The intonation of a synthetic voice in digital storytelling.

Does a different variation of pitch intonation, when telling a story, affect narrative transportation towards the story world?

Linda Hoeijmakers

SNR: 2050941

Master Thesis

Communication and Information Sciences

Specialization: New Media Design

Tilburg School of Humanities and Digital Sciences

Tilburg University, Tilburg

Supervisor: Dr. T.O. Lentz

Second Reader: Dr. S.R. Ashby

January 2022

Abstract

This thesis examined to what extent the variation of a synthetic voice's intonation affects the transportation to a digital story world, focusing on 'pitch' as a prosodic feature. When the variation of intonation is changed to a more robotic voice, people will have a lower level of narrative transportation than when the variation of intonation is more human-like. A between-subjects experiment was conducted on three different intonation levels: standard intonation, reduced intonation, and no intonation. Eighty-six participants listened to a story, told by a synthetic voice, in either one of the conditions. The data did not provide evidence of a significant effect on transportation and the different variations of intonation. A possible explanation for the data not providing a significant effect is that the focus was on one prosodic feature, namely, pitch. Another prosodic feature might have a significant effect on the data. In a future study, the prosodic features 'loudness' and 'duration' could be examined in detail to determine if one could significantly affect the variation of intonation on narrative transportation.

Keywords: intonation, synthetic voice, narrative transportation, digital storytelling, Text-To-Speech systems, Human-Machine Interaction

Table of contents

Introduction

With text-to-speech (TTS), a machine transforms a written text into a spoken text (Gálvez et al., 2020; Maro et al., 2018; Stan & Lőrincz, 2016). A synthetic voice is increasingly used in products because the voices and techniques are being improved – making the voice more natural. Furthermore, with the new technique of New Neural Voices (NTTS), people find it harder to distinguish a synthetical voice from a human one (Baiardi, 2021; Gálvez et al., 2020; Ronanki, 2019). However, a study by Gregg (2020) reported that the quality of a synthetic voice is lower than the quality of a human voice. People could hear the difference between a synthetic voice and a human voice and described the synthetic voice as too robotic.

Moreover, Montaño and Alias (2016) reported that researchers who conducted studies on the intonation of a synthetic voice used different approaches regarding the intonation but were not transparent about the approach they used. Through these approaches, researchers focus on one prosodic feature in detail or in general – with the features being: pitch, loudness, and duration (Montaño & Alías, 2016; Velner et al., 2020).

TTS can be used in storytelling by having a synthetic voice to narrate the story. Products in which a synthetic voice can tell a story are story chatbots and (interactive) digital narratives, such as gamebooks in which the user can choose the direction of the story at key points (Baiardi, 2021; Ligthart et al., 2020).

A told story consists of expressiveness and specific acoustic cues that differ from a normal conversation and lead to a specific storytelling speaking style. However, the storytelling speaking style is complex and hard to reproduce (Montaño & Alías, 2016; Sarkar et al., 2014). Less is known about the difference of implementing a synthetic voice in a ‘normal’ versus ‘storytelling’ speaking style. In their study, Sarkar and colleagues (2014) reported that other researchers argue whether a synthetic voice could have a storytelling speaking style. Therefore,

they provided a prosody rule-set for implementing a storytelling speaking style in a synthetic voice.

When a machine tells a story, the intonation is an element that helps to understand the meaning of a text (Aarts & McMahon, 2006). However, the machine only knows the intonation of individual words. Furthermore, intonation differs between culture and language. A text can come across as angry or joyful depending on the native language of the listener (Aarts & McMahon, 2006; Gálvez et al., 2020; Ronanki, 2019; Stan & Lőrincz, 2016). The machine cannot understand the context of a written text, making it harder for the machine to implement a correct prosodic pattern.

When a synthetic voice is used for a spoken story, the voice needs to have a specific level of naturalness. When this level is not met, the listener could feel eerie. The feeling of eeriness is called the Uncanny Valley Effect (Velner et al., 2020). The machine has to produce a human-like sounding voice not to have an eerie feeling while people listen to a story. Hence, a machine finds it hard to implement the correct variation of intonation in a spoken story, making it curious to investigate whether different variations of intonation enhance the spoken text.

Furthermore, using synthetic voices in spoken stories is less time-consuming and less costly than using a human voice (Fernández-Torné & Matamala, 2015; Montaño & Alías, 2016). A human voice is mainly used within storytelling, but some stories use a synthetic voice (Baiardi, 2021; Ligthart et al., 2020; Sarkar et al., 2014). However, people perceive the intonation of a synthetic voice as more flat and duller (Hodari et al., 2019).

This thesis investigated one specific aspect of the quality of narration, namely, narrative transportation. With narrative transportation, the story world feels more real than the actual world around the reader or listener. People find the experience of narrative transportation enjoyable and get more engaged with the story (Green, 2014; Green & Jenkins, 2014; Kuijpers et

al., 2014). Furthermore, writing quality influences the level of transportation. Higher writing quality leads to more narrative transportation than lower writing quality (Green, 2014). However, whether speech quality can influence transportation when a story is spoken instead of written remains unanswered.

Narrative transportation can happen in all kinds of stories – such as gamebooks (Green & Jenkins, 2014). In a gamebook, the reader interacts with the story as they decide its outcome by making decisions at specific key points. According to Kuijpers and colleagues (2014) and Green and colleagues (2008, 2014, 2021; 2014), people who are more engaged with a story tend to have more transportation than less engaged people. When the readers are not engaged, they will not decide the story's outcome. Therefore, a gamebook is a story for which engagement is essential.

However, according to Gregg (2020), a human voice provides a higher level of narrative transportation than a synthetic voice. In addition, a study by Appel and colleagues (2021) examined the effect of intonation on narrative transportation. They found that facial expressions (from a robotic storyteller) have an illusionary effect on intonation and lead to transportation. There may be an effect of intonation on narrative transportation, but Appel and colleagues (2021) did not investigate this.

Overall, there is a possibility for an effect between the intonation of a synthetic voice and the transportation towards the story world. On the one hand, some studies mentioned above say that a synthetic voice could lead to transportation, while other studies mentioned above say a synthetic voice could not have a storytelling speaking style (Baiardi, 2021; Gregg, 2020; Hodari et al., 2019; Ligthart et al., 2020; Sarkar et al., 2014). Furthermore, the intonation of a synthetic voice is found to be flat, dull, and too robotic (Hodari et al., 2019).

However, synthetic voices have been improved since the introduction of NTTS, making synthetic voices harder to distinguish from human ones (Baiardi, 2021). Researchers have not yet

investigated whether NTTS decreases or increases the transportation towards a story world and how the intonation is affected, thus illuminating the relevance of this thesis. Therefore, the research question will be “to what extent does the variation of a synthetic voice’s intonation, affects the transportation to a story world?”

Theoretical background

Text-To-Speech

Through the years, Text-To-Speech (TTS) has been improved and is implemented in more modern-day products such as phones and computers (Gálvez et al., 2020; Maro et al., 2018; Stan & Lőrincz, 2016). When users have a speech conversation with a machine, they prefer a natural conversation pattern. In a natural conversation pattern, the experience of human-machine interaction is the same as the experience of human-human interaction (Velner et al., 2020).

TTS examples in human-machine interaction are screen reading features on phones, virtual assistants (Siri, Alexa), and synthetic voices in customer service. When people navigate through their phones, they can enable screen reading. With screen reading, a TTS system reads the screen aloud, which is beneficial for people who are visually impaired or people who can, temporarily, not see their screen – for example, while driving (Morris et al., 2018).

In addition, TTS is used for non-interactive purposes, such as audio descriptions and audiobooks. In audio descriptions, the visuals from a film are described aloud (Fernández-Torné & Matamala, 2015). When something is described aloud, people could be more engaged and have a higher level of attention than when something is not described aloud. Furthermore, describing a story aloud helps listeners to keep their attention while their minds wander off (Gregg, 2020).

According to Fernández-Torné and Matamala (2015), humans produce most audio descriptions, and a synthetic voice produces only some audio descriptions. They stated that a

synthetic voice is less time-consuming and less costly to produce audio descriptions. However, because producing speech with a human voice is more costly and time-consuming, audio descriptions are not widely available for online videos and movies (Fernández-Torné & Matamala, 2015).

When a TTS produces audio descriptions, a video could automatically get an audio description, which is beneficial for the videos posted on social media. Audio descriptions and other digital products such as interactive digital stories, podcasts, and audiobooks could benefit from a synthetic voice narrating the story.

Intonation

When people want to show more emotion while speaking, they can do this by changing some of the prosodic features of which the intonation is made. Roughly said, the different features can be divided into three categories: pitch, loudness, and duration (Sarkar et al., 2014; Velner et al., 2020).

In current studies, not all prosodic features are examined in detail. According to Montaño and Alias (2016), Sarkar and colleagues (2014), and Velner and colleagues (2020), most research only looks at one of the three features in general, and only a few research them in detail – as it is complex to examine and implement the features in a speaking style.

Furthermore, intonation varies depending on the context and narrative of the text and differs per person, culture, and language. For example, the pitch differs from high to low, and the duration from fast to slow when spoken by someone else (Aarts & McMahon, 2006; Gálvez et al., 2020; Ronanki, 2019; Stan & Lőrincz, 2016).

The intonation of a synthetic voice

TTS systems are improved since the development of new voice techniques. One of these techniques is called New Neural Voices (NTTS), which makes the synthetic voice more natural

in intonation (Baiardi, 2021; Gálvez et al., 2020; Ronanki, 2019). Since the development of NTTS, people have found it harder to distinguish an artificial voice from a human one. However, even with NTTS, synthetic voices still have problems with showing humor, emotions, and laughter (Velner et al., 2020).

A human understands the context of a written text, but a machine does not, making it harder for the machine to provide a natural intonation pattern. Velner and colleagues (2020) researched the intonation of a synthetic voice. One of the aspects of their research was the naturalness of the voice. A robot can be human-like to some extent; if there is a mismatch, a voice could be perceived as eerie, and people do not want to listen anymore. A natural intonation would be perceived as less eerie and a reduced intonation as eerier. (Velner et al., 2020). The feeling of eeriness can be described as the Uncanny Valley effect.

With a natural intonation pattern, the conversation between the human and the synthetic voice is fluently and provides the same experience as a human conversation (Hodari et al., 2019; Velner et al., 2020). According to Hodari and colleagues (2019), TTS systems have problems producing different versions of a text. Because machines have difficulty producing a natural intonation pattern, the possibility for the Uncanny Valley effect is present (Velner et al., 2020).

However, in the study by Velner and colleagues (2020), the extent to which the intonation variation leads to the Uncanny Valley effect was not defined. Additionally, in their study, it is unclear whether the intonation of a synthetic voice can be perceived in the same way as the intonation of a human and if a synthetic voice can adhere to a natural intonation pattern.

In addition, Yang and colleagues (2017) researched the intonation of a synthetic voice. They reported that a female synthetic voice is stereotyped as more emotional, empathetic, and warm than a male voice. Therefore, according to their study, a female voice can be used in situations where an emotional, empathetic, and warm voice is preferred.

TTS in storytelling

TTS can be used for storytelling in interactive digital narratives – such as story chatbots, audiobooks, podcasts, and gamebooks. Experiences and values are shared through a story from a storyteller to a listener (Farmer, 2004; Whittingham et al., 2013). Human voices are primarily used for audiobooks. However, some platforms examine whether a synthetic voice can produce an audiobook, such as Google, which released multiple audiobooks produced by a synthetic voice on their platform for audiobooks (Baiardi, 2021). Furthermore, people find listening to a story told by a synthetic voice an enjoyable experience – showing the benefits for a synthetic voice to be used in stories (Baiardi, 2021)

Furthermore, Ligthart and colleagues (2020) investigated the interaction between an interactive storytelling robot and a child. The child interacted with the robot while the robot told the story. Unfortunately, they did not describe how the robot's voice was generated and reported that interaction positively affected the engagement with the story.

Storytelling in digital narratives

A spoken story has its own speaking style, which consists of expressiveness and specific acoustic cues. Because of the storytelling speaking style, the listener is engaged with the story and feels entertained. Analyzing the storytelling speaking style is complex, and it is mainly researched in human storytelling (Montaño & Alías, 2016).

Furthermore, the storytelling speaking style is hard to reproduce as some prosodic features differ between an everyday speaking style and a storytelling speaking style (Sarkar et al., 2014). According to Sarkar and colleagues (2014), other studies mentioned that synthetic voices do not provide the same storytelling style as humans.

A synthetic voice's intonation in storytelling

A study by Montaño and Alias (2016) provided some insights into research regarding the intonation of a synthetic voice. They reported that studies use different approaches towards intonation. These approaches are about the prosodic features Velner and colleagues (2020) mentioned in their study: pitch, loudness, and duration. With every approach, the focus will be on a different prosodic feature instead of a combination of features.

In addition, Sarkar and colleagues (2014) provided a prosody rule-set for (Indian speaking) synthetic voices to have a storytelling speaking style. The rule-set transformed an everyday speaking style to a storytelling speaking style by changing different prosodic features. For example: changing the pitch from low to high and duration from fast to slow (Sarkar et al., 2014). Their rule-set was explicitly made for Indian languages. It is uncertain if the rule-set can be used for storytelling in other languages because intonation differs per language (Aarts & McMahon, 2006; Gálvez et al., 2020; Ronanki, 2019; Stan & Lőrincz, 2016).

Furthermore, Fernandez-Torne and Matamala (2015) researched a synthetic female and synthetic male voice in audio descriptions. The female synthetic voices, used for audio descriptions, scored higher on intonation and naturalness than the male counterparts (Fernández-Torné & Matamala, 2015). However, it is unclear to what detail Fernandez-Torne and Matamala (2015) examined intonation and on which prosodic feature was focused.

Additionally, Hodari and colleagues (2019) mentioned that a synthetic voice is perceived as flatter and duller than a human voice. Their study has not tested the effects of this perceived flatness during storytelling. Therefore, it is uncertain whether an artificial voice can provide the same experience while telling a story – even with the inclusion of the rule-set from Sarkar and colleagues (2014). At last, it is uncertain which prosodic feature Hodari and colleagues (2019) focused on within their study.

Narrative transportation

When transported towards a story world, readers believe that the story world is more accurate than the real world. The readers' attention is drawn towards the story rather than to themselves or their surroundings. Readers who pay less attention to the story are less transported than readers who pay more attention. People find the experience of transportation enjoyable and tend to be more engaged with the story (Green, 2014; Gregg, 2020; Green & Jenkins, 2014; Kuijpers et al., 2014). Furthermore, writing quality influences transportation. When the quality is high, people could have a higher level of transportation than when the quality is low (Green, 2014; Gregg, 2020).

Additionally, the type of story, and their degree of interaction, impact the level of transportation. An interactive story has more transportation because it requires a higher degree of thinking and a higher degree of engagement than a non-interactive story. For example, when people read an interactive story such as a gamebook, they could reach a higher level of transportation as it requires more interaction than a regular book (Green, 2008, 2014, 2021; Green & Jenkins, 2014; Kuijpers et al., 2014).

Transportation through a synthetic voice

Appel and colleagues (2021) investigated transportation to a story world. Their study focused on stories spoken by a synthetic voice and reported that a robot’s facial expressions could have an illusionary effect on intonation – leading to transportation. However, they did not examine whether there is more than an illusionary effect on intonation and transportation – making it an unanswered question whether there is more than an illusionary effect.

Furthermore, a study by Gregg (2020) investigated different narrative formats and their level of transportation. They found that researchers have underexamined how different format types affect transportation. One of the formats they researched was the difference between a

human and a synthetic voice. They reported that a human voice yields a higher level of transportation than a synthetic voice, yet both human and synthetic voices provide transportation towards a story (Gregg, 2020).

Moreover, Gregg (2020) stated that the quality of a synthetic voice is lower than the quality of a human voice in storytelling. People find a synthetic voice too robotic and hear the difference between a synthetic voice and a human one. However, synthetic voices are being improved, and newer synthetic voices use NTTS. Because of NTTS, people find it harder to distinguish a synthetic voice from a human one (Baiardi, 2021; Ronanki, 2019). It is possible that if an NTTS voice is used, people prefer a synthetic over a human one – contradictory to the findings of Gregg's (2020) study.

In addition, Gregg's (2020) study was focused on non-interactive stories. However, an interactive story provides a higher level of engagement and could have a higher level of transportation than a non-interactive story (Green, 2008, 2014, 2021; Green & Jenkins, 2014; Kuijpers et al., 2014). Following this line of thought, if a synthetic voice told an interactive story, the transportation level could be the same (or higher) than a non-interactive story told by a human or synthetic voice.

Summary

When a synthetic voice provides a natural level of intonation, people find it easier to engage with the story and find it more pleasant to listen to it (Aarts & McMahon, 2006; Gregg, 2020; Velner et al., 2020). Therefore, if a synthetic voice is human-like, the listener will get more engaged with the story than when a synthetic voice is robotic.

Hodari and colleagues (2019) reported that a synthetic voice is more flat and dull than a human voice. However, they did not describe the synthetic voice's intonation that is found flatter and duller. Also, their study did not describe which prosodic feature (pitch, duration, or

loudness) led to the intonation being found flatter and duller and in what detail the prosodic feature was examined.

Furthermore, according to Montaño and Alias (2016), studies regarding intonation used different approaches – namely, focusing on another prosodic feature. For example, when a study is focused on the pitch instead of duration, the outcome could differ from when the study is focused on duration instead of pitch. In this line of thought, focusing on another prosodic feature, in a study about intonation, may lead to a different outcome.

Moreover, in Hodari and colleagues' (2019) study, participants may find a synthetic voice human-like and not flat or dull if the focus was on another prosodic feature. Thus, illuminating the relevance of further research on intonation. Therefore, this thesis will examine the variation of intonation, focusing on the pitch as a prosodic feature.

In addition, engagement interacts with transportation. Participants who are more engaged with a story have a higher level of transportation than participants who are less engaged with a story (Green, 2008, 2014, 2021; Green & Jenkins, 2014; Kuijpers et al., 2014). In conclusion, a synthetic voice with human-like intonation could lead to a higher level of transportation than a reduced (combination of human-like and robotic intonation) or robotic intonation.

Not much research is conducted on the different prosodic features, making it complex to implement a storytelling speaking style. For this reason, pitch was chosen as a prosodic feature (Montaño & Alías, 2016; Sarkar et al., 2014; Velner et al., 2020). The regular version of a synthetic voice, produced directly by the NTTS system, will be used as a standard intonation. However, the standard variation of intonation still could be incorrect and have some flaws, as people have mixed meanings towards them – some find the voice dull/flat, and others find the voice human-like (Baiardi, 2021; Hodari et al., 2019).

Additionally, the pitch variation was altered to have a more robotic intonation. Two variations were manipulated, next to the standard intonation produced by the NTTS. The voice was reduced to simulate a non NTTS voice for the first manipulation. The voice was altered to resemble a voice without intonation for the second manipulation. When the intonation is removed completely, the voice could be found the least human-like and almost robotic as all the pitch variations were deleted. With a robotic voice, people would be the least engaged with the story and could lead to a lower level of transportation than a human-like voice (Green, 2008, 2014, 2021; Green & Jenkins, 2014; Kuijpers et al., 2014).

In conclusion, the three levels of intonation lead to three hypotheses, with the 'no intonation' condition being the least human-like, almost robotic, and the 'standard intonation' condition being the most human-like and least robotic.

H1: When a synthetic voice's intonation is standard in variation, the listener is more transported to the story world than when the intonation variation is reduced.

H2: When a synthetic voice's intonation is standard in variation, the listener is more transported to the story world than when all intonation variation is removed.

H3: When a synthetic voice's intonation is reduced in variation, the listener is more transported to the story world than when all intonation variation is removed.

Method

Design

A between-subjects experiment was conducted, with intonation variation as the independent variable and transportation as the dependent variable. The independent variable consists of three levels: 'standard intonation,' 'reduced intonation,' and 'no intonation.' Furthermore, transportation was measured using the transportation scale by Green and Brock (2000). The participants were randomly assigned to either one of the conditions.

There is a possibility of moderation by native language. Therefore, a potential confound was included as a control variable. At last, the dimensions of the transportation scale were examined in more detail to see if a particular dimension had peculiar scores.

Participants

The recruiting of participants went through network/snowball and volunteer sampling because participants were required to be 18 years or older and needed to understand English (Treadwell, 2017). The participants were recruited through social media, friends, and family. Furthermore, survey exchange websites such as SurveyCircle and SurveySwap and survey exchange groups on Facebook and LinkedIn were used to recruit participants.

Participants were not selected for their native language because they were mainly recruited in the Netherlands. Therefore, there was not an even distribution between the languages. The distribution per language can be seen in Appendix F.

A total of 98 participants were recruited. During the experiment, the data of six participants were removed from the experiment because they did not answer the questions seriously as all the questions had the same answer. Even the control questions (that were worded negatively) had the same answer as questions worded positively.

Furthermore, the data of one participant was removed as the code was incorrect. The participant said they ‘did not know’ the code, making it unclear if they had listened to the story. Finally, the data of five more participants were removed as they completed the experiment within 5 minutes. The minimum completion time of the experiment was 6 minutes, as the story was approximately 5 minutes long. For the remaining 86 participants, 28 participants were in the ‘standard intonation’ condition, 28 participants were in the ‘no intonation’ condition, and 29 were in the ‘reduced intonation’ condition.

Materials

Experiment

In the first version of the experiment, the story was provided through the website of UXPin and presented to the participants with a hyperlink. The participant was provided with the audio files and the questions through UXPin. Every page had one audio file and question. After answering the question, the participant went to the next page until the story ended. After completing the story, a completion code was shown, which the participants had to include in the survey.

However, participants found it hard to go back to the questions after the story finished. Only ten participants completed the story using UXPin, while it was shown to 22 participants. Therefore, the story was included in the Qualtrics environment, and the code was removed. Through Qualtrics, the participant was presented with the audio file and the question to determine the story's direction. When the participant had listened to the audio file and answered the question, the following audio file was presented until the story ended. Every audio file and question combination had a timer. Therefore, the participant only could go to the following audio file when the previous audio file was listened to completely.

Story

For the story, an existing, copyright-free gamebook was used written by Benjamin Smith-Donaldson called ‘The Morpheus Quest’ (Smith-Donaldson, n.d.). The gamebook has never been published as an actual book, only as an Android application and website.

In the original story, the story was presented through multiple texts. At the end of every text, the reader answers a question that decides what the character in the story must do next. One decision path, that leads to the end of the story, was called a story branch. The original story consists of 135 story branches. To let the story fit the experiment, some choices were removed.

The experiment consisted of six story branches. Every story branch had six, approximately 40 seconds long, audio files in which the participant chose the direction of the story. The story branches and text can be seen in Appendix A.

Audio and manipulation

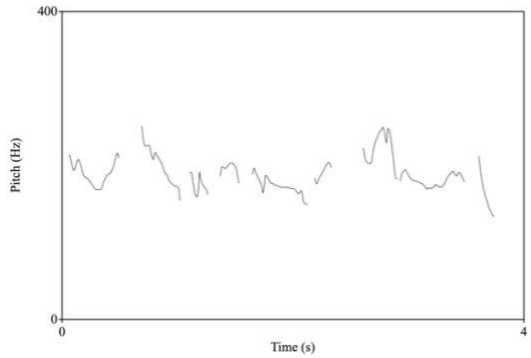
Microsoft Azure's Text-to-Speech was used to generate the spoken material. The voice used was an NTTS voice called: "English (United States) – Jenny (Newscast)." This voice was chosen as research indicated that a female synthetic voice had a more pleasant intonation, while telling a story, than a male synthetic voice (Fernández-Torné & Matamala, 2015). The resulting speech was used, without changes, and called the 'standard intonation.'

The program PRAAT was used to manipulate the voice for the other two conditions (Boersma & Weenink, 2021). Within PRAAT, the pitch points were changed to make a 'reduced intonation' and removed to make a 'no intonation' condition. To make the 'reduced intonation,' the pitch was stylized with 3.0 semitones to reduce the number of pitch points, and the mean Hertz per sentence was calculated. Furthermore, the distance of each pitch point to the mean was reduced by 50% to lower the variation in intonation.

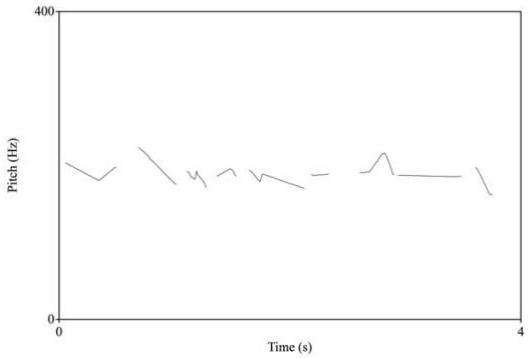
All original pitch points were removed for the 'no intonation' condition. Per sentence, the beginning Hertz and the ending Hertz point were calculated. A pitch point was placed at these Hertz points, drawing a line from high to low per sentence, resembling a sentence without intonation. A second rater checked all stimuli and agreed with all three conditions. The pitch tracks of the three conditions can be seen in figure 1.

Figure 1

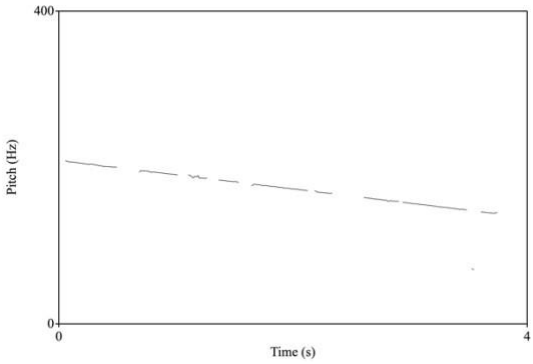
Pitch track for the different intonations per condition.



(1a) pitch track of the standard intonation.



(1b) pitch track of the reduced intonation.



(1c) pitch track of the no intonation condition.

Measurements

To measure transportation, the transportation scale by Green and Brock (2000) was used. This scale contains 15 questions about transportation to the story world. Due to experimenter error, one of the questions was not included in the experiment, namely, 'The events in the narrative are relevant to my everyday life.' Therefore, the questionnaire used contained 14 questions. Because the scale was widely used and researched, it offers good consistency and validity (Green, 2014).

The transportation scale contained three dimensions: cognitive engagement, mental imagery, and affective reactions. The mental imagery questions were specified for the story as all the participants could have a different story ending. The mental imagery questions were about items included in every story branch: ‘your character,’ ‘your colleagues,’ ‘Morpheus,’ and ‘your workplace.’ The altered transportation scale used in this study can be seen in Appendix B.

Procedure

The experiment and survey were created in Qualtrics and UXPin. Participants first had to sign a consent form, which can be seen in Appendix C. After giving consent, participants were presented with questions regarding their native language: ‘What is your native language?’ and ‘What language do you speak the most in your everyday life?’

A total of three versions of the story were made to provide each condition a story. The participant was presented with an audio file and a question, which decided the story’s direction. When the story finished, only the story through UXPin revealed a word that the participant needed to enter in the Qualtrics environment to determine whether the participant had completed the story, and all participants received the same word. The participants who experienced the story through Qualtrics were not provided with the word. These participants only could go to the following audio file if they had fully listened to the previous audio file.

Finally, the participants were presented with the transportation scale. After completing the scales’ questions, the participants were debriefed, and they indicated whether they wanted to read the complete thesis. Furthermore, they were asked whether they wanted to provide any feedback regarding the experiment. The debrief can be seen in Appendix D.

Analysis

A one-way ANOVA was conducted to analyze whether the variation of a synthetic voice’s intonation affects a participant’s transportation to a story world. The dependent variable,

called transportation, was calculated through the mean score of the items of the transportation scale. Every participant's mean score was calculated by adding all the answers and dividing them by the number of questions. Some questions in the questionnaire were worded negatively and are therefore scored in reverse. Those questions were re-coded, reversing negative to positive answers before calculating the mean score. The re-coded questions are marked in Appendix B.

A control check was performed on the language questions. The questions were about their native language and the language they speak the most. Both questions were asked because a native language can differ from the most spoken language. A mean transportation score was calculated for every language and was put in a histogram. Visual inspection of the histograms was used to see if the language's transportation differs from one another.

Furthermore, a control check was performed on the different dimensions of the transportation scale. The transportation scale consists of three dimensions: narrative engagement (questions 1, 3, 4), affective reactions (questions 5, 7, 11), and mental imagery (questions 12, 13, 14). Per condition, the mean of each dimension was calculated and made visible in a histogram. Visual inspection of the histogram was performed to see which dimension has the highest and lowest transportation scores.

Results

The mean transportation score consisted of 14 items on a 5-point Likert scale (e.g. 'While I was listening to the narrative, I could easily picture the events in it taking place), going from '1, strongly disagree' to '5, strongly agree'. Participants who did not fill in the questionnaire seriously were removed from the dataset. The scale's reliability was high $\alpha = 0.850$, meaning consistency between participants. Dropping items would not improve or decrease the reliability of the scale.

The ‘standard intonation’ condition had a mean transportation score of 2.92 (*SD* = 0.69). For ‘reduced intonation,’ the mean was 2.78 (*SD* = 0.55), and for ‘no intonation,’ the mean was 2.91 (*SD* = 0.64). There was significant skewness and kurtosis for ‘standard intonation’ (z-$_{skewness}$ = -1.10, z-$_{kurtosis}$ = 0.20) and for ‘no intonation’ condition (z-$_{skewness}$ = -0.35, z-$_{kurtosis}$ = 0.33). However, for the ‘reduced intonation’ condition (z-$_{skewness}$ = -2.75, z-$_{kurtosis}$ = 1.92), there was no significant skewness, which indicated that the normality assumption was not met. No bootstrap has been performed, as the one-way ANOVA is robust against this violation, but results must be interpreted carefully. The variance ratio had a score of 1.57, indicating that homogeneity was met.

The overall ANOVA was not significant, indicating that the data does not provide evidence of an effect between transportation and the different variations of intonation $F(2, 82) = 0.49$, $p = 0.616$. The effect size was calculated ($\omega^2 = -0.01$), representing a lower than a small-sized effect. Therefore, the hypotheses (H1, H2, and H3) cannot be supported.

For explorative purposes, a Post Hoc Tukey-HSD was conducted. The outcome of the Post Hoc Tukey-HSD test between ‘standard intonation’ and ‘reduced intonation’ was $M_{diff} = 0.15$, $p = 0.649$, *BCa* 95% CI [-0.25, 0.54]. The effect-size was $d = 0.02$, which represents a small-sized effect. Furthermore, the outcome of the data between ‘standard intonation’ and ‘no intonation’ was $M_{diff} = -0.01$, $p = 0.997$, *BCa* 95% CI [-0.39, 0.41], with an effect-size of $d = 0.001$ representing a smaller than a small-sized effect. At last, the outcome of the data between ‘reduced intonation’ and ‘no intonation’ was $M_{diff} = -0.13$, $p = 0.697$, *BCa* 95% CI [-0.53, 0.26]. The effect-size was $d = 0.02$, which represents a small-sized effect.

Control check

Native language

Participants reported 15 different native languages and three combinations of languages (Dutch/English, Dutch/German, and Arabic/English).

Overall, visual inspection of the histogram indicated that participants, with different native languages, had somewhat comparable transportation scores to one another. Two native languages provided a lower mean score than the other native languages, namely, 'Dutch/German' with a mean of 1.57 and 'Chinese' with a mean of 2.25. Furthermore, three native languages had a higher mean score than the other native languages, namely, 'Filipino' with a mean of 3.57, 'Bahasa Malaysia' with a mean of 3.36, and 'Italian' with a mean of 3.57. However, the differences between the native languages were minimal.

The confound of the native language was not examined further because the data did not provide a significant effect. The histogram for all the languages can be seen in Appendix E.

Transportation dimensions

Another control check was performed on the three different dimensions of the transportation scale (cognitive engagement, affective reactions, and mental imagery). Visual inspection of the histogram indicated that the participants scored the highest on cognitive engagement in all three conditions. With the 'no intonation' condition being the highest (M = 3.50, SD = 0.77), and the 'reduced intonation' the lowest of the three conditions (M = 3.30, SD = 0.80).

Furthermore, the affective reactions had the lowest scores in all three conditions. Participants in the 'no intonation' condition scored the highest on affective reactions (M = 2.69, SD = 0.77), and participants in the 'reduced intonation' condition scored the lowest (M = 2.44, SD = 0.69).

Finally, for the mental imagery, participants in the ‘standard intonation’ condition had the highest score (*M* = 2,99, *SD* = 0.94), and participants in the ‘reduced intonation’ condition had the lowest score of the three conditions (*M* = 2.79, *SD* = 0.84). The histograms for the transportation dimensions can be seen in Appendix G.

Feedback participants

Participants were asked if they had any feedback regarding the experiment. At first, one participant pointed out that a question was missing in the scale. Unfortunately, this participant was one of the last participating which resulted in leaving the question out of the experiment.

Second, one participant pointed out that they did not like the fixed options after each segment, decreasing their interest in the story. The participant who mentioned this was provided with the ‘no intonation’ condition, and their transportation score was 2.57. Another participant mentioned that the voice was too robotic, which distracted them. This participant was provided with the ‘reduced intonation’ condition, and their mean score was 3.21. Also, another participant mentioned the voice being too robotic. They suggested to “have a smoother ‘robot’ voice.” The condition presented to this participant was the ‘reduced intonation’ condition, and their transportation score was 2.93.

Additionally, two other participants negatively mentioned the ‘reduced intonation’ condition. One of them mentioned that they found the voice annoying, which caused them not to be transported towards the story, while the other mentioned the voice being too robotic and would find it more pleasant if it was a human voice. They both mentioned that they would be more transported if a human voice was used. The latter participant had a transportation score of 2.50, while the other had a score of 2.57.

Furthermore, it was mentioned that the synthetic voice sounded natural, not “robot-like,” by one participant. They were provided with the ‘standard intonation’ condition with a

transportation score of 3.43. At last, one participant said they found it funny and enjoyed the experiment. They were provided with the ‘no intonation’ condition, and their transportation score was 2.86.

Overall, it could be found that the participants who provided a negative comment had a lower transportation score than the participants who mentioned something positive about the experiment.

Discussion

This thesis investigated whether the variation of a synthetic voice’s intonation affects the transportation towards a story world. The data indicated no significant effect of a synthetic voice’s variation of intonation on transportation towards a story world. Therefore, it might be possible that there is no effect between one and the other. However, the data did not provide a conclusive answer.

Furthermore, the data did not indicate a confounding effect between the different native languages and their level of transportation. A possibility for not having a confounding effect is that the data indicated no significant effect in general. Another possibility is that the different languages were not equally distributed between the conditions, or participants were not recruited for their language.

Additionally, the three dimensions of the transportation scale (cognitive engagement, affective reactions, and mental imagery) were examined in more detail. The 'no intonation' condition provided the highest transportation score for two out three dimensions (cognitive engagement and affective reactions). There is a possibility that participants in the ‘no intonation’ condition were more engaged and had more affective reactions towards the story than participants in the other two conditions.

At last, participants had the opportunity to provide feedback about the experiment. Participants who provided negative feedback about the experiment, for example, finding the voice too robotic, had a lower transportation score than participants who provided positive feedback, for example, finding the voice enjoyable.

Previous research

Narrative transportation

Synthetic voices have improved since the introduction of NTTS (Baiardi, 2021; Gálvez et al., 2020; Ronanki, 2019). The data indicated that participants felt transported towards the story world, and a synthetic voice may suffice for storytelling. The grand mean of all conditions was 2.87 in a range from 1 to 5, which was above the median. However, the outcome is not in line with previous research by Hodari and colleagues (2019), in which they mentioned that people find synthetic voices dull and flat.

Furthermore, the outcome differs from Sarkar and colleagues' (2014) research, who report that artificial voices could not enforce a storytelling speaking style. A reason for this is that, in the current thesis, a specific prosodic feature was examined, which could lead to the findings. Another possibility is that synthetic voices were improved since Sarkar and colleagues (2014) conducted their study.

Intonation and narrative transportation

A more natural intonation could lead to a higher level of engagement than a less natural-sounding intonation. A higher level of engagement leads to a higher level of narrative transportation than a lower level of engagement (Green, 2008, 2014, 2021; Green & Jenkins, 2014; Kuijpers et al., 2014). However, the outcome of the data did not show that a variation in intonation led to a difference in narrative transportation.

Furthermore, the transportation scale consists of different dimensions, one of them being cognitive engagement (Green & Brock, 2000). However, the data showed that the ‘no intonation’ condition scored higher on cognitive engagement than ‘standard intonation,’ which is not in line with research regarding intonation by Aarts and McMahon (2006). Their study mentioned that a better variation of intonation leads to more engagement than a worse variation of intonation.

In addition, the outcome of the data did not indicate that a more human-like intonation could lead to a higher level of narrative transportation than a robotic intonation. A study by Appel and colleagues (2021) reported that robotic storytellers' facial expressions have an illusionary effect on intonation leading to people being transported towards a story world. The outcome of their study could be in line with the outcome of this thesis’ finding that intonation only has an illusionary effect on narrative transportation.

At last, participants who provided a negative comment had a lower transportation score than participants who provided a positive comment. The negative comments were about the voice being too robotic, which caused them to be less transported towards the story world. Combining an artificial voice and being less transported towards the story aligns with Sarkar and colleagues’ (2014) study. They mentioned that a robotic voice could not provide a storytelling speaking style. Furthermore, according to Gregg (2020), people find a synthetic voice robotic, aligning with the meaning of participants who found the voice being too robotic. However, it is not in line with the meaning of participants who found the voice human-like.

Naturalness in a synthetic voice

Another aspect of intonation is the naturalness of the voice. When a voice is unnatural, people would be less engaged and would be less transported towards the story than with a more natural voice (Green, 2008, 2014, 2021; Green & Jenkins, 2014; Kuijpers et al., 2014). The

outcome of the data showed that the three conditions had a similar transportation score, indicating that all three conditions could have some naturalness within the voice.

When a voice does not sound natural, people will find it eerie and find it harder to transport towards the story world. The eeriness of the voice can be described as the Uncanny Valley effect (Velner et al., 2020). However, a study by Velner and colleagues (2020) did not define to what extent a synthetic voice would sound unnatural and eerie.

On the one hand, there is a possibility that none of the intonations would be defined as eerie because the transportation score is somewhat the same for the different intonation types. On the other hand, it might be possible to find an eeriness when looking at individual scores as some participants provided negative feedback regarding the voice. Mainly, participants who provided negative feedback for the 'reduced intonation' condition (a combination of the standard and the robotic voice) could have found the voice eerie as it was not entirely realistic.

Limitations

Due to the lack of research, examining multiple prosodic features simultaneously and reproducing them into a storytelling speaking style is complex. In the current study, the manipulation of intonation was focused on one prosodic feature, namely, pitch. Therefore, it is possible that only manipulating pitch does not lead to a significant effect within the data.

Secondly, in the first version of the experiment, the story was provided through UXPin. However, more than 50% of the participants, who heard the story through UXPin, did not complete the survey after listening to the story – resulting in the story being implemented in Qualtircs. It remains unknown why participants did not complete the survey. However, the story within Qualtrics was not tested thoroughly as the story within UXPin because the survey was already published during that time.

Finally, participation exchange groups were used to gather participants. Through these exchange groups, people could get credits when completing surveys. It could be possible that participants were not focused enough while listening to the story or just wanted to finish to get the completion credits. The outcome of the test could be different when conducted in a lab where it is monitored whether participants were listening attentively. Therefore, the use of participation groups in combination with an online experiment could be seen as a limitation.

Future research

Intonation consists of multiple prosodic features. All the components of intonation can be placed into three categories: pitch, duration, and loudness. In the current thesis, the focus was on the prosodic feature 'pitch.' In a future study, it could be researched if manipulating another prosodic feature could affect the level of transportation towards a story world. Furthermore, the other prosodic features could be researched more in detail to examine how the features affect the intonation of a synthetic voice – as it is currently underexamined.

Secondly, the possibility of the Uncanny Valley effect within a synthetic voice needs further research. When the voice is unnatural, people could find the voice eerie, leading to less engagement with a story (Velner et al., 2020). A future study needs to examine to what extent a voice is perceived as natural and to what variation of intonation a voice is perceived as eerie or unnatural.

Research has shown that a female voice is stereotyped as emotional, empathetic, and warm. Furthermore, a female voice scores higher on intonation and naturalness than a male one (Fernández-Torné & Matamala, 2015; Yang et al., 2017). Therefore, a female voice was used to produce the story, but it was not examined whether a male counterpart could be used for storytelling. A future study could examine how gender differences play a role in a synthetic voice's intonation.

Furthermore, intonation differs between person, language, and culture (Aarts & McMahon, 2006; Gálvez et al., 2020; Ronanki, 2019; Stan & Lőrincz, 2016). However, participants were not selected for their language or culture. In a future study, participants could be recruited for their language or cultural difference to examine whether different languages lead to a difference in narrative transportation.

At last, only the transportation scale was used to measure narrative transportation during the experiment. One element of the transportation scale is cognitive engagement, measured by three questions (Green & Brock, 2000). However, engagement in the story is not measured thoroughly with only three questions. In the thesis, it remains unanswered if participants were entirely focused on the story. A future study could measure engagement towards the story using Busselle and Bilandzic's (2009) narrative engagement scale – with engagement being measured on four levels: narrative understanding, attentional focus, emotional engagement, and narrative presence. Furthermore, the narrative engagement scale is used to measure specific levels of engagement with a story (Busselle & Bilandzic, 2009; Green, 2021).

Implications

A synthetic voice can enable listeners to be transported to a story world. The data reported that all variations of intonation enable transportation towards the story world – as the grand mean was higher than the median. The outcome of the data indicated that a synthetic voice could be used to tell stories, saving time and money for producing audiobooks, podcasts, and other spoken stories. In addition, synthetic voices can be used for audio descriptions in online videos and other movies, adding screen reading to more products, and providing spoken interaction to digital products (Fernández-Torné & Matamala, 2015; Morris et al., 2018).

However, narrative transportation is not essential for screen reading and spoken interaction in digital products, but engagement is necessary. More engagement leads to a higher

level of transportation than a lower degree of engagement. Therefore, if a synthetic voice can lead to narrative transportation, the synthetic voice can also lead to engagement as the two are linked together (Green, 2008, 2014, 2021; Green & Jenkins, 2014; Kuijpers et al., 2014).

Conclusion

This thesis aimed to examine to what extent the variation of a synthetic voice's intonation affects the narrative transportation towards a story world. The results indicated that the variation of intonation did not affect the perceived transportation to the story world. Because the data did not indicate a significant effect, the possible confound of native language and the dimensions of the transportation scale were not investigated further. However, the grand mean of the different levels of intonation was higher than the median of the levels of intonation. The data indicated that participants, in all three conditions, felt transported towards the story world. The finding that participants felt transported in all conditions confirms that synthetic voices have been improved over the years, and people could find a synthetic voice engaging – because engagement is linked to narrative transportation. In the thesis, the focus was on the pitch as a prosodic feature. Focusing on another prosodic feature might result in a different outcome, possibly leading to a significant effect between the levels of intonation and narrative transportation. Future research is required, focusing on another prosodic feature, to examine the variation of intonation and the effect on narrative transportation. The thesis indicated that a synthetic voice could be used in products for which people need to be engaged or transported – such as audiobooks, podcasts, and audio descriptions.

References

Aarts, B., & McMahon, A. (2006). The Handbook of English Linguistics. In B. Aarts & A. McMahon (Eds.), *The Handbook of English Linguistics* (Vol. 89, Issue 6, pp. 433–457). Blackwell Publishing. https://doi.org/10.1002/9780470753002

Appel, M., Lugrin, B., Kühle, M., & Heindl, C. (2021). The emotional robotic storyteller: On the influence of affect congruency on narrative transportation, robot perception, and persuasion. *Computers in Human Behavior*, *120*(February), 106749. https://doi.org/10.1016/j.chb.2021.106749

Baiardi, L. (2021). *How to create Audiobooks with Text-to-Speech*. Woman in Voice. https://medium.com/women-in-voice/how-to-create-audiobooks-and-podcasts-with-text-to-speech-ff5a29dce8f2

Boersma, P., & Weenink, D. (2021). *Praat: doing Phonetics by Computer* (6.1.54). http://www.praat.org/

Busselle, R., & Bilandzic, H. (2009). Measuring Narrative Engagement. *Media Psychology*, *12*(4), 321–347. https://doi.org/10.1080/15213260903287259

Curry, C., & O'Shea, J. D. (2012). The Implementation of a Story Telling Chatbot. *Advances in Smart Systems Research*, *1*(1), 45.

Farmer, L. (2004). Using technology for storytelling: Tools for children. *New Review of Children's Literature and Librarianship*, *10*(2), 155–168. https://doi.org/10.1080/1361454042000312275

Fernández-Torné, A., & Matamala, A. (2015). Text-to-speech vs. Human voiced audio descriptions: A reception study in films dubbed into Catalan. *Journal of Specialised*

Translation, *24*, 61–88.

Gálvez, R. H., Gravano, A., Beňuš, Š., Levitan, R., Trnka, M., & Hirschberg, J. (2020). An empirical study of the effect of acoustic-prosodic entrainment on the perceived trustworthiness of conversational avatars. *Speech Communication*, *124*(May), 46–67. https://doi.org/10.1016/j.specom.2020.07.007

Green, M. C. (2008). Transportation Theory. *The International Encyclopedia of Communication*, 1–5. https://doi.org/10.1002/9781405186407.wbiect058

Green, M. C. (2014). Transportation into narrative worlds: The role of prior knowledge and perceived realism. *The Effects of Personal Involvement in Narrative Discourse: A Special Issue of Discourse Processes*, *6950*, 247–266. https://doi.org/10.1207/s15326950dp3802_5

Green, M. C. (2021). Transportation into Narrative Worlds. In *Entertainment-Education Behind the Scenes* (pp. 87–101). Springer International Publishing. https://doi.org/10.1007/978-3-030-63614-2_6

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*(5), 701–721. https://doi.org/10.1037/0022-3514.79.5.701

Green, M. C., & Jenkins, K. M. (2014). Interactive Narratives: Processes and Outcomes in User-Directed Stories. *Journal of Communication*, *64*(3), 479–500. https://doi.org/10.1111/jcom.12093

Gregg, P. B. (2020). Text to Speech: Transportation-Imagery Theory and Outcomes of Narrative Delivery Format. *Journal of Radio & Audio Media*, *00*(00), 1–18. https://doi.org/10.1080/19376529.2020.1801689

Hodari, Z., Watts, O., & King, S. (2019). *Using generative modelling to produce varied intonation for speech synthesis*. 239–244. https://doi.org/10.21437/ssw.2019-43

Kuijpers, M. M., Hakemulder, F., Tan, E. S., & Doicaru, M. M. (2014). Exploring absorbing reading experiences. *Scientific Study of Literature*, *4*(1), 89–122. https://doi.org/10.1075/ssol.4.1.05kui

Ligthart, M. E. U., Neerincx, M. A., & Hindriks, K. V. (2020). Design Patterns for an Interactive Storytelling Robot to Support Children's Engagement and Agency. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 409–418. https://doi.org/10.1145/3319502.3374826

Maro, M. Di, Cutugno, F., & Falcone, S. (2018). *Prosodic analysis in human-machine interaction*. https://doi.org/10.17469/O2104AISV000013

Montaño, R., & Alías, F. (2016). The role of prosody and voice quality in indirect storytelling speech: Annotation methodology and expressive categories. *Speech Communication*, *85*, 8–18. https://doi.org/10.1016/j.specom.2016.10.006

Morris, M. R., Johnson, J., Bennett, C. L., & Cutrell, E. (2018). Rich representations of visual content for Screen reader users. *Conference on Human Factors in Computing Systems - Proceedings*, *2018-April*. https://doi.org/10.1145/3173574.3173633

Ronanki, S. (2019). *Prosody generation for text-to-speech synthesis* [University of Edinburgh]. https://era.ed.ac.uk/handle/1842/36396

Sarkar, P., Haque, A., Dutta, A. K., Reddy, G. M., Harikrishna, M. D., Dhara, P., Verma, R., Narendra, P. N., Sunil, B. K. S., Yadav, J., & Rao, K. S. (2014). Designing prosody rule-set for converting neutral TTS speech to storytelling style speech for Indian languages:

Bengali, Hindi and Telugu. *2014 Seventh International Conference on Contemporary Computing (IC3)*, 473–477. https://doi.org/10.1109/IC3.2014.6897219

Smith-Donaldson, B. (n.d.). *The Morpheus Quest*. Retrieved October 12, 2021, from http://www.morpheus-quest.com/

Stan, A., & Lőrincz, B. (2016). Generating the Voice of the Interactive Virtual Assistant. *IntechOpen*, 13.

Treadwell, D. (2017). *Introducing Communication Research* (Vol. 3). SAGE Publications Inc.

Velner, E., Boersma, P. P. G., & De Graaf, M. M. A. (2020). Intonation in robot speech: Does it work the same as with people? *ACM/IEEE International Conference on Human-Robot Interaction*, *3*, 569–578. https://doi.org/10.1145/3319502.3374801

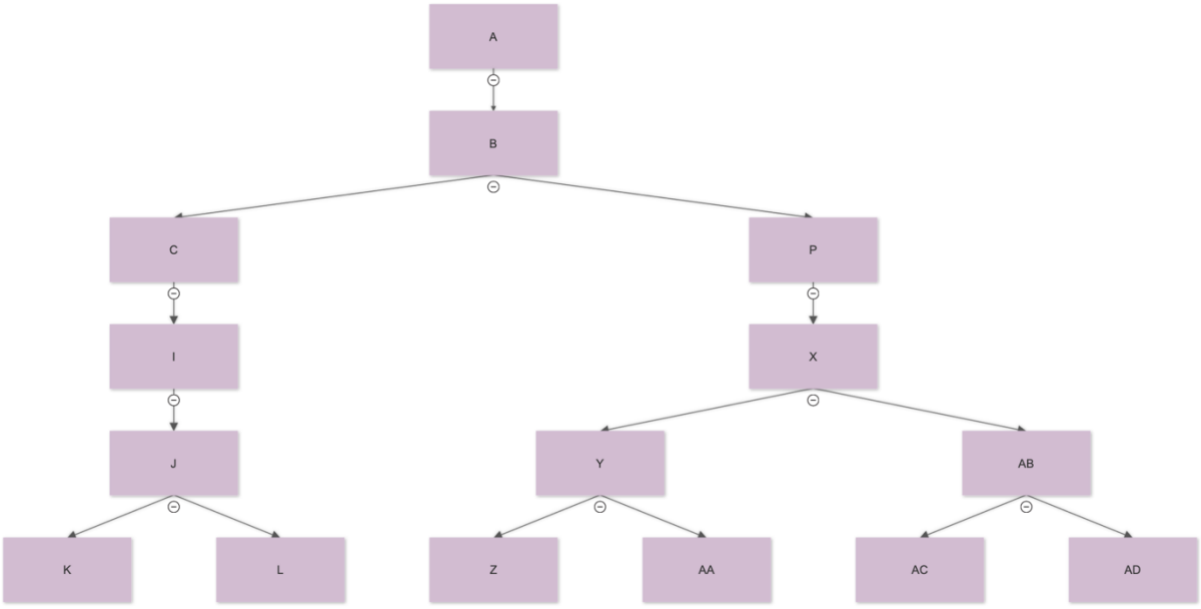
Whittingham, J., Huffman, S., Christensen, R., & McAllister, T. (2013). Use of audiobooks in a school library and positive effects of struggling readers' participation in a library-sponsored audiobook club. *School Library Research*, *16*(Benson 2011).

Yang, Y., Ma, X., & Fung, P. (2017). Perceived emotional intelligence in virtual agents. *Conference on Human Factors in Computing Systems - Proceedings*, *Part F1276*(March 2016), 2255–2262. https://doi.org/10.1145/3027063.3053163

Appendices

Appendix A

Branching structure of the story (gamebook)



Text of the narrative

Prologue text:

As you're drifting to sleep, a dream comes to you... A pale face materializes out of the darkness. A body forms with it, shifting in and out of the void. The man gazes at you with eyes as deep as eternity. "Some call me Morpheus. Some say Oneiros. I am dream incarnate and keeper of stories. Everyone has their own story, and you are no different. I have come to you to ask that you accept my story of you, so that I might at last see how it ends. "So, mortal, will you be my theatre for the night?"

A: Yes

You awake suddenly to the irritatingly abrupt sound of your alarm clock. You look over at the clock, noting the familiar time, and slap the top of it to silence the rude cacophony. It's 7:30, and

you have to be at work in an hour and a half. The morning routine is always the same: Drag yourself out of bed, empty the bladder, shower (sometimes) and breakfast which usually consists of some multi-colored flakes in a bowl of milk. You put on your shoes, your coat, get your keys, and you're off! But disaster strikes about a block from home. You've left your cell phone on the kitchen counter. You look at the clock on your dashboard. You don't have much time, but you have a little. What will you do?

B: Forget it. You're not expecting any calls**.**

You decide against going back for your phone. It's a nuisance, anyway, always bothering you with phone calls and messages or distracting you with videos and memes. it does feel a little stressful not to have something you keep so close so regularly, but it never killed you to be without one before. You flip on the radio; you haven't used it in weeks, since your music library now lives on your phone, but it still works. Unfortunately, you only have a few choices of songs from what's currently playing on the stations you can find.

C: Bee Gees – Night Fever

You pass through the various stations a couple times, revolving through them without finding much you really care for. Eventually, you find yourself pausing in the search as the chorus of 'Night Fever' fills the car. You spend a moment bopping your head with the rhythm when you glance to your side and see none other than Barry Gibb. The man's unexpected appearance prompts you to take in your surroundings, and you see Robin and Maurice Gibb in the back seat. Completely baffled, you fumble for words and eventually muster, "Who are you people?" Barry smiles at you and says, "We're The Bee Gees!" and the trio start singing along with the song. "Oh," you say. You turn the radio off, but The Bee Gees are still in the car.

I: Take them to work with you.

You step through the office door, with The Bee Gees following closely behind, determinedly strutting along and generally making known that cool is now in residence. “Well, you can tell by the way I use my walk, I’m a woman’s man, no time to talk,” says Barry meaningfully to an office temp. Typical, trust a celebrity to horn in on your action. You’ve been plucking up courage for WEEKS now, but you might as well forget it. Meanwhile, Maurice is playing a set of bongos that he has conjured up from somewhere. You see the boss coming towards you.

J: Try to shove the Bee Gees out of the nearest exit.

Of course your boss is coming to talk to you on the one morning you’re accompanied to work by a mysterious and funky musical trio. Of course he would. You actually hear him first, his voice which echoes around the corner and shoes which tap just a little too loudly on the tiles. Out of primal instinct, you react fast enough to deliver a fierce shove to the three Bee Gee bros. They’re tumbling out the door as your boss rounds the corner, but he doesn’t notice; he’s talking to that tiny earpiece, the one small enough to make it look like he’s talking to himself. Maybe he actually is talking to himself. At length, the boss notices you. Then he notices his watch. He notices you again, then glances back and forth between you and the watch a couple more times for good measure. “Hold on, Janice,” he says. He looks at you with haughty corporate disdain and says, “Steven, right?” “That’s not even close to my name,” you answer. “Right,” he says, “anyway, Steven, listen; we take punctuality very seriously here. Now, if my watch is right, and the person I paid seven thousand dollars to buy it from assured me it would be, then you are four minutes late.”

K: Apologize and get directly to work.

You mutter a sad apology to the man who obviously owns your life and scurry off to the giant hamster wheel which is your livelihood. Although you’re in a dream, the financial insecurity and meek servitude is so engrained into your subconscious that you spend the whole thing toiling

away as if you were awake. Eventually, you do wake up, and it's probably not so different. Continuing to live a life of mediocrity can feel like the end.

L: Call the Bee Gees back in to coordinate a powerful combo attack.

Your heart starts to pound. Your eyes focus on him. It's only four minutes! This kind of rampant disrespect can only be met with unbridled force! "Bee Gees!" you call, "come to me!" In a shattering of glass and brick, the entire front entry blasts open to reveal the trio of brothers framed before a glittering disco ball. "Prepare yourself, The Boss, for we are about to unleash our ultimate attack!" He should be shuddering with fear but instead The Boss...laughs? "Fool," he says, "who do you think sent the Bee Gees to you?" With that, he pulls away his business suit to reveal a suit of denim and leather beneath. "It was I, The Boss!" You gasp. You could have sworn the Bruce Springsteen option a couple steps back was a joke, but this has become deadly serious! "Now, Bee Gees, let us eliminate this man, and let his tardiness die with him!" Before you can react, the world is full of lights and music, pounding bass and moving rhythm guitar, disco melodies with rock and roll licks. Your assault by an incredible barrage of music which...frankly, it just doesn't sound very good together. In fact, it sounds awful together. It's so bad that it's killing you. That was probably the point, eh? Really, what a way to go.

P: Bruno Mars – 24K Magic

You settle on something a little poppy and catchy, some feel-good music, and let the dial rest on Bruno Mars's '24K Magic.' As it turns out, this radio station is only playing Bruno Mars at the moment, so next you listen to that other Bruno Mars song, then another. Eventually, you arrive at work, music still blaring. As you turn off your car and step out, the voice of a coworker surprises you from nearby. "Hey," comes the voice. You turn to look; it's April, the sales analyst. She says, "Were you really blasting Bruno Mars?"

X: Um… no?

You immediately start to formulate a lie to hide today's musical choice. Ultimately, you don't come up with much. "No," you tell her, "No I definitely was not listening to Bruno Mars. That's, like, so lame." "Oh," she says, "well, you don't have to be a jerk about it." Well, if you had a shot with her, there it goes.

Y: Forget it and get to work.

You give her a nervous smile, unsure of how else to reply, and walk in. Maybe she feels your shame and can accept that as some sort of apology, but you've got work to do either way. You quickly head to your desk, log into the computer, and don the ceremonial headdress of headphones and mic. As soon as the system starts, you hear the tone of a call being connected. You need to answer but...what do you say? What do you even do for a living?

Z: You're a vacuum cleaner sales person.

"Thank you for calling Vacuum Sales Monkeys Selling Sub-Par Equipment Inc., how can I help you?" you say. There is a long string of nearly incoherent shouting as your reply. "Yes, ma'am," you say, "I understand that I'm trash and using up precious oxygen that would be better spent on actual humans. Yes, ma'am, I do have trouble sleeping at night because of how I take advantage of innocent consumers." This continues for about an hour until your self-esteem is...well, normal. At the end of the call, she's the proud owner of four new vacuum cleaners and you have been convinced that the best thing you can do for humanity as a whole is kill yourself. Although you don't opt for suicide, withstanding the same abuse in dreams that you're faced with in waking life is too much for you, and you wake up dreading the coming work day. Maybe a new job would shake things up.

AA: You're a technical support trouble-shooter.

"Oh, God," Morpheus says as he materializes from the ether of your dream. "You do that forty hours a week?" You affirm that you do, in fact, work full time. "You poor pitiful soul,"

Morpheus commiserates. With a simple merciful wave, Morpheus dissolves your mind's ties to reality, letting you live forever in dreams where you'll never have to diagnose another router connectivity issue again. It's not really mercy; Morpheus doesn't know how to set up his wireless printer.

AB: Tell her you were just embarrassed and apologize.

"I'm sorry," you say, then explain the masculine stigma against modern pop music. Although she disapproves of the initial lie, April does understand that people's biases are defined in no small part by their opinions on your preferences. You nod and say, "...huh?" She continues on a discussion about intersectional feminism and how the patriarchy hurts everybody, regardless of gender.

AC: Tell her that you completely agree.

"I agree one-hundred percent," you tell her, "Men are just trash." "Well," she says, "thanks for the cosign, but that's really not what I'm saying at all." "Okay," you say with a smile, "sure! You're right about that, too!" "Are you even listening?" she asks. "Great point," you answer, "that's just what I was thinking." She sighs and walks away from you. "Sure thing," you agree to nobody, having not noticed your coworker's departure, "that's exactly it." After a few more empty affirmations, you start to wake up. Great job, that's definitely the best way to do things.

AD: Disagree vehemently.

"Hold on," you say, interrupting her. "I'm sorry, but I don't think you understand the root of the problem." She pauses to hear your mansplaining. "And what's that, exactly?" she asks. "Women are always trying to work and vote and do this and that, but what none of you understand is, your place is in the kitchen and with the children. You just don't have the same mental and physical capabilities of the superior sex, so it's just sad to see you trying to fit into society like you have a function." She blinks in disbelief. "You're joking, right?" "Nope," you say, "women are just

inferior to men." You flash her a satisfied smile and die from acute chauvinism poisoning. This is a real health issue threatening millions of Americans.

Appendix B

Transportation scale by Green & Brock (2000). Note, all the items with an (R) need to be re-coded.

1. While I was listening to the narrative, I could easily picture the events in it taking place.
2. While I was listening to the narrative, activity going on in the room around me was on my mind. (R)
3. I could picture myself in the scene of the events described in the narrative.
4. I was mentally involved in the narrative while listening to it.
5. After finishing the narrative, I found it easy to put it out of my mind. (R)
6. I wanted to learn how the narrative ended.
7. The narrative affected me emotionally.
8. I found myself thinking of ways the narrative could have turned out differently.
9. I found my mind wandering while listening to the narrative. (R)
10. The events in the narrative have changed my life.
11. While listening to the narrative I had a vivid image of my own character.
12. While listening to the narrative I had a vivid image of my colleagues.
13. While listening to the narrative I had a vivid image of Morpheus.
14. While listening to the narrative I had a vivid image of my workplace.

Appendix C

Consent form presented to the participant.

Dear participant,

Thank you for participating in the study. This study is part of a thesis research by Linda Hoeijmakers at Tilburg University. The supervisor for this study is Tom Lentz. In this study, you will be asked to listen to an interactive story told by a computer-generated voice. After the story, it will be measured how transported you were to the story. The interactive story will be presented through videos and questions. The original story is written by Benjamin Smith-Donaldson. It is a fictional story that contains some swearing words and provocative language.

In this story, you could make decisions about the story at certain points. **It is important to use headphones to listen to the story.**

After the experiment, the last questions will be asked. These questions will measure the transportation towards the interactive story. The study should take 5-10 minutes to complete.

All data collected for this experiment is anonymous and will be stored on a private iCloud account, which is only accessible by the researcher of this study and the supervisor. The researcher will be responsible for the data and will make sure confidentiality is met. Furthermore, no questions will be asked that could identify you as a person. At last, the results of this study will no longer be stored than necessary. Afterward, all data will be deleted.

Please respond to each question completely, honestly, and carefully. Remember that all information will be treated with complete confidentiality and is used for this study purposes only. If there are any questions regarding this study beforehand, please contact the researcher at: l.hoeijmakers@tilburguniversity.edu.

By clicking the button below, you acknowledge:

- Your participation in the study is voluntary.

- You are 18 years or older.
- You are aware that you may choose to terminate your participation at any time for any reason.

Appendix D

Debrief of the experiment

The real purpose of this study was to determine whether a variation in intonation could affect the transportation towards a story. The story was told by a computer-generated voice through Microsoft Azure.

A total of three conditions were used, one you have participated in. The difference between the conditions was the amount of variation in the intonation of the computer-generated voice. All participants received the same story, but your decisions in the story determined your story path. The use of interactivity was chosen as interactivity could lead to higher transportation.

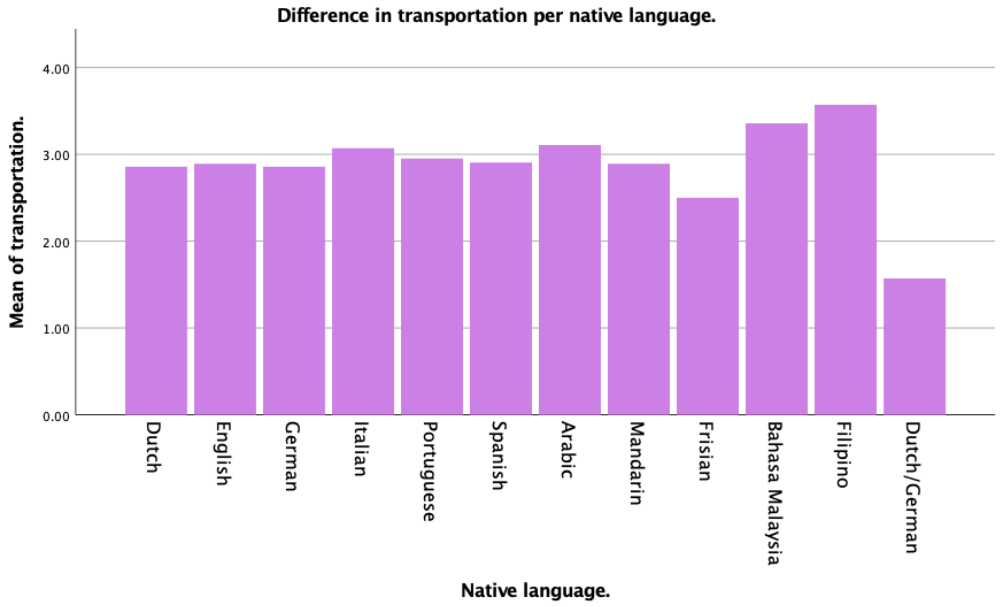
You have been asked to fill in your native language and the language you speak the most, before the story. After the story, questions regarding transportation were asked. These questions were to determine your level of transportation towards the story.

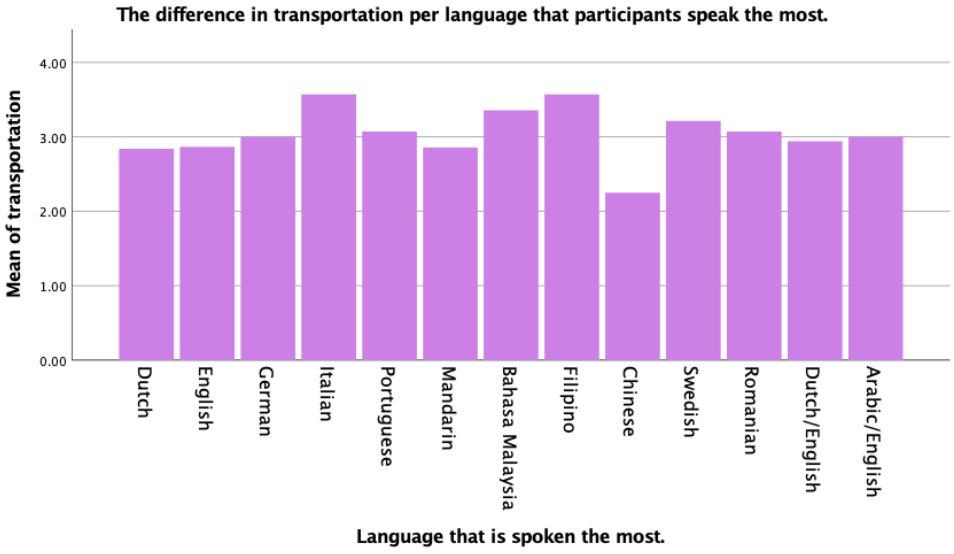
All data collected is anonymous, and will not be used for anything else than this thesis. If you have provided your e-mail address, it will only be used to provide the results and send the thesis. Afterward, the e-mail address will be removed. Furthermore, the e-mail address will not be linked to the answers provided in the experiment.

If you still have questions regarding the survey or this debriefing, please do not hesitate to contact: l.hoeijmakers@tilburguniversity.edu

Appendix E

A histogram for the mean transportation score per language category.





Appendix F

Language distribution per participant

Question: What is your native language?

Dutch	47 Participants
English	18 Participants
German	3 Participants
Italian	3 Participants
Portuguese	3 Participants
Spanish	3 Participants
Arabic	2 Participants
Mandarin	2 Participants
Frisian	1 Participants
Bahasa Malaysia	1 Participants
Filipino	1 Participants
Dutch/German	1 Participants

Note. Native language (N = 85):

Question: What language do you speak the most?

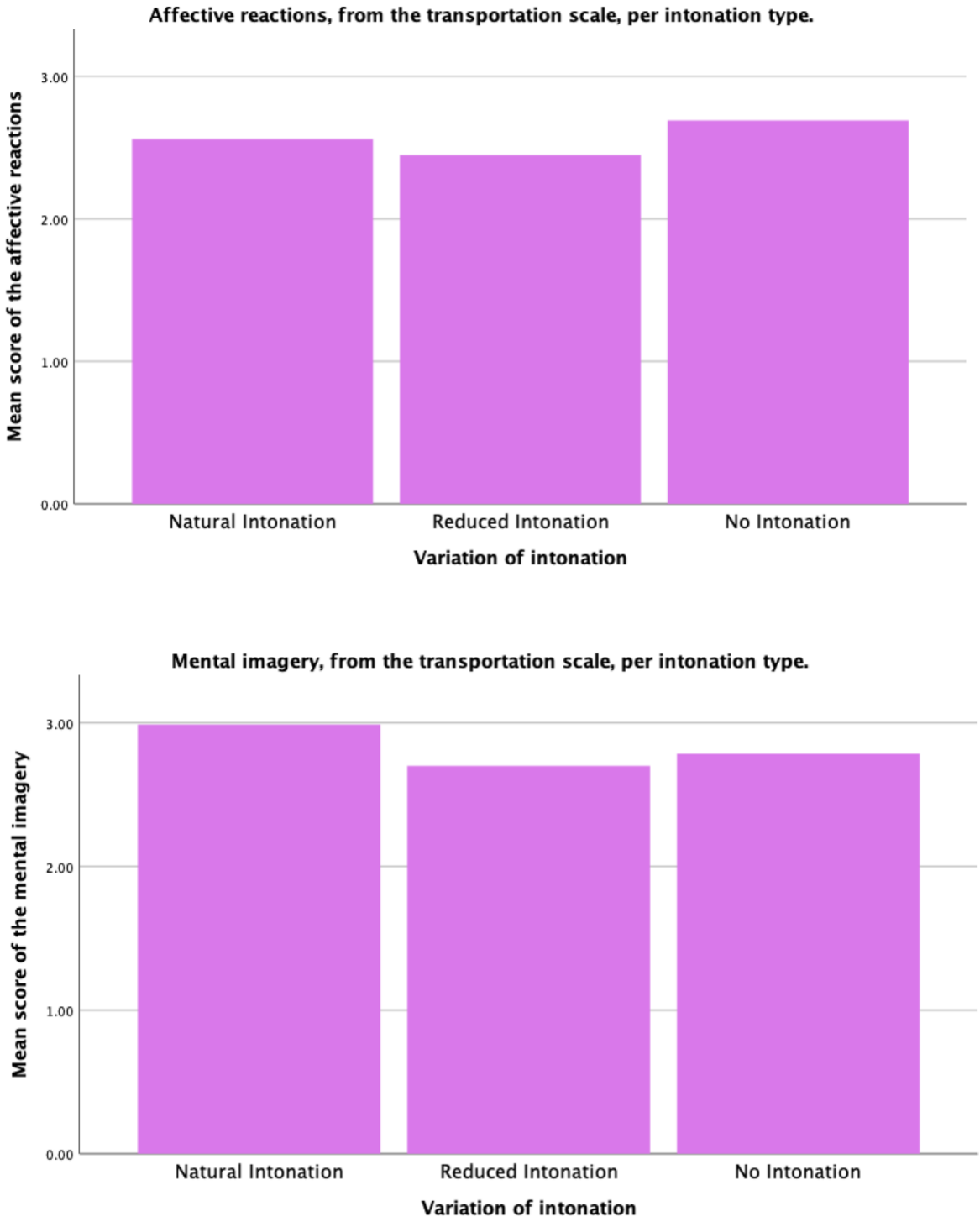
Dutch	39 Participants
English	29 Participants

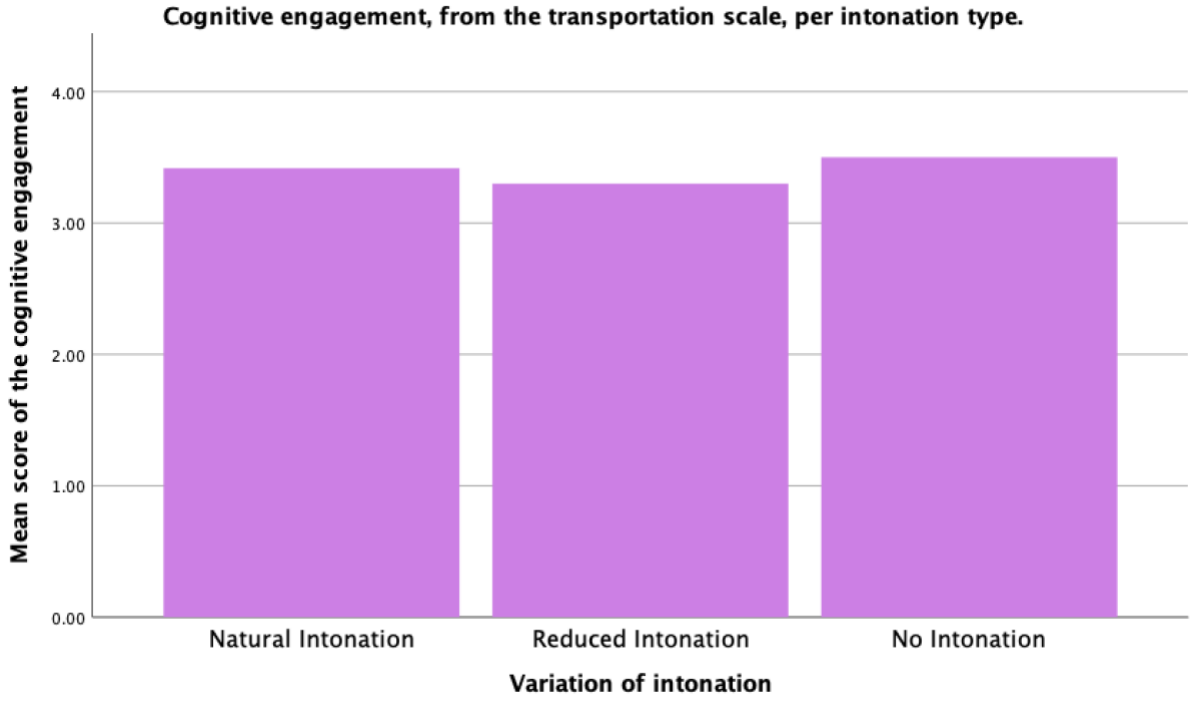
German	1 Participants
Italian	1 Participants
Portuguese	1 Participants
Mandarin	1 Participants
Bahasa Malaysia	1 Participants
Filipino	1 Participants
Chinese	2 Participants
Swedish	1 Participants
Romanian	1 Participants
Dutch/English	6 Participants
Arabic/English	1 Participants

Note. Language that is spoken the most (N = 85):

Appendix G

A histogram of the different transportation dimensions divided by intonation type.




Cognitive engagement, from the transportation scale, per intonation type.
Mean score of the cognitive engagement
4.00
3.00
2.00
1.00
0.00
Natural Intonation
Reduced Intonation
No Intonation
Variation of intonation