Recall Disruption Produced by Noise-Vocoded Speech:

A Study of the Irrelevant Sound Effect

Josh Dorsi

State University of New York at New Paltz

Abstract

The Irrelevant Sound Effect (ISE) is the finding that serial recall performance is impaired under complex auditory backgrounds such as speech as compared to white noise or silence (Colle & Welsh, 1976). Much of the current research investigates the role of changing-state complexity of the background stimuli in ISE (e.g., Jones & Macken, 1993). This study investigated whether speech-specific qualities of the irrelevant background have an effect on the ISE. This was done using noise-vocoded speech, an acoustic transformation that removes many of the acoustic properties of speech while preserving the speech intensity profile. Experiment 1 compared serial recall accuracy resulting from white noise and noise-vocoded speech backgrounds and found that noise-vocoded speech is more disruptive. Noise-vocoded speech preserves the intensity profile of nature speech with a number of amplitude channels; each channel matches the average intensity for the corresponding channel in natural speech. Experiment 2 systematically varied the resolution of noise-vocoded speech by adjusting the number of these channels. These results show that ISE varies based on the number of channels in noise-vocoded speech, but this change in disruption is not consistent across channel conditions. Results demonstrate that changing state complexity alone is not a sufficient explanation of ISE.

Recall Disruption Produced by for Noise-Vocoded Speech:
A Study of the Irrelevant Sound Effect

The Irrelevant Sound Effect (ISE) is the finding that irrelevant background sound, such as speech, impairs serial recall of visually presented items (Colle & Welsh, 1976). This phenomenon was first discovered when Colle and Welsh (1976) presented strings of eight letters visually to participants. These serial lists were presented either in silence, or along with a continuous speech background. The task for participants was to report the visually presented items in the order in which they had been presented. Though participants were instructed to ignore the background speech, there was a reliable decline in serial recall accuracy for trials with a speech as compared to a silent background.

One early explanation of ISE came from the working memory model proposed by Baddeley and Hitch (1974). According to this model, working memory consists of a visuo-spatial sketch pad that is responsible for storing visual information, a phonological store that is responsible for storing verbal information, and a central executive that controls these subsystems. By this account of ISE, the presented visual items enter the visuo-spatial sketchpad automatically but are transferred into the phonological store by sub-vocal articulation. From the phonological store, serial items can be rehearsed in the phonological loop. Termed the phonological loop hypothesis, this model offers an explanation for the serial recall disruption associated with irrelevant speech. Speech is assumed to have preferential access to the phonological loop, and thereby interferes with the deliberate entering of visual serial items.

A number of early studies supported the phonological loop hypothesis. For example Salamé and Baddeley (1983) found that speech caused more serial recall disruption than did white noise. This finding was consistent, even when the intensity of the backgrounds was increased, and when the speech was presented in a language not spoken by the participants (Salamé & Baddeley, 1983). Furthermore, nonsense words cause disruption equal to that of real words suggesting that disruption is not dependent on the semantic content of the background speech (Salamé & Baddeley, 1982; but see Buchner, Rothermond, Wentura, & Mehl, 2004). ISE has also been found when serial items are presented to participants during the silent intervals between irrelevant speech tokens. The authors interpreted this finding as an indication that ISE is not the result of irrelevant speech drawing attention away from the serial items (Salamé & Baddeley, 1982). Collectively, these results support the assumption that ISE is the result of speech automatically entering the phonological loop and interfering with the addition and rehearsal of serial items.

Another key finding that supported the phonological loop hypothesis is the effect of articulatory suppression on ISE. ISE is eliminated when participants engage in articulatory suppression, such as articulating a consonant repeatedly, during serial item presentation (Salamé & Baddeley, 1982). According to the phonological loop hypothesis, articulatory suppression removes the difference in recall accuracy between silent and speech backgrounds by engaging the phonological loop, thus blocking both the irrelevant speech and serial items from rehearsal. Accordingly, under conditions of articulatory suppression, the overall amount of serial recall disruption is significantly increased, but the amount disruption between silent and speech backgrounds is equal (Salamé & Baddeley, 1982).

A more recent theory, termed the feature model, has been proposed as an account of ISE (Neath, 1999). The feature model proposes two memory stores, primary memory which processes “cues”, and secondary memory which holds more detailed representations (Nairne, 1990). Items in secondary memory and cues in primary memory are composed of features, or segments of information (Nairne, 1990). Information is entered into secondary memory after it is processed by primary memory. Importantly, information can reach primary memory through a number of modalities (e.g. visual stimuli, auditory stimuli, sub-vocal articulation, etc.), but only modality-independent features, features that are associated with an item irrespective of the modality by which it entered primary memory, move into secondary memory (Neath, 2000). Recall involves identifying and retrieving the appropriate items from secondary memory; these items are identified based on the number of features they have in common with the respective cue in primary memory.

Items in primary memory degrade rapidly, as a function of interference from (overlap with) other cues, as well as time, but secondary memory is considered robust against such types of degradation (Neath, 2000). As primary memory cues degrade, they become less defined and can be linked to multiple items in secondary memory. Serial information is stored in primary memory as a series of positions held by a group of cues. As cues in primary memory degrade so do their serial position information such that a cue, which was identified as belonging in position three, may later appear to belong in positions two or four. Serial recall involves selecting each cue and calculating the likelihood that that cue belongs to each serial position. Irrelevant speech increases the number of cues in primary memory, amplifying the interference and leading to faster degradation. This interference can result in order errors in the selection of serial items from secondary memory.

The strength of this account is that, like the phonological loop hypothesis, it can explain many of the key findings reported in the literature. Accordingly, the feature model shares many of the same weaknesses of the phonological loop hypothesis. The most notable weakness of these accounts is their inability to address the pattern of recall disruption found for non-speech auditory backgrounds. Neath suggests that disruption caused by non-speech items may be epiphenomenal to ISE, a result of diffused attention resulting from a dual process. This explanation for non-speech disruption is insufficient for two reasons. The first weakness of the feature model concerns parsimony: the feature model of ISE invokes two separate processes of disruption, attention for non-speech and memory for speech backgrounds, while other models invoke a single process. The second weakness of the feature model is that the explanation that it does offer for the disruption caused by non-speech stimuli, does not address the variability of disruption across different backgrounds. This pattern of disruption, and theories that address it are discussed in detail below.

The finding that non-speech backgrounds, such as music, cause serial recall disruption challenged these speech-specific accounts of ISE[1] (Salamé & Baddeley, 1989). A central assumption of these accounts is that information contained in speech gives it unique disruptive qualities relative to other stimuli. Accordingly, serial recall disruption associated with non-speech auditory backgrounds challenges the assumption of speech-specificity in ISE. In fact, a range of non-speech background sounds such as, pure tones (Jones & Macken, 1993), traffic noise (Hygge, Boman, & Enmarker, 2003), and animal calls (Schlittmeier, Weiβgerber, Kerber, Fastl, & Hellbrück, 2012), have been found to cause serial recall disruption. The number of non-speech auditory backgrounds that have been found to cause serial recall disruption has led many researchers to investigate acoustic (as opposed to informational) qualities of auditory stimuli that might drive ISE.

Much of ISE research focuses on identifying acoustic properties of background, speech and non-speech, that cause serial recall disruption. One property of the acoustic signal that is commonly claimed to play a central role in serial recall disruption is *changing state complexity*, a term that is roughly defined as the number distinct acoustic and perceptual elements present in the auditory signal (Jones, 1992; Jones & Macken, 1993). This claim comes from a number of studies, which have found ISE with non-speech stimuli (e.g. Jones & Macken 1993; Beaman & Jones 1997, 1998). When taken together the results of these studies suggest a positive relationship between the amount of recall disruption and the number of changing states in the auditory background. For instance, LeCompte, Neely, and Wilson (1997) found that a single steady state tone causes less disruption than a pair of tones. Jones, Macken, and Murray (1993) found that a single repeating tone was less disruptive when the silent intervals between each repetition were masked by white noise creating the perception of a single, uninterrupted tone glide with bouts of noise. From these results, the authors concluded that while the tone with noise condition contained more acoustic variability it had fewer perceptual elements, thus causing less serial recall disruption. Given these findings (and others) recent theories of ISE, including the changing state hypothesis (Jones & Macken, 1993) and the primacy model (Page & Norris 2003) assume that recall disruption is directly related to the number of changes in state (percept) of the irrelevant sound. While these theories are both distinct from one another, they share the claim that recall disruption is a function of the changing state complexity of the auditory background.

Following the discovery that simple tones also produce serial recall disruption, Jones (1993) proposed the changing-state hypothesis as an explanation of the ISE. According to this hypothesis, information is entered into memory as a series of streams, which are systematically divided into sub-streams (Jones & Tremblay, 2000). In ISE visual serial items are streamed voluntarily, and are partitioned on the basis of the items themselves, while auditory items are streamed automatically and are partitioned on the basis of physical disparity between components (each change of state) of the stimulus (Jones & Tremblay, 2000). The changing-state hypothesis accounts for many of the key findings associated with ISE (e.g. Jones & Macken 1993; Beaman & Jones 1997, 1998). For example, this hypothesis can account for why ISE is more pronounced in serial recall, as opposed to free recall (LeCompte, 1996). Under the changing state hypothesis, serial order of to-be-remembered items is processed concurrently with that of the auditory stimuli resulting in competition for memory space between two serial processes. This hypothesis also includes an explanation for how non-speech items cause serial recall disruption; specifically, all auditory stimuli enters memory as a stream of states and this stream competes for memory space with streams generated for serial items (Jones, 1993). By extension, auditory stimuli with many changes in state result in more items streamed in memory, therefore reducing the amount of space available for serial items and resulting in decreased serial recall accuracy. Because of its explanatory power, and parsimony, the changing state hypothesis has become one of the most popular explanations of ISE.

Sharing many of the strengths, and a few of the central assumptions of the changing state hypothesis, the primacy model is another recent account of ISE. The primacy model proposes that serial recall is a function of the primacy effect, which is the finding that errors in recall are lower for the earlier serial positions as compared to positions in the middle of a serial list[2] (Page & Norris, 1998). The primacy model is a connectionist model, wherein each item from a to-be-remembered list corresponds to a node of activation forming a chain of activation nodes that represent the serial list. Under this model, item nodes receive progressively less activation, and serial order is recalled based on the relative amount of activation in each node. Complete chains of nodes, representing to-be-remembered lists activate in regular intervals, but each cycle receives less overall activation as a result of limited cognitive resources. As activation declines, so do the differences in activation between each node, making distinguishing node position more difficult (Page & Norris, 1998).

According to the primacy model, irrelevant sound forms its own chain of activation nodes, parallel to the chain representing to-be-remembered items (Page & Norris, 2003). Under this model, each change in state of the irrelevant sound creates its own node in the chain representing the background sound. Accordingly, sounds with greater changing state complexity form chains with more nodes and therefore require more activation. The additional chain created for irrelevant sound thus diffuses the resources available for the serial item list, resulting in increased serial recall disruption.

The theoretical focus on changing state complexity of irrelevant sound in ISE is not unwarranted, as there are numerous studies to support the claim that recall disruption corresponds with the number of acoustic changes in states of the irrelevant sound (Hughes, Tremblay, & Jones, 2005; Jones, Alford, Macken, Banbury, & Tremblay, 2000; Macken, Mosdell, & Jones, 1999). While there is a body of data that supports the role for changing state complexity in the ISE, this body does not support the assumption that changing state complexity is the only cause of ISE. The focus on only changing state complexity disregards other acoustic variables that may contribute to ISE.

Schlittmeier et al. (2012) attempted to model the relative effect that changing state complexity of the background has on ISE. To this end, the authors examined data collected from 40 different studies of ISE. These reviewed studies involved serial recall accuracy measures associated with over 70 different irrelevant background sounds, including foreign and native language speech, office noise, animal calls, and pure tones (Schlittmeier et al., 2012). From this review, it was found that speech resulted in the greatest amount of ISE.

It is intriguing that despite the breadth of irrelevant sound examined by the Schlittmeier et al. (2012) no sounds were found to induce greater ISE than speech. The authors speculate that speech may hold some characteristics, in addition to its changing state complexity, that increases its ISE relative to non-speech sounds. The authors do not, however, specify the nature of these speech-specific characteristics. Despite this, Schlittmeier et al. do indicate that attention, in the form of the attention drawing nature of speech, may contribute to ISE, a conclusion that is supported by ERP data reported by Little, Heritage Martin and Thomas (2010).

To investigate this question, Viswanathan, Dorsi, and George (in press) compared the serial recall accuracy associated with two different irrelevant sound backgrounds, sinewave speech and selectively-reversed sinewave speech. Sinewave speech preserves the spectro-temporal information of natural speech with a series of time-varying sinusoids that track formant centers (Remez, Rubin, Pisoni, & Carrell, 1981). Selectively-reversed sinewave speech is sinewave speech in which some but not all formants are temporally reversed relative to the others (Viswanathan et al., in press).

By using sinewave speech and selectively-reversed sinewave speech backgrounds, Viswanathan et al. were able to compare the recall accuracy associated with backgrounds that were equal in acoustic complexity but varied in their fidelity to the structure of speech. Specifically, selectively-reversed sinewave speech contained the same number of changing states as sinewave speech, but did not maintain the natural speech intra-formant organization of sinewave speech. This study showed that serial recall disruption was greater for the sinewave condition, relative to the selectively-reversed sinewave condition. This finding suggests that ISE is responsive to a speech-like frequency organization even when changing state complexity is held constant.

A weakness of all theories of ISE discussed so far is that the locus of ISE is placed entirely in one domain, be it speech-specific qualities or changing state complexity (see Table 1). I make the claim that speech-specific information contained in the acoustic signal may cause recall disruption in addition to disruption caused by changing state complexity. Based on the findings of Schlittmeier et al. (2012) and Viswanathan et al. the present study examined ISE with the broad goal of identifying properties of the acoustic signal that cause serial recall disruption, besides changing state complexity. The results of Viswanathan et al. (in press) show that a speech specific formant organization effects ISE. The present study investigates if a speech-specific intensity profile in the acoustic signal will also cause an ISE.

In this study, I used noise-vocoded speech as irrelevant sound during a series of serial recall tasks. Noise-vocoded speech is a manipulation of natural speech that is generated by dividing speech into logarithmically spaced amplitude channels, mapping the intensity variation within each channel, and then applying these channels and their intensity variations to white noise (Shannon, Zeng, Kammath, Wygonski, & Ekelid, 1995; Davis, Johnsrude, Hervais-Adelman, Taylor& McGettingan, 2005). Despite lacking many of the qualities of natural speech, noise-vocoded speech can still be intelligible as speech. Because of the way it is synthesized, noise-vocoded speech contains the intensity but not the frequency information of natural speech, and it is commonly assumed that noise-vocoded speech is intelligible primarily as a result of intensity variation (e.g. Shannon, Zeng, Kamath, Wygonski, Ekelid, 1995; but see Roberts, Summers, & Bailey, 2011).

Intelligibility of noise-vocoded speech is related to the amplitude of its channels. For example, Shannon, Zeng, and Wygonski (1998) found that noise-vocoded speech was optimally intelligible when its channel amplitudes logarithmically increased, that is, when each channel had amplitude that was exponentially greater than the preceding channel. The authors found that noise-vocoded speech with channels of equal amplitudes, while still intelligible, required more practice before subjects achieved speech percepts. Important to this study, the intensity variations preserved in the amplitude channels of noise-vocoded speech preserve the intensity profile of natural speech. Because noise-vocoded speech preserves the intensity, but not the frequency, profile of speech, it is an ideal manipulation to investigate if speech-specific intensity distributions in irrelevant sound will cause an ISE.

Experiment 1

The purpose of Experiment 1 was to investigate whether noise-vocoded speech backgrounds would produce ISE. My hypothesis was that ISE would be sensitive to acoustic properties of speech, and that some of these properties, namely the intensity profile of speech, would be preserved in noise-vocoded speech but not white noise. From this hypothesis, I predict that noise-vocoded speech backgrounds would result in a greater serial recall disruption than white noise backgrounds. This experiment used a within subjects measure to compare serial recall accuracy associated with either white noise or 6 channeled noise-vocoded speech backgrounds.

Method

Participants

Fifteen students from the State University of New York at New Paltz received course credit for their participation. All subjects were native English speakers and reported normal hearing and normal or corrected vision.

Materials

Noise-vocoded backgrounds were generated from natural speech tokens used by Viswanathan et al. (in press); these tokens were: *bowls*, *boy*, *day*, *dog*, *go*, *than*, and *view*. Each token was divided into six logarithmically spaced amplitude channels (channel characteristics shown in Table 2) these channel ranges were used by Davis et al. (2005) in a study of noise-vocoded speech intelligibility, and represent an exponential increase in channel amplitude, while remaining within the range of 50 and 8000hz. These amplitude channels were applied to white noise and were constrained to match the intensity distribution of the analogues channels from the natural speech samples.

Irrelevant sound tokens were arranged into eight lists, four white noise and four noise-vocoded in which the items were randomly ordered. Each list was repeated once to create a 14 item, 1600ms long background. Auditory backgrounds were presented through headphones at 70db and lists were matched on duration and intensity. Background conditions were presented in a randomized order across participants.

The irrelevant sounds were presented concurrent with the visually presented to-be-remembered targets: L R T S M K F also used by Viswanathan et al. (in press). Visual targets were presented on a computer screen for 1000ms, with a 500ms interval between items. Target items were generated into random order lists, before being presented to participants during the procedure.

Procedure

At the start of each session, participants received on-screen instructions and verbal directions from an experimenter before beginning the experiment. Directions informed participants that they would be visually presented with a series of letters, and may hear sounds through their headphones. Participants were instructed to ignore any sounds they heard and focus on the visual presentation. Participants were informed that the goal of the task was to report the presented letter sequence in the correct order. Directions further explained that following each presentation, subjects would be prompted by the computer to type the previously shown letter sequence into the computer. A copy of the script for verbal directions can be found in Appendix A, and a copy of the onscreen instructions in Appendix B.

Participants initiated each serial recall task by pressing the space bar on their keyboard (specified by on-screen directions), improving participant preparedness for each trial of the experiment. Participants were prompted to type in their response by a blinking cursor in the upper left corner of the computer screen, 1000ms following the presentation of the last visual item.

Results

This experiment measured the serial recall accuracy of participants who were presented with visual serial items concurrent with irrelevant background sounds. Irrelevant sound was presented in the form of noise-vocoded speech, or white noise. Serial recall accuracy was measured as reporting an item in its correct serial position. Data collected from this study were submitted to a 2 (Background condition) by 7 (Serial position) repeated measures analysis of variance (ANOVA). It was found that average serial recall accuracy for the noise-vocoded speech background was ($M = .54$) while the white noise background was ($M = .61$). This difference was statistically significant, $F(1, 14) = 11.06$, $p < .005$, $\eta^2_p = .44$, indicating that the noise-vocoded background was significantly more disruptive than the white noise background. A main effect for serial position, $F(6, 84) = 48.44$, $p < .005$, $\eta^2_p = .78$, indicated that a letter's position affected its serial recall accuracy. The interaction between background and serial position was also significant $F(6, 84) = 2.53$, $p < .05$, $\eta^2_p = .15$, suggesting that the impact of background was not equal for all serial positions. The relationship between serial recall accuracy for white noise and noise-vocoded speech is displayed in Figure 1.

Discussion

Overall, the data collected from Experiment 1 supports the hypothesis that noise-vocoded speech contains acoustic qualities which result in ISE. The results of this experiment do not necessarily implicate speech-specificity as contributing to ISE; rather, they show that some quality of noise-vocoded speech made it more disruptive than white noise. These data, however, justify additional investigation into what precisely about noise-vocoded speech causes ISE.

Experiment 2

To generate noise-vocoded speech natural speech is divided into different amplitude channels (Loizou, Dorman, & Tu, 1999). The number of channels that natural speech is divided into, and that noise-vocoded speech is ultimately generated from, can vary (Loizou et al., 1999). More importantly the intelligibility of noise-vocoded speech is strongly related to the number of channels it is composed of, such that three channeled noise-vocoded speech is generally unintelligible, while 10 or greater channels are readily intelligible (Davis et al., 2005). Noise-vocoded speech is often perceived as non-speech to the naive listener, but this perception changes over the course of repeated exposure, such that noise-vocoded speech becomes intelligible (Loizou et al., 1999). To increase the number of channels, but remain in the amplitude range of 50-8000 Hz, the amplitude of each channel must be reduced. Increasing the number of channels of noise-vocoded speech increases the resolution of each channel by reducing amplitude of each channel and therefore the amount of intensity variation each channel must average. Accordingly, additional channels not only result in improved intelligibility but also improved resolution of the respective natural speech sample.

Increasing the number of channels in noise-vocoded speech increases its complexity linearly. Conversely, intelligibility of noise-vocoded speech reaches a ceiling at nine channels, and it is possible that there are no gains in speech fidelity past this ceiling (Loizou et al., 1999). The goal of Experiment 2 was to determine if increasing the number of channels in noise-vocoded backgrounds would result in different degrees of serial recall disruption consistent with the predictions of changing state models of ISE.

Method

Much of the method for Experiment 2 was the same as the method from Experiment 1. The main difference in Experiment 2 is that it incorporated noise-vocoded speech backgrounds of varying channel numbers, and these different channel conditions were presented as a between subjects factor. Additionally, in order to examine the role of speech-specific qualities of noise-vocoded speech, which may influence ISE, I added an intelligibility measure[3] to Experiment 2. This measure was designed to afford insight into the question of how the presence of a speech like intensity profile (enough of a speech intensity profile to achieve intelligibility) would affect ISE. Additionally, this intelligibility measure allowed me to explore the effects of noise-vocoded speech's perceptual bistability[4] on ISE, that is, whether ISE depends on the participants' perception of noise-vocoded speech as speech.

Participants

Participants were 87 students from the State University of New York at New Paltz. Participants received course credit or five dollars cash for their participation. All subjects were English speakers and reported normal hearing and normal or corrected vision.

Materials

Experiment 2 compared the serial recall accuracy associated with noise-vocoded speech backgrounds composed of three, nine, and twelve channels. The frequency ranges for the channels of all noise-vocoded conditions are shown in Table 2. These channel ranges were generated from the range utilized by Davis et al. (2005) of 50.00hz to 8000.00hz. The exponential increase in channel ranges which is necessary for ideally intelligible noise-vocoded speech (Shannon, Zeng, & Wygonski, 1998) was obtained by using the following formula:

$$(50.00\text{hz} * X)^{\text{number of channels}} = 8000.00\text{hz}$$

For the intelligibility measure, 14 noise-vocoded words were prepared. All noise-vocoded conditions had the same corpus of words, but words were synthesized into noise-vocoded speech with the number of channels consistent with the noise-vocoded speech used in the serial recall task. The 14 words included the 7 words used in the serial recall task, as well as 7 novel words. The novel words were: *wish*, *while*, *save*, *time*, *note*, *eat*, and *die*. These words were matched to the set used in the serial recall task, in average length, spoken frequency, and concreteness. The inclusion of novel words afforded me the added flexibility to explore if intelligibility found in participants would generalize to novel noise-vocoded speech, and if the presence of the generalization would additionally influence ISE.

Procedure

This experiment utilized the same serial recall task, serial targets, and background tokens as Experiment 1. Participants were assigned to one of four noise-vocoded background conditions: 3 channels, 6 channels, 9 channels, and 12 channels. All noise-vocoded backgrounds were compared to white noise backgrounds within subjects, while different noise-vocoded backgrounds were compared between subjects. The purpose of the between subjects comparison was to guard against the demonstrated hysteresis effects in noise-vocoded speech intelligibility[5] (Hervais-Adelman, Davis, Johnsrude, Taylor, & Carlyon, 2011).

Results

Experiment 2 measured the serial recall accuracy of participants who were presented with visual serial items concurrent with irrelevant background sounds. All participants were exposed to irrelevant sound in the form of white noise, and noise-vocoded speech. Data from one participant in the 12-channel condition were eliminated from statistical analysis because this participant did not perform the serial recall task at all. The number of channels that noise-vocoded backgrounds were composed of varied across participants. Serial recall accuracy was measured in the same way as was done in Experiment 1; however, in order to determine the relative differences in recall disruption between noise-vocoded speech backgrounds, ISE was measured in differences scores. Difference scores were calculated, for each subject, as the difference in serial recall accuracy between white noise and noise-vocoded speech conditions, calculated by subtracting the serial recall accuracy associated with noise-vocoded conditions from the accuracy associated with white noise conditions at each serial recall position.

The mean difference scores of serial recall disruption for each condition are displayed in Figure 2. Difference scores were submitted to a 4 (Channel number) by 7 (Serial position) mixed factors analysis of variance (ANOVA). A main effect for serial position, $F(6, 228) = 5.02$, $p < .001$, $\eta^2_p = .06$, indicated that serial recall accuracy varied in part as a result of the letter's serial position. Of importance to this study, results found a significant main effect for background condition, $F(3, 83) = 3.92$, $p < .001$, $\eta^2_p = .29$, indicating that at least one noise-vocoded background caused significantly different ISE from the others. Post hoc tests were performed on all comparisons to identify the locus of this effect. Using a Tukey HSD, the comparison between the 3 channel ($M = .01$) and the 9 channel ($M = .09$) noise-vocoded speech backgrounds was significant ($p < .05$) as was the comparison between the 3 and 12 channel (M = .08) conditions ($p < .05$). No other comparisons were significant ($p > .05$). In short, we find an increase in serial recall disruption between the 3 and 9 channel conditions, but not between the 6 and 12, indicating that the change in ISE may not be following the regular increase in channel number between our noise-vocoded conditions. The interaction between background and serial position was not significant $F(18, 498) = 1.21$, $p = .24$, $\eta^2_p = .06$.

Results show that something besides number of channels (changing state complexity) may be influencing ISE. One possibility is that the speech fidelity of backgrounds contributes to the amount of ISE. The intelligibility measure was presented to each participant following the serial recall task portion of the experiment. Participants listened to 14 words in noise-vocoded speech (seven novel and seven previously heard), and reported what they heard for each. Intelligibility was measured as the proportion of the seven words correctly identified, averaged across participants. These proportions are displayed in Figure 3, which illustrates that the intelligibility varied across channel conditions.

It is important to note that while the intelligibility measures differed by channel number, the overall intelligibility was quite low, and the differences across channels were small. The original intent of the intelligibility measure was to investigate if there would be any differences in ISE associated with differences in intelligibility; however, the intelligibility for all the groups was so small it seems unlikely that such an analysis would be informative. With this in mind, no analysis was conducted that compared ISE between participants with and those without speech percepts of noise-vocoded speech.

Discussion

This experiment was conducted with the assumption that changing state complexity would directly correspond with the number of channels in noise-vocoded speech. With this assumption our results are not consistent with the predictions of a changing state model of ISE. This experiment found that the number of additional channels was not the essential difference between groups with significantly different amounts of serial recall disruption. Specifically, it was found that a difference of six channels was sufficient to cause an increase in ISE between the 3 and 9 channel conditions, but not between the 6 and 12 channel conditions. The exact cause for the pattern of ISE across the conditions of the experiment is not clear from these results.

One possible explanation for the pattern of ISE described above, could be the addition of speech fidelity in the intensity profile of certain conditions. To investigate this possibility, Experiment 2 employed an intelligibility measure. A prediction for this experiment was that there would be a sizeable group of participants, in the 6, 9, and 12 channel conditions, with speech percepts of noise-vocoded speech following the conclusion of the serial recall portion of the experiment. While the proportion of intelligibility appeared to increase across conditions, it was minimal. Such results make it difficult to interpret the role of intelligibility in ISE; however, it does seem that the pattern of ISE does not correspond directly with the intelligibility of the background.

General Discussion

The results of Experiment 1 show that noise-vocoded speech does in fact cause ISE. Noise-vocoded speech resulted in a reliable difference in serial recall accuracy compared to the disruption associated with white noise backgrounds. The question that remained after this experiment concerned the cause of this recall disruption. Changing state models of ISE would assume that noise-vocded speech has more complexity than white noise backgrounds. The fundamental difference between noise-vocoded speech and white noise is the intensity profile, which is preserved from natural speech in the different channels of noise-vocoded speech. Accordingly, the changing state complexity of noise-vocoded speech, must be (at least in part) related to the channels of noise-vocoded speech. To explore the effect of number of channels on ISE Experiment 2 was conducted.

The measures used in Experiment 2 calculated ISE as the difference in serial recall accuracy between white noise and noise-vocoded speech. These difference scores were compared between conditions with different numbers of channels. The noise-vocoded backgrounds were composed of 3, 6, 9, and 12 channels, representing a regular increase in channel composition. It was assumed that the changing state complexity of these conditions would increase regularly with the channels in the noise-vocoded speech. The ISE for these four conditions does not reflect a regular increase in serial recall disruption. What was found is that recall accuracy varies between specific, non-adjacent conditions, namely, between 3 and 9, as well as between the 3 and 12 channel conditions.

A weakness of this study is that the role of complexity of noise-vocoded speech in ISE was not adequately addressed. As is discussed above, changing state complexity is not explicitly defined in the literature, so this is a weakness that is present in most studies of ISE. Presumably changing state complexity increases with the number of channels in noise-vocoded speech; however, if this is the case, the pattern of ISE found in this study does not strictly correspond with the increase in complexity of the background. One possibility is that the change in complexity associated with the change in channel number may be small and not substantial enough to cause a change in ISE. On this line of reasoning, the increase in ISE between the 3 and 9, and the 3 and 12 conditions may have been found because the difference in channel number was larger than say between the 3 and 6 channel conditions. One way of investigating this possibility could be to expand the range of channels across conditions. This could be done by adding conditions in the interval already established in experiment 2 (e.g. adding 15, 18, and 21 channel conditions), or by increasing the interval between conditions (e.g. 3, 9, 15 channel conditions). If changing state complexity does correlate with the number of channels in noise-vocoded speech than either of these conditions could increase the amount of ISE possibly causing significant differences between conditions.

This explanation alone, does not account for why the difference between the 6 and 12 channel conditions was not significant, despite having the same difference in channel number as the 3 and 9 channel conditions. An alternative explanation is that the changing in changing state complexity across channels was not linear, as I assumed it would be. For example, the amount of change in changing state complexity could have declined as more channels were added to noise vocoded speech. Such a pattern of change in changing state complexity would be curvilinear as opposed to linear. If this is the case, the results of Experiment 2 could be explained entirely by changing state complexity with no need to invoke a role for speech-fidelity.

A third possibility is that increasing the number of channels in a noise-vocoded background does not actually increase changing state complexity at all. This possibility is difficult to address because the typical measure of changing state complexity has been the degree of ISE. For example, Jones, Macken, and Murrey (1993) showed that a tone glide interrupted by regular intervals of silence was more complex than the same tone glide with white noise masking the silence intervals. The authors indicate that while the white noise condition had more acoustic energy, the masking of the silence intervals gave participants the perception of a steady (uninterrupted) tone, thus this condition was said to have fewer perceptual changes in state. This explanation conflicts with the finding that speech is the most disruptive stimulus as speech is *perceived* as a small collection of phonemes not a complex array of changing acoustic states. Collectively, there is little explanation as to what changing state complexity truly is. The lack of clarity in theoretical the construct of “changing-state complexity”, make the results of the present study difficult to interpret, and emphasize the need for future studies to clarify the nature of this construct.

An alternative explanation of ISE is that serial recall disruption is partly caused by the speech fidelity of the background. This study attempted to investigate the contribution of speech fidelity to ISE by measuring the intelligibility each noise-vocoded background. The purpose of the intelligibility measure following the serial recall task was to assess if participants who gained speech percepts of noise-vocoded speech during the experiment would have a different pattern of ISE compared to participants who did not have speech percepts. I predicted that this kind of comparison would be possible because prior research has shown that participants who were previously unaware of its speech nature achieved speech percepts of noise-vocoded speech over the course of several exposures (Davis et al., 2005). Specifically, it was found that participants attending to 6 channel noise-vocoded speech with no prior training reached a 60% transcription accuracy with 10 sentences worth of exposure (Davis et al 2005). The present study used noise-vocoded backgrounds consisting of 14 words, presented 24 times, over the course of the serial recall task.

I predicted that, participants would gain, at least, a modest level of accuracy for the transcription of noise-vocoded based on the results reported by Davis et al. (2005). I made this prediction because the participants of this study, both in number of words and in amount of time, were exposed to substantially more noise-vocoded speech than the participants of Davis et al. (2005). The very low intelligibility scores found in the present study suggest a role for attention in the intelligibility of noise-vocoded speech. While our participants received more overall exposure to noise-vocoded speech, they were instructed to ignore all auditory stimuli; conversely, participants in Davis et al. (2005) were instructed not only to attend to noise-vocoded speech, but to attend to the speech nature of noise-vocoded while they attended to it. This direction of attention to the speech nature of noise-vocoded speech could make a substantial difference in its intelligibility. This insight is important to future studies of noise-vocoded speech intelligibility.

Given the role of attention in the intelligibility of noise-vocoded speech, it is unlikely that the differences in intelligibility that naturally occur during a typical test of ISE would be substantial enough to make any meaningful comparisons between subjects, as is the case for the results of Experiment 2. This is because central to the irrelevant sound effect is that recall is disrupted by *irrelevant* sound, that is, sound which participants are instructed to ignore. Because it seems that attention to noise-vocoded speech is needed for gains in intelligibility, participants will likely need to be assigned to intelligible and non-intelligible groups. In this way, participants can be allowed to gain intelligibility of noise-vocoded speech prior to being instructed to ignore noise-vocoded speech backgrounds during a serial recall task.

It is likely that any comparisons between participants based on intelligibility would have found null results. Tremblay et al. (2000) conducted an experiment of ISE using sinewave speech, in which participants were randomly assigned to either have speech or non-speech percepts of sinewave speech. This experiment found no difference between the two sinewave conditions, suggesting that intelligibility of a background does not influence ISE. Conversely, results reported by Viswanathan et al. (in press) indicate that, irrespective of intelligibility, the presence of a speech-specific frequency profile does affect ISE. This pattern of findings suggests that intelligibility is not an ideal measure of speech fidelity. In fact, several studies of sinewave and noise-vocoded speech show that listeners can shift from non-speech percepts (an unintelligible signal) to speech percepts (an intelligible signal) for the same stimuli (e.g. Davis et al., 2005; Remez et al., 1981; Shannon et al., 1995; also see foot note 4 on perceptual bi-stability).

This study was not concerned with the impact of intelligibility on ISE, rather it used intelligibility as a proxy to investigate the effect of speech fidelity on ISE. Intelligibility is likely not to be a sufficient measurement of speech fidelity, this point was made clear by Tremblay et al. (2000), who compared the ISE produced by sinewave speech between participants with, and those without intelligible percepts of the background. While Tremblay et al. did not find a difference in ISE based on intelligibility, Viswanathan et al (in press) did find a difference based on the presence of a speech like formant organization, supporting the assumption that speech fidelity contributes to ISE. Taken together, these studies illustrate that intelligibility, does not precisely correspond to the presence of speech fidelity, making it an imperfect measure of the construct of interest to this study. However, the possibility of intelligibility can be indicative of the presence of speech fidelity. Rather than comparing groups of *participants* based on the presence of speech percepts of noise-vocoded speech, we can compare ISE between different noise-vocoded *conditions*, based on the presence of intelligibility. The intelligibility measure used in this study did not specify to participant that the background was speech, but rather participants were asked to report their general impressions of the auditory stimuli. That participants correctly identified words in the noise-vocoded speech indicates that it contained acoustic features that distinguished it as speech as opposed to non-speech. This assumption is supported by the findings that many of the participants that did not correctly identify the words during the intelligibility measure did identify the stimuli as speech (e.g. responding “Balls” when listening to “Bowls”). Therefore, I assume that speech fidelity was present at the 6 channel condition, and persisted with the 9 and 12 channel conditions.

One possible explanation for the pattern of results reported for Experiment 2 is that the pattern of ISE abated after the 3 channel condition. This change in ISE across conditions could be the result of the presence of a speech like-intensity profile that was present in the 6, 9, and 12 channel conditions but not the 3 channel condition. Three channel noise-vocoded speech was never intelligible as speech in this study, nor has it been reported to be intelligible in any of the known literature. Conversely, our study did find limited amounts of intelligibility in the 6, 9, 12 channel conditions. The presence of intelligibility indicates that speech information was present in the stimuli. The argument here is that this presence of speech information could boost ISE. If the addition of three channels of noise-vocoded speech only increased ISE modestly, such that the difference between adjacent conditions would be insufficient to cause a significantly different ISE, but the addition of speech information could cause a step up in ISE resulting in a disparity on either side of this step.

For the present study, I suggest that experiment 2 may have failed to find a difference between the 6 and 12 channel condition because the increase in changing state complexity between these conditions was too small to cause a significant change in ISE. If a speech intensity profile contributes to ISE then a difference resulting from speech fidelity should not be expected between the 6, 9, and 12 channel conditions, as this profile will be present in all of them. The lack of an effect may be the result of the linear increase in channel number being an insufficient increase of changing state complexity effect serial recall disruption. One way of investigating this hypothesis is to amplify the effect of changing state complexity, by expanding the range of channels in noise-vocoded speech backgrounds (see above). Another way of investigating this possibility would be to examine the role of speech fidelity independent of complexity. A clear conclusion that can be drawn from this study is that changing state complexity must be better clarified, if its role in ISE is to be understood. Additionally, when considering speech backgrounds, both the information contained in, as well as the complexity of, the acoustic signal should be investigated carefully so that the relative contribution of each can be investigated independently, as they relate to ISE.

References

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior, 15*, 17–32.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *The psychology of Learning and Motivation*, *8,* 47-89.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Beaman, P. C., Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: tests of the changing-state hypothesis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23 (2), 459-471

Beaman, C. P., & Jones, D. M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *The Quarterly Journal of Experimental Psychology: Section A*, *51*(3), 615-636.

Buchner, A., Rothermund, K., Wentura, D., & Mehl, B. (2004). Valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory & Cognition*, *32*(5), 722-731.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, H., Taylor, K., & McGettigan (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134* (2), 222-241

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., Taylor, K., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance, 37* (1), 283-295

Hughes, R. W., Tremblay. S., Jones, D, M. (200[5]). Disruption by speech of serial short-term memory: The role of changing-state vowels. *Psychonomic Bulletin and Review, 12*, 886-90.

Hygge S., Boman E., & Enmarker I. (2003). The effects of road traffic noise and meaningful irrelevant speech on different memory systems. *Scandinavian Journal of Psychology, 44*,13–21.

Jones, D. M. (1993). Objects, streams and threads of auditory attention. *In Attention: Selection, Awareness and Control, ed*. AD Baddeley, L Weiskrantz, pp. 87–104. Oxford, UK: Clarendon

Jones, D. M., Alford, D, Macken, W. J., Banbury S. P., & Tremblay S. (2000). Interference from degraded auditory stimuli: Linear effects of changing-state in the irrelevant sequence. *Journal of the Acoustical Society of America*, *108*(3), 1082–1088.

Jones, D. M., & Macken, W. J. (1993). Irrelevant Tones Produce an Irrelevant Speech Effect: Implications for Phonological Coding in Working Memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 19 (2), 369-381.

Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition*, *21*(3), 318-328.

Jones, D. J., & Tremblay, S. (2000). Interference in memory by process or content? A reply to neath (2000). *Psychonomic Bulletin & Review. 7* (3), 550-558.

Larsen, J. D., Baddeley, A., & Andrade, J. (2000). Phonological similarity and the irrelevant speech effect: Implications for models of short-term verbal memory. *Memory*, *8*(3), 145-157.

LeCompte, D. C. (1996). Irrelevant speech, serial rehearsal, and temporal distinctiveness: a new approach to the irrelevant speech effect. *Journal of experimental psychology. Learning, memory, and cognition*, *22*(5), 1154.

LeCompte, D. C., Neely, C. B., & Wilson, J. R. (1997). Irrelevant speech and irrelevant tones: the relative importance of speech to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition 23* (2) 472-483

Loizou, P. C., Dorman, M., & Tu, Z. (1999). On the number of channels needed to understand speech. *Journal of the Acoustical Society of America*, *106* (4) 2097-2103.

Little, J. S., Martin, F. H., & Thomson, R. H. (2010). Speech versus non-speech as irrelevant sound: Controlling acoustic variation. *Biological psychology*, *85*(1), 62-70.

Macken, W. J., Mosdell, N., & Jones, D. M. (1999). Explaining the irrelevant-sound effect: Temporal distinctiveness or changing state? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25* (3), 810-814.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251-269.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, *7*(3), 403-423.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic*

Bulletin & Review, 7, 403-423.

Page, M. P.A., & Norris, D. (1998). The primacy model: a new model of immediate serial recall. *Psychological Review*. *105*, 761–81.

Page, M. P. A., & Norris, D. G. (2003). The irrelevant sound effect: What needs modelling, and a tentative model. *Quarterly Journal of Experimental Psychology Section A*, *56*(8), 1289-1300.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. Science, 212 (4497), 947-950.

Roberts, B., Summers, R. J., & Bailey, P. J. (2011) The intelligibility of noise-vocoded speech: spectral information available from across-channel comparison amplitude envelopes. *Proceedings of the Royal Society: Biological Science 278* (1711) 1595-1600.

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of verbal learning and verbal behavior*, *21*(2), 150-164.

Salamé, P., & Baddeley, A. (1983) Differential effect of speech and noise on short term memory. In G. Rossi (Ed.) Procedings of the fourth international congress on noise as a public health problem (pp. 751-758). Milan: Centro Ricerch e studi amplifon.

Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology 41* (1) 107-122

Schlittmeier, S. J., Weiβgerber, T., Kerber, S., Fastl, H., & HellbrÜck, J. (2012). Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. *Attention, Perception, and Psychophysics 74*. 194-203.

Shannon, R. V., Zeng, F., Kammath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270* (5234) 303-304

Shannon, R. V., Zeng, F., & Wygonksi, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *Journal of the Acoustical Society of America*, *104* (4), 2467-2476.

Viswanathan, Dorsi, & George (in press). The role of speech-specific properties of the background in the Irrelevant Sound Effect. *Quarterly Journal of Experimental Psychology*.

Author Note

I thank Dr. Annie Olmstead for her extra help in the writing of the thesis. Thanks also to Dr. Matt Davis for his assistance in the design of the noise-vocoded speech used in this study. This research was supported by NIDCD grant R15DC011875-01 to Dr. Navin Viswanathan.

Figure Captions

Figure 1 depicts the serial recall accuracy for noise-vocoded speech and white noise across serial position (Experiment 1). Noise-vocoded speech results in significantly worse serial recall than does white noise.

Figure 2 displays the average difference score calculated as the serial recall accuracy associated with noise-vocoded speech subtracted from the serial recall accuracy associated with white noise. Difference scores were calculated for each background condition. Error bars represent standard error. The horizontal axis represents each noise-vocoded speech condition, and the vertical axis represents the average difference in recall accuracy between white noise and noise-vocoded speech.

Figure 3 shows the average proportion of words that were correctly identified for each background condition. Error bars represent standard error. The horizontal axis represents each noise-vocoded speech condition, and the vertical axis represents the proportion of words recalled. Note that even in the most intelligible condition (12 channels) intelligibility is still quite low (less then 0.05).

Figure 1.

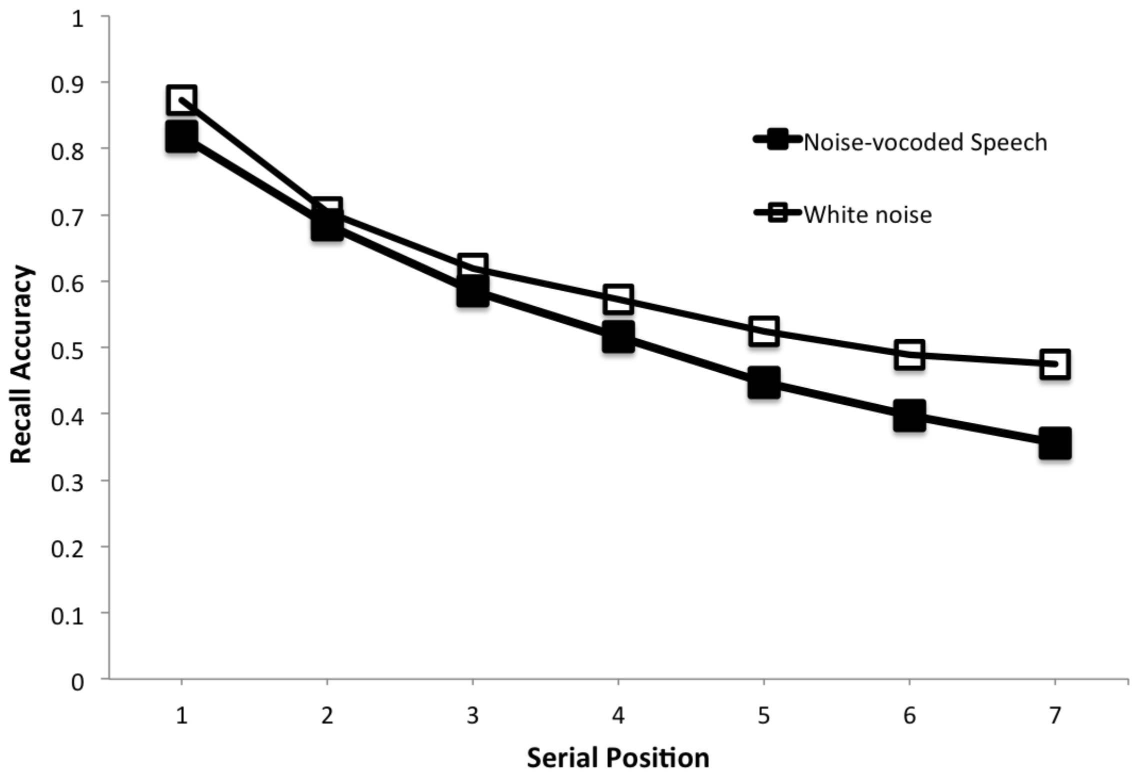
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Recall Accuracy
1
2
3
4
5
6
7
Serial Position
Noise-vocoded Speech
White noise

Figure 2.

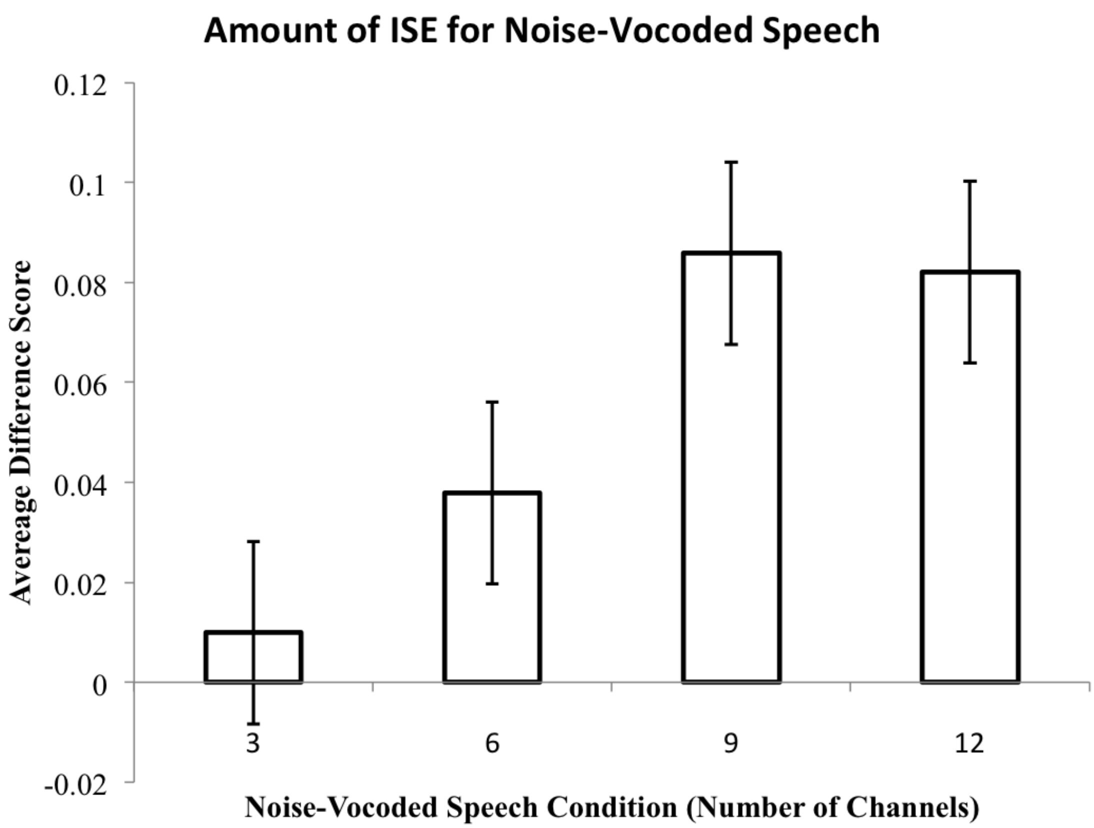
Amount of ISE for Noise-Vocoded Speech
Avereage Difference Score
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
3
6
9
12
Noise-Vocoded Speech Condition (Number of Channels)

Figure 3.

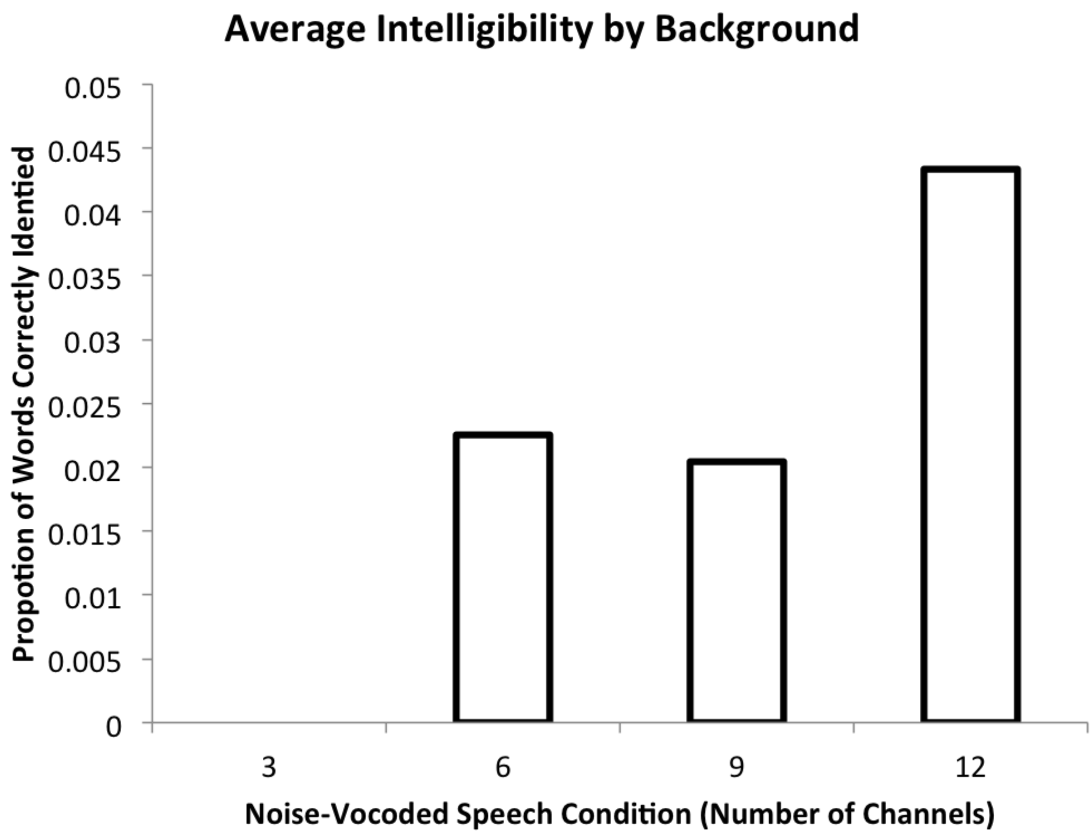
Average Intelligibility by Background
Propotion of Words Correctly Identied
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
3
6
9
12
Noise-Vocoded Speech Condition (Number of Channels)

Table Captions

Table 1 Summarizes the theories related to ISE. The phonological loop hypothesis and the feature model are part of the broad assumption that ISE is driven by unique qualities of speech. Alternatively, the primacy model and changing state hypothesis assume that ISE is entirely a product of the number of changes in state in the irrelevant sound.

Table 2 displays the frequency range of each channel for each background condition. Channel numbers ranging from 1-12 are shown in the top row. The left column shows the noise-vocoded conditions used in this study, 3, 6, 9, and 12 channeled noise-vocoded speech. All channel ranges occur within the range of 50-8000hz, meaning the channel 1 of all conditions has a lower limit of 50hz, and the highest limit for the highest channel is always 8000hz.

Table 1

	Information Models		Complexity Models	
	Phonological loop hypothesis	Feature model	Primacy model	Changing state hypothesis
Mechanism of Disruption	Speech prevents rehearsal of serial items	Speech accelerates decay of serial cues	Number nodes diffuses activation resources	Sub-streams compete for space in memory
Explanation for speech related disruption	Speech has preferential access to the phonological loop	Speech crowds memory with cues increasing the rate of decay	Speech has many changes in state, resulting in many nodes	Speech has many changes in state, resulting in many sub-streams
Explanation for non-speech disruption	Non-speech gains access to the store preferentially based on *how* speech like it is	Non-speech disruption is epiphenomenal, the result of split attention	Each change in state of irrelevant sound results in a node of activation	Each change in state of irrelevant sound results in a sub-stream in memory

Table 2

Conditions	Channels											
	1	2	3	4	5	6	7	8	9	10	11	12
3	221	1202	6526									
6	179	329	603	1104	2025	3710						
9	38	67	117	206	361	635	1116	1962	3448			
12	26	40	62	93	143	218	333	509	775	1185	1807	2759

Channel ranges shown in Hz

Appendix A

Read parts in *italics* aloud.

Thank you for participating in our experiment.

For this experiment you will wear these head phones, point to the headphones, *and you will watch this screen*, point to the computer screen.

Be sure to pay attention to the computer screen, and read all instructions carefully.

In this experiment you will be shown a series of letters on the computer screen. Your task is to report this items in the order in which they were presented. You will report these letters by typing on the keyboard and pressing enter when you are done. You will report these items shortly after the presentation ends.

During the presentation you may hear some sounds through your headphones, do your best to ignore these sounds and focus on the letters being shown on the screen.

Do you have any questions?

When this portion of the experiment is done, the computer will prompt you to come get me and then we will begin the next segment of the experiment.

During this part of the experiment you will listen to some sounds through the headphones. After each sound, you will be asked to just write what you heard.

Answer any questions they may have, don't describe the sounds as anything other than sounds. Press the "Z" key so that they may continue.

Appendix B

At the center of the screen at the beginning of the experiment.

Thank you for participating in our study.
For this experiment you will be shown a series of letters on the screen.
Be sure to pay attention to these letters, you will be asked to report.
When you report these letters be sure that your report them in the CORRECT ORDER.
You may hear sounds through your headphones, please ignore these and focus on the letters being shown to you.
Press SPACE to begin.

At the bottom of the report screen following serial presentations.

Press ENTER when done.

At center of the screen following report screen.

Press SPACE to go on.

At center of the screen following the final report screen.

This portion of the experiment is complete.
Please contact the experimenter to continue.

At center of the screen after the experimenter unlocks the program.

In this part of the experiment you will hear some sounds in your headphones.
After listening to each sound write what you heard.
Press SPACE to continue.

At center of the screen following presentation of intelligibility test words.

What did you hear?
Press ENTER when done.
Press SPACE to go on

Footnotes

[1] Larsen, Baddeley and Andrade (2000) suggest that non-speech items may enter the phonological store in a gradient of preference towards speech. The claim being that non-speech items can access the loop with different degrees of preference that are dictated by how similar the item is to speech. The finding that pure tones cause ISE raises the question of how speech-like something must be to gain access to the phonological loop. Surely the sub-vocal articulations that add serial items to the phonological loop have more in common with speech than pure tones.

[2] Page and Norris (2003) explain that recency effects can also be explained by the primacy model. The explanation is that, as the relative position of serial items becomes less defined, the likelihood that items will be recalled in earlier or later positions adjacent to their own increases. The recency effect occurs because the last item of a list has only one adjacent position with which it can be confused.

[3] Due to a programmer's error in the experiment, data was not collected from every participant's intelligibility measure. Specifically, I analyzed intelligibility data from eight participants in the 3 channel condition, 19 participants in the 6 channel condition, six participants in the 9 channel condition, and eight participants in the 12 channel condition.

[4] I use the term *perceptually bistable* for lack of a better term. When listening to noise-vocoded speech, the same signal can be heard as speech and non-speech by the same

listener, however, once a speech percept is achieved, listeners do not revert back to the non-speech percept.

[5] It is commonly acknowledged that noise-vocoded speech has a hysteresis effect. While listeners may not have speech percepts of noise-vocoded speech of low channel ranges, having a speech percept of noise-vocoded speech with many channels subsequently makes speech with fewer vocoded channels more likely to be intelligible. For more information on this effect I direct the reader to: http://www.mrc-cbu.cam.ac.uk/people/matt.davis/vocode/.