varieties of Phonological Competence* (pp 213–239). Mouton de Gruyter: Berlin.

Corina, D. P., Poizner, H., Bellugi, U., Feinberg, T., Dowd, D., & O'Grady-Batch, L. (1992). Dissociation between linguistic and nonlinguistic gestural systems: A case for compositionality. *Brain and Language*, 43(3), 414–447.

Coulter, G. R. (1982). On the nature of ASL as a monosyllabic language. Paper presented at the *Annual Meeting of the Linguistic Society for America*, San Diego, CA.

Cowan, N. (1995). Sensory memory and its role in information processing. In G. Karmos, M. Molnár, V. Cspe, I., Czigler, J.E. Desmedt (Eds.), *Perspective of Event-Related Potentials Research, EEG Supplement 40* (pp.21-31). New York: Elsevier.

Crone, N. E., Hao, L., Hart, J., Boatman, D., Lesser, R. P., Irizarry, R., & Gordon, B. (2001). Electrographic gamma activity during word production in spoken and sign language. *Neurology*, 57(11), 2045-2053.

Czigler, I., Winkler, I., Pató, L., Várnagy, A., Weisz, J., & Balázs, L. (2006). Visual temporal window of integration as revealed by the visual mismatch negativity event-related potential to stimulus omissions. *Brain Research*, 1104(1), 129–140.

Davis, S., & McCroskey, R. (1980). Auditory fusion in children. *Child Development*, 51, 75-80.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208(4448), 1174-1176.

Deiber, M. P., Missonnier, P., Bertrand, O., Gold, G., Fazio-Costa, L., Ibañez, V., & Giannakopoulos, P. (2007). Distinction between perceptual and attentional processing in working memory tasks: a study of phase-locked and induced oscillatory brain dynamics. *Journal of Cognitive Neuroscience*, 19(1), 158–172.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(4), 499-533.

DeKeyser, R. M. (2005). What Makes Learning Second-Language Grammar Difficult? A Review of Issues. *Language Learning*, 55(S1), 1–25.

DeMatteo, A. (1977). Visual imagery and visual analogues in American Sign Language. In L. Friedman (Ed.), *On the other hand: New perspectives on American Sign Language* (pp. 109-136). New York: Academic Press.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, 9, 719–721.

Dolata, J. K., Davis, B. L., & MacNeilage, P. F. (2008). Characteristics of the rhythmic organization of vocal babbling: Implications for an amodal linguistic rhythm. *Infant Behavior and Development*, 31(3), 422–431.

- Dudis, P.G. (2011). Response: Some observations on form-meaning. In G. Mathur & D.J. Napoli (Eds.), *Deaf Around the World* (pp. 83-95). Oxford University Press.
- Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A. (2000). A neural basis for general intelligence. *Science*, 289(5478), 457-460.
- Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. *Science*, 209(4461), 1140-1141.
- Elbers, L. (1982). Operating principles in repetitive babbling: a cognitive continuity approach. *Cognition*, 12(1), 45-63.
- Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America*, 66, 651-653.
- Elliott, L. L., & Katz, D. R. (1980). Children's pure-tone detection. *Journal of the Acoustical Society of America*, 67, 343-344.
- Embick, D., & Noyer, R. (2007). Distributed morphology and the syntax/morphology interface. In G. Ramchand & C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces* (pp.289-324). Oxford University Press.
- Emmorey, K. (1995). Processing the dynamic visual-spatial morphology of signed languages. In L.B. Feldman (Ed.), *Morphological Aspects of Language Processing: Crosslinguistic Perspectives* (pp.29-54). Mahwah, NJ: Lawrence Erlbaum Associates.
- Emmorey, K., & Corina, D.P. (1990). Lexical recognition in sign language: Effects of phonetic structure and morphology. *Perceptual and Motor Skills*, 71, 1227-1252.
- Emmorey, K., & Corina D. (1993). Hemispheric specialization for ASL signs and English words: Differences between imageable and abstract forms. *Neuropsychologia* 31(7), 645- 653.
- Emmorey, K., & Kosslyn, S. M. (1996). Enhanced image generation abilities in deaf signers: A right hemisphere effect. *Brain and Cognition*, 32(1), 28-44.
- Emmorey, K., Bellugi, U., Friederici, A., & Horn, P. (1995). Effects of age of acquisition on grammatical sensitivity: Evidence from on-line and off-line tasks. *Applied Psycholinguistics*, 16, 1-23.
- Emmorey, K., Corina, D.P., & Bellugi, U. (1995). Differential processing of topographic and referential functions of space. In K. Emmorey & J. Reilly (Eds.), *Language, Gesture, and Space* (pp. 43-62). Mahwah, NJ: Lawrence Erlbaum Associates.

- Emmorey, K., Klima, E., & Hickok, G. (1998). Mental rotation within linguistic and non-linguistic domains in users of American Sign Language. *Cognition*, 68(3), 221-246.
- Emmorey, K., Kosslyn, S. M., & Bellugi, U. 1993. Visual imagery and visual-spatial language: Enhanced imagery abilities in deaf and hearing ASL signers. *Cognition*, 46, 139-181.
- Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals. *Psychological Science*, 19, 1201–1206.
- Emmorey, K., McCullough, S., & Brentari, D. (2003). Categorical perception in American Sign Language. *Language & Cognitive Processes*, 18, 21-45.
- Emmorey, K., Mehta, S., & Grabowski, T. J. (2007). The neural correlates of sign versus word production. *Neuroimage*, 36(1), 202–208.
- Emmorey, K., Petrich, J., & Gollan, T. (2009). Simultaneous production of American Sign Language and English costs the speaker but benefits the perceiver. In Paper Presented at *the 7th International Symposium on Bilingualism*, Utrecht, The Netherlands.
- Emmorey, K., Thompson, R., & Colvin, R. (2009). Eye gaze during comprehension of American Sign Language by native and beginning signers. *Journal of Deaf Studies and Deaf Education*, 14(2), 237.
- Engel, A.K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, 2, 704–716.
- Engle, R.W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science*, 11(1), 19-23.
- Fallon, M., Trehub, S. E., & Schneider, B. A. (2000). Children's perception of speech in multitalker babble. *Journal of the Acoustical Society of America*, 108, 3023-3029.
- Fénelon, V. S., Casasnovas, B., Simmers, J., & Meyrand, P. (1998). Development of rhythmic pattern generators. *Current Opinion in Neurobiology*, 8(6), 705–709.
- Fenlon, J., Denmark, T., Campbell, R., & Woll, B. (2008). Seeing sentence boundaries. *Sign Language & Linguistics*, 10(2), 177–200.
- Féry, C. & van de Vijver, R. (2004). *The syllable in optimality theory*. Cambridge University Press.

- Figuerola, V. (2009). Representaciones fonológicas en el procesamiento del lenguaje: modalidad de input, restricciones temporales y correlatos neurofisiológicos. Unpublished doctoral dissertation, Pontificia Universidad Católica De Chile.
- Figuerola, V., Howard, M., Idsardi, W., & Poeppel, D. (2009). Rate and local reversal effects on speech comprehension. Abstract in The Neurobiology of Language Conference, October 2009, Chicago, IL.
- Fiorentino, R. (2006). Lexical Structure and the Nature of Linguistic Representations. Doctoral dissertation, University of Maryland, College Park.
- Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive processes*, 22(7), 953–1000.
- Fischer, S. D., Delhorne, L. A., & Reed, C. M. (1999). Effects of rate of presentation on the reception of American Sign Language. *Journal of Speech, Language, and Hearing Research*, 42(3), 568-582.
- Flege, J. E., MacKay, I.R.A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 106(5), 2973-2987.
- Foulds, R. A. (2004). Biomechanical and perceptual constraints on the bandwidth requirements of sign language. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 12(1), 65–72.
- Foulke, E. (1971). The perception of time compressed speech. In D. Horton & J. Jenkins (Eds.), *Perception in language* (pp.79-107). Pittsburgh, PA: Pittsburgh University Press.
- Foulke, W., & Sticht, T. G. (1969). Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin*, 72(1), 50–62.
- French, N. R., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America*, 19, 90-119.
- Friedman, L. (1974). On the physical manifestation of stress in the American Sign Language. Unpublished manuscript, University of Calironifa, Berkeley.
- Fries, P., Nikolic, D., & Singer, W. (2007). The gamma cycle. *Trends in Neurosciences*, 30(7), 309–316.
- Furman, O., Dorfman, N., Hasson, U., Davachi, L., & Dudai, Y. (2007). They saw a movie: Long-term memory for an extended audiovisual narrative. *Learning & Memory*, 14(6), 457 -467.

- Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallet, D., Dahlgren, N. (1993). Darpa, TIMIT, Acoustic-phonetic continuous speech corpus. (NISTIR Publication No. 4930). Washington, DC: US Department of Commerce.
- Gathercole, S.E. and Baddeley, A.D. (1993). *Working Memory and Language*. Erlbaum.
- Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the association of computational linguistics* (pp. 199–206), Philadelphia, PA.
- Ghez, C. & Krakauer, J. (2000). The organization of movement. In E.R. Kandel, J.H. Schwartz, T.M. Jessel (Eds.), *Principles of Neuroscience*. New York: McGraw-Hill.
- Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, 66(1-2), 113–126.
- Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, 56(6), 1127–1134.
- Goldin-Meadow, Susan (1993). When does gesture become language? A study of gesture used as a primary communication system by deaf children of hearing parents. In K.R. Gibson & T. Ingold (Eds.), *Tools, Language and Cognition in Human Evolution* (pp.63–85). Cambridge University Press.
- Green, D.M. (1971). Temporal auditory acuity. *Psychological Review*, 78, 540-551.
- Green, K. P., & Miller, J. L. (1985). On the role of visual rate information in phonetic perception. *Perception & Psychophysics*, 38(3), 269–276.
- Greenberg, S. (1996) Understanding speech understanding: towards a unified theory of speech perception. In W.A. Ainsworth & S. Greenberg (Eds.), *Proceedings of the ESCA Tutorial and Advanced Research Workshop on the Auditory Basis of Speech Perception* (pp.1-8). Keele University, UK.
- Greenberg, S., & Arai, T. (2001). The relation between speech intelligibility and the complex modulation spectrum. In the *7th International Conference on Speech Communication and Technology*, Scandinavia (pp. 473– 476).
- Greenberg, S., Hollenback, J. and Ellis, D. (1996) Insights into spoken language gleaned from phonetic transcription of the switchboard corpus. *Proceedings of the International Conference on Spoken Language Processing*, pp. S24-27.
- Grosjean, F. (1979). A study of timing in a manual and a spoken language: American Sign Language and English. *Journal of Psycholinguistic Research*, 8(4), 379 – 405.

- Grosjean, F. (1981). Sign and word recognition: A first comparison. *Sign Language Studies*, 32, 195-219.
- Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In Proceedings of the North American Association of Computational Linguistics.
- Halle, M. & Stevens, K. N. (1959). Analysis by synthesis. In W. Wathen-Dunn & L.E. Woods (Eds.) *Proc. Seminar on Speech Compression and Processing, Vol. 2*, paper D7.
- Halle, M. & Stevens, K. N. (1962). Speech recognition: a model and program for research. Reprinted in Halle, 2002.
- Halle, M. (2002). From memory to speech and back: papers on phonetics and phonology 1954–2002. Berlin, Germany: Mouton de Gruyter.
- Heiman, G. W., & Tweney, R. D. (1981). Intelligibility and comprehension of time compressed sign language narratives. *Journal of Psycholinguistic Research*, 10(1), 3–15.
- Henry, W. G. (1966). Recognition of time compressed speech as a function of word length and frequency of usage. Unpublished doctoral dissertation, Indiana University.
- Hickok, G., Bellugi, U., & Klima, E. S. (1998). The neural organization of language: Evidence from sign language aphasia. *Trends in Cognitive Sciences*, 2(4), 129–136.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Holcombe, A. O. (2009). Seeing slow and seeing fast: two limits on perception. *Trends in Cognitive Sciences*, 13(5), 216–221.
- van der Hulst, H. (1993). Units in the analysis of signs. *Phonology*, 10, 109-241.
- Hwang, S.-O., Monahan, P. J., & Idsardi, W. J. (2010). Underspecification and asymmetries in voicing perception. *Phonology*, 27(2), 205–224.
- Hyde, M.B., & Power, D.J. (1991). Teachers' use of simultaneous communication: Effects on the signed and spoken components. *American Annals of the Deaf*, 136(5), 381-387.
- Jackson, C. (1989). Language acquisition in two modalities: The role of nonlinguistic cues in linguistic mastery. *Sign Language Studies*, 62, 1-21.

- Jaeger, T.F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62.
- Jantunen, T. (2010). On the role of transitions in SL or: What's wrong with the sign? Presentation given at the *10th Theoretical Issues in Sign Language Research Conference*, Purdue University, West Lafayette, IL.
- Jantunen, T. & Takkinen, R. (2010). Syllable structure in sign language phonology. In D. Brentari (Ed.), *Sign Languages* (pp.312-331). Cambridge University Press.
- Jensen, J. K., Neff, D. L., & Callaghan, B. P. (1987). Frequency, intensity, and duration discrimination in young children. *Asha*, 29, 88.
- Jensen, O., & Lisman, J. E. (2005). Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. *Trends in Neurosciences*, 28(2), 67–72.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Attention, Perception, & Psychophysics*, 14(2), 201–211.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1), 60–99.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207.
- Kabak, B., & Idsardi, W. J. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech*, 50(1), 23-52.
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132(1), 47-70.
- Kantor, R. (1978). Identifying native and second language signers. *Communication and Cognition*, 11, 39-55.
- Kimura, M., Schröger, E., Czigler, I., & Ohira, H. (2010). Human visual system automatically encodes sequential regularities of discrete events. *Journal of Cognitive Neuroscience*, 22(6), 1124–1139.
- Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 18, 686–706.

- Klein, W., & Dittmar, N. (1979). Developing grammars: The acquisition of German syntax by foreign workers (Vol. 1). Berlin: Springer.
- Klima, E. S., & Bellugi, U. (1979). *The signs of language*. Cambridge, MA: Harvard University Press.
- Klima, E. S., Tzeng, O. J. L., Bellugi, U., Corina, D., & Bettger, J. G. (1996). *From sign to script: effects of linguistic experience on perceptual categorization* (Tech. Rep. No. INC-9604). Institute for Neural Computation, University of California, San Diego.
- Kohlrausch, A., Püschel, D., & Alpei, H. (1992). Temporal resolution and modulation analysis in models of the auditory system. In M.E.H. Schouten (Ed.) *The Auditory Processing of Speech: From Sounds to Words* (pp.85–98). Berlin/New York: Mouton de Gruyter.
- Korte, A. (1915) Kinematoskopische Untersuchungen. *Zeitschrift fuer Psychologie*, 72, 194-296.
- Krentz, U. C., & Corina, D. P. (2008). Preference for language in early infancy: The human language bias is not speech specific. *Developmental Science*, 11(1), 1–9.
- Kroll, J. F., Bobb, S. C., & Wodnieka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, 9, 119–135.
- Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, 100(15), 9096-9101.
- Kurtzrock, G.H. (1957). The effects of time and frequency distortion upon word intelligibility. *Speech Monographs*, 24, 94.
- Kushalnagar, P., Hannay, H. J., & Hernandez, A. E. (2010). Bilingualism and Attention: A Study of Balanced and Unbalanced Bilingual Deaf Users of American Sign Language and English. *Journal of Deaf Studies and Deaf Education*, 15(3), 263-273.
- Ladefoged, P. (2005). Vowels and consonants: An introduction to the sounds of languages (Vol. 1). Wiley-Blackwell Publishing.
- Lahiri, A. & Reetz, H.(2002). Underspecified recognition. In C. Gussenhoven, N. Werner, & T. Rietveld (Eds.) *Laboratory Phonology 7* (pp.637-676). Berlin: Mouton de Gruyter.

- Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology*, 94(3), 1904-1911.
- van de Laar, V., Kleijn, W. B., & Deprettere, E. (1997). Perceptual entropy rate estimates for the phonemes of American English. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 3, 1719-1722.
- Lehtonen, M., Monahan, P. J., & Poeppel, D. (2011). Evidence for Early Morphological Decomposition: Combining Masked Priming with Magnetoencephalography. *Journal of Cognitive Neuroscience*, (Early Access), 1-14.
- Levitt, A., & Wang, Q. (1991). Evidence for language-specific rhythmic influences in the reduplicative babbling of French- and English-learning infants. *Language and Speech*, 34(3), 235-249.
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.) *Advances in neural information processing systems (NIPS)*, Vol. 19 (pp. 849-856). Cambridge, MA: MIT Press.
- Lieberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4(5), 187-196.
- Lieberman, A.M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.
- Liddell, S.K. (1978). Non-manual signs and relative clauses in American Sign Language. In P. Siple (Ed.) *Understanding language through sign language research* (pp.59-90). New York: Academic Press.
- Liddell, S.K. (1984). THINK and BELIEVE: Sequentiality in American Sign Language. *Language*, 60, 372-392.
- Liddell, S.K. (1990). Structures for representing handshape and local movement at the phonemic level. In S. Fischer & P. Siple (Eds.) *Theoretical Issues in Sign Language Research* (pp.37-65). Chicago:Chicago University Press.
- Liddell, S. K. (2000). Blended spaces and deixis in sign language discourse. In D. McNeil (Ed.), *Language and gesture* (pp.331-357). Cambridge University Press.
- Liddell, S.K. (2003). Sources of meaning in ASL classifier predicates. In K. Emmorey (Ed.) *Perspectives on classifier constructions in sign language* (pp.199-220). Mahwah, NJ: Lawrence Erlbaum Associates.

- Liddell, S.K., & Johnson, R.E. (1986). American Sign Language compound formation processes, lexicalization, and phonological remnants. *Natural Language and Linguistic Theory*, 8, 445-513.
- Liddell, S.K. & Johnson, R. (1989). American Sign Language: the phonological base. *Sign Language Studies*, 64, 195-277.
- Lillo-Martin, D. (1991). Universal Grammar and American Sign Language: Setting the Null Argument Parameters. *Studies in Theoretical Psycholinguistics*. Dordrecht: Kluwer.
- Lillo-Martin, D. (1999). Modality effects and modularity in language acquisition: the acquisition of American Sign Language. In W.C. Ritchie & T. K. Bhatia (Eds.) *Handbook of Language Acquisition* (pp.531-567). San Diego, CA: Academic Press.
- Lisker, L. (1975). Is it VOT or a first-formant transition detector. *Journal of the Acoustical Society of America*, 57(6), 1547–1551.
- Lisker, L., Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word* (20) 384–422.
- Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Press.
- Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6), 1001-1010.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21(4), 499–546.
- MacNeilage, P. F., & Davis, B. L. (2001). Motor mechanisms in speech ontogeny: phylogenetic, neurobiological and linguistic implications. *Current Opinion in Neurobiology*, 11(6), 696–700.
- MacWhinney, B. (2006). Emergent fossilization. In Z. Han & T. Odlin (Eds.), *Studies of fossilization in second language acquisition* (pp. 134-156). Clevedon, UK: Multilingual Matters.
- Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, 19(23), 1994–1997.
- Manin, D. (2006). Experiments on predictability of word in context and information rate in natural language. *Journal of Information Processes*, 6(3), 229–236.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition*, 6(2), 97–116.

Marmor, G. S. & Petitto, L. A. (1979). Simultaneous communication in the classroom: How well is English grammar represented? *Sign Language Studies*, 3, 99-136.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Masataka, N. (1992). Motherese in a signed language. *Infant Behavior and Development*, 15(4), 453–460.

Masataka, N. (2003). *The onset of language*. Cambridge University Press.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. (1996). Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of America*, 100, 1777–1786.

Mathur, G., & Rathmann, C. (2011). Two types of nonconcatenative morphology in signed languages. In G. Mathur & D.J. Napoli (Eds.), *Deaf Around the World* (pp.54-82). Oxford University Press.

Mayberry, R. I. (1993). First-Language acquisition after childhood differs from second-language acquisition: The case of American Sign Language. *Journal of Speech and Hearing Research*, 36, 51-68.

Mayberry, R. I. (2007). When timing is everything: Age of first-language acquisition effects on second-language learning. *Applied Psycholinguistics*, 28(3), 537–549.

Mayberry, R. I., & Eichen, E. (1991). The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition. *Journal of Memory and Language*, 30, 486-512.

Mayberry, R. I., & Fischer, S. D. (1989). Looking through phonological shape to lexical meaning: The bottleneck of non-native sign language processing. *Memory & Cognition*, 17(6), 740–754.

Mayberry, R. I., & Lock, E. (2003). Age constraints on first versus second language acquisition: Evidence for linguistic plasticity and epigenesis. *Brain and Language*, 87, 369-383.

Mayberry, R. I., del Giudice, A. A., & Lieberman, A. M. (2011). Reading achievement in relation to phonological coding and awareness in deaf readers: A meta-analysis. *Journal of Deaf Studies and Deaf Education*, 16(2), 164-188.

- Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, 40(3), 686-693.
- McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America*, 77, 678-685.
- McGurk, H., & McDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-747.
- Meador, D., Flege, J. E., & Mackay I. R. A. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition*, 3, 55-67.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2), 143-178.
- Meier, R. P. (1987). Elicited imitation of verb agreement in American Sign Language: Iconically or morphologically determined? *Journal of Memory and Language*, 26(3), 362-376.
- Meier, R.P. (2002). Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech. In R.P. Meier, K. Cormier, & D., Quinto-Pozos (Eds.) *Modality and structure in signed and spoken languages* (pp.1-25). Cambridge University Press.
- Meier, R. P. (2006). The form of early signs: Explaining signing children's articulatory development. In M. Marschark, B. Schick, & P. Spencer (Eds.), *Advances in sign language development by deaf children* (pp.202-230). Oxford University Press.
- Meier, R.P. (2008). Channeling language: Review of Wendy Sandler & Diane Lillo-Martin (2006). *Natural Language and Linguistic Theory*, 26, 451-466.
- Meier, R.P., Cormier, K., Quinto-Pozos, D. (2002). *Modality and structure in signed and spoken languages*. Cambridge University Press.
- Meier, R. P., & Newport, E. L. (1990). Out of the hands of babes: On a possible sign advantage in language acquisition. *Language*, 66, 1-23.
- Meier, R. P., & Willerman, R. (1995). Prelinguistic gesture in deaf and hearing infants. In K. Emmorey & J. Reilly (Eds.), *Language, gesture and space* (pp.391-409). Hillsdale: Lawrence Erlbaum Associates.

- Merigan, W. H., & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience*, 16(1), 369–402.
- Miller, G. A. (1951). *Language and communication*. New York: McGraw-Hill.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 2(3), 217–228.
- Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of speech. *Journal of the Acoustical Society of America*, 22, 167–173.
- Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of experimental Psychology*, 41(5), 329-335.
- Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Attention, Perception, & Psychophysics*, 25(6), 457–465.
- Mills, J. H. (1975). Noise and children: A review of literature. *Journal of the Acoustical Society of America*, 58, 767-779.
- Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, 27, 272-277.
- Mirus, G., Rathmann, C., & Meier, R. (2001). Proximalization and distalization of sign movement in adult learners. In V. Dively, M. Metzger, S. Taub, & A.M. Baer (Eds). *Signed languages: Discoveries from international research* (pp.103-119). Washington, DC: Gallaudet University Press.
- Mitchell, R.E. and Karchmer, M.A. (2002). Chasing the mythical ten percent: parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies*, 4, 128–163.
- Morford, J. P., & MacFarlane, J. (2003). Frequency Characteristics of American Sign Language. *Sign Language Studies*, 3(2), 213–25.
- Morford, J. P., Wilkinson, E., Villwock, A., Piñar, P., & Kroll, J. F. (2010). When deaf signers read English: Do written words activate their sign translations? *Cognition*, 118(2), 286-292.
- Morford, J., & Mayberry, R. (2000). A reexamination of “Early Exposure” and its implications for language acquisition by eye. In C. Chamberlain, J. Morford, & R.

Mayberry (Eds.), *Language acquisition by eye* (pp.111-128). Mahwah, NJ: Lawrence Erlbaum.

Müller, M. M., Gruber, T. & Keil, A. (2001). Modulation of induced gamma band activity in the human EEG by attention and visual information processing. *International Journal of Psychophysiology*, 38, 283–299.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Attention, Perception, & Psychophysics*, 58(3), 351–362.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility. *Psychological Science*, 15(2), 133-137.

Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Nabelek, A. (1988). Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss. *Journal of the Acoustical Society of America*, 84, 476–484.

Napoli, D. J., & Sutton-Spence, R. (2010). Limitations on simultaneity in sign language. *Language*, 86(3), 647-662.

Nespor, M., & Sandler, W. (1999). Prosody in Israeli sign language. *Language and Speech*, 42, 143–176.

Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex*, 2(3), 244-258.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science*, 14, 11-28.

Newport, E. L., & Meier, R. P. (1985). *The acquisition of American Sign Language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95, 1085-1099.

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *Journal of the Acoustical Society of America*, 87, 2705-2715.

- Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, 59(2), 441–449.
- Oller, K., Wieman, L., Dole, W., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, 3, 1-12.
- Olsho, L. W., Schoon, C., Sakai, R., Turpin, R., & Sperduto, V. (1982). Auditory frequency discrimination in infancy. *Developmental Psychology*, 18(5), 721–726.
- Ostry, D. J., & Munhall, K. G. (1985). Control of rate and duration of speech movements. *Journal of the Acoustical Society of America*, 77(2), 640–648.
- Oyama, S. (1976). A sensitive period in the acquisition of a non-native phonological system. *Journal of Psycholinguistic Research*, 5, 261-285.
- Padden, C.A. (1988). Interaction of morphology and syntax in American Sign Language. New York, NY: Garland.
- Padden, C.A. (1991). The acquisition of fingerspelling by deaf children. In P. Siple & S. Fischer (Eds.), *Theoretical issues in sign language research* (pp.191-210). Chicago, IL: University of Chicago Press.
- Padden, C. A. (2000). Simultaneous interpreting across modalities. *Interpreting*, 5(2), 171–187.
- Padden, C.A., & LeMaster, B. (1985). An alphabet on hand: The acquisition of fingerspelling in deaf children. *Sign Language Studies*, 47, 161-172.
- Padden, C. A., & Perlmutter, D. M. (1987). American Sign Language and the architecture of phonological theory. *Natural Language & Linguistic Theory*, 5(3), 335–375.
- Pandey, P. C., Kunov, H., & Abel, S. M. (1986). Disruptive effects of auditory signal delay on speech perception with lipreading. *Journal of Auditory Research*, 26(1), 27-41.
- Parasnis, I., Samar, V. J., Bettger, J. G., & Sathe, K. (1996). Does Deafness Lead to Enhancement of Visual Spatial Cognition in Children? *Journal of Deaf Studies and Deaf Education*, 1(2), 145-152.
- Pearson, D. E. (1981). Visual communication systems for the deaf. *IEEE Transactions on Communications*, 29, 1986-1992.
- Perlmutter, D. M. (1990). On the segmental representation of transitional and bidirectional movements in ASL phonology. In S. Fischer & P. Siple (Eds.)

Theoretical Issues in Sign Language Research (pp.67-80). Chicago: Chicago University Press.

Perlmutter, D. M. (1992). Sonority and syllable structure in American Sign Language. *Linguistic Inquiry*, 23(3), 407–442.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47(3), 329–342.

Petitto, L. A. (1987). On the autonomy of language and gesture: Evidence from the acquisition of personal pronouns in American sign language. *Cognition*, 27(1), 1–52.

Petitto, L. A., & Marentette, P. F. (1991). Babbling in the manual mode: Evidence for the ontogeny of language. *Science*, 251, 1493–1496.

Petitto, L. A., Katerelos, M., Levy, B. G., Gauna, K., Tétreault, K., & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *Journal of Child Language*, 28(2), 453–496.

Petitto, L. A., Zatorre, R. J., Gauna, K., Nikelski, E. J., Dostie, D., & Evans, A. C. (2000). Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. *Proceedings of the National Academy of Sciences*, 97(25), 13961–13966.

Petitto, L., Holowka, S., Sergio, L., & Ostry, D. (2001). Language rhythms in baby hand movements. *Nature*, 413, 35.

Petitto, L., Holowka, S., Sergio, L., Levy, B., & Ostry, D. (2004). Baby hands that move to the rhythm of language: Hearing babies acquiring sign languages babble silently on the hands. *Cognition*, 93, 43–73.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, 28(1), 96–103.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication*, 41, 245–255.

Poeppel, D., Idsardi, W. J., van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society London B*, 363, 1071–86.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences*, 1(2), 56–61.

Portnoff, M. (1981). Time-scale modification of speech based on short-time Fourier analysis. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29(3), 374–390.

Pyers, J. E., & Emmorey, K. (2008). The Face of Bimodal Bilingualism. *Psychological Science*, 19(6), 531-535.

Quadros, R.M., Lillo-Martin, D., & Chen Pichler, D. (2010). Two languages but one computation: Code-blending in bimodal bilingual development. Presentation given at the *10th Theoretical Issues in Sign Language Research Conference*, Purdue University, West Lafayette, IL.

Quinto-Pozos, D. (2010). Rates of fingerspelling in American Sign Language. Poster given at the *10th Theoretical Issues in Sign Language Research Conference*, Purdue University, West Lafayette, IL.

R Development Core Team. (2005). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at <http://www.r-project.org>.

Ramsey, C. (1989). Language planning in deaf education. In C. Lucas (Ed.), *The sociolinguistics of the deaf community* (pp.123-146). San Diego, CA: Academic Press.

Rathmann, C., & Mathur, G., (2010). Two types of nonconcatenative morphology in signed languages. Presentation given at the *10th Theoretical Issues in Sign Language Research Conference*, Purdue University, West Lafayette, IL.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.

Reed, C. M., & Durlach, N. I. (1998). Note on information transfer rates in human communication. *Presence*, 7(5), 509–518.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, 27(3), 465-485.

Rosen, R. (2004). Beginning L2 production errors in ASL lexical phonology. *Sign Language Studies*, 7, 31-61.

Rosen, S., 1992. Temporal information in speech: acoustic, auditory, and linguistic aspects. *Philosophical Transactions of the Royal Society B*, 336, 367–373.

- Rosenzweig, M. R., & Postman, L. (1957). Intelligibility as a function of frequency of usage. *Journal of Experimental Psychology*, 54(6), 412-422.
- Ross, J. R. (1967). Constraints on Variables in Syntax. Doctoral dissertation, MIT.
- Saberi, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature*, 398, 6730.
- Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, 19(4), 499–526.
- Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge University Press.
- Schroeder, C. E., & Lakatos, P. (2009a). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1), 9–18.
- Schroeder, C. E., & Lakatos, P. (2009b). The gamma oscillation: master or slave? *Brain Topography*, 22(1), 24–26.
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, 12(3), 106–113.
- Selkirk, E. O. (1986). Phonology and syntax: The relation between sound and structure. Cambridge, MA: MIT Press.
- Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science*, 12(4), 323–328.
- Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, 305(5691), 1779-1782.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(10), 623–656.
- Shannon, C.E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, 30(1), 50–64.
- Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303–304.
- Shiffrar, M., & Freyd, J. J. (1990). Apparent motion of the human body. *Psychological Science*, 1(4), 257-264.

- Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 18(1), 555–586.
- Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416(6876), 87-90.
- Stevens, K. N. & Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. In W. Wathen-Dunn (Ed.) *Models for the perception of speech and visual form* (pp.88-102). Cambridge, MA: MIT Press.
- Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111(4), 1872-1891.
- Stilp, C. E., Kiefte, M., Alexander, J. M., & Kluender, K. R. (2010). Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. *Journal of the Acoustical Society of America*, 128(4), 2112-2126.
- Stokoe, W. C. (1960). Sign language structure: An outline of the visual communication systems of the American Deaf. *Studies in linguistics, occasional papers* 8. Silver Spring, MD: Linstok Press.
- Sperling, G., Landy, M. S., Cohen, Y., & Pavel, M. (1985). Intelligible encoding of ASL image sequences at extremely low information rates. *Computer Vision, Graphics, and Image Processing*, 31, 335-391.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.
- Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7(5), 1074–1095.
- Supalla, S. J. (1991). Manually Coded English: The modality question in signed language development. In P. Siple & S.D. Fischer (Eds.), *Theoretical issues in sign language research* (pp. 85-109). Chicago: University of Chicago Press.
- Supalla, S. J., & McKee, C. (2002). The role of Manually Coded English in language development of deaf children. In R. Meier, K. Cormier, & D. Quinto-Pozos (Eds.) *Modality and Structure in Signed and Spoken Languages* (pp.143-165). Oxford University Press.
- Supalla, T. R. (1982). Structure and acquisition of verbs of motion and location in American Sign Language. Doctoral dissertation, University of California, San Diego.

- Supalla, T.R., & Newport, E. (1978). How many seats in a chair? The derivation of nouns and verbs in American Sign Language. In P. Siple (Ed.) *Understanding language through sign language research* (pp.181-214). New York: Academic Press.
- Tallal, P., Miller, S. and Fitch, R.H. (1993) Neurobiological basis of speech: a case for the pre-eminence of temporal processing. *Annals of the New York Academy of Sciences*, 682, 27–47.
- Tartter, V. C., & Knowlton, K. C. (1981). Perception of sign language from an array of 27 moving spots. *Nature*, 298, 676-678.
- Theunissen, F., & Miller, J. P. (1995). Temporal encoding in nervous systems: a rigorous definition. *Journal of Computational Neuroscience*, 2(2), 149–162.
- Tweney, R. D., Heiman, G. W., & Hoemann, H. W. (1977). Psychological processing of sign language: Effects of visual disruption on sign intelligibility. *Journal of Experimental Psychology: General*, 106(3), 255-268.
- Van Rullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, 7(5), 207–213.
- Viemeister, N. F., & Wakefield, G. H. (1991). Temporal integration and multiple looks. *Journal of the Acoustical Society of America*, 90, 858-865.
- Vihman, M. M. (1996). Phonological development: The origins of language in the child. Wiley-Blackwell.
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, 438(7067), 500–503.
- Wallace, A. B., & Blumstein, S. E. (2009). Temporal integration in vowel perception. *Journal of the Acoustical Society of America*, 125, 1704-1711.
- Wang, Y., Trezek, B. J., Luckner, J., & Paul, P. V. (2008). The role of phonology and phonologically related skills in reading instruction for students who are deaf or hard of hearing. *American Annals of the Deaf*, 153(4), 396–407.
- Wanner, E. & Gleitman, L.R. (1982). *Language acquisition: the state of the art*. Cambridge University Press.
- Warren, R.M. (1999). *Auditory Perception*. Cambridge University Press.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598–607.

- Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 43(2), 230-246.
- 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(1), 49-63.
- Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, 46(3), 233-251.
- Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, 349-355.
- Wertheimer, M. (1912). Experimentelle studien uber das Schen von Beuegung. *Zeitschrift fuer Psychologie* (61) 161-265.
- Wightman, F., Allen, P., Dolan, T., Kistler, D., & Jamieson, D. (1989). Temporal resolution in children. *Child Development*, 611-624.
- Wilbur, R. B. & Nolen, S. B. (1986) Duration of syllables in ASL. *Language & Speech*, 29(3), 263- 280.
- Wilbur, R. B. (1999). Stress in ASL: Empirical evidence and linguistic issues. *Language & Speech*, 42, 229-250.
- Wilbur, R. B., & Allen, G. D. (1991). Perceptual evidence against internal structure in American Sign Language syllables. *Language and speech*, 34(1), 27-46.
- Wilbur, R. B., & Petersen, L. (1998). Modality interactions of speech and signing in simultaneous communication. *Journal of Speech, Language, and Hearing Research*, 41(1), 200-212.
- Wilbur, R. B., & Zelaznik, H. N. (1997). Kinematic correlates of stress and position in ASL. Paper presented at *The Annual Meeting of the Linguistic Society of America*, Chicago, IL.
- Wilbur, R.B. (1986). Why syllables? An examination of what the notion means for ASL research. Oral paper presented at the *Conference on Theoretical Issues in Sign Language Research*, Rochester, NY.
- Wilbur, R.B. (2000). Phonological and prosodic layering of non-manuals in American Sign Language. In K. Emmorey & H. Lane (Eds.), *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima* (pp.215-243). Mahwah, NJ: Lawrence Erlbaum Associates.

- Wilbur, R.B. (2009). Effects of varying rate of signing on ASL manual signs and nonmanual markers. *Language and Speech*, 52, 245-285.
- Wilcox, S. (1992). *The phonetics of fingerspelling*. Philadelphia: John Benjamins.
- Wilson, M. (2001). The impact of sign language expertise on perceived path of apparent motion. In M.D. Clark & M. Marschark (Eds.), *Context, Cognition, and Deafness* (pp.38-48). Washington, DC: Gallaudet University Press.
- Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172), 676-679.
- Wilson, M., & Emmorey, K. (1997). A visuospatial “phonological loop” in working memory: Evidence from American Sign Language. *Memory and Cognition*, 25, 313-320.
- Wilson, M., Bettger, J. G., Niculae, I., & Klima, E. S. (1997). Modality of language shapes working memory: Evidence from digit span and spatial span in ASL signers. *Journal of Deaf Studies and Deaf Education*, 2, 150-160.
- Wingfield, A., Lombardi, L., & Sokol, S. (1984). Prosodic features and the intelligibility of accelerated speech: Syntactic versus periodic segmentation. *Journal of Speech and Hearing Research*, 27(1), 128-134.
- Woll, B. (2001). The sign that dares to speak its name: Echo phonology in British Sign Language. In P. Boyes Braem & R. Sutton-Spence (Eds.), *The Hands are the Head of the Mouth* (pp.87-90). Hamburg, Germany: Signum.
- Yabe, H., Tervaniemi, M., Sinkkonen, J., Huottilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, 35(5), 615–619.
- Yost, W. A., Popper, A. N., & Fay, R. R. (1993). *Human psychophysics*. Springer.
- Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10(7), 301–308.
- Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Attention, Perception, & Psychophysics*, 67(3), 531–544.
- Zeng, F.G., Nie, K., Stickney, G.S., Kong, Y.Y., Vongphoe, M., Bhargave, A., Wei, C.G., and Cao, K. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings from the National Academy of Sciences*, 102, 2293–2298.
- Zipf, G. K. (1935). *The psycho-biology of language*. Oxford, England: Houghton, Mifflin.