August 2015

Effect of Contextual Speech Rate on Speech Comprehension

David Michael Weintraub
University of Nevada, Las Vegas, weintra2@unlv.nevada.edu

EFFECT OF CONTEXTUAL SPEECH RATE ON SPEECH COMPREHENSION

By

David M. Weintraub

Bachelor of Arts - Psychology
University of Nevada, Las Vegas
2009

Master of Arts - Psychology
University of Nevada, Las Vegas
2012

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy - Psychology

Department of Psychology
College of Liberal Arts
The Graduate College

University of Nevada, Las Vegas
August 2015

UNLV | GRADUATE COLLEGE

Dissertation Approval

The Graduate College
The University of Nevada, Las Vegas

May 12, 2015

This dissertation prepared by

David M. Weintraub

entitled

Effect of Contextual Speech Rate on Speech Comprehension

is approved in partial fulfillment of the requirements for the degree of

Doctor of Philosophy - Psychology
Department of Psychology

Joel Snyder, Ph.D.
Examination Committee Chair

Erin Hannon, Ph.D.
Examination Committee Member

Jefferson Kinney, Ph.D.
Examination Committee Member

Steve McCafferty, Ph.D.
Graduate College Faculty Representative

Kathryn Hausbeck Korgan, Ph.D.
Graduate College Interim Dean

ABSTRACT

EFFECT OF CONTEXTUAL SPEECH RATE ON SPEECH COMPREHENSION

By

David M. Weintraub

Joel S. Snyder, Ph.D., Advisory Committee Chair
Associate Professor of Psychology
University of Nevada, Las Vegas

Despite an extensive history of study, the effects of phonetic context are only known to affect small units of speech (e.g., formant transitions, function words). Critical aspects of speech perception, however, occur at larger scales. The series of experiments reported here investigated the effects of contextual speech rate on perception of a large unit of speech, namely sentences. In particular, there was an effect of relative rate on sentence comprehension – the rate of a sentence compared to the average rate of all other sentences within the same conversation-length period of speech – such that relatively slow sentences were better comprehended than relatively fast sentences (Experiment 1); however, differences in perceptual learning between the relatively slow and the relatively fast rates accounted for the effect of relative rate (Experiment 2). The results of these studies, therefore, do not support an effect of contextual speech rate on sentence comprehension. Finally, based on the results of a modified version of Experiment 1 in which context sentences were replaced with non-speech sounds (i.e., 1-channel noise vocoded speech), exposure to temporal information was not sufficient for generalization of perceptual learning (Experiment 3). These experiments are a novel investigation into both the effects of phonetic context on sentence comprehension, and the efficacy of non-speech sounds on generalization of perceptual learning

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

It is well known that temporal information in speech is critical for speech perception. Our ability to comprehend noise-vocoded speech (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), comprised of severely degraded spectral information and preserved temporal information, is a compelling illustration of this point. In contrast, time-compressed speech, comprised of preserved spectral information and severely degraded temporal information, is difficult to comprehend (Dupoux & Green, 1997; Ghitza & Greenberg, 2009).

Despite their importance, however, temporal cues often vary within and between speakers in a way that precludes precise one-to-one mappings with their intended phonological representations (Gay, 1978). Auditory and/or speech-related perceptual systems overcome this temporal variability by adjusting to the distribution of temporal information within their current context in order to process (and perceive) incoming temporal information in a context-dependent manner (Repp, 1982). In this way, speech is perceived relative to its surrounding context. For example, exposure to slow speech rates causes subsequent temporal information to be perceived as relatively fast and, in contrast, exposure to fast speech rates causes the same temporal information to be perceived as relatively slow (Ainsworth, 1974; Summerfield, 1981). Such contrastive context effects on perception are, in fact, a general phenomenon observed across all sensory modalities (Treisman & Williams, 1984; Warren, 1985). Consider the rate of an initial formant transition that distinguishes between [wa] and [ba] – such that syllables with faster initial formant transitions are more likely to be perceived as [ba]. Syllables with an ambiguous-rate initial formant transition, such that they are just as likely to be perceived as [wa] or [ba] when heard in isolation, are more likely to be perceived as [ba] when they are embedded within slow speech and non-speech contexts (Miller & Liberman, 1979; Pisoni, Carrell, & Gans, 1983; Wade & Holt, 2005). Here, exposure to slow contexts increase the perceived rate of the ambiguous initial formant transition and, consequently, increase the likelihood syllables are perceived as [ba]. Similar effects of contextual speech rate on the perception of temporal information in speech are abundant and are shown to influence perception of cues such as voice onset time (Sawusch & Newman, 2000; Summerfield, 1981), vowel duration (Ainsworth, 1974), and gemination (Pickett & Decker, 1960). These effects are not isolated to the perception of such small units of speech, but also influence the perception of larger units of speech. Function words (e.g., *are*, *or*), for example, are less likely to be heard when they are preceded by slow speech contexts (Baese-Berk et al., 2014; Dilley & Pitt, 2010). For example, during the phrase "*Deena doesn't have any leisure or time*", listeners are less likely to report hearing the word *or* when the rest of the sentence is spoken at a slow rate, compared to when the same *or* is embedded within the same sentence spoken at a fast rate. Here, exposure to slow contexts decrease the perceived duration of the segment containing the word *or* (i.e., "-*sure or t*-"). Consequently, shorter perceived durations of the segment "-*sure or t*-" decrease the likelihood the segment contains an additional syllable corresponding to *or*. Similar effects occur when rhythm, as opposed to rate, suggest the presence of an additional syllable (Morrill, Dilley, McAuley, & Pitt, 2014).

The effects of contextual speech rate occur on multiple timescales. The effects described above were studied at short timescales such that the affected speech was influenced by the rate of speech and non-speech that was either immediately adjacent or within the same sequence. At long timescales, in contrast, the perception of function words is also sensitive to the average rate of a conversation-length period of speech (Baese-Berk et al., 2014). In particular, as reviewed above, slow sentences decrease the likelihood a function word is heard; however, the size of this effect also depends on the relative rate of the sentence – rate of the sentence compared to the average rate of all other sentences within the same block of speech. The effect of relative rate does not influence perception of function words immediately and, instead, requires several minutes of exposure.

A compelling theory, which has recently developed much interest, argues that speech comprehension relies on entrainment of neural activity within auditory cortex to low-frequency amplitude changes in speech, a cue for rate (Ding & Simon, 2014; Giraud & Poeppel, 2012; Peelle & Davis, 2012; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008). In particular, high-excitability peaks of neural activity entrain to high-amplitude portions of encoded speech, and entrainment to earlier parts of speech persists into subsequent parts (Lakatos et al., 2013; Peelle & Davis, 2012; Schroeder et al., 2008). Neural entrainment may explain the effects of contextual speech rate. Consider the example phrase, "*Deena doesn't have any leisure or time*", in which the context (i.e., "*Deena doesn't have any lei…*") is spoken at a slow rate and the target (i.e., "…*sure or ti*…") is spoken at a relatively fast rate. Neural activity entrains at a correspondingly slow rate during the context, and continues at a slow rate during the relatively fast target. The result is fewer high-amplitude peaks of neural activity during the target, compared to if the context is spoken at faster rates, which decreases the likelihood listeners hear the word *or* (Baese-Berk et al., 2014; Dilley & Pitt, 2010).

Effects of phonetic context on speech perception are not limited to speech rate. Spectral context, in particular, is well known to influence speech perception. Consider the frequency of a second formant (f_2; i.e., second amplitude peak in a speech spectrum) that distinguishes between [u] and [e] – such that vowels with higher f_2 frequencies are more likely to be heard as [e]. Vowels with an ambiguous f_2 frequency, such that they are just as likely to be perceived as [u] or [e] when heard in isolation, are more likely to be perceived as [e] when they are embedded within low-frequency speech and non-speech contexts (Holt, Lotto, & Kluender, 2000; Lindblom & Studdert-Kennedy, 1967). Here, adaptation to low-frequency contexts increased the likelihood that neural populations selective for high frequencies encoded ambiguous f_2 frequencies (Holt, 2005; Huang & Holt, 2012; Sjerps, Mitterer, & McQueen, 2011). Consequently, ambiguous f_2 frequencies were more likely to be perceived as relatively high, and vowels they were contained within were more likely to be perceived as [e]. As with contextual speech rate, similar effects of spectral context on speech perception are abundant and are shown to influence perception of cues such as place of articulation (Lotto & Kluender, 1998; Mann & Repp, 1981) and formant structure (Huang & Holt, 2012; Ladefoged & Broadbent, 1957). A central theoretical interest of these studies has been whether or not general-auditory processing, as opposed to speech-specific processing, governs the effects of phonetic context. In favor of a general-auditory processing account, non-speech contexts (e.g., simple sine-wave tones) are sufficient to influence perception of speech targets [see (Holt, 2005; Huang & Holt, 2012; Sjerps et al., 2011) for a more detailed discussion].

Despite an extensive history of study, the effects of phonetic context have only been shown to affect small units of speech (e.g., formant transitions, function words). Critical aspects of speech perception, however, occur at larger scales. Sentence comprehension, for example, is an important aspect of speech perception and communication. It is, therefore, imperative to understand how it is affected by context. The current experiments tested whether the effects of phonetic context, contextual speech rate in particular, generalized to a larger unit of speech, namely sentences. In particular, Experiments 1 and 2 tested whether sentence comprehension was sensitive to relative rate – sentence rate compared to the average rate of all other sentences within the same block – at long timescales. Experiment 3 tested whether temporal information of non-speech context sounds was sufficient to modify sentence comprehension, which may infer some of the underlying neural processes.

CHAPTER 2

EXPERIMENT 1

Experiment 1 tested the hypothesis that sentences spoken at relatively slow rates are better comprehended than sentences spoken at relatively fast rates. Sentences were presented in two blocks of trials, which differed in their average rate. In short, the average rate of sentences in the *fast context* block was faster than the average rate of sentences in the *slow context* block. Both blocks included sentences spoken at the same intermediate rate, which was relatively slower than the average rate of sentences in the fast context block; in contrast, the same intermediate rate was relatively faster than the average rate of sentences in the slow context block. Based on a *contrastive effect* account, during the fast context block, adaptation to fast rates was expected to cause the intermediate rate to be perceived as relatively slow, which would manifest as better comprehension. In contrast, during the slow context block, adaptation to slow rates was expected to cause the same intermediate rate to be perceived as relatively fast, which would manifest as poorer comprehension. Finally, based on the results of a previous study (Baese-Berk et al., 2014), an effect of relative rate may not occur immediately, instead emerging after several minutes of speech exposure.

Methods

Participants

Forty undergraduates (25 females, mean age = 22.20 years, age range = 18 - 39 years), with reported normal hearing, from the University of Nevada, Las Vegas Psychology subject pool participated after giving written informed consent according to the guidelines of the University's Office for the Protection of Research Subjects.

Participants were excluded from participation if they began learning English after the age of 10 years.

Stimuli and Procedures

Figure 1 displays the basic stimulus design. Speech stimuli included 150 meaningful sentences, taken from the DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (Garofolo et al., 1993), recorded by an English female speaker (digitized at 22050 Hz, 16 bit resolution). The final word of each sentence was spliced out at zero crossings, where there was no energy in the waveform. Sentences, without final words, ranged in length from 1152 to 2548 ms (M = 1950.32 ms; SD = 240.04 ms) and included 4 to 10 words (M = 6.79; SD = 1.08). The intensities of all sentences were normalized to the same root-mean-square value using the Scale Intensity function in Praat (Boersma & Weenink, 2014). Stimuli were presented using a custom interface written in Presentation (Neurobehavioral Systems, Inc., Albany, CA), generated using an SB X-Fi sound card (Creative Technology, Ltd.), and delivered via Sennheiser HD 280 headphones at around 70 dB SPL (Sennheiser Electronic Corporation, Old Lyme, CT). Behavioral responses were recorded by Presentation, and stored for off-line analysis.

The rates of spoken sentences were modified using the Pitch-Synchronous Overlap and Add (PSOLA) function in Praat (Moulines & Charpentier, 1990). In short, the PSOLA function compressed (or expanded) sentences to a percentage of their original duration, such that smaller percentages corresponded to faster rates. Sentences were presented in two blocks of trials, which differed in their average rate. The durations of sentences in the *fast context* block were modified to 25%, 30%, or 35% of their original durations. The durations of sentences in the *slow context* block were modified to 35%, 90%, or 110% of their original durations. Both blocks included sentences compressed to 35%. Importantly, in the fast context block, 35% was slower than the average rate (i.e., 30%); however, in the slow context block, 35% was faster than the average rate (i.e., 78%).

A computer monitor, with a light grey background, placed directly in front of participants remained blank during the speech stimuli. At the end of each spoken sentence, a printed word appeared on the center of the computer monitor for 1 s. For half the trials, the printed word was the original ending to the spoken sentence. For the remaining trials, the printed word was not the original ending (i.e., a randomly selected word). At the end of each trial, participants responded whether the printed word was the original ending to the spoken sentence, was not the original ending, or if they did not know using the ‘1’, ‘2’, and ‘3’ buttons on the computer keypad, respectively. The inter-trial interval (i.e., the silent duration between the offset and onset of adjacent sentences) was 2 s.

The study was conducted in a quiet room. As mentioned above, trials were presented in two blocks, which differed in their average rate. Within a block, each speech rate (i.e., 25%, 30%, 35% and 35%, 90%, 110%) was presented 25 times. Speech rate and whether or not the printed word was the original ending to the spoken sentence were randomized. No sentence and printed word was repeated and, therefore, each trial was unique. For half the participants, the fast context was presented before the slow context block. For the remaining participants, the slow context was presented before the fast context. Furthermore, for half the participants within a given presentation order, the list of sentences compressed to 35% in the fast context block was the same list of sentences compressed to 35% in the slow context block for the remaining participants. Participants were randomly assigned to one of four presentation types. Prior to the start of the experiment, participants were given 5 practice trials of unmodified sentences and, subsequently, 6 practice trials of sentences modified from 40% to 130% of their original duration presented in random order.

Data Analysis

Speech comprehension was measured as performance on the task in which participants judged whether the printed word was or was not the original ending to the spoken sentence. Trials in which participants correctly identified whether the printed word was the original or was not the original ending were considered correct. Incorrectly identified trials and trials in which participants responded “I don’t know” were considered incorrect. In order to examine an effect of speech rate on comprehension, correct performance was averaged for each rate, with the 35% rate from the fast and slow context blocks collapsed together, and for each participant separately. These averages were entered into a 1-factor (speech rate: 25%, 30%, 35%, 90%, 110%) repeated-measures analysis of variance (ANOVA). In order to examine an effect of relative rate on comprehension and whether the size of the effect increased over time, correct performance at the 35% rate was averaged across trials within the early (i.e., first 30 trials) and late (i.e., last 30 trials) phases of each block separately. These averages were entered into a 2 (relative rate: slow, fast) x 2 (block phase: early, late) repeated-measures ANOVA. As part of a planned comparison, correct performance at the 35% rate was entered into a paired-sample t-test, for early- and late-phase trials separately, to test whether average performance differed depending on relative rate. In order to examine an effect of block phase on the remaining speech rates, correct performance at each non-target speech rate (i.e., 25%, 30%, 90%, and 110%) was averaged across trials within the early (i.e., first 30 trials) and late (i.e., last 30 trials) phases of each block separately. These averages were entered into a 4 (speech rate: 25%, 30%, 90%, 110%) x 2 (block phase: early, late) repeated-measures ANOVA. For all ANOVAs, presentation type was entered as a between-subjects factor to ensure presentation order and/or sentence list wasn't driving any main effects or interactions. *P*-values less than .05 were considered statistically significant, and when appropriate Greenhouse-Geisser corrected *p*-values were reported.

Results and Discussion

The results of Experiment 1 are plotted in Figure 2. Sentence comprehension, measured as performance on the comprehension task, was better at slower speech rates. More importantly, comprehension at the 35% rate was better during the fast context block (i.e., when it was relatively slower than the average rate) than during the slow context block (i.e., when it was relatively faster than the average rate).

Effect of Speech Rate. There was a significant main effect of speech rate on comprehension, $F(4,144) = 235.91$, $p < .001$, $\eta_p^2 = .87$, such that comprehension was better at slower rates, and this effect did not differ between the presentation type groups, $F(12,144) = 1.63$, $p = .15$, $\eta_p^2 = .12$. The size of the effect was robust such that average proportion correct on the comprehension task jumped from .26 at the 25% rate to .89 at the 110% rate.

Effect of Relative Rate. The main effect of relative rate on comprehension was not significant, $F(1,36) = 1.93$, $p = .17$, $\eta_p^2 = .05$; there was, however, a significant interaction between relative rate and block phase, $F(1,36) = 5.23$, $p < .05$, $\eta_p^2 = .13$, which did not differ between the presentation type groups, $F(3,36) = .56$, $p = .64$, $\eta_p^2 = .169$, such that comprehension at the relatively slow 35% rate (i.e., during the fast context block) was better than comprehension at the relatively fast 35% rate (i.e., during the slow context block), and this difference did not emerge until comparing trials from the late block phases. Consistent with this interpretation, within the late block phases, comprehension at the relatively slow 35% rate was significantly better than comprehension at the relatively fast 35% rate, $t(39) = 3.20$, $p < .01$ (Figure 2, red); in contrast, within the early block phases, this difference was negligible, $t(39) = -.43$, $p = .67$ (Figure 2, blue). The difference in relative rate was primarily caused by the relatively slow 35% rate, such that average proportion correct increased from .52 at the early block phase to .63 at the late block phase.

Effect of Block Phase. As suggested above, there was a significant main effect of block phase on the target speech rates (i.e., relatively slow 35%, relatively fast 35%), $F(1, 36) = 4.72$, $p < .05$, $\eta_p^2 = .12$, which was driven by the significant relative rate x block phase interaction reported above, and did not differ between the presentation type groups, $F(3,36) = .66$, $p = .58$, $\eta_p^2 = .05$. The main effect of block phase on all remaining non-target speech rates (i.e., 25%, 30%, 90%, and 110%) was not significant, $F(1,36) = .01$, $p = .78$, $\eta_p^2 = .00$, and neither was the interaction between block phase and speech rate, $F(3, 108) = .98$, $p = .40$, $\eta_p^2 = .03$, such that comprehension was similar between the early and late block phases. These results suggest that only for the relatively slow 35% rate was comprehension significantly better during the late block phases than during the early block phases.

Summary. These results provide the first preliminary evidence that sentence comprehension, a large unit of speech, is sensitive to relative rate – sentence rate compared to the average rate of all other sentences within the same block – at long timescales.

CHAPTER 3

EXPERIMENT 2

In Experiment 1, sentence comprehension was sensitive to relative rate. An important distinction, however, is whether a *contrastive effect* or *perceptual learning* drove the effect. Based on a contrastive effect account, during the fast context block, adaptation to fast rates caused perception at the 35% rate to be relatively slow, which manifested as better comprehension. In contrast, during the slow context block, adaptation to slow rates caused perception at the 35% rate to be relatively fast, which manifested as poorer comprehension. Critical to this account is the relative slowness (or fastness) of the 35% rate during the fast context (or slow context).

The results of Experiment 1, however, may be explained by a perceptual learning account. Perceptual learning is a stimulus-specific improvement of perception following repeated exposure (Banai & Amitay, 2012). Exposure to sentences at a 38% rate, for example, improves comprehension of subsequent sentences spoken at similar rates (Dupoux & Green, 1997). Importantly, perceptual learning of impoverished speech is more likely to generalize to phonetically similar speech (Borrie, McAuliffe, & Liss, 2012; Dupoux & Green, 1997). It is of note, then, that in Experiment 1 the 35% rate was more similar to average rate of sentences in the fast context block. Assume that, within each block, perceptual learning occurred at the average rate of sentences (i.e., fast context: 30%; slow context: 78%) and was more likely to generalize to similar rates (Figure 3). Perceptual learning was more likely to generalize to the 35% rate in the fast context block, compared to the same rate in the slow context block, because this rate was more similar to the average rate of sentences in the fast context (difference marked with arrowheads). This account disregards the relative slowness of the 35% rate during the fast context as contributing to its comprehension, and suggests instead the similarity of the 35% rate to the average rate of sentences in the fast context as the main contributing factor to its improved comprehension.

In Experiment 2, the 35% rate was as similar to the average rate during the fast context as it was to the average rate during the slow context; however, the 35% remained relatively slow during the fast context block. In this case, generalization of perceptual learning from non-target rates to target rates should be matched between blocks. Based on a perceptual learning account, therefore, comprehension at the 35% rate should be similar during the fast context and the slow context blocks. In contrast, based on a contrastive effect account, comprehension at the 35% rate should be better during the fast context block (i.e., when it was relatively slower than the average rate) than during the slow context block (i.e., when it was relatively faster than the average rate).

Methods

Participants

Forty undergraduates (33 females, mean age = 20.70 years, age range = 18 - 33 years), with reported normal hearing, from the University of Nevada, Las Vegas Psychology subject pool participated after giving written informed consent according to the guidelines of the University's Office for the Protection of Research Subjects. Participants were excluded from participation if they began learning English after the age of 10 years.

Stimuli and Procedures

Similar to Experiment 1, the durations of sentences in the fast context block were modified to 25%, 30%, or 35% of their original durations; however, unlike Experiment 1, the durations of sentences in the slow context block were modified to 35%, 40%, or 45% of their original durations. Critically, in both blocks, the difference between the average rate of sentences (i.e., fast context: 30%; slow context: 40%) and the 35% rate was 5%. In a pilot study (n = 15), using the same procedures as Experiment 1 (Figure 1), speech comprehension was measured for sentences compressed to 30%, 35%, and 40%. The average difference in proportion correct between the 30% and the 35% rate was .24 and, similarly, the average difference between the 35% and the 40% rate was .25 (Figure 4). Therefore, the difference between the average rate of sentences and the 35% rate, as measured both phonetically and perceptually, was matched between the slow and the fast context blocks. All other aspects of the stimuli and procedures were similar to Experiment 1.

Data Analysis

All aspects of the data analysis were similar to Experiment 1.

Results and Discussion

The results of Experiment 2 are plotted in Figure 5. Sentence comprehension was better at slower speech rates. Importantly, however, comprehension at the 35% rate was similar during the fast context block (i.e., when it was relatively slower than the average rate) and the slow context block (i.e., when it was relatively faster than the average rate).

Effect of Speech Rate. There was a main effect of speech rate on comprehension, $F(4,144) = 201.93$, $p < .001$, $\eta_p^2 = .85$, such that comprehension was better at slower rates. The size of the effect was robust such that average proportion correct on the comprehension task jumped from .32 at the 25% rate to .82 at the 45% rate. The effect differed between the presentation type groups, $F(12,144) = 3.25$, $p < .01$, $\eta_p^2 = .21$; however, the effect was linear and qualitatively similar across all groups.

Effect of Relative Rate. The main effect of relative rate on comprehension was not significant, $F(1,36) = .37$, $p = .55$, $\eta_p^2 = .01$, and neither was the interaction between relative rate and block phase, $F(1,36) = .07$, $p = .80$, $\eta_p^2 = .002$. That is, comprehension at the relatively slow 35% rate was not significantly better than comprehension at the relatively fast 35% rate in either the early, $t(39) = -.17$, $p = .87$, or late block phases, $t(39) = -.62$, $p = .54$.

Effect of Block Phase. There was a significant main effect of block phase on the target speech rates (i.e., relatively slow 35%, relatively fast 35%), $F(1, 36) = 4.72$, $p < .05$, $\eta_p^2 = .12$, which did not differ between the presentation type groups, $F(3,36) = .25$, $p = .86$, $\eta_p^2 = .02$, such that comprehension improved from the early to the late block phases. There was a similar main effect of block phase on all remaining non-target speech rates (i.e., 25%, 30%, 40%, and 45%), $F(1,36) = 10.08$, $p < .01$, $\eta_p^2 = .22$, which did not interact with the main effect of speech rate, $F(3, 108) = .33$, $p = .81$, $\eta_p^2 = .01$, nor did it differ between the presentation type groups, $F(3,36) = .77$, $p = .52$, $\eta_p^2 = .22$. These results suggest that comprehension improved from the early to the late block phases, and this effect was similar for all speech rates.

In Experiment 1, any comprehension improvement from the early to the late block phases was largely exclusive to the relatively slow 35% rate. In Experiment 2, however, comprehension of all speech rates improved across block phases. This discrepancy may have been caused by stimulus differences between the experiments. Experiment 2, compared to Experiment 1, tested speech rates that were more similar to each other, and all of which were faster than usual (i.e., below 50% compression rate). Exposure to the first block of trials may have facilitated an effect of block phase on all speech rates in a second block of trials 1) when both blocks of trials shared more similar rates and/or 2) when all rates in the first or second block were faster than usual. Indeed, in Experiment 2, differences in comprehension between the early and the late block phases were qualitatively larger in the second block of trials (M = .07) than the first block of trials (M = .02), when collapsed across speech rates within each block.

Summary. In Experiment 2, the difference between the average rate of sentences and the 35% rate, as measured both phonetically and perceptually, was matched between the slow- and fast-context blocks. Importantly, in this case, a perceptual learning account uniquely predicts a null effect of relative rate – as opposed to a contrastive effect account, which predicts better comprehension at relatively slow rates. Furthermore, perceptual learning implies that comprehension should improve across block phases. The results of Experiment 2 are consistent with a perceptual learning account, such that there was a null effect of relative rate and significant improvement across block phases. These results support the interpretation that perceptual learning accounts for the effect of relative rate observed in Experiment 1.

CHAPTER 4

EXPERIMENT 3

Despite the absence of a contrastive effect of relative rate, the paradigm used in Experiment 1 is, nonetheless, useful to study generalization of perceptual learning. Perceptual learning at an average rate was more likely to generalize to a similar target rate (i.e., relatively slow rate), compared to a less-similar target rate (i.e., relatively fast rate). The difference in comprehension between the two target rates provides a measure of generalization of perceptual learning, and this measure may be used in testing which features of the non-target speech were critical for generalization to occur. Experiment 3 adopted this logic, and tested whether exposure to temporal information at non-target rates was sufficient for generalization of perceptual learning to a similar target rate. To this end, Experiment 3 tested a modified version of Experiment 1 in which non-target speech were replaced with non-speech sounds (i.e., 1-channel noise vocoded speech), which nonetheless conveyed the same rate information as the speech stimuli used in Experiment 1.

Experiment 3, by extension, tested whether *general-auditory processing* or *speech-specific processing* govern generalization of perceptual learning. For clarification, general-auditory processing refer to neural processes within auditory brain regions that respond to several classes of sound; speech-specific processing refer to neural processes within brain regions that respond selectively to speech sounds (Chan et al., 2014; Mesgarani, Cheung, Johnson, & Chang, 2014). Importantly, a speech-specific processing account uniquely predicts a lack of generalization of perceptual learning from non-speech non-target sounds to target speech – as opposed to a general-auditory processing account which predicts preservation of generalization of perceptual learning.

Evidence from previous studies at least partially favors a general-auditory processing account. Perceptual learning of time-compressed speech generalizes from incomprehensible speech (i.e., speech in a foreign language) to comprehensible speech (i.e., speech in a native language) (Mehler et al., 1993; Pallier, Sebastian-Galles, Dupoux, Christophe, & Mehler, 1998; Sebastian-Galles, Dupoux, Costa, & Mehler, 2000). Lexical access and/or speech comprehension, therefore, does not appear to be necessary for generalization of perceptual learning; however, rhythmic information appears to be important. In particular, perceptual learning of time-compressed speech in a native language benefits from exposure to incomprehensible time-compressed speech in a foreign language, but only when the foreign language falls under a similar rhythmic class as the native language. For example, for monolingual Spanish speakers, exposure to time-compressed Catalan speech facilitates subsequent perceptual learning of time-compressed Spanish speech, because both are syllable-timed languages. In contrast, for monolingual English speakers, exposure to time-compressed French speech, another syllable-timed language, does not facilitate perceptual learning of time-compressed English speech, a stressed-timed language. In these studies, speech sounds constituted the incomprehensible non-native speech, which may have recruited speech-specific processes. It remains unclear, then, whether exposure to temporal information (i.e., rate and rhythm) using non-speech sounds – which presumably do not recruit speech-specific processes – is sufficient for generalization of perceptual learning.

Consider the possibility that the effect of relative rate, as observed in Experiment 1, actually reflects a contrastive effect. Evidence again favors a general-auditory processing account. In particular, the rate of non-speech sounds (e.g., pure tones) influences the perceptual categorization of ambiguous syllables (Pisoni et al., 1983; Wade & Holt, 2005); however, it remains unclear whether similar general-auditory processes governed the effect of relative rate on sentence comprehension. Indeed, Repp (1982) theorized that speech-specific processes govern phonetic context effects. Based on this account, non-speech sounds are not sufficient to facilitate the effect of relative rate on sentence comprehension, because they presumably do not recruit speech-specific processes.

Methods

Participants

Forty undergraduates (31 females, mean age = 21.33 years, age range = 18 - 38 years), with reported normal hearing, from the University of Nevada, Las Vegas Psychology subject pool participated after giving written informed consent according to the guidelines of the University's Office for the Protection of Research Subjects. Participants were excluded from participation if they began learning English after the age of 10 years.

Stimuli and Procedures

The stimuli and procedures of Experiment 3 were similar to Experiment 1 with the following exceptions. As in Experiment 1, the durations of sentences in the fast context block were modified to 25%, 30%, or 35% of their original durations. The durations of sentences in the slow context block were modified to 35%, 90%, or 110% of their original durations. 35%-compressed sentences were presented as is; however, the remaining sentences (i.e., fast context: 25%, 30%; slow context: 90%, 110%) were noise vocoded into 1 spectral channel (Shannon et al., 1995). Noise vocoding was performed using a custom written script in Praat. In short, the amplitude envelope of each sentence was extracted and used to amplitude modulate white noise. Noise vocoding preserved much of the same temporal information present in the original sentences, including information pertaining to speech rate (Figure 6); however, none of the meaningful spectral information was preserved. Consequently, 1-channel noise-vocoded sentences were unintelligible and did not sound like speech. The same sentence comprehension task, as used in Experiments 1 and 2, was used in Experiment 3, including during noise-vocoded trials. The purpose of using the same comprehension task was to keep Experiment 3 as similar as possible to Experiment 1, for comparison purposes. It is worth mentioning that, in previous studies, performing a comprehension task on time-compressed incomprehensible speech (e.g., speech in a foreign language) did not disrupt generalization of perceptual learning to time-compressed comprehensible speech (e.g., speech in a native language) (Pallier et al., 1998; Sebastian-Galles et al., 2000). Therefore, able performance on a comprehension task with non-target speech does not appear to be necessary for generalization of perceptual learning to time-compressed target speech.

Data Analysis

All aspects of the data analysis were similar to Experiment 1.

Results and Discussion

The results of Experiment 3 are plotted in Figure 7. As expected, sentence comprehension of noise-vocoded context sentences was poor, and comprehension of target sentences remained high. Importantly, however, comprehension at the 35% rate was similar during the fast context block (i.e., when it was more similar to the average rate) and the slow context block (i.e., when it was less similar to the average rate).

Effect of Speech Rate. There was a main effect of speech rate on comprehension, $F(4,144) = 376.77$, $p < .001$, $\eta_p^2 = .91$; however, this effect was likely caused by the large difference between the 35% rate and the remaining noise-vocoded sentences. Indeed, after removing the 35% rate from the analysis, the main effect of speech rate was not significant, $F(3,108) = .69$, $p = .56$, $\eta_p^2 = .02$. These results were expected given that the noise-vocoded sentences were incomprehensible and, therefore, comprehension was expected to be poor regardless of speech rate. The effect of speech rate did not differ between the presentation type groups when the 35% rate was included in the analysis, $F(12,144) = .28$, $p = .88$, $\eta_p^2 = .02$, and when it was not included in the analysis, $F(9,108) = .144$, $p = .21$, $\eta_p^2 = .11$.

Generalization of Perceptual Learning. In the current analysis, generalization of perceptual learning was measured as a difference in comprehension between target speech rates (i.e., relatively slow 35%, relatively fast 35%). Comprehension was similar between target speech rates, $F(1,36) = .49$, $p = .49$, $\eta_p^2 = .01$, regardless of whether comparing trials within the early or late block phases, $F(1,36) = .80$, $p = .38$, $\eta_p^2 = .02$. That is, comprehension at the relatively slow 35% rate was not significantly better than comprehension at the relatively fast 35% rate in either the early, $t(39) = -.64$, $p = .53$, or late block phases, $t(39) = .05$, $p = .96$.

In Experiment 1, average proportion correct at the 35% rate was .56. In Experiment 3, however, average proportion correct at the same rate was much higher at .69. It remains unclear why this difference occurred. Nonetheless, a few possible explanations are provided below – in addition to the possibility that participants in Experiment 3 were simply better at comprehending sentences at the 35% rate. It may be that, in Experiment 3, participants exerted less cognitive effort (i.e., attention, decision-making processes) during noise vocoded sentences, which caused participants to have more available cognitive resources during target sentences. Alternatively, participants may have actually exerted more cognitive effort during noise vocoded sentences, given they were more difficult to comprehend, which carried over to target sentences. Both interpretations suggest that more cognitive effort was exerted to target sentences, which would presumably facilitate their comprehension. Finally, it is interesting to consider that the clarity of target sentences "popped out" from degraded noise-vocoded sentences. Hearing low-quality noise-vocoded sentences may have caused subsequent target sentences to be perceived as more salient and/or clear. These interpretations remain speculation, and the available data are not sufficient to test them.

Effect of Block Phase. There was a significant main effect of block phase on target speech rates, $F(1,36) = 7.97$, $p < .01$, $\eta_p^2 = .18$, which did not differ between the presentation type groups, $F(3,36) = .25$, $p = .86$, $\eta_p^2 = .02$, such that comprehension improved from the early to the late block phases. The main effect of block phase on all remaining non-target speech rates (i.e., 25%, 30%, 90%, and 110%) was not significant, $F(1,36) = 2.64$, $p = .11$, $\eta_p^2 = .07$, and neither was the interaction between block phase and speech rate, $F(3, 108) = 1.10$, $p = .34$, $\eta_p^2 = .03$, such that comprehension was similar between trials within the early and late block phases. Again, the latter results were expected given that the noise-vocoded sentences were incomprehensible and, therefore, comprehension was expected to be poor regardless of block phase. These results suggest that comprehension improved from the early to the late block phases, and this effect was similar for the relatively slow and relatively fast 35% rates.

Summary. In Experiment 3, sentences at target speech rates (i.e., relatively slow 35%, relatively fast 35%) were presented as is, and sentences at the remaining non-target rates were noise vocoded into 1-channel. Consequently, noise-vocoded sentences were unintelligible, but contained much of the same temporal information present in the original sentences. The results of Experiment 3 reveal that exposure to temporal information at non-target rates was not sufficient for generalization of perceptual learning, measured as a difference in comprehension between target rates. These results, by extension, support a speech-specific processing account of generalization of perceptual learning, to the extent that comprehension was similar between target speech rates – as opposed to a general-auditory processing account, which predicts better comprehension at relatively slow target rates.

CHAPTER 5

GENERAL DISCUSSION

Despite an extensive history of study, the effects of phonetic context are only known to affect small units of speech, such as formant transitions (Miller & Liberman, 1979; Pisoni et al., 1983; Wade & Holt, 2005) and function words (Baese-Berk et al., 2014; Dilley & Pitt, 2010). Critical aspects of speech perception, however, occur at larger scales. The series of experiments reported here investigated the effects of contextual speech rate on the perception of a large unit of speech, namely sentences. The experiments, in particular, tested whether sentence comprehension was sensitive to relative rate (Experiment 1) – the rate of a sentence compared to the average rate of all other sentences within the same conversation-length period of speech, whether the effect of relative rate can be explained by a *perceptual learning* account (Experiment 2), and whether exposure to temporal information was sufficient for generalization of perceptual learning (Experiment 3).

These experiments were particularly designed to test a *contrastive effect* of relative rate. Based on this account, over the course of a conversation length-period of speech, listeners were expected to adapt to the average speech rate. Adaptation to fast speech rates was expected to cause perception of intermediate rates to be relatively slow, which would manifest as better comprehension.

Relative Rate: Limited to Small Units and/or Perceptual Categorization?

The results reported here do not support a contrastive effect of relative rate on sentence comprehension. In Experiment 1, comprehension at relatively slow rates was better than comprehension at relatively fast rates, which at first seems consistent with a contrastive effect account. The results of Experiment 2, however, favor a perceptual learning account of the effect of relative rate. This account assumes that, within each block, perceptual learning occurred at the average speech rate and was more likely to generalize to similar speech rates (Borrie et al., 2012; Dupoux & Green, 1997). Importantly, in Experiment 1, the relatively slow rate was more similar to the average speech rate during the fast context block, compared to how similar the relatively fast rate was to the average speech rate during the slow context block. Consequently, perceptual learning was more likely to generalize to the relatively slow rate than the relatively fast rate, which may have caused the difference in comprehension between them. Indeed, after minimizing expected differences in perceptual learning, as in Experiment 2, the effect of relative rate was insignificant.

Note that the difference in rate between the fast-context average rate and the slow-context average rate was smaller in Experiment 2, compared to Experiment 1. In particular, whereas the difference in average rate between the fast and the slow context blocks was 10% in Experiment 2, the difference was 48% in Experiment 1. A contrastive effect of relative rate presumably requires that the difference in average rate between contexts be sufficiently large. If so, in Experiment 2, the difference in average rate between contexts may have been too small to facilitate an effect of relative rate. In contrast, in Experiment 1, the difference in average speech between contexts may have been sufficiently large to facilitate an effect of relative rate. This heeds warning against a perceptual learning account, because a contrastive effect of relative rate may have occurred in Experiment 2 had there been a larger difference in average speech rate between contexts.

The effects of contextual speech rate, including relative rate, may be limited to perception of small units of speech. This conclusion is somewhat unexpected, given that an effect of relative rate persists for long durations across a conversation-length period of speech (Baese-Berk et al., 2014). Similarly, in the non-speech domain, the perceived rate of a short tone sequence is sensitive to its relative rate – the rate of a tone sequence compared to the average rate of all other tone sequences within the same block of trials, and the size of this effect increases across extended durations (Jones & McAuley, 2005; McAuley & Miller, 2007). Taken together, effects of relative rate, in both the speech and non-speech domains, persist for long durations; however, at least in the speech domain, the perceptual units affected by relative rate may only occur over short durations.

Alternatively, the effects of contextual speech rate may influence categorization of lexically ambiguous speech, but do less to improve comprehension of degraded speech. Indeed, it has already been argued that the effects of contextual speech rate, including those reviewed in the Introduction, should be conceptualized as influencing perceptual categorization of speech (Holt & Lotto, 2010). Contextual speech rate may, for example, increase the likelihood an ambiguous syllable is categorized as [ba], opposed to [wa] (Miller & Liberman, 1979; Pisoni et al., 1983; Wade & Holt, 2005), or it may increase the likelihood an ambiguous phrase is categorized as "…*leisure or time*…", opposed to "…*leisure time*…" (Baese-Berk et al., 2014; Dilley & Pitt, 2010). It remains less clear, however, whether contextual speech rate influences comprehension of degraded speech, or whether an effect of contextual speech rate on perceptual categorization would improve comprehension of degraded speech. If we assume, for example, that the processes that mediate perceptual categorization are distinct from those that mediate comprehension, then an effect of contextual speech rate on perceptual categorization would not necessarily cause a change in comprehension as well.

It is possible that, despite achieving similar levels of performance on the comprehension task, participant's subjective experience at relatively fast rates may have been less intelligible, compared to subjective experience at relatively slow rates. Importantly, performance on the comprehension task may not be suited to capture subjective experience of speech intelligibility. A more suited method may be to directly measure participant's subjective impression of the intelligibility to speech stimuli. Alternatively, a measure of recognition memory for speech may reveal performance differences between the relatively slow and the relatively fast rates, given that recognition memory is better for intelligible speech than less-intelligible speech (Van Engen, Chandrasekaran, & Smiljanic, 2012). In addition, participants may have exerted more cognitive effort during trials spoken at relatively fast rates, compared to trials spoken at relatively slow rates. Pupil diameter is a well-known correlate of cognitive effort, and is known to vary as a function of sentence processing difficulty (Just & Carpenter, 1993). Pupillary responses may reveal differences in the level of cognitive effort participants exerted in order to comprehend speech at relatively slow rates, compared to speech at relatively fast rates.

Generalization of Perceptual Learning: A Speech-Specific Process?

Despite the absence of a contrastive effect of relative rate, the paradigm used in Experiment 1 is, nonetheless, useful to study generalization of perceptual learning. Perceptual learning at an average rate was more likely to generalize to a similar target rate (i.e., relatively slow rate), compared to a less-similar target rate (i.e., relatively fast rate). The difference in comprehension between the two target rates provides a measure of generalization of perceptual learning, and this measure may be used in testing which features of the non-target speech were critical for generalization to occur. Experiment 3 adopted this logic, and tested whether exposure to temporal information at non-target rates was sufficient for generalization of perceptual learning to a similar target rate, by noise vocoding non-target speech into 1 spectral channel.

Exposure to temporal information at non-target rates was not *sufficient* for generalization of perceptual learning. In particular, despite testing the same rates as Experiment 1, comprehension was the same between the similar target rate (i.e., relatively slow rate) and the less-similar target rate (i.e., relatively fast rate). Temporal information is, however, *necessary* for generalization of perceptual learning. Monolingual listeners, for example, generalized perceptual learning of time-compressed speech from a foreign language to a native language; however, this effect only occurred when the foreign and native languages were rhythmically similar (Mehler et al., 1993; Pallier et al., 1998; Sebastian-Galles et al., 2000). In this case, rhythmic information was necessary for generalization of perceptual learning.

The results of Experiment 3, by extension, support the interpretation that speech-specific processing, as opposed to general-auditory processing, governed generalization of perceptual learning. For clarification, general-auditory processing refer to neural processes within auditory brain regions that respond to several classes of sound; speech-specific processing refer to neural processes within brain regions that respond selectively to speech sounds (Chan et al., 2014; Mesgarani et al., 2014). Non-target speech presumably recruited general-auditory processes, given that noise vocoding eliminated many of their speech-like features; in contrast, at some level of processing, target speech was more likely to recruit speech-specific processes. It is possible that this neural separation occurred at a critical stage that prevented generalization of perceptual learning, and generalization relied on non-target speech and target speech to recruit similar speech-specific processes.

In Experiment 3, comprehension was similar between both target rates (i.e., relatively slow 35%, relatively fast 35%), which was interpreted as a lack of generalization of perceptual learning from non-target rates to a more similar target rate (i.e., relatively slow 35%); however, note that perceptual learning occurred at both target rates. It seems likely that exposure to target rates facilitates this learning; however, it is possible that perceptual learning at one or both target rates actually reflects generalization of perceptual learning from non-speech non-target rates – an interpretation consistent with a general-auditory processing account.

Perceptual learning of time-compressed speech occurs rapidly, such that exposure to as few as 10 to 20 sentences significantly improves comprehension; however, perceptual learning occurs much more slowly when time-compressed sentences are severely degraded (Adank & Janse, 2009; Dupoux & Green, 1997; Golomb, Peelle, & Wingfield, 2007). It should be noted, then, that non-target sentences were severely degraded in Experiment 3, given they were noise vocoded into 1 spectral channel. This heeds warning against a speech-specific processing account, because perceptual learning at non-target rates – including its generalization to target rates – may have required much more exposure to non-target sentences than was provided. In support of a general-auditory processing account, generalization of perceptual learning may have occurred had there been sufficient exposure to non-target sentences.

In Experiment 3, participants were asked to perform a sentence comprehension task on non-target sentences that were completely incomprehensible – as well as lacking much resemblance to speech. The purpose of using the comprehension task was to have participants attend to non-target sentences, and to maintain as much similarity between Experiments 1 and 3 as possible. It is possible, however, that participants quickly realized the comprehension task was nearly impossible during non-target sentences, which may have disrupted generalization of perceptual learning to target rates. Perhaps, for example, participants disengaged attention during non-target trials. It is worth mentioning that, in previous studies, performing a comprehension task on time-compressed incomprehensible speech did not disrupt generalization of perceptual learning to time-compressed comprehensible speech (Pallier et al., 1998; Sebastian-Galles et al., 2000). Therefore, able performance on a comprehension task with non-target speech does not appear to be necessary for generalization of perceptual learning to time-compressed target speech.

Neural Correlates of Speech Comprehension

Speech comprehension is thought to rely on entrainment of neural activity within auditory cortex to low-frequency amplitude changes in speech, and entrainment may explain an effect of contextual speech rate on perception of small units of speech – as reviewed in the Introduction (Ding & Simon, 2014; Giraud & Poeppel, 2012; Peelle & Davis, 2012; Schroeder et al., 2008). Neural entrainment may also explain many of the behavioral observations reported here. Consider the effect of relative rate reported in

Experiment 1, which may reflect either a contrastive effect or perceptual learning. A neural entrainment model assumes that, for both accounts, neural activity adapted to the average rate during each block, so as to optimize entrainment at this rate (Baese-Berk et al., 2014; McAuley & Miller, 2007). According to a contrastive effect account, adaptation to the slow average rate caused neural activity to under sample the relatively fast rate, which manifested as poorer comprehension. On the other hand, according to a perceptual learning account, entrainment to an average rate was more likely to generalize to similar rates. In particular, entrainment to the fast average rate (i.e., 30%) was more likely to generalize to the relatively slow rate (i.e., 35%), compared to generalization from the slow average rate (i.e., 78%) to the relatively fast rate (i.e., 35%). As a result, comprehension was better at the relatively slow rate than it was at the relatively fast rate.

Summary

In summary, the results do not favor a contrastive effect of phonetic context on sentence comprehension, a large unit of speech. The paradigm described here was, however, useful to study generalization of perceptual learning. In particular, it was inferred that speech-specific processing governs generalization of perceptual learning. Many of the interpretations suggested here, however, remain speculation, and warrant further investigation.

APPENDIX A

FIGURE CAPTIONS

Figure 1. Each trial consisted of a spoken meaningful sentence with the final word spliced out at zero crossing. Each sentence was followed by a printed word, displayed on a computer screen directly in front of participants, which was either the original or not the original ending to the spoken sentence. For example, the spoken sentence, “A pencil with black lead writes…” was be followed by “best” printed on the computer screen. Participants responded whether the printed word was the original ending to the spoken sentence (as in the above example), was not the original ending, or if they did not know.

Figure 2. Sentences were presented in two blocks of trials, which differed in their average speech rate. Comprehension at the 35% rate was better during the fast context block (solid lines) than during the slow context block (dashed lines). This difference was negligible during the first half of each block (blue), and did not emerge until the second half of each block (red). Error bars based on within-subject confidence intervals (Loftus & Masson, 1994).

Figure 3. Based on this account, comprehension at the 35% rate was better during the fast context block, compared to comprehension during the slow context block, because of the similarity of the 35% rate to the average speech rate during the fast context block.

Figure 4. Results of pilot experiment. The difference in comprehension between the 30% and 35% rate was similar to the difference between the 35% and 40% rate. Error bars based on within-subject confidence intervals (Loftus & Masson, 1994).

Figure 5. Comprehension at the 35% rate was similar during the fast context and slow context blocks. Comprehension improved, for all speech rates, from the early to the late

block phases. Error bars based on within-subject confidence intervals (Loftus & Masson, 1994).

Figure 6. The amplitude envelope of the example sentence, “The young girl gave no clear…” is plotted when the sentence was in its original form (solid line) and when the sentence was noise vocoded (dashed line). Both sentences were modified to 30% of their original durations. Importantly, despite large differences in intelligibility, the two sentences shared similar amplitude envelopes. Amplitude envelopes were extracted and plotted using the ‘Draw Intensity Contour’ function in Praat.

Figure 7. Comprehension at the 35% rate was similar during the fast context and slow context blocks, and improved from the early to the late block phases. Error bars based on within-subject confidence intervals (Loftus & Masson, 1994).

APPENDIX B

FIGURES

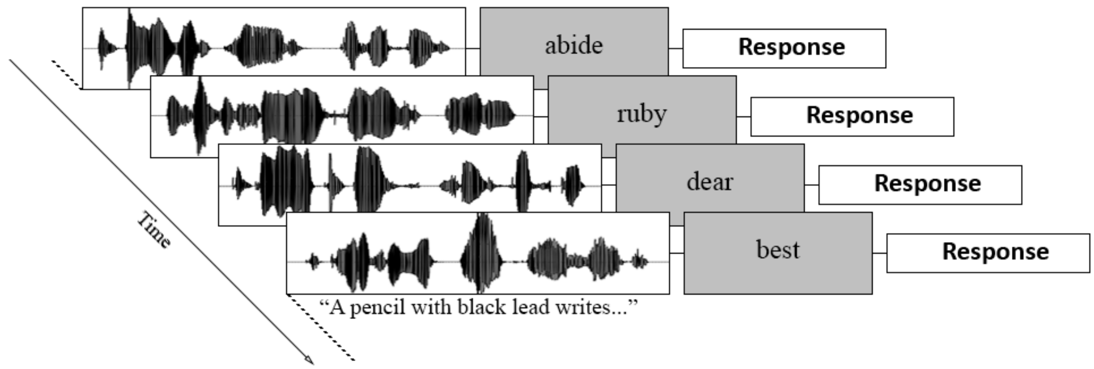

Figure 1. Basic stimulus design.

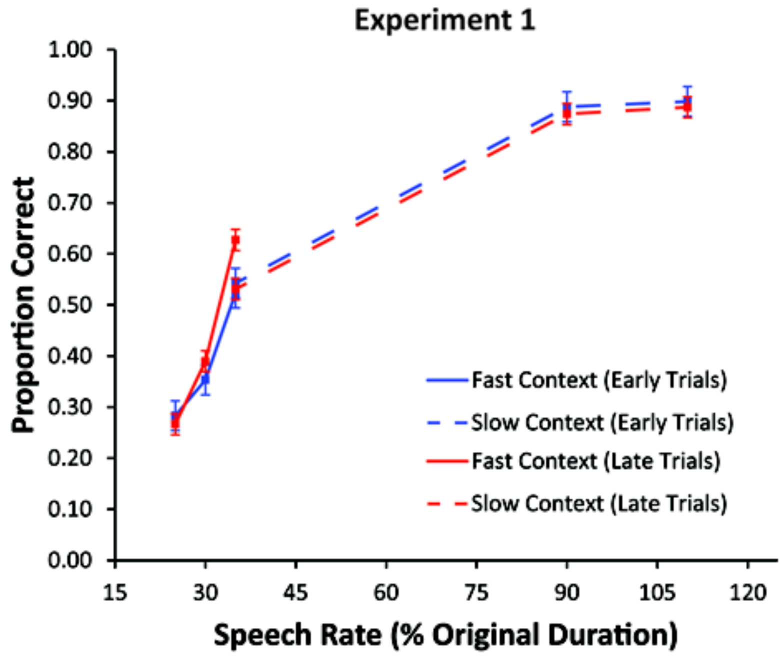

Figure 2. Results of Experiment 1.

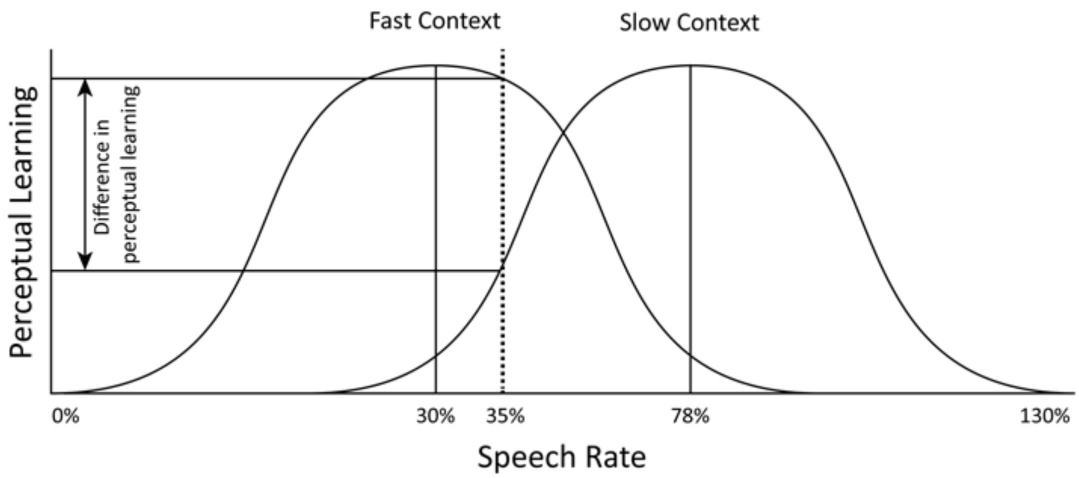

Figure 3. Perceptual learning account of the effect of relative rate.

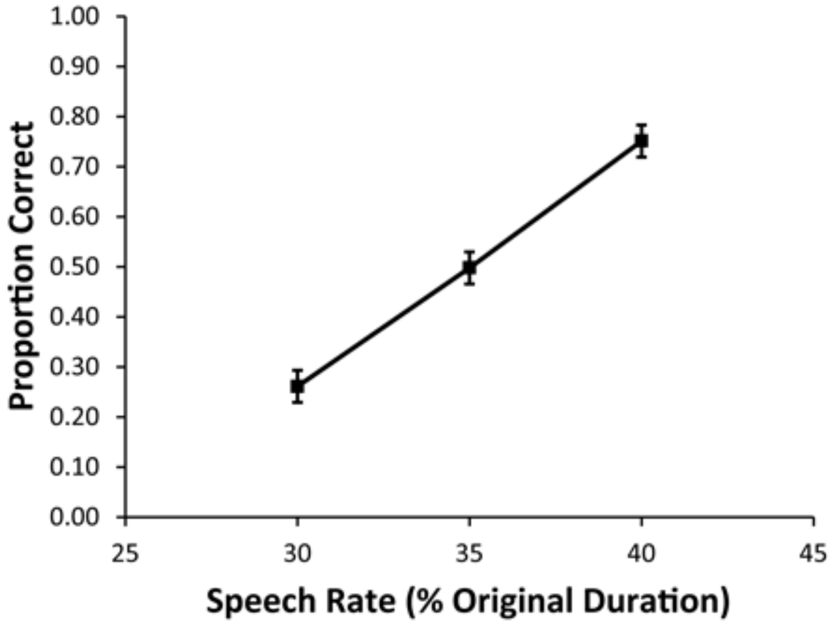

Figure 4. Results of pilot experiment.

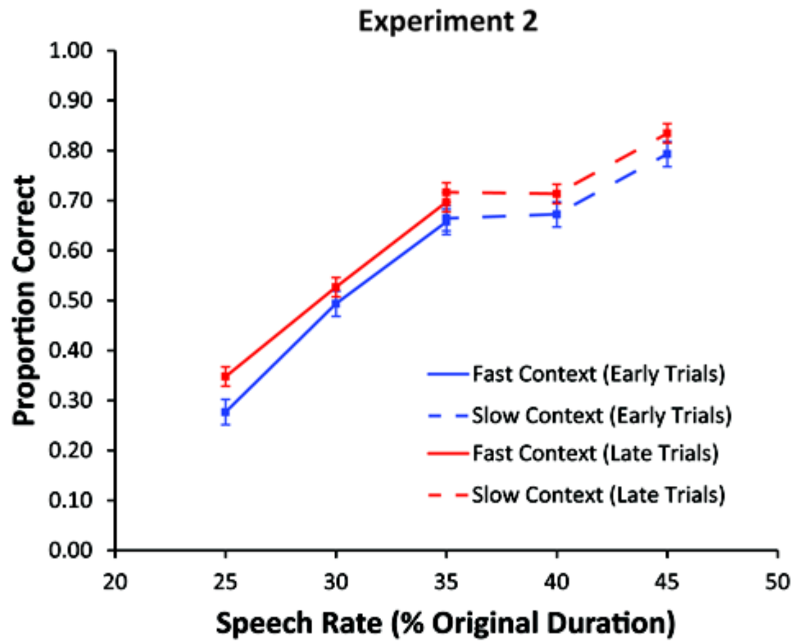

Figure 5. Results of Experiment 2.

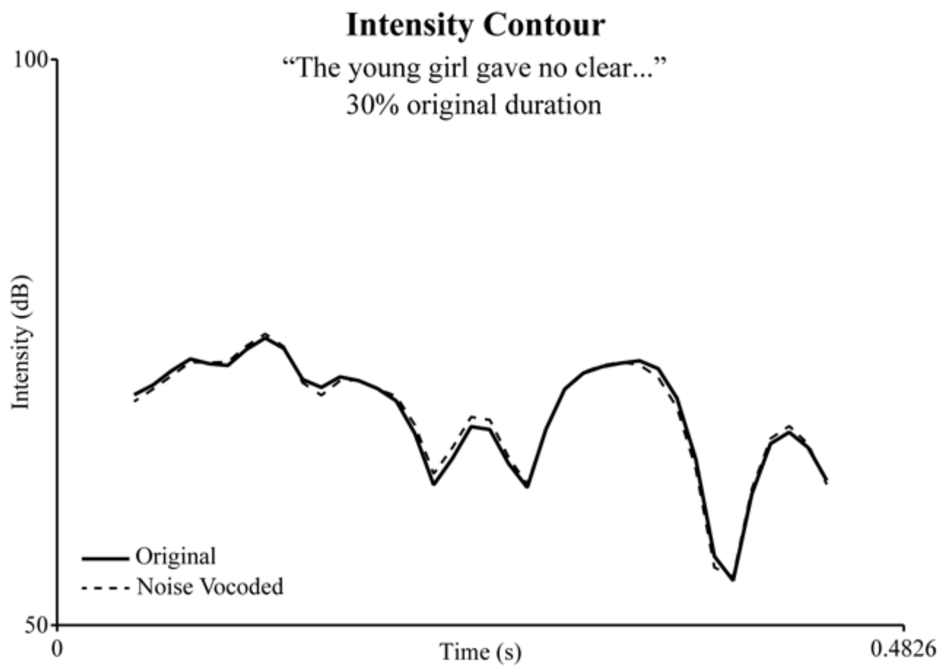

Figure 6. Original versus noise-vocoded amplitude envelope.

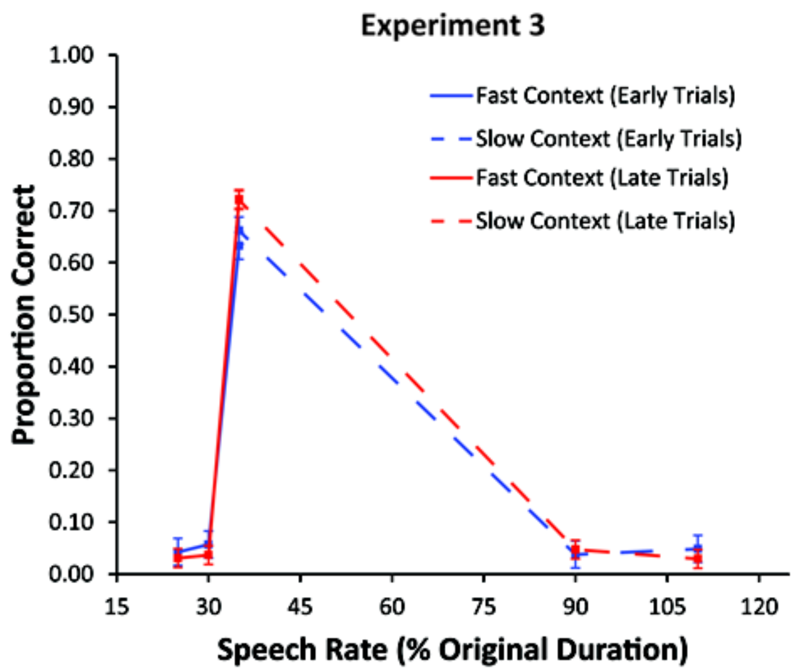

Figure 7. Results of Experiment 3.

APPENDIX C

IRB PROTOCOL APPROVAL

Social/Behavioral IRB – Expedited Review
Continuing Review Approved

NOTICE TO ALL RESEARCHERS:
Please be aware that a protocol violation (e.g., failure to submit a modification for <u>any</u> change) of an IRB approved protocol may result in mandatory remedial education, additional audits, re-consenting subjects, researcher probation, suspension of any research protocol at issue, suspension of additional existing research protocols, invalidation of all research conducted under the research protocol at issue, and further appropriate consequences as determined by the IRB and the Institutional Officer.

DATE: October 29, 2014

TO: **Dr. Joel Snyder**, Psychology

FROM: Office of Research Integrity – Human Subjects

RE: Notification of IRB Action
Protocol Title: **Neural Mechanisms of Auditory and Visual Processing in Healthy Adults**
Protocol #: 0710-2518
Expiration Date: October 28, 2015

Continuing review of the protocol named above has been reviewed and approved.

This IRB action will reset your expiration date for this protocol. The protocol is approved for a period of one year from the date of IRB approval. The new expiration date for this protocol is October 28, 2015. If the above-referenced project has not been completed by this date you must request renewal by submitting a Continuing Review Request form 30 days before the expiration date.

PLEASE NOTE:
Upon approval, the research team is responsible for conducting the research as stated in the protocol most recently reviewed and approved by the IRB, which shall include using the most recently submitted Informed Consent/Assent forms and recruitment materials. The official versions of these forms are indicated by footer which contains current approval and expiration dates.

Should there be *any* change to the protocol, it will be necessary to submit a **Modification Form** through ORI - Human Subjects. No changes may be made to the existing protocol until modifications have been approved by the IRB. Modified versions of protocol materials must be used upon review and approval. Unanticipated problems, deviations to protocols, and adverse events must be reported to the ORI – HS within 10 days of occurrence.

If you have questions or require any assistance, please contact the Office of Research Integrity - Human Subjects at IRB@unlv.edu or call (702) 895-2794.

Office of Research Integrity – Human Subjects
4505 Maryland Parkway • Box 451047 • Las Vegas, Nevada 89154-1047
(702) 895-2794 • FAX: (702) 895-0805 • IRB@unlv.edu

REFERENCES

Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *Journal of the Acoustical Society of America, 126*, 2649-2659. doi: 10.1121/1.3216914

Ainsworth, W. A. (1974). The influence of precursive sequences on the perception of synthesized vowels. *Language and Speech, 17*, 103-109.

Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science, 25*, 1546-1553. doi: 10.1177/0956797614533705

Banai, K., & Amitay, S. (2012). Stimulus uncertainty in auditory perceptual learning. *Vision Research, 61*, 83-88. doi: 10.1016/j.visres.2012.01.009

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer (Version 5.3.70). Retrieved from http://www.praat.org.

Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech Language and Hearing Research, 55*, 290-305. doi: 10.1044/1092-4388(2011/10-0349)

Chan, A. M., Dykstra, A. R., Jayaram, V., Leonard, M. K., Travis, K. E., Gygi, B., . . . Cash, S. S. (2014). Speech-specific tuning of neurons in human superior temporal gyrus. *Cerebral Cortex, 24*, 2679-2693. doi: 10.1093/cercor/bht127

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21*, 1664-1670. doi: 10.1177/0956797610384743

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience, 8*. doi: 10.3389/fnhum.2014.00311

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 914-927. doi: 10.1037/0096-1523.23.3.914

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., & Dahlgren, N. (1993). DARPA TIMIT: Acoustic-phonetic continuous speech corpus. NIST Technical Report.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America, 63*, 223-230. doi: 10.1121/1.381717

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*, 113-126. doi: 10.1159/000208934

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15*, 511-517. doi: 10.1038/Nn.3063

Golomb, J. D., Peelle, J. E., & Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *Journal of the Acoustical Society of America, 121*, 1701-1708. doi: 10.1121/1.2436635

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science, 16*, 305-312. doi: 10.1111/j.0956-7976.2005.01532.x

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics, 72*, 1218-1227. doi: 10.3758/App.72.5.1218

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America, 108*, 710-722. doi: 10.1121/1.429604

Huang, J. Y., & Holt, L. L. (2012). Listening for the norm: Adaptive coding in speech categorization. *Frontiers in Psychology, 3*. doi: 10.3389/Fpsyg.2012.00010

Jones, M. R., & McAuley, J. D. (2005). Time judgments in global temporal contexts. *Perception & Psychophysics, 67*, 398-417. doi: 10.3758/BF03193320

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indexes of sentence processing. *Canadian Journal of Experimental Psychology, 47*, 310-339. doi: 10.1037/h0078820

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29*, 98-104. doi: 10.1121/1.1908694

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron, 77*, 750-761. doi: 10.1016/j.neuron.2012.11.034

Lindblom, B. F., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America, 42*, 830-843. doi: 10.1121/1.1910655

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence-intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*, 476-490. doi: 10.3758/Bf03210951

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics, 60*, 602-619. doi: 10.3758/Bf03206049

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America, 69*, 548-558. doi: 10.1121/1.385483

McAuley, J. D., & Miller, N. S. (2007). Picking up the pace: Effects of global temporal context on sensitivity to the tempo of auditory sequences. *Perception & Psychophysics, 69*, 709-718. doi: 10.3758/BF03193773

Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., & Pallier, C. (1993). Understanding compressed sentences: The role of rhythm and meaning. *Annals of the New York Academy of Sciences, 682*, 272-282. doi: 10.1111/j.1749-6632.1993.tb22975.x

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science, 343*, 1006-1010. doi: 10.1126/science.1245994

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25*, 457-465. doi: 10.3758/Bf03213823

Morrill, T. H., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2014). Distal rhythm influences whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis. *Cognition, 131*, 69-74. doi: 10.1016/j.cognition.2013.12.006

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech snythesis. *Speech Communication, 9*, 453-467. doi: 10.1016/0167-6393(90)90021-Z

Pallier, C., Sebastian-Galles, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & Cognition, 26*, 844-851. doi: 10.3758/Bf03211403

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3*. doi: 10.3389/fpsyg.2012.00320

Pickett, J. M., & Decker, L. R. (1960). Time factors in perception of a double consonant. *Language and Speech, 3*, 11-17.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception & Psychophysics, 34*, 314-322. doi: 10.3758/Bf03203043

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental-evidence for a speech mode of perception. *Psychological Bulletin, 92*, 81-110. doi: 10.1037//0033-2909.92.1.81

Sawusch, J. R., & Newman, R. S. (2000). Perceptual normalization for speaking rate II: Effects of signal discontinuities. *Perception & Psychophysics, 62*, 285-300. doi: 10.3758/Bf03205549

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. doi: 10.1016/j.tics.2008.01.002

Sebastian-Galles, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception & Psychophysics, 62*, 834-842. doi: 10.3758/Bf03206926

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303-304. doi: 10.1126/science.270.5234.303

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention, Perception, & Psychophysics, 73*, 1195-1215. doi: 10.3758/s13414-011-0096-8

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1074-1095. doi: 10.1037/0096-1523.7.5.1074

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91*, 68-111. doi: 10.1037//0033-295x.91.1.68

Van Engen, K. J., Chandrasekaran, B., & Smiljanic, R. (2012). Effects of speech clarity on recognition memory for spoken sentences. *Plos One, 7*, e43753. doi: 10.1371/journal.pone.0043753

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics, 67*, 939-950. doi: 10.3758/Bf03193621

Warren, R. M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review, 92*, 574-584. doi: 10.1037//0033-295x.92.4.574

CURRICULUM VITAE

David M. Weintraub

EDUCATION

Fall 2009 – Summer 2012	University of Nevada, Las Vegas Master of Arts (Psychology)
Fall 2004 – Spring 2009	University of Nevada, Las Vegas Bachelor of Arts (Psychology)

PUBLICATIONS

Weintraub, D.M., & Snyder, J.S. (in press). Evidence for high-level feature encoding and persistent memory during auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*.

Ramage, E.M., **Weintraub, D.M.**, Vogel, S., Sutton, G., Ringdahl, E., Allen, D.N., & Snyder, J.S. (2015). Preliminary evidence for reduced auditory lateral suppression in schizophrenia. *Schizophrenia Research*, *162*, 269-275.

Weintraub, D.M., Metzger, B.A., & Snyder, J.S. (2014). Effects of attention to and awareness of preceding context tones on auditory streaming. *Journal of Experimental Psychology: Human Perception & Performance*, *40*, 685-701.

Snyder, J.S., & **Weintraub, D.M.** (2013). Loss and persistence of implicit memory for sound: Evidence from auditory stream segregation context effects. *Attention, Perception, & Psychophysics*, *75*, 1059-1074.

Ramage, E.M.*, **Weintraub, D.M.***, Allen, D.N., & Snyder, J.S. (2012). Evidence for stimulus-general impairments on auditory stream segregation tasks in schizophrenia. *Journal of Psychiatric Research*, *46*, 1540-1545. * = These authors contributed equally.

Weintraub, D.M., Ramage, E.M., Sutton, G., Ringdahl, E., Boren, A., Pasinski, A.C., Thaler, N., Haderlie, M., Allen, D.N., & Snyder, J.S. (2012). Auditory stream segregation impairments in schizophrenia. *Psychophysiology*, *49*, 1372-1383.

Snyder, J.S., Gregg, M.K., **Weintraub, D.M.**, & Alain, C. (2012). Attention, awareness, and the perception of auditory scenes. *Frontiers in Psychology*, *3*, 15.

Snyder, J.S. & **Weintraub, D.M.** (2011). Pattern specificity in the effect of prior Δf on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1649-1656.

Snyder, J.S., Holder, T., **Weintraub, D.M.**, Carter, O., & Alain, C. (2009). Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology*, *46*, 1208-1215.

ADDITIONAL TRAINING

Summer 2014	UC-Merced CHASE Summer School
Summer 2013	UC-Davis ERP Boot Camp
Fall 2012 – Spring 2015	Cleveland Clinic Luo Ruvo Center for Brain Health Job title: Research Intern

ASSOCIATIONS

Spring 2010 – Spring 2015	Association for Research in Otolaryngology (ARO)

HONORS AND AWARDS

Fall 2014 – Spring 2015	UNLV Foundation President's Graduate Research Fellowship Total: $23,000
Fall 2012	Edward Lovinger Scholarship Total: $2,000
Spring 2011	ARO Student Travel Award Total: $500
Spring 2010, 2011, 2012, 2015	Graduate & Professional Student Association Travel Grant Total: $2,250
Fall 2004 – Spring 2009	Millennium Scholarship Total: $10,000