2019-05-01

The Effects of Distracting Background Audio on Spontaneous Speech

Kacy Nicole Chapman
Brigham Young University

The Effects of Distracting Background Audio on Spontaneous Speech

Kacy Nicole Chapman

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Christopher Dromey, Chair
Shawn Nissen
Kathryn Cabbage

Department of Communication Disorders

Brigham Young University

ABSTRACT

The Effects of Distracting Background Audio on Spontaneous Speech

Kacy Nicole Chapman
Department of Communication Disorders, BYU
Master of Science

This study examined the changes that occur in spontaneous speech when speakers are distracted by background audio. Forty young adults answered open ended questions under five audio conditions (pink noise, movie dialogue, heated debate, classical music, and contemporary music) and a silent condition. Acoustic parameters assessed during the sessions included mean and variability of the fundamental frequency (F_0), mean and variability of intensity, speaking time ratio, and disfluency ratio. It was hypothesized that there would be significant increases in the mean and variability of F_0 as well as the mean and variability of intensity. There were statistically significant increases in mean and variability of intensity and mean F_0 across most audio conditions. There were no significant changes in variability of intensity in the pink noise condition and no significant changes in variability of F_0 in any audio condition. We hypothesized that the speaking time ratio would decrease in the presence of background audio compared to the silent baseline. Results demonstrate significant increases in speaking time ratio except for the classical music condition. It was expected that the disfluency ratio of speech production for each participant would increase in the presence of background audio, with informational masking demonstrating the most increase. Results revealed a significant increase in disfluency ratios across background audio conditions except for the pink noise and classical music conditions. Participants reported the heated debate and contemporary music to be the most distracting. These results have potential clinical implications regarding the type of environment where therapy is given, and what type of everyday situations might cause the most difficulties with fluency as well as the processing and production of speech.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

DESCRIPTION OF THESIS STRUCTURE

This thesis, *The Effects of Distracting Background Audio on Spontaneous Speech*, is written in a hybrid format. The hybrid format brings together traditional thesis requirements with journal publication formats, which allows for portions of this thesis to be published in articles citing the thesis author as a co-author. The introductory pages of this thesis reflect requirements for submission to the university. The body of the thesis conforms to length and style requirements for submission to a peer-reviewed journal in speech-language pathology. The annotated bibliography is included in Appendix A and Appendix B contains the research consent form. Appendix C contains a list of the monologue topics used during the experimental sessions. This thesis contains a reference list that includes references included in the journal-ready article.

Introduction

It is common to find ourselves attempting to communicate with others in less than ideal noise conditions. As researchers study the effects of various types of background noise on communication, more becomes known about a person's ability to modify their speech to improve intelligibility. Many of the studies have focused on a person's ability to perceive speech in distracting noise conditions, such as the Howell (2008) study, which focused on how intensity provides cues to distance if the sound is familiar (such as speech) and how auditory stream segregation helps listeners make use of information about harmonic structure to segregate sounds (Howell, 2008). The distracting noise environment that is present when we are listening to someone speak can be considered a form of masking. In this context, masking can usually be divided into two types: energetic and informational. Energetic masking refers to a listening situation in which competing noise overlaps in time and frequency in a way that parts of the speech signal become inaudible. Informational masking differs in that it describes a listening situation where the listener is unable to separate the target signal elements from the similar-sounding distracters (Brungart, Simpson, Ericson, & Scott, 2001). This means that there is involuntary processing of language that is unrelated to the intended signal. An example of this would be an attempt to listen to a friend speak while someone nearby is also talking loudly. This would make it hard to completely focus on the friend while involuntarily processing the other person's speech. The listener has to selectively attend to the message signal, while consciously attempting to avoid distraction from the linguistic components of the informational masking they are being exposed to. For this reason, informational masking is known to cause a greater cognitive load and requires higher-level processing (Meekings et al., 2016) to selectively focus on the intended signal without processing the distracting information. The idea that

informational masking causes a greater cognitive load means that it could also interfere with the listener's control of their own speech. In the current study, we attempt to investigate these effects on speech of processing distracting information while speakers are answering open ended questions.

The effect of masking on speech production has not been frequently studied, with few reports focusing on spontaneous speech tasks, which would have more natural ecological validity than rote speech tasks (Hazan & Baker, 2011). For example, one study found that fundamental frequency, intensity, and articulatory parameters (lip spreading, lip aperture, and inter-lip area) showed a general increase in magnitude as noise increased, and speakers would enhance these parameters in content words more than function words (Garnier, Dohen, Loevenbruck, Welby, & Bailly, 2006). One study reported changes in articulatory kinematics when speakers were exposed to qualitatively different types of noise that included both energetic and informational masking (Dromey & Scott, 2016). The participants in this study completed a sentence repetition task (prepared speech) rather than spontaneous speech tasks with communicative intent. The current work is motivated by the studies that preceded it, as it uses a variety of masking types to interfere with the completion of different speech tasks, in order to learn of the effects of masking on selected acoustic aspects of speech.

The Lombard effect is a predictable behavioral response where speakers involuntarily increase their vocal amplitude in the presence of noise. It can be measured acoustically, perceptually, and in other ways. Studies on the Lombard effect have laid the groundwork for further research on speech production in noise. Lombard speech demonstrates the flexibility humans have in their acoustic communication systems to allow their speech to be received appropriately under challenging noise conditions (Zollinger & Brumm, 2011). Specifically, it

means that in human communication, intensity of speech will increase as masking noise increases to ensure the speaker is able to hear their own voice. In the Howell (2008) research review, it was discussed that this voice level regulation could be due to a negative feedback mechanism (the brain sends signals to regulate voice level based on the incoming information) for listener perception. Fundamental frequency can also increase with the Lombard effect, which could help with speech intelligibility in a noisy environment (Howell, 2008). Although Lombard speech is considered a predictable phenomenon that can occur under noise conditions, studies still have found that it varies between genders as well as from speaker to speaker. One study (Junqua, 1993) found that inter-speaker variability due to an increase in vocal effort was found to be more significant for female speakers. In this same study, an acoustic analysis was performed at the phonemic level, which revealed that vowels were more emphasized while consonants were shortened and more distorted during Lombard speech. These findings have been supported in contemporary studies highlighting interspeaker variability in the Lombard effect (Marxer, Barker, Alghamdi, & Maddock, 2018). However, it leads to more questions such as the commonalities in Lombard speech among speakers and how it varies based on the type of noise condition a person is exposed to, and the type of speech tasks being performed (e.g., reciting sentences vs. having a conversation).

A past study examined Lombard speech in spontaneous sentence-level productions to observe any acoustic modifications to the utterances in noise. Specifically, the study examined acoustic modifications in noise and how word type affects those modifications (Patel & Schell, 2008). It was found that duration was enhanced for information-bearing content words relative to function words or content words conveying less information. It makes sense that when speakers are trying to convey meaning, they will put more vocal effort on key details. Elevated

pitch also served as a primary cue for marking information-bearing word types, especially agent words. In a similar study, speakers manipulated specific intensities in words to increase the contrastivity of adjacent syllables in the presence of noise. Specifically, they increased intensity contrastivity in their strong to weak syllabic productions to improve word recognition for the listener (Arciuli, Simpson, Vogel, & Ballard, 2014). Both of these studies provide us with information about how speakers acoustically modify their speech to improve intelligibility, with the Patel and Schell (2008) study being an important precursor to the present investigation as it specifically focuses on acoustic modifications in spontaneous speech.

A number of studies have examined the interference effects of informational versus energetic masking in different contexts ranging from speech perception to the ability to perform a cognitive task (Brungart, 2006; Cooke, Garcia Lecumberri, & Barker, 2008; Lidestam, Holgersson, & Moradi, 2014; Mattys, Carroll, Li, & Chan, 2010; Meekings et al., 2016). One recent study measured changes in articulatory kinematics under several different noise conditions to examine movement variability as a function of the masking type while performing a repetitive speaking task (Dromey & Scott, 2016). Speakers were only minimally influenced by auditory distractions in a controlled environment, using repetitive stimuli rather than a more naturalistic speaking task. Acoustic analysis, which will be used in the present study, demonstrated that speech rate was the only change that could be linked to the distracting influence of hearing another person's speech while talking. The repetitive and non-communicative task could have obscured possible interference effects from hearing another's speech, which is why it would be valuable to include spontaneous speech tasks instead of rote speech tasks to further analyze effects of different types of masking, as natural speech involves higher linguistic processing demands as the speaker has to generate spoken language while being exposed to distractors.

Acoustic features of speech have also been examined when speakers were presented with energetic and informational masking during noncommunicative and communicative tasks. Lu and Cooke conducted two previous studies investigating the effects of informational and energetic masking on speaker intelligibility and temporal characteristics (Lu & Cooke, 2008; Lu & Cooke, 2009). In the 2008 study the authors found that in noisy, compared to quiet backgrounds, there were increases in intensity, frequency, sentence duration, spectral center of gravity, pauses before talking, and the voiced-to-unvoiced ratio at the utterance level. The pause before speaking was longer in the single-talker background than in the other babble conditions. This study provided evidence that noise-influenced speech may be more intelligible than speech produced in quiet as speakers attempt to compensate for environmental conditions. However, the tasks used in this study lacked communicative intent, suggesting that speakers had less motivation to change their speech even when the masking noise was present. A task involving communicative intent could have brought to light more complete information on the masking effects. In their 2009 study, Lu and Cooke found little evidence to support the idea that speakers can modify their speech by possibly adopting active timing strategies in order to overlap least with intermitting masking (timing speech to occur when speakers perceive the least amount of distracting noise). Again, this could be due to the repetitiveness and lack of naturalness in the speech tasks.

These studies led to the Lu and Cooke 2010 study, which has several similarities to the present investigation. The researchers examined communicative, spontaneous speech tasks compared to rote tasks under different masking types. Competing speech (four speakers) was considered informational and speech modulated noise (speech-shaped noise modulated with the temporal envelope of speech) was considered energetic. This study revealed that speakers were

able to reduce temporal overlap (speaking at the same time as the main acoustic events in the masking) more efficiently when competing speech was present as compared to speech modulated noise. Speakers seemed to time their speech to reduce temporal overlap in an intelligible speech background more efficiently than in a modulated noise background. Talkers had longer and a greater number of pauses in the competing speech background, and very frequent short pauses in the speech modulated noise background. These findings were also more prominent in the communicative than noncommunicative tasks. The communicative tasks led to more contrast in vowel space (a greater distinction between vowel types) in the conversational task relative to the speaking alone task, which has been associated with an attempt to improve intelligibility. Word durations were shorter in the communicative tasks, and intensity was higher. This work not only suggests that temporal changes in speech were most significant under informational masking conditions, but that talkers adopt a "listening-while-speaking" (Lu & Cooke, 2010) strategy which suggests an ability to modify speech to help the conversational partner understand in noisy backgrounds. This study revealed how talkers can reduce informational masking effects, which significantly differed depending on task type and the intelligibility of the masking stimulus.

A recent fMRI study of how neural systems are recruited under different masking conditions revealed a dominant cortical effect of informational masking (the sensorimotor auditory portion of the cortex was activated) during speech production, suggesting that unattended speech is still processed (Meekings et al., 2016). It was found that the superior temporal cortex might act as an auditory error monitor during talking and that responses in the bilateral superior temporal gyrus were greatest when the participants spoke in noise as compared to the quiet condition. Also, the superior temporal gyrus suppressed activation when speaking in the quiet condition relative to the listening without speaking task. These responses were related

more to informational masking and this study supported the idea that it is not a reduced ability to self-monitor but rather the processing of unattended speech at a high cortical level that causes the intrusions into speech production. It would be valuable to further study the effects of various types of informational masking on speech during communicative tasks to see how each influences speech production in different ways.

Studies examining natural speech and masking have not focused specifically on informational versus energetic masking effects. However, the studies that do exist still contribute to the foundational understanding of how natural speech can be affected by a specific masking type. In one study from the past decade researchers studied how speakers adapted their speech based on the type of adverse listening condition (Hazan & Baker, 2011). During conversational speech, talkers heard their conversational partner through a three-channel vocoder that spectrally degraded the speech by removing much of the perceived pitch, or they heard their partner through babble speech. These two listening conditions were implemented to be compared to communicating with someone with a cochlear implant (vocoder condition) and someone in a noisy environment (babble condition). The results showed that in both spontaneous speech tasks, adverse listening conditions for the partner caused the speaker to adapt their speech to the situation so that they would be perceived correctly. It was found that both babble and vocoder noise conditions caused adaptations, but the babble condition resulted in greater changes in F_1 range and mean energy than vocoder noise. This provides evidence for the idea that the way the environment affects sound is not always detrimental and speaking clearly can help counterbalance poor acoustic characteristics of the environment. The tasks the speakers completed had communicative intent but did not allow for speakers to have flexibility in topics and syntax. However, the study moved away from examining non-communicative tasks and

more towards examining natural speech with communicative intent. Hazan and Baker (2011) found that conversational speech was more finely modulated than speech without communicative intent, as evidenced by speakers varying their strategies and matching their speech to the listeners' needs.

The previous studies on acoustic features of speech and how informational versus energetic masking affects these features have laid the groundwork for the present investigation. Only recently have studies examined the effects of masking on natural speech, and how a person implements strategies to ensure understanding by the listener. Unlike previous studies, the present study explored the effect of different types of noise conditions on speakers when conversing spontaneously and with communicative intent. The study was designed to examine how individuals modify their speech when presented with different types of intelligible informational masking. While previous studies have examined fewer types of masking and measured a limited range of acoustic parameters, the current study considers a speaker's fluency and other speech modifications.

Method

Participants

Twenty male and 20 female native speakers of American English participated in the study. The mean age was 24.6 years ($SD = 1.9$) for the men and 23.7 ($SD = 1.3$) years for the women. None of the participants reported any history of language, speech, or hearing disorders. Each passed a hearing screening bilaterally at 25 dB HL at 500, 1000, and 2000 Hz. Participants signed an Institutional Review Board-approved written consent to participate in the experiment.

Instrumentation

Each participant sat in a sound attenuating booth to provide an optimal environment to make high quality acoustic recordings and reduce auditory distractions outside of the presented stimuli. The participants were exposed to the experimental audio conditions through headphones, and their speech was recorded with a boom microphone approximately 50 cm from the mouth. A sound level meter was placed 100 cm from the mouth to allow a reference recording to subsequently calibrate speech intensity during acoustic analysis of the microphone signal. This signal was digitized with a FocusRite Scarlett 2i2 USB analog to digital converter at 44,100 Hz and Audacity software (Version: 2.3.1, Audacity Team, 2019). To establish the intensity level of the stimuli, the pink noise stimulus was perceptually matched to the loudness of masking noise from an audiometer at 75 dB HL. After the intensity level was established, all of the stimuli were equalized in amplitude to the pink noise using Audacity.

Procedures

The data for each participant were collected within a one-hour session. This experiment was part of a larger study that involved other speaking tasks. Participants were given a list of open-ended questions and were asked to circle 8-10 topics they felt comfortable speaking about. The questions that were used to prompt the monologues were verbally presented to the participants by the experimenter, and they were asked to answer at a comfortable speaking rate and loudness. Each participant answered these open-ended questions as a monologue while being exposed to the different audio stimuli. The listening conditions presented were as follows: a silent-baseline condition (ST), pink noise (PK), dialogue from a movie (MV), two speakers having a debate (DT), classical music (CL), and contemporary music (CO). The dialogue was taken from a contemporary movie in which the scene was intense and emotional, preceding a

violent act. In the debate stimulus, two sports commentators argued about the merits of two basketball players. The classical music was characterized by a wide dynamic range, and it included both instrumental and vocal components. The contemporary music stimulus was a lively and rhythmic song. Any pauses in the stimuli longer than 200 ms were removed to ensure continuity. In order to better understand the responses of the participants, they were asked about which experimental condition they found to be the most distracting. Answers were qualitatively assessed to learn more about the participants' subjective experience.

Data Analysis

The digital recordings were analyzed using the Praat software program (version 5.4; Boersma & Weenink, 2014). Pauses in between questions, nonspeech behaviors (coughing, laughing, etc.), and experimenter speech were removed from the recordings before analysis.

Acoustic parameters. Acoustic measures of connected speech, including characteristics of intensity, fundamental frequency (F_0), and the proportional amount of time participants spoke during their responses were computed and analyzed to quantify features of speech performance. F_0 was computed by taking the mean *(M)* and standard deviation *(SD)* under each experimental condition, with F_0 being manually edited when necessary to overcome tracking errors. Praat provided a voicing report with the *M* and *SD* in Hz. Sex differences were accounted for in F_0 variability by converting the *SD* in Hz into semitones with a spreadsheet equation. Intensity variables were computed using the *M* and *SD* of the monologue in each experimental condition. In order to avoid including pauses or nonspeech sounds in the intensity measurements, a dB floor was selected based on the intensity level of the softest speech sounds in each response. The Praat intensity listing was exported as a comma separated values file (csv) which was brought

into a Matlab application (version 9.0; The Mathworks, Inc., 2016) to compute the *M* and *SD* of the intensity above the dB floor.

Speaking time ratio was measured to analyze the proportion of time a participant was speaking versus the amount of time they were pausing. It was expressed as a proportion with 1.0 representing all speaking with no pauses, 0.75 would be 75% speaking and 25% pausing, and so forth. A Matlab application was used to measure this speaking time ratio, which recognized pauses as being longer than 200 ms. The application normalized the intensity of the file to 100 (arbitrary units). The threshold was ten percent of the normalized maximum, with amplitude values below being defined as pausing, and above being defined as speaking.

Fluency. Fluency characteristics (pausing, hesitations, sound repetitions, and fillers) were analyzed perceptually. Disfluencies were counted and then divided by the total number of words spoken in each experimental condition. The result was then multiplied by 100 to reveal the disfluency percentage. These measures were made in order to better understand how the experimental audio conditions might influence a speaker's conversational fluency.

Statistical Analysis

A repeated measure analysis of variance (ANOVA) was used to test for significant changes in the dependent measures under the different stimulus conditions. When the Mauchly test revealed violations of the assumption of sphericity, Huynh-Feldt corrections were then applied which then resulted in noninteger degrees of freedom. Concurrent contrasts compared the performance under the individual stimulus conditions against the silent baseline condition in order to determine which led to significant changes in the dependent measures. Speaker sex was included as a factor to evaluate potential interactions with the experimental condition or between-subject effects due to differences in performance between women and men.

Results

Descriptive statistics for the dependent measures can be found in Table 1. Significant differences across the experimental conditions are reported below, with the ANOVA details in Table 2 and the concurrent contrast results in Table 3.

Fundamental Frequency

Overall, there were significant changes in mean fundamental frequency across the conditions. The contrasts analysis testing revealed that the fundamental frequency significantly increased in each audio condition as compared to the silent condition. The standard deviations of the fundamental frequency in semitones did not change in the presence of the audio stimuli, with no significant contrasts compared to the silent baseline. Mean F_0 was significantly higher for women than for men.

Intensity

ANOVA testing revealed significant changes in intensity across the experimental conditions. Contrast analysis revealed an increase in intensity under each condition with $p < .001$ and effect size ranging from 0.277 to 0.779 for each contrast. ANOVA testing demonstrated a significant change in the standard deviation of intensity, with contrast analysis revealing significant differences in standard deviation of intensity across experimental conditions compared to the silent baseline except in the pink noise condition. There were no sex differences or interaction effects.

Speaking Time Ratio

There were significant changes in speaking time ratio. Contrast analysis revealed significant ratio increases for each audio condition compared to the silent baseline except for the classical music condition. There were no sex differences or interaction effects.

Table 1

Descriptive Statistics for the Acoustic Measures by Sex and Condition

Condition		Silent		Pink Noise		Movie		Debate		Classical		Contemporary	
	Sex	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
F_0	F	204	11.8	212	12.0	206	12.8	208	12.7	208	14.4	213	17.6
	M	113	16.1	125	17.2	119	16.6	122	17.2	122	19.5	130	26.3
STSD	F	2.52	0.48	2.56	0.56	2.62	0.64	2.55	0.53	2.56	0.60	2.70	0.57
	M	2.33	0.52	2.44	0.53	2.38	0.51	2.48	0.73	2.58	0.58	2.42	0.57
DB	F	68.9	4.53	72.5	3.56	70.3	3.63	71.0	3.83	70.1	3.80	71.7	3.92
	M	68.8	5.16	72.0	5.48	69.7	5.71	70.6	5.55	69.7	5.72	71.0	5.50
DBSD	F	5.19	1.10	6.21	1.08	5.41	1.08	5.34	0.96	5.49	1.06	5.95	0.91
	M	5.35	0.75	6.18	0.99	5.47	0.82	5.66	0.76	5.65	0.72	6.06	0.86
STR	F	0.74	0.09	0.80	0.06	0.79	0.09	0.78	0.11	0.76	0.07	0.80	0.09
	M	0.72	0.09	0.76	0.11	0.78	0.08	0.78	0.09	0.76	0.09	0.79	0.11
DSR	F	4.4%	2.8%	4.8%	2.3%	5.42%	3.1%	5.8%	2.8%	4.9%	2.6%	6.1%	3.2%
	M	5.2%	3.4%	5.3%	3.2%	5.92%	3.0%	6.9%	4.2%	6.0%	3.3%	6.7%	4.9%

Note. F = female; M = male; F_0 = mean fundamental frequency; STSD = semitone standard deviation; DB = mean intensity; DBSD = intensity standard deviation; STR = speaking time ratio; DSR = disfluency ratio

Table 2

ANOVA Results for Each Acoustic Measure

	df	df error	F	*p*	ES
F_0	5	190	16.757	< .001	.306
STSD	4.975	161.413	1.004	.417	.026
DB	5	190	52.348	< .001	.579
DBSD	5	190	29.742	< .001	.439
STR	5	190	6.489	< .001	.146
DSR	5	190	5.765	< .001	.132

Note. ES = effect size

Table 3

Contrast Statistics for Each Condition Compared to the Silent Baseline

	Pink Noise			Movie			Debate			Classical			Contemporary		
	F	*p*	ES	F	*p*	ES	F	*p*	ES	F	*p*	ES	F	*p*	ES
F_0	53.5	< .001	.59	12.89	.001	.25	21.89	< .001	.37	14.9	< .001	.28	29.6	< .001	.44
STSD	1.41	.243	.04	1.15	.291	.03	1.20	.281	.03	2.97	.093	.07	2.32	.136	.06
DB	134.3	< .001	.78	18.21	< .001	.32	52.1	< .001	.58	14.6	< .001	.28	61.8	< .001	.62
DBSD	74.7	< .001	.66	2.68	.11	.07	5.41	.025	.13	7.98	.007	.17	37.5	< .001	.5
STR	13.1	.001	.26	13.71	.001	.27	11.2	.002	.23	2.82	.101	.07	21.3	< .001	.36
DSR	0.50	.482	.01	4.33	.044	.10	15.5	< .001	.29	2.59	.116	.06	12.2	.001	.24

Note. degrees of freedom = 1, 38 for each stimulus condition

Fluency

There were significant changes in the disfluency ratios across the experimental conditions. All but two of the experimental conditions involved a significant increase in disfluency ratios; the exceptions were the pink noise and classical music conditions. There were no sex differences or interaction effects.

Participant Questionnaire

Results of the participant questionnaire can be found in Table 4. The heated debate and contemporary music were reported to be the most distracting for all participants out of the six experimental conditions.

Table 4

Participant Responses for the Most Distracting Stimulus Condition

	Silent	Pink Noise	Movie	Debate	Classical	Contemporary
Male	0%	0%	5%	55%	0%	50%
Female	0%	5%	25%	45%	0%	55%
Total	0%	3%	15%	50%	0%	53%

Note. Eight participants reported two stimulus conditions being equally distracting, causing the percentages to equal more than 100%.

Discussion

The purpose of the current study was to determine the effects of different listening conditions on speech production. The data revealed several significant effects on acoustic metrics and on fluency.

Fundamental Frequency and Intensity

It was expected that the intensity and F_0 would increase in each participant's speech due to the Lombard effect, where intensity and F_0 increase as background audio increases. The results showed that this occurred in the present study, even though participants had been instructed to talk at a comfortable conversational pitch and loudness. A rise in F_0 in the presence

of noise is consistent with previous work showing evidence that an increase in F_0 is directly associated with an increase in intensity (Garnier et al., 2006; Junqua, 1993; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). The most significant increases in mean F_0 relative to the silent baseline condition were under the pink noise, debate, and contemporary music conditions. Women had a higher average F_0 compared to men, which was expected because of anatomic differences in the larynx (Pernet & Belin, 2012). Both men and women demonstrated a higher average F_0 under the pink noise and contemporary music conditions. The pink noise condition also caused the most significant increase in intensity relative to every other experimental condition. It is possible that because the pink noise condition was presented first after the silent baseline, participants were adjusting to speaking with audio stimuli in the background whereas they became more accustomed to auditory distractions as the following conditions were presented. This finding suggests that the Lombard effect may be most pronounced when the energetic component of the masking is increased, as the pink noise caused the greatest increase in intensity by far, which is consistent with previous findings that intensity increased the most in speech shaped noise compared to speech modulated noise and competing speakers (Lu & Cooke, 2010). Because the Lu and Cooke (2010) study mostly examined the effects of different types of energetic masking, the present study was beneficial in comparing energetic and informational masking.

There were no significant differences in F_0 variability (STSD) across the experimental conditions. It was hypothesized that there might be an increase in variability as the background audio conditions became more distracting, but the changes were small and inconsequential. The reason for this remains unclear. The standard deviation of intensity changed significantly. Compared to the silent baseline condition, every experimental condition other than the movie

dialogue led to significantly greater variability in intensity, with the pink noise and contemporary music conditions demonstrating the largest effect sizes. A previous study found that speakers significantly manipulated specific intensities in words to increase contrastivity of adjacent syllables when exposed to background noise as compared to a silent baseline (Arciuli et al., 2014). These results support the suggestion that in the presence of background audio, speakers in the present study may have increased intensity on specific words in order to increase intelligibility. This could potentially explain the significant increase in variability of intensity across audio stimuli compared to the silent baseline.

Speaking Time Ratio

Compared to the silent baseline, all experimental audio conditions except for classical music were associated with significant increases in speaking time ratio. The largest effect size for the change from the silent baseline condition was for the contemporary music condition. It was hypothesized that speaking time ratio would decrease as a result of the presence of informational masking as compared to energetic masking or the silent baseline condition. We reasoned that when a person is distracted, they would be more likely to pause because of the involuntary linguistic distractions associated with an informational masking condition and would need to be more deliberate about what to say next.

Previous studies, although limited due to their use of rote speech tasks, have been consistent with this hypothesis (Lu & Cooke, 2008), with findings that sentence duration and the number of short pauses in rote speech tasks increased during the presence of masking compared to a silent baseline condition. However, the present results reflected a different effect. For male participants, the speaking time ratios increased in every audio condition compared to the silent baseline, with more of an increase in the informational masking conditions. For the female

participants, there was also an increase in speaking time ratio under each audio condition compared to the silent baseline. However, the pink noise and contemporary music conditions demonstrated the greatest increase with the heated debate and classical music demonstrating a milder increase. Two potential explanations can be suggested for the increase in speaking time ratio compared to the silent baseline. One is based on a previous study that found reduced sentence duration and number of short pauses when background noise was increased. The study suggested that an increase in speaking time ratio could be caused by a "sense of urgency on the part of the speaker, which occurs due to the persistent exposure of the environmental noise to the speaker's ears" (Varadarajan & Hansen, 2006, p. 938). The second possibility could be that specific audio stimuli were more familiar or distracting depending on each individual, causing different effects on their speaking time ratios.

Fluency and Participant Questionnaire

It was hypothesized that the disfluency ratio would increase more with informational masking as compared to energetic masking. It was also hypothesized that the disfluency ratio might be associated with each participant's opinion about which experimental condition they found the most distracting. The two experimental conditions where the most disfluencies were present were the heated debate and contemporary music, which were the two rated as the most distracting experimental conditions. Based on the responses of the participants, it could be speculated that the heated debate about sports was distracting due to the emotionally charged words and the constant arguing with minimal pausing. More male participants reported the heated debate as distracting, and many stated it was because it was familiar, and they wanted to listen to it. For the contemporary music, male and female participants reported the song was distracting because it was familiar, appealing, and they wanted to sing along. From these

responses, it can be speculated that informational masking that is familiar or interesting to the listener will cause the most disfluencies in speech production.

Limitations of Present Study

The experimental conditions were only presented at one dB level, which provided only limited amount of information on masking effects, whereas testing with a range of dB levels could allow for more insights on the effects of informational and energetic masking presented in different modes. Another limitation occurred when data were collected from the first two participants. The microphone was not attached to the right connection, causing some deficiencies in the audio quality of the recording. Although this was accounted for through dB calibration and removal of artifact noise, it was not an ideal circumstance. Another limitation of the present study was that the same sequence of audio conditions was presented to all of the participants. This could have led to a sequence effect in the results. The contemporary music had both energetic (consistently loud music) and informational components (the lyrics). This did not permit a distinct separation of each type of masking. The classical music had a wide dynamic range of intensity, and the softer parts could have been less engaging to the listener. Lastly, there could have been some variability in the intensity measures because participants were able move their heads during the recordings, which could have affected the mouth-to-microphone distance.

Directions for Future Research

Recordings of the experimental stimuli were time-aligned in a second channel along with the recordings of the participants' speech. While the present study focused solely on the acoustic characteristics and fluency of speech production, it would be beneficial in a follow-up study to analyze the timing of the participants speech relative to the stimuli being played to assess spectral and/or temporal overlap and build on previous studies investigating whether speakers

can time their productions during pauses in the background audio (Lu & Cooke, 2009). It would also be valuable to examine how individuals with mild to severe traumatic brain injuries would perform in a similar task, because of the difficulties in this population with selective attention.

Conclusion

The present study revealed several changes in acoustic measures of speech and fluency that occur due to the presence of different types of distracting background audio. There were significant changes in acoustic parameters including mean F_0, mean intensity, speaking time ratio, as well as an increase in disfluency in spontaneous speech across audio conditions compared to the silent baseline condition. The disfluency ratio was related to what the speakers' rated as the most distracting audio stimuli, with informational masking that is emotional or familiar being the most distracting. The study, however, revealed an increase in speaking time ratios relative to an increase in distracting background audio which was different than what was predicted. It would be beneficial for future studies to analyze speaking time ratios of spontaneous speech with background audio conditions to confirm these findings. The present study has possible clinical implications for the type of environment therapy can take place in, and what type of everyday situations will cause the most difficulties with fluency as well as processing of and production of speech.

References

Arciuli, J., Simpson, B. S., Vogel, A. P., & Ballard, K. J. (2014). Acoustic changes in the production of lexical stress during Lombard speech. *Language and Speech*, *57*, 149-162. doi:10.1177/0023830913495652

Audacity Team (2019). Audacity: Free audio editor and recorder (Version 2.3.1) [Computer program].

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer (Version 5.4) [Computer program].

Brungart, D. S. (2006). Informational and energetic masking effects in multitalker speech perception. In P. Divenyi (Ed.), *Speech separation by humans and machines* (pp. 261-267). doi: 10.1007/0-387-22794-6_17

Brungart, D., Simpson, B., Ericson, M., & Scott, K. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America, 109*, 1101–1109. doi:10.1121/1.1345696.

Cooke, M. P., Garcia Lecumberri, M. L., & Barker, J. P. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America, 123*, 414–427. doi:10.1121/1.2804952

Dromey, C., & Scott, S. (2016). The effects of noise on speech movements in young, middle-aged, and older adults. *Speech, Language, and Hearing, 19*, 131-139. doi:10.1080/2050571X.2015.1133757

Garnier, M., Dohen, M., Loevenbruck, H., Welby, P., & Bailly, L. (2006, December). *The Lombard effect: A physiological reflex or a controlled intelligibility enhancement?* Paper presented at the 7th International Seminar on Speech Production*,* Ubatuba, Brazil.

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America, 130*, 2139-2152. doi:10.1121/1.3623753

Howell, P. (2008). Effect of speaking environment on speech production and perception. *Journal of the Human-Environment System, 11*, 51-57. doi:10.1618/jhes.11.51

Junqua, J. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America, 93*, 510-524. doi:10.1121/1.405631

Lidestam, B., Holgersson, J., & Moradi, S. (2014). Comparison of informational vs. energetic masking effects on speechreading performance. *Frontiers in Psychology*, *5*, 1-7. doi:10.3389/fpsyg.2014.00639

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise*. Journal of the Acoustical Society of America, 124*, 3261-3275. doi:10.1121/1.2990705

Lu, Y., & Cooke, M. (2009). Speech production modifications produced in the presence of low-pass and high-pass filtered noise. *Journal of the Acoustical Society of America, 126*, 1495-1499. doi:10.1121/1.3179668

Lu, Y., & Cooke, M. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *Journal of the Acoustical Society of America, 128*, 2059-2069. doi:10.1121/1.3478775

Marxer, R., Barker, J., Alghamdi, N., & Maddock, S. (2018). The impact of the Lombard effect

    on audio and visual speech recognition systems. *Speech Communication, 100*, 58-68,

    doi:10.1016/j.specom.2018.04.006

MathWorks. (2016). MATLAB (Version 9.0) [Computer software].

Mattys, S. L., Carroll, L. M., Li, C. K., & Chan, S. L. (2010). Effects of energetic and

    informational masking on speech segmentation by native and non-native speakers.

    Speech Communication, 52, 887-899. doi:10.1016/j.specom.2010.01.005

Meekings, S., Evans, S., Lavan, N., Boebinger, D., Krieger-Redwood, K., Cooke, M., & Scott, S.

    K. (2016). Distinct neural systems recruited when speech production is modulated by

    different masking sounds. *Journal of the Acoustical Society of America, 140*, 8-19.

    doi:10.1121/1.4948587

Patel, R., & Schell, K. W. (2008). The influence of linguistic content on the Lombard effect.

    Journal of Speech, Language, and Hearing Research, 51, 209-220. doi:10.1044/1092-

    4388(2008/016)

Pernet C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization.

    Frontiers in Psychology, 3, 1-11. doi:10.3389/fpsyg.2012.00023

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects

    of noise on speech production: Acoustic and perceptual analyses. *Journal of the*

    Acoustical Society of America, 84, 917-928. doi:10.1121/1.396660

Varadarajan, V. S., & Hansen, J. H. L. (2006, September 17-21). *Analysis of Lombard effect*

    under different types and levels of noise with application to in-set speaker ID system.

    Paper presented at the International Conference on Spoken Language Processing,

    Pittsburgh, PA (pp. 937–940). Redhook, NY: Curran Associates, Inc.

Zollinger, S. A., & Brumm, H. (2011). The evolution of the Lombard effect: 100 years of

psychoacoustic research. *Communication and Social Behaviour Group, 148*, 1173–1198.

doi:10.1163/000579511X605759

APPENDIX A

Annotated Bibliography

Arciuli, J., Simpson, B. S., Vogel, A. P., & Ballard, K. J. (2014). Acoustic changes in the
production of lexical stress during Lombard speech. *Language and Speech*, *57*, 149-162.
doi:10.1177/0023830913495652

Objective: The purpose of this study was to evaluate the production of contrastive lexical stress
during Lombard speech. This was done through objective, quantitative measures of the acoustic
speech signal.
Methods: 25 female and 2 male young adults recited two sentences with at least one trisyllabic
word twice within each noise condition: a quiet condition, white noise, and multi-talker babble.
Overall vocal intensity, speaking rate, syllable duration, vowel intensity, and vowel F_0 were
analyzed.
Results: Vocal intensity significantly increased from quiet to both noise conditions. Speaking
rate findings were not significant. The effect of interaction between stress pattern and noise was
significant, but only for the quiet- noise comparison in regard to vowel F_0, vowel intensity and
syllable duration. Results showed that speakers manipulated specific intensities in words to
increase contrastivity of adjacent syllables. In F_0, contrastiveness increased from quiet
conditions to noise conditions for the strong to weak syllabic words, but the reciprocal would
occur for the weak to strong syllabic words.
Conclusions: The participants were found to alter their amount of contrastivity in the production
of lexical stress while speaking in the noise conditions. Specifically, they increased contrastivity
in their strong to weak syllabic productions in regard to vocal intensity.
Relevance to the current work: This work is believed to be the first of its kind, and no other
studies have investigated lexical stress in Lombard speech. It brought forth more evidence and
information on finer-grained effects of speaking in noise

Dromey, C., & Benson, A. (2003). Effects of concurrent motor, linguistic, or cognitive tasks on
speech motor performance. *Journal of Speech, Language, and Hearing Research, 46*,
1234-1246. 10.1044/1092-4388(2003/096)

Objective: The purpose of this study was to compare three different distractor tasks to assess
their influence on speech movements. The distractor tasks were meant to require different
processing resources, which could then cause different effects on motor performance.
Methods: 10 male and 10 female young adults participated in one speech-only task and three
speech tasks performed concurrently with either a cognitive, linguistic, or motor task. In the
speech-only task, the speaker would repeat an 8-word sentence 15 times. In the linguistic task,
the last word the speaker produced was a verb related to the target nouns presented via a
loudspeaker. In the cognitive task, the last word in the sentence was a number in a sequencing
task. The motor task dealt with following directions to attach a nut and washers to a bolt.
Results: Duration, displacement and velocity, correlation of upper and lower lip displacement,
and spatiotemporal index (STI) were all analyzed. Lower lip displacement and velocity
significantly decreased but there were no changes in the STI in the motor distractor task

compared to the speech-only task. The STI for the lower lip significantly increased in the linguistic distractor task, and men had greater increases in STI than women. The linguistic distractor task was also associated with a strong negative correlation between the upper and lower lip. For the cognitive distractor task, participants spoke significantly faster, STI for the lower lip increased, and men had a greater increase in STI than women; they also had a strong negative correlation between the upper and lower lip.

Conclusions: It was found that the motor distractor task did not affect labial kinematics in the same way the other two distractor tasks did, meaning its effect was qualitatively different on speech movements. Because of the observed gender interactions, it may be that men cannot divide their attention between linguistic and cognitive distractions and speech as efficiently as women. Ultimately, the finding that cognitive, linguistic, and motor demands have a significant effect on even normally speaking individuals could give more insight to clinical intervention to those who have communication disorders.

Relevance to the current work: The current work seeks to further understand the impact of auditory distractors on acoustic measures of speech.

Dromey, C., & Scott, S. (2016). The effects of noise on speech movements in young, middle-aged, and older adults. *Speech, Language, and Hearing, 19*, 131-139. 10.1080/2050571X.2015.1133757

Objective: The purpose of this study was to observe changes in articulatory kinematics when speakers were exposed to qualitatively different types of noise. This was done by measuring changes in articulatory kinematics under several different noise conditions to examine movement variability depending on the given noise condition.

Method: 30 male and 30 female (10 per age group: 20-30, 40-50, 60-70) American English speakers were asked to repeat a sentence in a block of 15 tokens under each noise condition. Measurements taken included sound pressure level, lip displacement and velocity, utterance duration, lip coordination, and movement stability. ANOVA was also used to evaluate the statistical significance of differences in the dependent variables in each noise condition.

Results: The lower lip spatiotemporal index, peak velocity of the closing gesture, velocity peak count, and intensity of each speaker varied significantly across all noise conditions. The utterance duration during the 1-talker noise condition was significantly shorter compared to silence. In regard to age and sex differences, men had stronger negative bilabial correlations than women and men had significantly higher velocities than women across all noise conditions.

Conclusions: Speakers seem to be only minimally influenced by auditory distractions in a controlled environment. Speech rate was the only change that could be linked to the distracting influence of hearing speech while talking. The repetitive task could have masked possible interference effects from hearing another's speech. More naturalistic tasks could allow these effects to be revealed.

Relevance to the current work: This study sets a foundation for the current work, as the current work further examines potential interference effects during a naturalistic task under different noise conditions.

Garnier, M., Dohen, M., Loevenbruck, H., Welby, P., & Bailly, L. (Dec 2006). The Lombard effect: A physiological reflex or a controlled intelligibility enhancement? *7th International Seminar on Speech Production,* Ubatuba, Brazil, December, 2006.

Objective: This study aimed to investigate whether speech in noisy environments caused articulatory and acoustic modifications or any changes specific to units within the utterance. *Methods:* A female native speaker of French spoke 33 short sentences in a subject-verb-object structure in silence, white noise, and "cocktail party" noise. French was evaluated as it contains prosodic differences between function words and content words as well as an F_0 declination across the utterance which could also act as an intelligibility cue. Audio and articulatory signals were recorded, including analysis of lip aperture, lip spreading, inter-lip area, and mean amplitude of articulatory movements.
Results: Acoustic and articulatory parameters showed a general increase in noise. It was found that all articulatory and durational amplitude parameters were more enhanced in noise for content word-final syllables than function word syllables or content word-initial syllables. Content word-initial syllables were found to be produced with more open articulation and longer in duration in noise rather than silence. Utterance declinations in intensity and F_0 were consistent across noisy and silent conditions.
Conclusions: For content words, the speaker could be trying to reinforce cues in the final position, and also trying to enhance durational and articulatory cues to prosodic hierarchy and word segmentation. These findings show that speech modifications in noise are not purely physiologic.
Relevance to the current work: The current work attempts to investigate speech modifications and the possible intelligibility enhancements that might occur in noise.

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, *130*, 2139-2152. doi:10.1121/1.3623753

Objective: The purpose of this study was to evaluate whether speech produced during spontaneous interactions when addressing a talker under different communication barrier conditions differs in acoustic and phonetic characteristics from speech produced without communicative intent under imaginary challenging conditions or with communicative intent under more ideal conditions. The study also examined acoustic-phonetic modifications made to offset the effects of a noisy environment, and if these modifications are attuned to the noise condition the conversation partner is in.
Methods: 20 male and 20 female speakers participated in a spontaneous and interactive speech task with another talker and a read speech task alone, both in good and challenging noise conditions. In the noise conditions, one talker heard the other through a three-channel noise vocoder and with simultaneous babble noise. Fundamental frequency, long-term average spectrum, word duration and vowel measures were all analyzed.
Results: The median F_0 was higher in read speech than spontaneous interaction speech and was higher in clear speech over conversational speech. Men showed a greater increase in F_0 range in the reading task than conversational speech in the interaction task than women. Women had more mid-frequency energy during the spontaneous interaction task than for the read task,

whereas the male speakers did not vary depending on task type. Mid-frequency energy was higher in clear speech than in conversational speech. There was a greater difference in F_1 range between conversational and clear conditions in read speech than for the interactive speech task. Women have a larger F_2 range than men in conversational speech, and men have a more reduced F_2 range in conversational speech. Women would often use a more expanded vowel range in conversational speech than men. Decrease in speech rate was greater in the read speech over the spontaneous interactive speech. Speakers made greater changes in their F_1 range and mean energy in the babble condition over the vocoder noise.

Conclusions: Speakers tend to read in a higher pitch than in typical speech. Talkers tend to vary their strategies for the listeners depending on the adverse listening condition in order to ensure the most efficient and effective communication. Conversational speech is more finely modulated than clear speech, showing that the speakers matched their speech to the listener needs.

Relevance to the current work: This work provided evidence that speakers adapt strategies based on the noise condition and based on the needs of the interlocutor.

Howell, P. (2008). Effect of speaking environment on speech production and perception. *Journal of the Human-Environment System, 11*, 51-57. doi:10.1618/jhes.11.51

Objective: This article reviews research on how the environment affects speech behavior. Specifically, it investigates how the environment affects characteristics of speech production and speech perception.

Method: The author discusses how the environment affects timing, frequency, and intensity in speech production. In regard to speech perception, the author reviews clear speech, masking, localization, and auditory stream segregation. Cognitive influences are also reviewed for speech production and comprehension.

Results: For sounds, time, frequency, and intensity properties depend on the room dimensions. For speech production, delayed auditory feedback (DAF) can disrupt speech production. A few previous studies investigated frequency shifted feedback (FSF), which resulted in fluency improvements. For intensity, the Lombard effect could be observed. Speakers can also produce more clear speech if they know their audience is hard of hearing, which improves intelligibility. In regard to speech perception, environmental sounds can act as maskers, intensity can provide cues to distance if the sound is familiar (such as speech). Auditory stream segregation helps describe how listeners make use of information about harmonic structure to segregate sounds.

Conclusions: Intensity and delay can be more disruptive to fluent speakers than frequency changes, although manipulations such as FSF, DAF, and masking of the voice can alleviate the fluency problems. The way the environment affects sound is not always detrimental. Speaking clearly can help counterbalance poor acoustic characteristics of the environment.

Relevance to the current work: This review of research discussed how the environment can affect speech production and perception. The current study aims to further study how noise in the speaker's environment can affect speech production.

Junqua, J. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America, 93*, 510-524. doi:10.1121/1.405631

Objective: This study aimed to evaluate the impact of the Lombard effect on automatic speech recognizers, to better comprehend the influence of the Lombard effect on speech perception, and to investigate the acoustic-phonetic changes that occur in Lombard speech. The work goes on further to evaluate the Lombard effect influence on speaker-dependent and speaker-independent recognition as well as listener intelligibility.

Methods: The study reviewed the Lombard effect, describing how a talker's vocal level increases and their articulatory movements change when noise begins, and returns to its former level when the noise stops. 10 speakers produced 49 words with no noise exposure and with white-Gaussian noise at 85 dB SPL. The study replicated previous studies but also focused on the influence of the context on each phoneme as well as the Lombard effect influence on female speakers.

Results: Findings were compatible with previous studies, except in the energy where others found an increase in Lombard speech which differed from this work. This could be explained by the male and female speaker variability as well as more phonemes being investigated. This work found that vowels were more emphasized while consonants were shortened and more distorted in Lombard speech. Intraspeaker variability due to an increase in vocal effort was found to be more important for female speakers. The Lombard effect influence is different for each individual. Acoustic analyses were performed on the speech, specifically measuring bandwidths, spectral slope, energy, formants, burst, norm of the cepstral coefficients, and other variables. The second formant of female speakers is higher than the masking in multitalker noise, possibly causing their speech to be more intelligible. Female speakers have more breathiness in Lombard speech, which this work's data supports could help with intelligibility as well. Male speakers have a lower consonant-to-vowel energy ratio in their Lombard speech, due to an increase in vocal effort.

Conclusions: Based on perceptual and acoustic analysis, it was found that male and female speakers differ in their Lombard speech. The Lombard effect is also variable from speaker to speaker, meaning the strategy for increasing intensity differs among speakers.

Relevance to the current work: The Lombard effect is a documented finding that sets a foundation for what happens to speakers in noisy backgrounds.

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America, 124*, 3261-3275. doi:10.1121/1.2990705

Objective: The main purpose of this study was to identify how noise-induced speech production changes can be affected by the degree of energetic masking and informational masking potential of the noise. Intelligibility based on the level of background noise and the number of talkers is also assessed in this study. To fulfill this last goal, this study used a "computational model of EM in an attempt to determine whether the acoustic changes produced by noise-induced speech result from an attempt to reduce the EM effect at the listener's ears" (Lu & Cooke, 2008, p. 3262).

Method: N-speaker babble noises were used as masking, with N ranging from 1 (a single speaker) to infinity (speech-shaped noise). This would mean informational masking would be the strongest when the N is small and the opposite for energetic masking. The N values used included 1,2,4,8,16 and infinity. Four males and four females were chosen and each initially produced the same set of 50 six-word sentences. From there, the talkers each produced a different set of 50 sentences. Background utterances would be introduced randomly during a sentence. Utterance-level, phoneme-level, correlation, and intelligibility were analyzed. Three experiments compared the intelligibility of utterances when they were presented in stationary noise, competing utterances, and matched and unmatched backgrounds.

Results: The most significant effects on the utterance-level from quiet to noise backgrounds were increases in intensity, frequency, sentence duration, spectral center of gravity, pauses before talking, and the voice-to-unvoiced ratio. N plateaued when it reached 8 talkers. Phoneme-level analysis led to findings in an increase in duration except for nonalveolar plosives and /f/, an increase in spectral center of gravity, and a flatter spectral tilt. The pause before speaking was longer in single-talker background than other babble conditions. The F_1 frequency and energy became more correlated in response to noise. The results from the two experiments showed that intelligibility increased with noise level, speech in an intense background with a single-competing talker is more intelligible than in a quiet background with the same talker, and that talkers don't modify their production in response to the details of a competing utterances.

Conclusions: It is suggested that noise-influenced speech may be more intelligible than speech produced in quiet because the articulatory manipulations lead to a release from energetic masking. This study also supports the idea that speakers compensate for environmental conditions, resulting in increased intelligibility. However, the tasks used in this study lacked communicative intent, meaning the speakers had little intent to change their speech even when the masking noise was present. Communicative intent could have shown more information on masking effects.

Relevance to the current work: This work discussed findings for the effect of the number of talkers rather than just background babble. The study further discussed the idea that informational masking could affect speech production as it causes competition for limited attentional resources. However, it acknowledged that informational masking effects of a competing talker were not found due to tasks with a lack of communicative intent. Evidence that communicative intent has a strong influence has been found in past studies.

Lu, Y., & Cooke, M. (2009). Speech production modifications produced in the presence of low-pass and high-pass filtered noise. *The Journal of the Acoustical Society of America, 126*, 1495-1499. doi:10.1121/1.3179668

Objective: This study aimed to investigate if speakers actively shift their spectral energy distributions to the areas least affected by the masking.

Methods: 4 females and 4 males produced identical set of 30 simple six-word sentences in each of five noise conditions, based on the grid used in previous studies (Lu and Cooke, 2008). Changes in speech production were measured in one full-band, two high-pass filtered, and two low-pass filtered noise conditions.

Results: In both low and high-pass filtered conditions, and for all parameters (root mean square energy, mean fundamental frequency, spectral center of gravity (COG), and mean first formant

frequency) the values increased. Similar effects were observed among most of the speakers across background noise conditions. There were no significant differences based on background noise type and gender. The increases in the high-pass conditions were significant compared to the quiet background. Increases were significantly smaller in the low-pass condition compared to the full-band condition. This study found that talkers modify their productions more in full-band noise, but not in low or high-pass conditions. The authors expected the parameters to increase more in the low-pass condition, but rather they were identical to the full-band noise. The authors also noticed no difference in "narrow- and wide-band low-pass conditions, where an active strategy would predict larger increases in the presence of wide-band low-pass noise in order to place spectral energy in the noise-free region" (Lu & Cooke, 2009). It was found that speakers were unable to modify their speech in high-pass filtering conditions, as the F_0, F_1, and spectral C0G shifted upward.

Conclusions: This work found little evidence to support the idea that speakers can modify their productions to place information in spectral regions least affected by noise. One explanation could be the increase in vocal effort, limiting the scope for variability of other parameters like fundamental frequency. With communicative tasks, speakers might possibly adopt active strategies to take advantage of noise-free regions for listeners.

Relevance to the current work: This work found little evidence for speakers adopting active strategies to place information-bearing words in noise-free regions when reciting sentences. However, the current study aims to explore how speakers adopt strategies in noise conditions under more realistic communicative settings to see if active strategies are adopted.

Lu, Y., & Cooke, M. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustical Society of America, 128*, 2059-2069. doi:10.1121/1.3478775

Objective: Modifications of speech production are well understood for energetic masking but informational masking (like competing speech) they are not as well known. The purpose of this study is to examine the effect of informational and energetic maskers on speech production by talkers speaking alone or in pairs, with communicative and non-communicative tasks.

Methods: 4 males and 4 females participated in the study and were tested under four conditions: quiet, speech-shaped noise, speech-modulated noise, and competing speech. They were then asked to solve Sudoku puzzles either alone or in pairs, describing their progress in each setting. The authors measured the standard Lombard speech metrics which consist of changes in mean F_0, spectral tilt, intensity and duration. They also measured temporal overlap between speech and background as well as vowel space dispersion to examine intelligibility, to see the different responses from speech and non-speech noise conditions.

Results: The Lombard effects found in this study were consistent with previous studies, which found increases in parameter values in energetic maskers (competing speech and speech modulated noise conditions) compared to a quiet background. It was also found that there was no additional effect of communication on speech level, F_0, and spectral tilt depending on the quiet to noisy background. Both tasks produced an overlap reduction in speech-modulated noise and competing speech maskers, with greater reduction in competing speech maskers. The communicative task led to significantly smaller overlap in the backgrounds of speech modulated noise, as well as longer pauses. Talkers had longer and a greater number of pauses in competing

speech background, and very frequent short pauses in the speech modulated noise background. The authors believed that the talkers' can time their speech to reduce temporal overlap in an intelligible speech background more efficiently than in a modulated noise background. These findings were also stronger in the communicative tasks. The communicative tasks led to more contrast in vowel space for two of the four noise conditions. Word durations were smaller in the communicative tasks, and intensity was larger.

Conclusions: Speech-shaped noise brought about the most increases in the Lombard effect, showing that speech-shaped noise could be a more effective energetic masker than competing speakers and speech-modulated noise. The authors were still unsure of what perceptual processes helped to reduce overlap but speculated that intelligible maskers can help talkers predict pauses. It was also found that vowel space expansion to reduce informational masking effects was stronger in communicative tasks.

Relevance to the current work: This work suggested that talkers adapt a "listening-while-speaking" technique to help the other conversational partner understand in noisy backgrounds. It found that the Lombard effect does not change depending on the task type. However, the work did bring forth more information about how talkers reduce informational masking effects, which significantly differed depending on task type and masking type.

Meekings, S., Evans, S., Lavan, N., Boebinger, D., Krieger-Redwood, K., Cooke, M., & Scott, S. K. (2016). Distinct neural systems recruited when speech production is modulated by different masking sounds. *The Journal of the Acoustical Society of America, 140*, 8-19. doi:10.1121/1.4948587

Objective: The purpose of this study was to examine neural underpinnings and activations that occur when someone is speaking in noise. These maskers were meant to exist on a continuum of strongly energetic to strongly informational.

Methods: 7 females and 9 males were asked to read sentences aloud, read silently in their head, and listen to noises in different acoustic environments consisting of intelligible speech, rotated speech, speech modulated noise, continuous white noise, and quiet. While performing the tasks, an fMRI was taken of each participant and analyzed using Statistical Parametric Mapping (SPM 8).

Results: Acoustic parameters were assessed, including mean intensity, median F_0, spectral center of gravity, mean harmonic-to-noise ratio, mean duration, and spectral standard deviation. Using a linear mixed model for analysis, no significant effects were found except in intensity and spectral standard deviation based on the masking conditions. Intensity increased as the energetic component of masking increased, and spectral standard deviation decreased in the speech modulated condition compared to the intelligible speech condition. The perception of the maskers was associated with activation of the dorsolateral temporal lobes. However, speech production was associated with activation in the sensorimotor auditory cortical fields. The bilateral postcentral gyri were activated significantly more in the speaking tasks rather than the listening task. The superior temporal gyrus was mostly activated throughout the speaking and listening tasks. No significant task hemisphere interaction was found, showing that there was no significant lateralization of brain response to speech in noise versus quiet conditions. Activation in the bilateral superior temporal cortices and the left middle temporal gyrus showed activation which was greatest during talking over speech and decreased as informational masking

increased. In the right superior temporal gyrus, analysis revealed that all the conditions were significantly different to each other. In the left superior temporal gyrus, all conditions were significantly different except for speech modulated and white noise.

Conclusions: It was found that the superior temporal cortex might act as an auditory error monitor during talking. Responses in the bilateral superior temporal gyrus were greatest when the participants spoke in noise as compared to the quiet condition, where the superior temporal gyrus suppressed activation when speaking in the quiet condition relative to the listening task. It was also found that these responses were related more to informational masking. This information makes the hypothesis that the suppression effect is a feedback response hard to sustain. It is possible that there is "considerable central processing of 'unattended' information (consistent with information masking accounts) and also that there is considerable competition between activated lexical items when a talker is speaking" (Meekings et al., 2016, p.17). This data suggests a dominant cortical effect of informational masking during speech production, meaning that unattended speech is still processed.

Relevance to the current work: This work emphasizes the effect that informational content can have on speakers when they are attempting to speak at the same time and gives a clearer image of the neural workings while someone speaks in noise.

Patel, R., & Schell, K. W. (2008). The influence of linguistic content on the Lombard effect. *Journal of Speech, Language, and Hearing Research, 51*, 209-220. 10.1044/1092-4388(2008/016)

Objective: The purpose of this study was to investigate Lombard speech in spontaneous sentence-level productions to see if acoustic modifications in noise were evenly applied to all words or if just the content words were enhanced relative to function words. Prosodic cues used to convey contrasts were also an area of interest.

Methods: 8 male and 8 female speakers played an interactive computer game which promoted spontaneous communication with a listener via headphones. 60 dB and 90 dB of multitalker noise was played in the background. The speaker had to direct the listener of what to do in the video game, and data were collected on their accuracy of the listener's direction following. Acoustic modifications were also analyzed.

Results: There were statistically significant effects for syllable duration based on word type and noise type, with location words showing the greatest duration increase from baseline noise type, and function words showing the least. Agent and object word type followed in duration change behind location. Change in peak F-0 from baseline to each noise condition was statistically significant, and differences of change in peak F_0 were found between agent word type and function words. The change in peak intensity in function words was significantly higher in both noise conditions compare to location and object modifiers.

Conclusions: This study provided evidence that duration is enhanced for information-bearing content words relative to function words and content words holding less information. Peak F_0 also served as a primary cue for marking information-bearing word types, especially agent words.

Relevance to the current work: This work is similar in its purpose to the current work in that natural speech is of most interest. The current work seeks to study speech modifications as well as linguistic content when noise is presented in the background.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America, 84*, 917-928. doi:10.1121/1.396660

Objective: The purpose of this study was to examine acoustic-phonetic changes in speech under high levels of noise, like an aircraft cockpit. The study investigated changes that take place in the distribution of spectral energy over time, like modifications in the patterns of vowel formant frequencies or short-term spectral changes in speech sounds produced in noise

Methods: Two male speakers produced a set of utterances in quiet, 80, 90, and 100 dB SPL of white noise.

Results: A consistent increase in amplitudes and word duration was seen as the masking levels increased, which was present in every word. There was a significant effect of noise on F_0 for each speaker, showing that F_0 in the quiet condition was much lower than any of the noise conditions. There was a relative increase in spectral tilt at high frequencies, but was not completely due to increases in F0 as an increase in noise led to a decrease in spectral tilt for one speaker, and for the other a decrease in spectral tilt occurred with no change in F_0. Fundamental frequency, amplitude, and duration increased in the presence of noise. The vowel formant F_1 frequency for one speaker increased as the noise did, but did not increase as much in the second speaker. For the first speaker, noise had an influence on F_1 independent of its influence on F_0. For this speaker, the range of F_2 frequencies was reduced in the presence of noise, but not in the second speaker. Vowels had a relatively flat spectra with most of their total energy in the higher frequency regions and vowels produced in quiet had steeper spectra with little energy present in high-frequency regions.

Conclusions: This study found many differences in the acoustic characteristics of speech produced in noise compared to a quiet condition. Talking in noise also affected the prosodic aspects of speech. It is possible that these changes in spectral, temporal, and prosodic properties helped to improve intelligibility.

Relevance to the current work: The main goal of this experiment was similar to the current work, except in the fact that this study used repetitive utterances whereas the current work will use more natural and spontaneous speech. However, both aim to examine speech modifications and acoustic parameters that could be meant to increase intelligibility.

Webster, J. C., & Klumpp, R. G. (1962). Effects of ambient noise and nearby talkers on a face-to-face communication task. *Journal of the Acoustical Society of America, 34*, 936-941. doi:10.1121/1.1918224

Objective: This study investigates the interaction between physically introduced ambient noise as well as the speech levels of nearby talkers, to see the effect it has on the performance of the talker-listener pairs during a communication task.

Method: 5 talkers were seated shoulder-to-shoulder across from 5 listeners in the same manner. The talkers would rapidly read monosyllabic words one at a time and the listener would repeat the work back. If the listener was incorrect, the talker would repeat the word 3 times or until the listener's response was correct. The results of 3 pairs were scored. The different sessions either took place in quiet environments, or under ambient thermal noise levels of 65, 75, and 85 dB.

Results: The speech levels of the listeners and talkers did not differ. In the lowest level of ambient noise, the amount of speech-level shift was about 5 dB. In the highest ambient-noise conditions, the speech level approaches the maximum possible sustained level. While the over-all speech level increased as the noise did, the voice spectrum slowly shifted to the higher frequencies. The rate of word delivery was analogous with increasing communicators, but reciprocal in its behavior compared to the increase in ambient noise.

Conclusions: The finding that additional communicators in low noise situations, and noise above 75 dB with few communicators both promote more errors in the talking task has also been observed in previous studies. This study also found that the speech level of the talkers increased at a rate of 5 dB for each increase of 10dB in the thermal noise, as well as a rate of 5 dB for every time the number of communicators around them doubled. The talkers' rate of word delivery decreased with increased ambient noise, and increase when there were more communicators in the room.

Relevance to the current work: This work provides further evidence of the Lombard effect, and additional evidence that rate and perceptual errors are dependent on the type and intensity level of the environment.

Zollinger, S. A., & Brumm, H. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Communication and Social Behaviour Group, 148*, 1173-1198. doi:10.1163/000579511X605759

Objective: This paper reviews literature and studies on how humans and other animals have noise-induced vocal amplitude changes. It goes over the biological and scientific history of the Lombard effect by describing the evolution of its study, as well as the biological evolution of the effect itself.

Conclusions: This review discusses the origins of the Lombard effect, and how it has transformed and influenced different disciplines. Mammals and birds were also studied, showing that they also exhibited the Lombard effect in noise. The Lombard effect helps to show plasticity of animal behaviors depending on their changing environments, but also shows the flexibility of acoustic communication systems to enable signals to be received appropriately under challenging circumstances. In certain species, other changes, such as frequency, can occur alongside the Lombard effect which could also help with recognition of the vocalizations in a noisy environment. Other animal types would need to be studied to establish the Lombard effect as a derived trait in mammals and birds. It could be possible that the Lombard effect has been a trait for millions of years. The Lombard effect has transformed into modern comparative psychoacoustics in present day.

Relevance to the current work: The current work aims to further the study on the Lombard effect and other possible parameters that are affected by noise. This work sets a foundation for further research that can continually transform the use and beneficence of the Lombard effect.

APPENDIX B

Consent Form

Consent to be a Research Subject

Introduction

This research study is being conducted by Professor Christopher Dromey at Brigham Young University to determine how speech changes with background noise. You were invited to participate because you are a native speaker of English and have no history of speech or hearing disorders. Two graduate students, Kacy Chapman and Camille Cowley, will assist with the study.

Procedures

If you agree to participate in this research study, the following will occur:
- You will be shown a list of potential topics to talk about, and given time to select at least 8
- you will wear lightweight headphones and sit in a sound booth in the Taylor Building at BYU
- you will be audio recorded while talking about your selected topics under several noise conditions
- you will also read aloud a few sentences from a computer screen
- the background noise conditions include silence, noise similar to radio static, recordings of people speaking, and recordings of music
- total time commitment will be less than 60 minutes in a single visit to the lab

Risks/Discomforts

It is possible that you may experience fatigue during the recording session. Therefore, you may take a break at any time during the experiment if you need a rest.

Benefits

There are no direct benefits to you as a research participant. However, it is hoped that the information obtained from this study will increase our understanding of the impact of noise on speech, which may have potential benefits for the way speech disorders are treated in the clinic.

Confidentiality

The research data will be kept on a password-protected computer and only the researcher will have access to the data. All identifying information will be replaced with anonymous subject codes and the data will be kept in the researcher's locked office. Only aggregate data will be reported in conference presentations or publications that are derived from this study.

<u>Compensation</u>

You will receive $10 cash for your participation; compensation will not be prorated.

<u>Participation</u>

Participation in this research study is voluntary. You have the right to withdraw at any time or refuse to participate entirely without jeopardy to your class status, grades, or standing with the university.

<u>Questions about the Research</u>

If you have questions regarding this study, you may contact Christopher Dromey at 801-422-6461 or dromey@byu.edu for further information.

<u>Questions about Your Rights as Research Participants</u>

If you have questions regarding your rights as a research participant, please contact the IRB Administrator at (801) 422-1461; A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu.

Statement of Consent
I have read, understood, and received a copy of the above consent and desire of my own free will to participate in this study.

Name (Printed):_____   Signature:_____   Date:_____

APPENDIX C

Monologue Topics

Personal
- Would you quit if your values did not match your employer's?
- If you could be rich, famous, or influential, which would you choose and why?
- How would you define faith?
- How do you define wealth?
- Do you believe people make happiness or stumble across it?
- Which is more important, talent or hard work?
- Are you an introvert or an extrovert?
- What are the pros and cons of each?

Media
- Are antidrug and antismoking ads effective?
- What video game would you like to redesign?
- Do social media campaigns stimulate real change?
- Should people be allowed to obscure their identities online?
- Is TV stronger than ever, or becoming obsolete?
- What ideas do you have for a reality show?
- What is your opinion about violence on television and in video games?
- What artists of today are destined for the Rock and Roll Hall of Fame?

Generations
- What is the difference between your generation and my generation and why?
- Is your generation more self-centered than earlier generations?
- Are young people generally more selfish than their parents and grandparents?
- How will our current culture be remembered in history books?
- Do children today have good manners?
- Does age make you more aware of and caring for others?
- Should adults try to teach young people lessons or should they leave them alone to find out about things themselves?
- Should parents continue to financially support their children after the children are 18?
- Is modern culture ruining childhood?

Local Issues
- If you could expand the Trax system, what changes would you make?
- What do you see as the pros and cons of the proposed rebuilding of the Salt Lake airport?
- Is it important to shop at locally owned businesses?
- What could be done about Salt Lake's homeless population?
- What are the pros and cons of the Sugarhouse trolley?

Social
- What has caused the obesity epidemic in America?
- Should people get plastic surgery?
- Should rich people have to pay more taxes?
- What is your opinion about cloning?
- What are the ethical implications of eating meat?
- Are children of illegal immigrants entitled to a public education?
- Should welfare recipients be required to take drug tests?
- If you were a philanthropist, what groups would you finance and why?
- When should juvenile offenders receive life sentences?
- Should women soldiers be in combat?
- What is your opinion about legalizing marijuana?
- Are we losing the art of listening?
- Do attractive people have advantages others don't?
- What are the most important changes in the world since the year 2000?

Education and Related
- Is online learning as good as face-to-face learning?
- How necessary is a college education?
- Should cash-strapped schools cut arts education?
- Should guns be permitted on college campuses?
- What do you think about home-schooling vs. public vs. private school?
- How would you make over the university system?
- Whose fault is it if a child is failing in school?
- Should parents/grandparents give cash rewards to kids for good test scores?
- Should university students be required to take drug tests?
- Should junk foods and soda-pop be sold in elementary school or high school vending machines?
- How well do you think standardized tests measure people's abilities?