Assessing and Improving listening skills: a test of two theories

Clint Bowers[1], Talib Hussain[2], Katelyn Procci[3]
[1]University of Central Florida, clint.bowers@ucf.edu
[2]Raytheon BBN Technologies, thussain@bbn.com
[3]University of Central Florida, katelynprocci@gmail.com

Abstract

Many important tasks depend upon the ability of personnel to be able to extract information from verbal communication in suboptimal conditions. However, there is little guidance in how best to train people to improve this skill, specifically regarding the most effective combination of human or synthesized speech with or without text captions. In this study, we examined two competing theories, the cognitive theory of multimedia learning versus resilient listening. One-hundred and nineteen U.S. Navy recruits (53% male, average age of 21.5 years) participated in this study. The results indicated that games with degraded auditory conditions did not improve listening abilities in a transfer condition. Games using recorded human voices resulted in the best performance.

Keywords: *training, game, synthesized speech, multimedia learning*

1. Introduction

Communication is an important component of many critical tasks, and humans must be able to extract information in challenging environmental conditions. It is often the case that critical information, such as orders, locations, and questions, are delivered in noisy or degraded conditions. This is especially true in emergency situations where accurate reception of information is of paramount importance, and effective performance relies on the ability of operators to extract this information. Interestingly, however, there has been little research about how best to improve humans' skill in this regard. The present work describes a study that sought to test potential training interventions that were derived from contemporary theories of skill training.

Good communication has both a sender and a receiver, a message that can be understood using knowledge common to both parties, a shared mental model of communication structures and strategies as appropriate for different situations, effective listening, and a feedback system, which ensures that the message was correctly interpreted. In many critical situations, such as military command or emergency response, communication is an unambiguous, highly structured, hierarchical system that requires accuracy, consistency, and brevity to be effective [1]. Further, messages that are laconic yet complete reduce mental load, thus improving performance [2] [3] [4]. However, the ultimate effectiveness of communication depends on the listener to discern the information, even in environments that are suboptimal. Most communication training is targeted at the speaker, with the goal of satisfying the information requirements described above. However, there is also a need, we suggest, to improve the skills of the listener to obtain the information.

Specifically, we seek to develop simulation environments that can be useful in improving listener's abilities to extract information from imperfect verbal exchanges. Simulations are extremely useful for training in general, including for communication skills (e.g [5] and [6]). Live training exercises are the traditional method of communications training, but compared to the financial and personnel resources needed for this, computer-based simulations are inexpensive interventions and accessible practice environments. However, key design issues remain as how to best deliver communications training in simulations. This communication skill is active – it involves both learning the proper format for sending messages as well as the ability to receive and correctly comprehend them. This can be achieved in a computer-based training simulation in a number of ways, from the low fidelity of text-based inputs and outputs to incorporating high-fidelity spoken audio and speech recognition programs. Along with the level of fidelity, these methods vary in terms of both cost, complexity, effectiveness, and return on investment. As useful as simulations are for training, empirically determining the best practices for delivering communications training in a simulation is vital.

Besides the advantages of cost and accessibility, simulation-based training allows us to create experimental conditions to test new interventions. For example, using synthesized audio allows manipulations of speech characteristics assumed in various theories of human performance. Some researchers have suggested that any synthesized audio increases demands on working memory (e.g. [7]), and according to the work of Mayer, this additional cognitive load will damage learning outcomes if not structured properly (e.g. [8]). On the other hand, this additional cognitive load may be a benefit rather than an obstacle to be overcome. Stemming from the idea that more difficult conditions encourage deeper processing and thus better learning [9], training under higher levels of cognitive load may help to develop something we call resilient listening skills and improve communications training outcomes. The present study manipulates human and synthesized speech in a game-based training simulation, using U.S. Navy recruits, to investigate their comparative benefits for communication performance. The goal of the study was to investigate whether such a training regimen could help workers who must perform in impoverished communication conditions.

2. Cognitive Load and Training Communications Skills

The brain has limits with regard to mental processing capability, and this has been addressed in the education literature. For example, Mayer’s work focuses on increasing declarative knowledge through managing cognitive load. His theory of multimedia learning states that when information is presented in two different modalities simultaneously, such as an animation and text, the information is processed in the brain by splitting the different presentation modalities into two separate channels. This reduces the amount of cognitive load, ultimately resulting in enhanced recall and retention [8]. Pairing different modalities is effective in improving learning outcomes. For example, one study found that those who viewed an instructional animation paired with spoken narration outperformed those who had the animation paired with text [10]. This is one of the reasons why simulations are so effective. Simulations are basically informational, interactive animations, and when paired with verbal information, this has the potential to enhance their efficiency.

Whether additionally including text will improve recall and retention in learning is unclear. For example, in a sample of college students playing a game to learn about botany, there was improved recall in those who received narration alone and narration integrated with text compared to text alone, however those in the two narration conditions performed equally well [11]. This suggests that narration is an important part of the game experience, but the inclusion of text captions was irrelevant. A different study found that auditory and text explanations were superior to auditory explanations alone with respect to learning if there were no other visual distractions on-screen [12]. In an engaging virtual training environment with the potential for multiple interactions occurring on-screen, including verbal narration while engaging in other tasks as a part of the immersive education seems pertinent, although the relative effectiveness of including text is unclear.

Mayer's work, however, may not apply to the training of skills. Perhaps training communications skills in a simulation under different levels of cognitive load may better prepare people for actual tasks in real-world, complex environments. Altering the level of cognitive load can be achieved through the way communication is presented in an interactive training simulation.

One way to provide spoken narration in a training simulation is to use a speech synthesis program. Doing so, rather than recording audio clips using live human voice actors, saves both time and money, thus improving on return-on-investment, especially when the spoken training content may need to be updated repeatedly. Importantly, the use of synthetic speech during training increases cognitive load. Studies have found that the use of synthetic voices may lead to poorer learning when compared to interventions that used natural speech, apparently due to the additional working memory resources needed to process the synthetic speech, leading to faulty encoding and decreased recall (e.g. [13], [7], [14], [15]).

It may be, however, that this increased cognitive load increases transfer to environments with impoverished auditory environments. If people are trained on communication protocols and to extract important information using synthetic speech, despite the difficulties of increased load and improper encoding, they may develop resilient listening skills. For example, it has been demonstrated that training with synthetic speech allows participants to perform better in a transfer task that uses synthetic speech [16]. This is related to the idea of disfluency, a body of research that suggests the harder a task is to perform, the more deeply the information is processed in an attempt to overcome the additional load [9]. For example, one set of studies found that students who read instructional materials with a font not commonly used were able to comprehend more information than those with standard, supposedly easier to read fonts [17]. The additional mental effort needed to process the information apparently improved learning. While this finding does not directly pertain to skills, applying the same theory to the communication training problem, using synthetic voices may yield improved training outcomes. Training under this additional load and becoming accustomed to extracting information from difficult audio samples may result in a more deeply-trained skill. As communication during a military emergency occurs under great stress and degraded listening conditions, the development of resilient listening skills may further improve their ability to extract information and communicate more effectively under these suboptimal conditions.

This study investigated the effect of training under different levels of cognitive load by including narration and text in game-based training interventions as well as determine how the use of synthetic speech affects training outcomes specific to damage control communication skills. Using synthetic speech may increase cognitive load, which may damage learning and training outcomes [18]. However, one might also suggest that training under these more difficult circumstances will result in the development of resilient listening skills, which may improve performance in the real world. If synthetic speech improves the resilient listening skills of trainees compared to those trained under other conditions, this suggests that increases in cognitive load improve skills-based training.

The theory of situated cognition, in which learning should occur in relevant contexts [19], suggests that audio is better than text for this training exercise since it is a verbal communication skill. We hypothesize that audio training should result in an increased ability to extract critical information in degraded auditory environments than text training. We also hypothesize, in line with our resilient listening theory, that training with synthetic speech will result in greater recall of critical communications under strained, more realistic conditions in which it is more difficult to hear, which represents what recruits may face during an actual damage control situation.

Hypothesis 1a: Those with narration, either human or synthetic, will recall more critical information elements than those with text only for the normal audio clips.

Hypothesis 1b: Those in the synthetic speech condition will recall more critical information elements than those in the human voice only and text only conditions for the distorted audio clips.

Another view encourages cognitive load management for teaching declarative knowledge [20]. If this same holds true for training other skills, it suggests that pairing multiple modalities (e.g., audio and text) will result in the best outcomes. This may only work in this specific instance as this training intervention does not have any other visual information on the screen other than text input options (i.e., no actual animations). We also hypothesize that synthetic speech without text, given its increases in cognitive load without multimodal management by using captions, will result in less recall from communications under conditions in which it is difficult to clearly hear. Including this set of hypotheses will clarify whether it is additional cognitive load that results in skill training improvements or if it is cognitive load management that is key.

Hypothesis 2a: Those in the narration and text conditions will recall more critical information elements than those in the text-only and narration-only conditions across all audio clips.

Hypothesis 2b: Those in the synthetic voice only condition will recall fewer critical information elements than all other conditions in extracting information from the distorted audio clips.

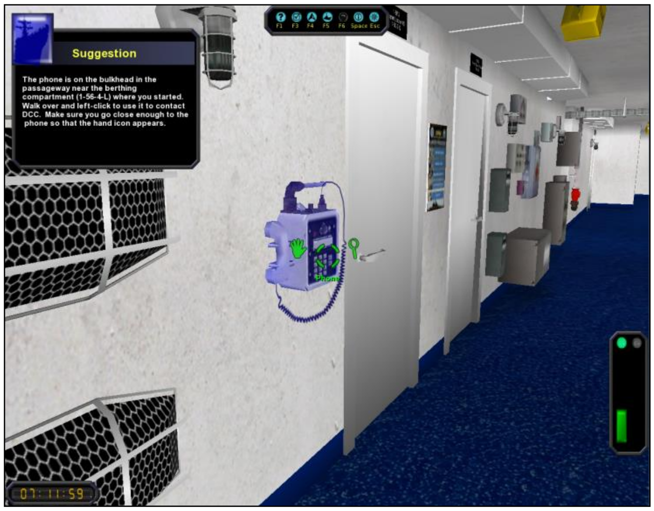
3. Method

3.1. Participants

Our sample consisted of 119 U.S. Navy recruits in their fifth week of basic training at Recruit Training Command at Naval Station Great Lakes in Waukegan, IL. Of these participants, 62 were male and 55 were female (two did not specify either sex), and were an average of 21.5 years old (*SD* = 3.44). Also, 35 (29.4%) were Caucasian, 14 (11.8%) were Hispanic, 12 (10.1%) were African-American, 4 (3.4% were Asian), 2 (1.7%) were some other race/ethnicity, and 52 (43.7%) did not respond. Participants indicated they spent 12.2 hours per week playing games before enlisting (*SD* = 18.6). Participants were not compensated.

3.2. Measures and Equipment

*Game-based training intervention***.** All participants were required to play the *Virtual Environments for Ship and Shore Experiential Learning* (VESSEL) *Damage Control Trainer* (DCT) as a part of their routine training. It is a game-based training environment used to train U.S. Navy recruits on damage control procedures (how and what to communicate to commanders), cognitive skills for effective decision making and adaptive thinking, situational awareness, and communication protocols. It is a single-player game from the first-person perspective, runs on a desktop computer, and the player uses a mouse and keyboard. The setting of the game is the detailed interior of a simulated Arleigh-Burke class destroyer. The player interacts with the virtual world through mouse clicks to use tools, open doors, and inspect objects. The player also is able to interact with other non-playing characters via Integrated Voice Communication System (IVCS) through the use of pop-up dialogue boxes and choosing between response alternatives (See Figure 1).



Suggestion

The phone is on the bulkhead in the passageway near the berthing compartment (1-56-4-L) where you started. Walk over and left-click to use it to contact DCC. Make sure you go close enough to the phone so that the hand icon appears.

Figure 1. IVCS on bulkhead.

The narrative involves a sailor who has recently been assigned to this vessel on a support mission in the Middle East. The game features three missions, each beginning with a briefing to establish a narrative and to list mission and learning objectives. The player progresses through the mission by completing specific tasks. The game, through instructional logic, provides instruction as needed as well as provides feedback on actions in real-time. An after-action review is provided at the conclusion of each mission.

The missions increase in difficulty and complexity over time. The first mission is a tutorial, the second mission focuses on undertaking standard duties such as securing compartments while teaching the recruit navigation skills, and the third mission begins with the ship vessel colliding with another, creating a damage control situation. The player is sent to investigate and eventually repair a leaking pipe (see Figure 2). The player not only practices the appropriate decision-making and technical skills for a flooding control situation, but the player also learns how to communicate correctly with Damage Control Central, which are the individuals who make the higher-level decisions during a General Quarters situation, and with their Leading Petty Officer.

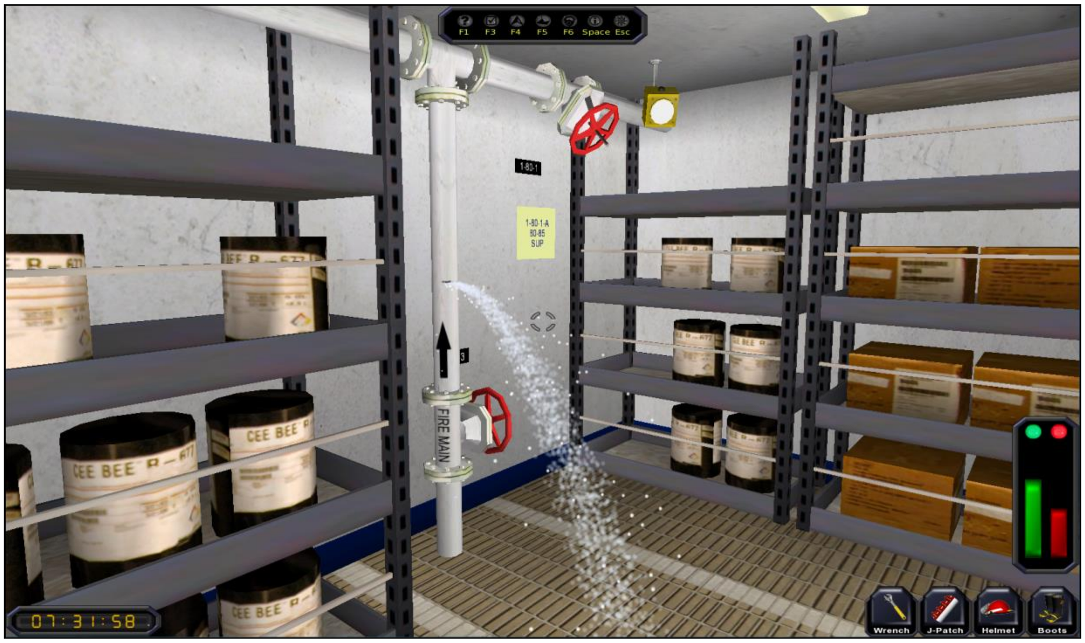
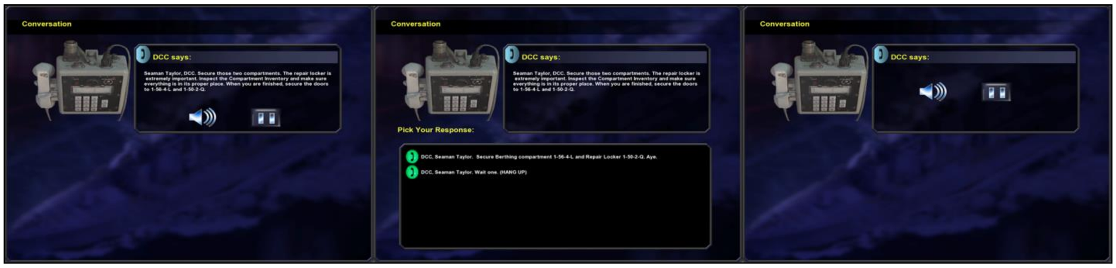


Figure 2: Leaking pipe in Mission 3.

Validation studies of the DCT have had promising results. Specific to the training problem presented here, comparing groups of recruits who only had traditional training and those who had training supplemented by playing the DCT, communication errors as tested in a capstone evaluation were reduced up to 80%. These included specific communication behaviors such as whether they identified themselves when initiating contact, reporting that they were ready to enter a compartment, and if they used repeat-backs correctly [21], [22], [23].

The DCT, however, only used text-based communications. Thus, it offered an ideal platform for examining the impact of using audio to further improve communication skills. The game was augmented to present the communications from Damage Control Center in five different formats. These formats consisted of combinations of audio clips recorded from a live human voice actor or audio clips generated with a synthesized speech program (see Figure 3). To prepare for this study, we investigated a variety of speech synthesis platforms before deciding to use the Loquendo™ text-to-speech system.



Seaman Taylor, DCC. Secure those two compartments. The repair locker is extremely important. Inspect the Compartment Inventory and make sure everything is in its proper place. When you are finished, secure the doors to 1-56-4-L and 1-50-2-Q.

DCC, Seaman Taylor. Secure Berthing compartment 1-56-4-L and Repair Locker 1-50-2-Q. Aye.

DCC, Seaman Taylor. Wait one. (HANG UP)

Figure 3: Communication interface for text and audio, text only, and audio only conditions.

*Game play***.** Using self-reported numbers, we collected basic DCT gameplay information including how many times the participant played each training mission and how long it took.

*Demographics***.** We also collected basic demographic information from the participants, which included their age, sex, and ethnic background. Additionally, we asked them to report, on average, how many hours a week they spent playing video games.

*Information extraction***.** In order to test the amount of information that recruits were able to comprehend from communications from the Damage Control Center we developed an information extraction exercise that tested their active listening skills. Participants were presented with 15 audio clips using human voice actors, all of which featured hypothetical instructions from the Damage Control Center such as, "Seaman Taylor, DCC. Secure those two compartments. The repair locker is extremely important. Inspect the compartment inventory and make sure everything is in its proper place. When you're finished, secure the doors to 1-56-4-Lima and 1-50-2-Quebec." For each audio clip, participants were instructed to write down any information they believed to be important. For each audio clip, essential items were identified and used as a basis for scoring. Participants received one point for each critical item they identified. There were 47 total information elements that could have been identified.

The audio clips were edited to differ in comprehension difficulty. The control audio clips featured human audio that was clearly spoken (Control). Other audio clips, also recorded using human voice actors, were edited to introduce background static and white noise (Noise), which simulates the potential quality of the communications during a damage control situation. This was accomplished by overlaying pink noise (at -2.5db). Finally, a third group of audio clips was edited to increase the tempo of speech (Speed), which also increased the audio clip's difficulty by increasing the speech rate by 50%.

Audio clips were balanced by their different types to include roughly equal numbers of male and female speakers, with three female and two male audio clips for both Noise and Speed and three male and two female audio clips for Control. Preliminary analysis of the entire dataset did not find any difference between extraction scores between the male and female audio clips across the different audio clip subscales, $F(4,95) = 2.064$, $p = .092$, therefore there was no further distinction made between the male and female audio clips during analysis.

3.3. Procedure

Participants played through three missions in the DCT. Importantly, the second mission introduced the player to communication protocols and the third mission had participants actively participating in a flooding control exercise, which required them to use communication and shipboard navigation skills as well as damage control procedures specific to keeping the compartment watertight.

When recruits reported for their required training using the DCT, they were randomly assigned to one of five conditions as they entered the room. Based on their condition, they played a different version of the DCT that varied how communication protocols were practiced. Only the conversation text was manipulated, while the other forms of text in the game such as the briefings, help screens, feedback, and demerits, stayed consistent across conditions. In the game, players interact with the Damage Control Center during a shipboard emergency. This requires the player to pick up telephone-like device and make accurate reports by selecting the appropriate responses, which are presented on-screen as text options. To succeed, players must understand and use the correct communications protocol. The game presented the communications from the Damage Control Center in the five different formats across five different conditions. For the synthesized conditions, both male and female voices were generated, and for the human voice condition, both male and female voice actors were used. We also alternated by condition whether on-screen text captions were present. The specifics of the conditions are listed below in Table 1.

Table 1. Description of experimental conditions

Condition	Abbrev	Description	*n*
Human Audio Only	HO	Audio clips featuring a live human voice actor without text-based captions displayed on-screen	27
Human Audio + Text	HT	Audio clips featuring a live human voice actor with text-based captions displayed on-screen	23
Synthetic Audio Only	SO	Audio clips generated synthetically without text-based captions displayed on-screen	22
Synthetic Audio + Text	ST	Audio clips generated synthetically with text-based captions displayed on-screen	23
Text Only (control)	TO	No audio clips with text-based captions displayed on-screen	24

Recruits recorded their own game performance on a sheet of paper that we provided. After playing the game, an experimenter addressed the recruits and informed consent for the data collection portion of the study was obtained. If participants agreed, they completed the information extraction exercise and their data was retained. Participation was optional and anonymous.

Several one-way ANOVAs were conducted to examine for pre-intervention differences across conditions. There were no differences across conditions with respect to sex, age, race, hours per week spent playing games, or degree of comfort with games.

4. Results

4.1. Intervention Effectiveness

Throughout the course of our analysis, a number of mixed-model ANOVAs were conducted. -The main effect of clip type on extraction scores was examined for the four main hypotheses to assess the effectiveness of the intervention. The data satisfied the assumptions for analysis by ANOVA without need for transformation.

Hypothesis 1a: Narration vs. Text Only - Those with narration, either human or synthetic, will identify more critical information elements then those with text only for the normal audio clips.

It was hypothesized that participants trained with any type of voice narration would outperform those who trained with only text. The one-way ANOVA was significant, $F(2,72) = 3.929$, $p = .024$, $eta^2 = .10$, a moderate effect where 10% of the variance in total pieces of information correctly extracted was due to condition. A Tukey test was conducted to determine which groups differed significantly. Across all types of audio clips, those in the Human Only condition were able to extract more total pieces of information than those in the Text Only condition. The Synthetic Only condition did not significantly differ from either of the two other conditions. See Table 2 for means, standard deviations, and the confidence intervals.

Our hypothesis was partially supported in that those who were exposed to audio training extracted more information from audio clips compared to those who only had text-based training. Although synthetic audio training did not differ significantly from either human audio or text-based training, it seems that human audio is superior overall given the relative gain over text-only training.

Hypothesis 1b: Synthetic Audio Only vs. Human Audio Only and Text Only - Those in the synthetic speech only condition will identify more critical information elements than those in the human voice only and text only conditions for the distorted audio clips.

If training of skills is improved by increasing cognitive load, we hypothesized that those in the synthetic audio only condition would recall and report more critical information elements than those in the human voice only and text only conditions for extracting information from the difficult clips. To test this, we conducted a 3x3 mixed model ANOVA where the between-subjects factor was condition (Synthetic Only, Human Only, and Text Only) and the within-subjects factor were the three types of audio clips (Speed, Control, and Noise). There was not a significant main effect for condition, $F(1, 55) = 2.241$, $p = .116$, $eta^2 = .075$, nor was there a significant interaction between condition and audio clip, $F(4,110) = 0.470$, $p = .758$, $eta^2 = .017$. We did not find that training was significantly associated with the amount of information extracted from the more difficult audio clips and did not find any support for the hypothesis of resilient listening.

Hypothesis 2a: Narration and Text vs. Narration Only and Text Only - Those in the narration and text conditions will identify more critical information elements those in the text-only and narration-only conditions across all audio clips.

If managing cognitive load is a better method for training a skill, we hypothesized based on Mayer's principles that those in the narration and text combined conditions would obtain more information from audio excerpts than those in the narration-only and text-only conditions across all audio clips with respect to total amount of information correctly extracted, regardless of how difficult information was to extract from the different types of audio clips. To test for this, we conducted a one-way ANOVA. The ANOVA was not significant $F(4,118) = 2.155$, $p = .079$, $eta^2 = .07$. Even though the ANOVA was nonsignificant, the size of the effect suggests it would be helpful to evaluate this relationship in a study with greater statistical power. There remains the possibility that there is a smaller effect where human audio only outperforms text-only while there do not seem to be any other differences across the other conditions (see Table 2).

Table 2. Total extraction scores by condition

Condition	*N*	Mean	Std. Dev	95% CI	Extracted
HO	27	16.41	7.32	13.51-19.30	34.91%
HT	23	13.91	6.57	11.07-16.75	29.60%
SO	22	13.32	8.75	9.44-17.20	28.34%
ST	23	14.17	6.90	11.19-17.16	30.15%
TO	24	10.54	6.28	7.89-13.19	22.42%

Notes: HO = human speech audio only; HT = human speech audio with text captions; SO = synthetic speech audio only; ST = synthetic speech audio with text captions; TO = text captions only; Total possible score is 47.

Hypothesis 2b: Synthetic Audio Only vs. All Other Conditions - Those in the synthetic voice only condition will identify fewer important information elements than all other conditions.

If managing cognitive load is a superior method for training a skill, we had also hypothesized that the synthetic voice only condition would be least able to identify and recall important information. A 5x3 mixed model ANOVA was used to evaluate this hypothesis. The between-subjects factor was condition and the within-subjects factor was the three types of audio clips: Speed, Control, and Noise. There was no main effect for condition, $F(1,88) = 1.397$, $p = .242$, $eta^2 = .06$, and the interaction between clip and condition was also nonsignificant, *F*(8, 176) = 1.300, *p* = .246, *eta*^{2} = .056. The training condition did not affect recruits' ability to extract information from audio clips of varying difficulties.

4.2. Further Examination of the Multimedia Theory of Learning

Given the complexity of Mayer's principles and the pattern of results indicating poor support for either theory, this suggested that cognitive load did not matter as much as using human audio to train a listening skill. Hence, we decided to look a bit deeper into the data from the perspective of these principles with a series of comparisons within the major divisions in the conditions to focus on training under different audio conditions with and without supplemental text.

Hypothesis 3a: Human Audio Only vs. Human Audio and Text – Those in the human audio and text condition should identify more information elements than those in the human audio only condition.

In accordance with Mayer, we hypothesized that all of those in the human audio and text (HT) condition would outperform those in the human audio only condition in extracting information across all conditions due to its optimal management of cognitive load. We conducted a 2x3 mixed model ANOVA where the between-subjects factor was condition and the within-subjects factor was the three types of clips. There was a nonsignificant main effect for condition, *F*(1,40) = 0.317, *p* = .576, *eta*^{2} = .008, which suggests that there was no additional gain when text was added. The interaction was not significant, *F*(2,80) = 0.175, *p* = .840, *eta*^{2} = .004. Again, there were no training effects related to human audio, with or without text, with respect to the amount of information identified from the difficult audio clips.

Hypothesis 3b: Synthetic Audio Only vs. Synthetic Audio and Text – Those in the synthetic audio and text condition identify more important information elements than those in the synthetic audio only condition.

We similarly hypothesized that scores in the synthetic audio with text conditions would be higher than those in the synthetic audio only conditions, indicating more information identified across all audio clips. We conducted a 2x3 mixed model ANOVA where the between-subjects factor was condition and the within-subjects factor was the three types of clips. The main effect of condition (*F*(1,30) = 0.023, *p* = .881, *eta*^{2} = .001) and the interaction (*F*(2, 60) = 2.242, *p* = .115, *eta*^{2} = .07) were not significant. There were no training effects related to synthetic audio with or without text with respect to being able to extract more information from the difficult audio clips.

5. Discussion

Communication is an extremely important element many critical tasks, especially those found in emergency response, and there is a need to improve training of these skills. Computer-based simulation training interventions are a known effective and promising tool [24]. However questions remained as to how such interventions should be designed for communication skills. There also are outstanding questions on the role cognitive load might play in training skills.

Two competing theories were used to explain the role of cognitive load in learning. The question was whether using technology to increase the cognitive load associated with listening would result in a greater ability to identify and recall critical information in a transfer condition. Specifically, our study sought to determine the effects of using synthetic speech, human recordings of speech and adding text captions on the amount of information identified in various training scenarios. There were no training effects with respect to different types of voices (i.e., natural human speech and synthesized speech) in combination with text when the task varied in difficulty. We also found that including text-based captions during training did not influence training outcomes. These results suggest that using human audio, regardless of whether text is ultimately included, is better than using synthetic audio for training listening skills in simulations. See Table 3 for a summary of the results.

Table 3. Summary of results

Hypothesis	Outcome
H1a: Narration vs. TO. Those with narration, either human or synthetic, will outperform those with text only for the normal audio clips.	Partially Supported - Audio training, specifically human audio, is superior
H1b: SO vs. HO and TO.* Those in the synthetic speech only condition will outperform those in the human voice and text conditions for the distorted audio clips.	Not Supported
H2a: Narration and Text vs. Narration Only and TO. Those in the narration and text conditions will outperform those in the text-only and narration-only conditions across all audio clips.	Not Supported - Trend present where human audio was best
H2b: SO vs. All. Those in the synthetic voice only condition will perform worse than all other conditions in extracting information from the distorted audio clips.	Not Supported
H3a: HO vs. HT. Those in the human audio and text condition should perform better than those in the human audio only condition in extracting information from the distorted audio clips.	Not Supported - Text is irrelevant
H3b: SO vs. ST. Those in the synthetic audio and text condition should perform better than those in the synthetic audio only condition in extracting information from the distorted audio clips.	Not Supported - Text is irrelevant

Notes: *Indicates this hypothesis was driven by the resilient listening theory.

5.1. Limitations

While this study offers the advantage that the participants are the actual end-users of the training, this characteristic also imposes shortcomings to the generalizability of the results. Recruits are, by their nature, task novices. These participants had not experienced an actual (or even simulated) naval emergency. As such, many of the words and concepts were novel to them. Therefore, it should be noted that these results are likely only applicable to people who are very early in the skill acquisition cycle. It may well be that participants at this level are least able to tolerate increases in cognitive load. The utility of the interventions might be more positive for those with greater level of skill.

A second limitation is that the participants had little time with the training intervention. This represents a trade-off between internal and external validity. While greater levels of exposure to the training might have created larger effects, the reality of the current military training environment is that brief training is the only feasible application of this approach. If greater levels of training are required to create an effect, the intervention simply could not be adopted due to the many other training demands for this group.

Finally, it should be noted that this study used a serious game to imposed a relatively high degree of cognitive load. It might be that the results of training would be more robust under lower loads.

6. Conclusions

This study tested whether Mayer’s guidelines for cognitive load management applied to skills such as listening, or whether training under more cognitive load improved outcomes. It appears that neither was true, at least in the present setting. For skills, higher fidelity seems to be more important [25]. While this study examined fidelity in part with the manipulation of speech generation, and found that human voices were optimal, further examination of multimedia learning theory and skill acquisition as it relates to fidelity is warranted. Specifically, further study of whether these results generalize across different methodologies should be explored.

There are other theories that may apply to this problem. For example, more native to the theory of situated learning, the present study did not train these recruits under actual stressful conditions (e.g., with background noise and static). A future study could examine how training using noisy audio clips affects performance. Practicing under stress then transferring to a non-stress environment is another training method, referencing the faster than real-time training literature [26], in which individuals are more efficiently trained under conditions that are worse than what they will actually experience. Stemming from work in the 1940s and 1950s, faster-than-real-time training for pilots, also known as FTRTT, found that training at a pace that was faster than the actual task resulted in more transfer [27]. This has generalized to other skills outside of the aviation domain. For example, Worm et al. [28] found that both experienced and novice trainees learned more using accelerated training than those trained at real-time speed in a drinking water treatment plant simulator.

Further, researchers should explore the possibility that synthetic speech did not add to mental overload, an issue we did not assess directly. Possibly, there is no difference in the brain processing capacity required to identify critical information from synthetic speech and natural speech. Future research in the area of simply identifying words versus complex comprehension would help understand the possible relationship between the two.

Also, factors such as reading level and preference should be considered. Participants were asked about their gaming experience and comfort level, but their reading experience could determine how well they extract information when text is present. Possibly, those who prefer reading will ignore narration in order to quickly scan text. If this is the case, then Mayer’s theories do not apply to the participants in this group.

Acknowledgements

We would like to thank the VESSEL team that helped with this study, both directly and indirectly: John Ostwald, Kerry Moffitt, Bruce Roberts, David Diller, and Brian Krisler from Raytheon BBN Technologies; Curtiss Murphy, Chris Rodgers, Brad Anderegg, and David Guthrie at Alion Science and Technology, Kelly Pounds at IDEAS; Susan Coleman and Ellen Menaker at Intelligent Decisions Systems, Inc.; and Allysa Singer, James Bohnsack, Benjamin Repkay, and Corina Lechin from the UCF RETRO Lab. We also would like to thank Mr. John Drake at the Naval Service Training Command and the staff at the U.S. Navy Recruit Training Command for their support and participation. This material is based upon work supported by the Office of Naval Research under Contract number N00014-08-C-0030, under the management of Dr. Ray Perez.

References

[1] Bowers, C. A., Jentsch, F., Salas, E., & Braun, C. C. (1998). Analyzing communication sequences for team training needs assessment. Human Factors: The Journal of the Human Factors and Ergonomics Society, 40(4), 672-679. http://dx.doi.org/10.1518/001872098779649265

[2] Achille, L. B., Schulze, K. G., & Schmidt-Nielson, A. S. (1995). An analysis of communication and use of military terms in Navy team training. Military Psychology, 7(2), 95-107. http://dx.doi.org/10.1207/s15327876mp0702_4

[3] Cannon-Bowers, J. A., & Salas, E. (1998). Team performance and training in complex environments: Recent findings from applied research. Current Directions in Psychological Science, 7(3), 83-87. http://dx.doi.org/10.1111/1467-8721.ep10773005

[4] Jacobson, S. K. (2009). Communication skills for conservation professionals. Washington, DC: Island Press.

[5] Raybourn, E. M. (2005). Adaptive thinking and leadership training for cultural awareness and communication competence. Interactive Technology and Smart Education, 2(2), 131 - 134. http://dx.doi.org/10.1108/17415650580000038

[6] Johnsen, K., Dickerson, R., Raij, A., Lok, B., Jackson, J., Shin, M., Hernandex, J., Stevens, A., & Lind, D. S. (2005). Experiences in using immersive virtual characters to educate medical communication skills. Proceedings of the IEEE Virtual Reality 2005 (pp. 179-186, 324). Piscataway, NJ: IEEE. http://dx.doi.org/10.1109/VR.2005.1492772

[7] Luce, P. A., Feustel, T. C., &Pisoni, D. B. (1983). Capacity demands in short-term memory for synthetic and natural speech. Human Factors, 25(1), 17-32.

[8] Mayer, R. E. (2002). Multimedia learning. Psychology of Learning and Motivation, 41, 85-139. http://dx.doi.org/10.1016/S0079-7421(02)80005-6

[9] Oppenheimer, D. M. (2008). The secret life of fluency. Trends in Cognitive Sciences, 12(6), 237-241. http://dx.doi.org/10.1016/j.tics.2008.02.014

[10] Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. Journal of Educational Psychology, 91(2), 358-368. http://dx.doi.org/10.1037/0022-0663.91.2.358

[11] Moreno, R. & Mayer, R. E. (2002a). Learning science in virtual reality multimedia environments: Role of methods and media. Journal of Educational Psychology, 94, 598-610. http://dx.doi.org/10.1037/0022-0663.94.3.598

[12] Moreno, R. & Mayer, R. E. (2002b). Verbal redundancy in multimedia learning: When reading helps listening. Journal of Educational Psychology, 94, 156-163. http://dx.doi.org/10.1037/0022-0663.94.1.156

[13] Lai, J., Wood, D., & Considine, M. (2000, April). The effect of task conditions on the comprehensibility of synthetic speech. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 321-328). ACM. http://dx.doi.org/10.1145/332040.332451

[14] Waterworth, J. A., & Thomas, C. M. (1985). Why is synthetic speech harder to remember than natural speech? CHI 85’ Proceedings (pp. 201-206). New York: ACM.

[15] Hongpaisanwiwat, C. & Lewis, M. (2003). Attentional effect of animated character. In M. Rauterberg, M. Menozzi, & W. Janet (Eds.), Human-Computer Interaction INTERACT '03 (pp. 423-431). Amsterdam: IOS Press.

[16] Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. Human Factors: The Journal of the Human Factors and Ergonomics Society, 27(4), 395-408.

[17] Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favor the bold (and the italicized): Effects of disfluency on educational outcomes. Cognition, 118, 111-115. http://dx.doi.org/10.1016/j.cognition.2010.09.012

[18] Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. Learning and instruction, 12(1), 1-10. http://dx.doi.org/10.1016/S0959-4752(01)00014-7

[19] Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge: University of Cambridge Press. http://dx.doi.org/10.1017/CBO9780511815355

[20] Clark, R. C., Nguyen, F., & Sweller, J. (2011). Efficiency in learning: Evidence-based guidelines to manage cognitive load. Wiley. http://dx.doi.org/10.1002/pfi.4930450920

[21] Hussain, T. S., Menaker, E., Pounds, K., Bowers, C., Cannon-Bowers, J. A., Murphy, C., Koenig, A., Wainess, R., & Lee, J. (2009, Nov. – Dec.). Designing and developing effective training games for the US Navy. Paper presented at I/ITSEC 2009, Orlando, FL.

[22] Hussain, T. S., Bowers, C., Blasko-Drabik, H. & Blair, L. (in press). Validating cognitive readiness on team performance following individual game-based training. In H. F. O’Neil, R. S. Perez & E. L. Baker (Eds.), Teaching and measuring cognitive readiness. New York: Springer.

[23] Bowers, C., Hussain, T., Roberts, B., Cannon-Bowers, J., & Blair, L. (under review). Preparing to practice: The use of a game-based simulation as a pre-training intervention for recruits. Simulation & Gaming.

[24] Steadman, R. H., Coates, W. C., Huang, Y. M., Matevosian, R., Larmon, B. R., McCullough, L., & Ariel, D. (2006). Simulation-based training is superior to problem-based learning for the acquisition of critical assessment and management skills. Critical care medicine, 34(1), 151-157. http://dx.doi.org/10.1097/01.CCM.0000190619.42013.94

[25] Schmidt, R. A., & Lee, T. D. (2011). Motor control and learning: A behavioral emphasis (5th ed.). Champaign, IL US: Human Kinetics.

[26] Guckenberger, Dutch and Stanney, Kay M. (1995): Poor Man's Virtual Reality. In: Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting 1995. p. 1063.

[27] Koonce, J. M. (2002). Human factors in the training of pilots. London: Taylor & Francis. http://dx.doi.org/10.1201/9780203164587

[28] Worm, G. I. M., van der Wees, M., de Winter, J. C. F., de Graaf, L., Wieringa, P. A., & Rietveld, L. C. (2012). Training and assessment with a faster than real-time simulation of a drinking water treatment plant. Simulation Modelling Practice and Theory, 21(1), 52-64. http://dx.doi.org/10.1016/j.simpat.2011.09.007