The effect of empty pause manipulations on the perception of competency-based trustworthiness of native and non-native speech; using different levels of implied expertise

Ma thesis: Leiden University

Tessa van der Meijden - 1821296

Linguistics - Applied linguistics

Supervisor: Nivja de Jong

Second reader: Willemijn Heeren

Date: 12-07-2021

Abstract

Trust is an important aspect of daily-life communication. We decide whether to trust someone based on just a face or a voice. Trust is part of credibility and is invariably connected to expertise. Previous research on credibility, expertise and trustworthiness has shown the impact of disfluencies on all three aspects. However, it has often focused only on native speech, and never used the duration of empty pauses as a fluency measure. The current study investigated the effect of empty pause manipulations in both native and non-native speech on perceived competency-based trustworthiness in two scenarios, differing in the level of expertise. Recordings of a native and a non-native speaker were manipulated phonetically, which resulted in two conditions: long empty pauses (LongPauses) and no empty pauses (NoPauses). The perceived competency-based trustworthiness of the speakers was measured using a 5-point likert-scale, based on McCroskey's (1966) authoritativeness scale. Results indicated a significant effect for scenario and an interaction effect between scenario and nativeness. Post-hoc t-tests showed that the native speaker was perceived as significantly less trustworthy than the non-native speaker in the non-expert scenario. Second post-hoc t-tests showed that the native speaker was perceived as significantly more competency-based trustworthy in the expert scenario, than in the non-expert scenario. This same effect was not found for the non-native speaker. Follow-up questions showed that listeners in the native group indicated that the implied expertise in the scenarios did affect their judgements of the speaker, while the listeners in the non-native group indicated it did not.

Keywords:

Competency-based trustworthiness, communication, perceived competency-based trustworthiness, native speech, non-native speech, speech fluency, empty pause durations, authoritativeness scale, implied expertise, expert scenario, non-expert scenario

Acknowledgements

I would like to thank my supervisor Nivja de Jong for her advice and feedback during the process of my thesis. I could not have completed my thesis, as it is now, without her help and support.

I would also like to thank my native and non-native speaker for voluntarily participating in my research.

Content

1. Introduction

When we meet new people, whether it is for work, or in our social groups, we make a decision based on their face, attitude, body language and speech in just a couple of minutes whether they are trustworthy or not. Through the use of social media this has become harder, since someone can pretend to be someone else or to have a different personality than they have in person. Someone can even pretend to be an expert on a topic, and try to convince other people to adopt their opinion, while he or she is actually no expert on that specific topic. Especially now, with all the news around COVID-19, there are so many people that claim to be an expert on the topic. Imagine listening to a podcast in which certain aspects of COVID-19 are discussed between speakers who all imply to be experts on the topic. How do you decide whether to trust the speakers and their opinions on the topic? What are the features that affect the way you perceive the speaker's trustworthiness, on which you decide whether to accept their opinion? Is it only based on what they say or also on how smoothly they bring the message? And does it matter whether the speaker is implied to be an expert on the topic, or not? Does the fact that the speaker is introduced as an expert affect the expectations you have of that speaker?

The speakers, who talk about these expert topics online, in this case COVID-19, can be from all over the world and have totally different language backgrounds. Both native and non-native speakers are involved in these discussions. To be able to transfer their opinion, some speakers may have to speak a second language. Speaking a second language goes paired with less fluent delivery of speech (e.g., Cucchiarini, Strik, & Boves, 2000). Does the fluency of the speech delivery make the speaker more or less trustworthy? Is a speaker who uses their native language on such a topic therefore more likely to be trustworthy?

If speech fluency, nativeness and implied expertise do affect the ability of native and non-native speakers to be trustworthy, this could have an impact on the communicative act of trustworthiness on individual levels, but also on the transfer of expert information on the broader national and international level. This research explores the effect of the fluency of speech on trustworthiness, in terms of the effect of the duration of empty pauses on the perceived competency-based trustworthiness of speakers in two scenarios, in which different levels of expertise are implied. The study compares this effect in L1 and L2 speech in order to determine whether speech fluency affects the ability to be perceived as trustworthy for both native and non-native speakers. This is relevant since the majority of people are now confronted daily with expert opinions on different topics, on for example social media, in

both their L1 and L2, and have to decide whether they trust the speaker's expert opinion on the relevant topic.

1.1 Thesis outline

This thesis is split up into 8 sections, including this introduction. Section 2 includes the literature review, which summarises all the relevant theories and past research in the field. Section 3 contains the methodology of the experiment, which includes the design of the study and the experiment setup. Section 4 contains the results, which include the statistical tests of a quantitative research and a summary of some follow-up questions. Section 5 contains the discussion, which includes the interpretation of the results in relation to the literature and the limitations of this study. Section 6 is the conclusion, which includes a summary of the key findings and suggestions for further research. Section 7 contains the full list of references. Lastly, section 8 contains the appendices, which include additional materials.

2. Literature review

The literature review starts with an explanation of the concept of speech fluency, the relationship between utterance and perceived fluency, and the native and non-native influences on both. Furthermore, the aspects of trustworthiness, and how we perceive this in speech is discussed. Finally, the research question is introduced.

2.1 Speech fluency

The level of fluency can vary, in the broad sense, which includes the general proficiency in the language, and in the narrow sense, which include the speed, pausing, and smoothness of delivery (Lennon 1990). People often refer to the broad sense of fluency in daily life, e.g., someone claims to be 'fluent' in English. The narrow sense of fluency can be defined as a process in which little attention and effort is needed to produce fluent speech (Bergmann, Sprenger & Schmid, 2015), but also as 'an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently' (Lennon, 1990, p. 391). The narrow sense of fluency is what I am concerned with in the current study.

Narrow fluency can further be divided into three different aspects: cognitive fluency, perceived fluency and utterance fluency. The first aspect, cognitive fluency, includes the underlying processes for the production of utterances, and is difficult to test. Perceived

fluency refers to the subjective impression of spoken fluency, as experienced by the listener, and can be tested by listener-ratings. Perceived fluency can be tested on multiple levels from which two are message perception (perception of what is said) and speaker perception (perception of the source). Utterance fluency refers to the objective and acoustic measures of fluency and can be measured by among other things the number and duration of silent and filled pauses and the number of repetitions and repairs (Segalowitz, 2010). Utterance fluency is what I manipulate in this study, and can further be divided into three types: speed fluency, repair fluency and breakdown fluency. Speed fluency is the rate and density of speech delivery. Repair fluency relates to the number of corrections and repetitions in speech. And breakdown fluency is the extent to which a continuous speech signal is interrupted (De Jong, 2018; Bosker et al., 2012). In the current thesis the duration of silent pauses will be manipulated, which belongs to breakdown fluency. Silent pauses can be defined as empty pauses and consist of no speech, but can include a breath or a cough. The effect of these manipulations on the perceived fluency is what I am concerned with in this study. The perceived fluency will be indirectly processed in the trustworthiness ratings. This will be tested on the level of speaker perception, and not on message perception.

2.1.1 The relationship between utterance fluency and perceived fluency

Riggenbach (1991) was one of the first to investigate the relationship between objective measures (utterance fluency) and subjective ratings of fluency (perceived fluency). Objective measures of pauses (silent and filled) and speech rate have been found to be significant predictors of perceived fluency (Cucchiarini, Strik & Boves, 2002; Derwing, Rossiter, Munro & Thomson, 2004; Kormos & Denes, 2004; Rossiter, 2009). Before Derwing et al. (2004) used untrained native listeners to rate perceived fluency; most researchers had used trained assessors to rate L2 speakers (Riggenbach, 1991; Lennon, 1990). Since then, more studies showed that untrained listeners are able to give accurate ratings on fluency too (Cucchiarini, Strik & Boves, 2002; Derwing et al., 2004; Rossiter, 2009). Rossiter (2009) used both trained and untrained listeners in her experiment to compare their ability to rate perceived fluency in L2 speech. She found their fluency ratings to be highly intercorrelated; the same features of fluency affected novice native speakers, as well as expert native speakers, and highly proficient non-native speakers. Untrained listeners have even been found to be able to focus on specific aspects of fluency, depending on the amount of instructions they receive (Bosker et al., 2013; Kormos & Denes, 2004).

However, de Jong (2018) points out that some of the above-mentioned studies have

not taken into account the issue of multicollinearity (e.g., Bosker et al., 2013; Derwing et al., 2004; Kormos & Denes, 2004), which is why they have to be interpreted with caution. Bosker et al. (2014b) used an experimental setup with which they bypassed this issue. They manipulated different aspects of fluency in native and non-native speech. However, they manipulated only one measure of fluency at once, while they kept the other potential confounding factors constant. By doing this they could confidently conclude from their results that an effect was caused by the specific measure they manipulated (De Jong, 2018). Using this experimental-design they could even study different properties of the same acoustic phenomenon, such as the duration and number of empty pauses. In one of their experiments they manipulated the duration of empty pauses with durations of more than 250ms (Bosker et al., 2014b). 250ms is shown to be a good cut-off point for L2 research (De Jong & Bosker, 2013). Bosker et al. (2014b) created three conditions in their experiment, from which one included no empty pauses (NoPauses), one included short empty pauses (ShortPauses) and the third included long empty pauses (LongPauses). In the NoPauses condition all empty pauses of more than 250ms were reduced to durations of 150ms or less. The durations of the empty pauses in the ShortPauses condition were manipulated to durations between 250ms and 500ms. Lastly, the durations of the empty pauses in the LongPauses condition were manipulated to durations between 750ms and 1000ms. They chose these durations for the different conditions, because research has shown that silent pauses of 250-1000ms are very common in native (Campione & Véronis, 2002) as well as non-native speech (De Jong & Bosker, 2013). Since a number of zero silent pauses within spontaneous speech sounds unnatural, they decided to leave the duration of one silent pause with a breath in the middle of the recordings as it was. Their results showed that the NoPauses condition was rated as more fluent than the ShortPauses and LongPauses conditions, and that the ShortPauses condition was rated as more fluent than the LongPauses condition. Their results also indicated native speakers to be rated as more fluent in general, than non-native speakers. However, the manipulation of silent pauses did not affect the perceived fluency of the native speaker any differently from that of the non-native speaker.

In the experiment of Bosker et al. (2014b) the non-native speaker was rated as less fluent than the native speaker in general, for reasons other than the duration of the empty pauses. It could have been affected by other fluency measures, or by the foreign accent of the non-native speaker. Rossiter (2009) found fluency ratings to be negatively affected by the pronunciation of the speaker. However, increased fluency did not reduce the perception of accentedness in the study of Derwing et al. (2004), probably because those judgments were

based on other linguistic phenomena. In response to Rossiter (2009), Pinget et al. (2014) studied the relationship between fluency ratings and foreign accent, and found only a weak correlation. Other measures of accent could have potentially predicted some variance of accent ratings (e.g., segmental and suprasegmental measures). Based on these studies we can conclude that perceived fluency and perceived foreign accent can be judged as separate constructs.

This section shows that there is a relationship between utterance fluency and perceived fluency; certain disfluencies have been shown to affect the perceived fluency of a speaker (e.g., Cucchiarini, Strik & Boves, 2002; Derwing et al., 2004). It also shows that non-expert listeners are able to give accurate ratings on L2 fluency, and that they can make a distinction between the concepts of foreign accent and fluency of speech. The current study is concerned with the relationship between speech fluency and the perception of trustworthiness in native and non-native (accented) speech. It is therefore preferable to use an experimental setup similar to that of Bosker et al. (2014b). This experimental design bypasses the problem of multicollinearity and enables me to compare the fluency of native and non-native speech by manipulating only one variable, while keeping all other variables constant.

2.1.2 Utterance and perceived fluency of native and non-native speech in relation to other linguistic aspects

Native and non-native speakers differ in the production (utterance fluency) of their speech. L2 speakers, whether the L2 is learned after fully learning the L1 or whether the speaker is bilingual, have both been found to be much more disfluent than native speakers (Bergmann et al., 2015). One of these disfluencies relates to the use of pauses; L2 speakers tend to pause more within their speech than L1 speakers do (Matzinger et al., 2020). These pauses (filled and empty) in L2 speech tend to occur more often within constituents, which are words or a group of words that function as one unit, than pauses in L1 speech do (De Jong 2018). L2 speakers also tend to pause more often within clauses (Kahng, 2014; Tavakoli, 2010) and within Analysis of Speech units[1], than L1 speakers do (Skehan & Foster, 2012; De Jong, 2016). Fulcher (1996) distinguished eight categories as possible explanations for L2 learners to pause; some only occur between the above-mentioned units (e.g., end-of-turn pauses) and some can occur within the above-mentioned units (e.g., content planning hesitation, expressing lexical uncertainty). The pauses close to syntactic boundaries, between

[1] An analysis of speech unit (AS) is “a single speaker’s utterance consisting of an independent clause or subclausal unit, together with any subordinate clause(s) associated with it" (Foster et al., 2000, p. 365).

constituents or clauses, are often related to conceptualizing and are therefore weighed as less disfluent than pauses that occur within clauses and constituents. However, whether listeners are able to distinguish between these two types of pauses is not yet investigated (De Jong, 2018).

Although disfluencies are often prescribed to and studied in non-native speech, native speakers also produce disfluencies in their natural speech (Lennon, 1990; Riggenbach, 1991). Both L1 and L2 speakers differ in the different sorts and rates of specific fluency aspects (McDougall & Duckworth, 2017). Everyone has their own idiolect, which is determined by the personality of the speaker and their speaking style. This means that features, such as filled and empty pauses, could also be related to a person's speaking style, and not only their fluency level (De Jong et al., 2015).

Native and non-native speakers not only differ in the production of speech, but also the perception (perceived fluency) of both is different. As mentioned in 2.1.1 the effect of the manipulations in Bosker et al. (2014b) on the perceived fluency, did not differ between the native and non-native speakers. This means that the manipulation of silent pauses did not affect native speech any differently from non-native speech. Branum (2019) on the other hand investigated the effect of speech rate manipulations of both native and non-native speech on the perception of persuasiveness, and found that the fluency manipulations in his study did affect the perception of both speakers differently; the native speaker was affected by the manipulations, while the same effect for the non-native speaker was not found.

Some other aspects of speech are shown to only affect the fluency perceptions of a non-native speaker. First of all, the L1 of the speaker can have a large impact on the listener's judgment of the comprehensibility of the speaker, which consists of among other things fluency (Growther et al., 2015). Further, L2 speakers who utilize a broad pitch range and pause in syntactically appropriate places within their speech are found to be perceived as more fluent than L2 speakers who have a limited pitch range and pause within constituents and clauses (Wennerstrom, 2000; in Derwing et al., 2004). An important factor can be the type of communication the L2 speaker is in. L2 speakers are perceived as more fluent in a conversation with a native speaker, than in a conversation with a non-native speaker or in a monologue (Ejzenberg, 2000; as cited in Derwing et al., 2004). It is also shown that the overall perception of non-native speech places greater demands on the neural systems underlying speech perception, than the perception of native speech does (Yi et al., 2014). However we are found to be able to adapt our predictive strategies when we know that we are communicating with a non-native speaker (Bosker et al., 2014a). Our brains can even adapt

to the effect it will normally show when a mismatch occurs. P600 (syntactic mismatch) and N400 (semantic mismatch) effects in an EEG[2] signal are found when a native speaker produces disfluencies in the form of a grammatical or semantic mistake. These same effects are however not found when a non-native speaker with a foreign accent produces similar disfluencies (Hanulikova et al., 2012). This shows that we create different expectations based on the nativeness of the speaker.

It can be concluded from the research in this section that there are both similarities and differences in native and non-native speech fluency, in terms of production on the speaker's part, and processing and perception on the listener's part. Non-native speakers produce more pauses, which occur more often in grammatically inappropriate places (De Jong, 2018). These disfluencies in non-native as well as native speech can also be described to the personality and speaking style of a speaker. There are disfluencies, such as speech rate and duration of pauses, which are shown to affect the perception of native and non-native speech in the same way (Bosker et al. 2014b), and in a different way (Branum, 2019). The perception of non-native speech places greater demands on the neural systems, although we are able to adapt our predictive strategies as soon as we know that we are listening to a non-native speaker (Bosker et al., 2014a; Hanulikova et al., 2012). In this current study the effect of such a fluency measure, on the perception of competency-based trustworthiness in both native and non-native speech will be tested.

2.2 Credibility, competence and trustworthiness

Researchers on the topic of trustworthiness have never accepted one specific definition of trustworthiness (e.g., Jones, 2012; Hardin, 2002). It originates from Aristotle's rhetorical proof, which consisted of: ethos, logos and pathos. Ethos refers to the credibility of a speaker, which can be divided in: competence, trustworthiness, and dynamism. Trustworthiness is thus a part of credibility together with competence, and is described as the degree to which the audience perceives a speaker to present credible and accurate information in a non-manipulative way. It is both based on the content of the speech as on the personality of the speaker (Author Removed At Request Of Original Publisher, 2016). Competence is defined as the perception of the speaker's expertise on the topic that they discuss. A competent speaker should know the content of their speech and be able to deliver that content

[2] Electroencephalography is a record of the electric signal, generated by the active brain cells. P600 and N400 effects are event-related potentials (ERP) that occur when a mismatch is received and can be measured with EEG (Hagoort, 2003).

effectively (Author Removed At Request Of Original Publisher, 2016). Competence (expertise) and trustworthiness are the most important factors of credibility (McGinnies & Ward, 1980), and are directly linked to one another.

According to Hardin (2002) interest is the most important aspect of trustworthiness; " … it is often only the interest we have in maintaining particular relationships that makes it in our interest to be trustworthy" (Hardin, 2002, p. 42). To maintain the relationship between the truster and the trustee it is important to meet the expectations of the other party. It is easier to trust the other party if you know that the interests of the other party will cause them to live up to your expectations. The overlapping interests can only be a subset of the other party's total interests, but that can be enough to maintain a trustworthy relationship (Hardin, 2002).

Jones (2012) compared earlier definitions of trustworthiness as defined by other researchers (e.g., Hardin, 2002; Baier, 1986; Hume, 1978). She concluded that both an excess and a deficiency can undermine the act of trustworthiness. Someone can be overly prone, such that they think the other is counting on them when they are not (excess of trustworthiness). Someone can also be insufficiently prone, such that they think the other is not counting on them, while they actually are (deficiency of trustworthiness). Besides being insufficiently prone to thinking others are not counting on them, someone can also mistake what others are counting on them for. It is often not a specific action, but some rather vaguely specified broader object, which typically includes some kind of social ability that extends beyond the capacity to respond to the other. It takes attunement to others, to grasp what they expect from you. This shows the importance of social knowledge in the concept of trustworthiness and why it can be hard for someone with a totally different cultural background to be trustworthy (Jones, 2012).

These studies show that trustworthiness is part of credibility together with competence, which means that they are closely related (Author Removed At Request Of Original Publisher, 2016). A competency-based trustworthy speaker is perceived as an expert on the topic, who presents credible and accurate information in a non-manipulative and effective way. Trustworthiness itself is a construct between a truster and a trustee, which is based on the interests of both parties (Hardin, 2002). The cultural background of a party can affect the trustworthy relationship, by means of differences in social knowledge (Jones, 2012). In the current study a native and a non-native speaker will be rated on their perceived competency-based trustworthiness.

2.2.1 Credibility, competence and trustworthiness in speech

We automatically form first impressions of the personality of a speaker based on only their voice. These impressions include traits such as credibility, competence, and trustworthiness (Boehme, 2014; Belin et al., 2017; Oleskiewicz et al., 2017). Most research on trustworthiness in speech has therefore often been conducted on acoustic or facial and acoustic measures (e.g., Boehme, 2014; Belin et al., 2017). Using other measures, such as fluency measures, it has often been part of research on credibility or expertise in speech, in which the difference is made between powerful and powerless speech styles (McGinnies & Ward, 1980; Erickson et al., 1978; Hosman & Stiltanen, 2011; Smith, Stiltanen & Hosman, 1988).

A powerless speech style is characterized by the use of linguistic features, such as intensifiers, hedges, hesitation forms and questioning, and slower speech rate. The powerful style is characterized by the less frequent use of those linguistic features (Smith et al., 1988). Erickson et al. (1977) found that a powerful speech style resulted in greater attraction to a witness in their experiment. The powerful speech style also resulted in greater perceived credibility of the speaker. The effect was even stronger when the participant and the speaker were of the same gender, then when they were of the opposite sex. Hosman and Stiltanen (2011) found tag questions and argument quality to affect the perception of the quality of the speaker, the message and the cognitive responses to it. The interaction between the three also directly affected the perception of the speaker's power and credibility (Hosman & Stiltanen, 2011). Although hedges and tag questions have never been considered as features of disfluency, hesitations do overlap with breakdown fluency (De Jong, 2018).

McGinnies and Ward (1980) found that the trustworthiness of a source is more important than the expertise of that source for them to be persuasive. They indicated expertise to undoubtedly affect the credibility of the speaker, but that the trustworthiness of the speaker may be more important in certain circumstances. Smith et al. (1988), found that when speakers with high expertise speak disfluent, they are perceived as less authoritative and to have lower expertise (competence), than when they speak fluently. This suggests that speakers create expectations on how speakers, varying in expertise, will speak. Elsbach and Elofson (2000) investigated the perception of competency-based trustworthiness in native speakers. They tested whether the difficulty degree of the language affected the perception. They used items based on McCroskey's (1966) authoritativeness scale in a 5-point likert-type scale. They found that easy-to-understand language causes greater perceptions of competency-based trustworthiness in a speaker, than difficult-to-understand language does.

This means that the packaging of the message is just as important, if not more important than the message itself, for a speaker to be perceived as competency-based trustworthy (Elsbach & Elofson, 2000). Schulze and Pishwa (2015) connected the concepts of trust and expertise. They found that trust is invariably a consequence of expertise: greater competence of a speaker leads to more positive thoughts and attitudes towards that speaker, which in turn lead to increasing trust.

As mentioned in 2.2 the cultural background of one of the parties can affect the feeling of trustworthiness (Jones, 2012). People tend to associate positive beliefs and feelings with their own group (in-group), and tend to have more negative feelings against people that are members of the out-group (Williams, 2001). This can have a direct effect on the perception of trustworthiness of the out-group members. According to the social identity theory[3], levels of trustworthiness should be higher between the same language-based in-group members. This results in lower levels of trustworthiness between language-based in- and out-groups (Barner-Rasmussen & Björkman, 2007).

This section shows how we perceive credibility, expertise, and trustworthiness in speech. Disfluencies, such as hesitations, tag questions, and number of pauses, in native speech are shown to be associated with negative traits, such as lower credibility, lower expertise, and lower competency-based trustworthiness (Erickson et al., 1977; Hosman & Stiltanen, 2011; Smith et al., 1988; Elsbach & Elofson, 2000). Perceived expertise of a speaker is affected by disfluent speech, which shows that speakers create expectations on how speakers, varying in expertise, will speak (Smith et al., 1988). The perceived expertise directly affects the perception of trustworthiness (Schulze & Pishwa, 2015). Not only disfluencies affect the perception of trustworthiness; non-native speakers are often labelled as language-based out-group members, which cause them to be perceived as less trustworthy (Barner-Rasmussen & Björkman, 2007).

2.3 Research question

Bosker et al. (2014b), investigated the perceived fluency of native and non-native speakers, and showed that the manipulations of empty pauses did affect both the native and the non-native speakers in the same way. The studies on other measures, such as the perception of expertise and trustworthiness (e.g., Erickson et al., 1978; Hosman & Stiltanen, 2011; Smith et al., 1988), found effects for the difference between powerless and powerful

[3] Social identity theory is a social psychological analysis of the role of self-conception in intergroup relations and group membership (Hogg, 2020).

speech. These studies used all sorts of fluency measures, but never the duration of empty pauses, and only looked at the effects on native speech, and not on non-native speech. Branum (2019) had his participants rate the perceived persuasiveness of a native and a non-native speaker and did find the speakers to be affected differently by the manipulated fluency; only the native speaker was affected by the fluency manipulations. Smith et al. (1988) found that speakers create an expectation of how someone will speak, based on the implied expertise. When these expectations are met, the speaker is rated as having higher expertise. The perceived expertise directly affects the perceived trustworthiness of a speaker (Schulze & Pishwa, 2015). Barner-Rasmussen & Björkman (2007) found non-native speakers to be labelled as language-based out-group, which causes them to be perceived as less trustworthy in general.

In the current thesis the perception of competency-based trustworthiness of speakers will be investigated, as has already been done in similar research (e.g., Erickson et al., 1978; Hosman & Stiltanen, 2011; Smith et al., 1988). However not all measures of fluency will be used, but only a specific one: duration of empty pauses, as already shown to affect perceived fluency by Bosker et al. (2014b). As in the research of Branum (2019), I will not only look at the effect on the perception of native speakers, but also on the perception of non-native speakers. I will also investigate what the effect of implied expertise is on the perception of native speakers, as already shown in Smith et al. (1988), and on the perception of non-native speakers.

Based on the above-mentioned studies there are two possible expectations for the effect of fluency manipulations on the perception of the speakers in this study. Based on Bosker et al. (2014b) it is possible that the manipulations of empty pauses will affect the competency-based trustworthiness of the native and non-native speaker in the same way. Based on Branum (2019) it is possible that the empty pause manipulations will only affect the competency-based trustworthiness of the native speaker, and not the non-native speaker. Further, based on the study of Barner-Rasmussen & Björkman (2007) I can expect the non-native speaker to be perceived as less competency-based trustworthy than the native speaker in all conditions, since the concept of language-based in-group and out-group can largely affect the perception of trustworthiness. For the different scenarios, I can expect the participants to create expectations based on the implied expertise, as in Smith et al. (1988), which will cause them to rate the speakers as more competency-based trustworthy in the expert scenario if their expectations are met.

These factors are all combined in the current study. This is done by using one 5-point likert-scale, applied twice, on the perception of competency-based trustworthiness of the speakers. The items in these likert-scales are based on the authoritativeness scale created by McCroskey (1966), which has been shown to be a reliable measure of credibility or competency-based trustworthiness (Wanzenried & Powell, 1993; Elsbach & Elofson, 2000). The research design based on Bosker et al. (2014b) is able to capture the elements of fluency, native and non-native speech, implied expertise, and the perceived competency-based trustworthiness of speech.

In order to investigate the effect of speech fluency on the perception of competency-based trustworthiness of native and non-native speech in an expert and non-expert scenario, I will address the following two research questions:

RQ 1: How do empty pause manipulations of native and non-native speech affect the perception of competency-based trustworthiness in native speakers?

RQ 2: Do such perceptions differ with respect to the level of implied expertise?

3. Methodology

3.1 Participants

A total number of 80 participants indicated their competency-based trustworthiness ratings in the online Qualtrics survey. All participants in the experiment were native speakers of Dutch. Most participants reported to have normal hearing, except for two participants. The mean scores for both scenarios differed for one participant less than 1 standard deviation from the group mean and for the other less than 2 standard deviations from the group mean, so they were included in the analysis. The participants were between the ages of 17 and 75, with a mean age of 32.60. From all participants 50 were female and the other 30 participants were male. The participants were divided into four groups based on the four separate blocks within the survey (NoPauses native, LongPauses native, NoPauses non-native, LongPauses non-native). The survey randomly chose one of the four blocks for every participant, which eventually resulted in four groups consisting of 20 randomly assigned participants. The participants were recruited via personal connections of the experimenter, and via two online platforms: Surveycircle (*SurveyCircle – The Largest Community for Online Research*, z.d.) and Surveyswap (*SurveySwap | Find Survey Participants Today*, z.d.).

8 of the participants were later asked to answer 5 follow-up questions via email, from which there were two participants of each group. These participants were recruited via the Qualtrics survey; at the end of the survey participants were asked whether I could contact them for more extensive questions about the recordings they had heard in the survey.

The speakers in the recordings were both female. The age of the native speaker was 22 and of the non-native speaker 28. The native speaker's L1 was Dutch, which was the only language she grew up with. The non-native speaker's L1 was Slovak, which was also the language she spoke while growing up. The non-native speaker was 19 when she started learning Dutch. Her last obtained certificate of her proficiency in Dutch based on the Common European Framework of Reference for Languages (CEFR) was C1, received from the language centre at the Leiden University. These two speakers were chosen, since differences in pitch, caused by gender and age, have been shown to affect the perceptions of trustworthiness (Belin et al., 2017). In the experiment of Belin et al. (2017) they found the perfect f0 for trustworthiness ratings. However the exact f0 was only based on the word 'hello', therefore, their exact f0 could not be used as the baseline in the current experiment. The pitch of both speakers could however be placed somewhere within the average pitch range for women as determined by earlier researchers (Borkowska & Pawlowski, 2011; Childers & Wu, 1991; Feinberg et al., 2008); the native speaker had an average pitch of 239.5Hz and the non-native speaker of 202Hz in their original recordings.

3.2 Materials

Four separate blocks were created within one survey, one for every condition. In total there were eight different speech fragments used in the survey made from four recordings. Every participant heard each scenario once within the survey. They indicated ratings on a 5-point likert-scale (strongly disagree - strongly agree) about each fragment. The materials section is divided into four parts: the recordings, the scenarios, the manipulations of fluency and the ratings for trustworthiness.

3.2.1 Recordings

Two speakers created the recordings: one native and one non-native speaker of Dutch. They both recorded two fragments for the experiment, with each a duration of around 30 seconds. The mean duration of the non-native recordings was 35.38 seconds and of the native recordings 44.70 seconds. Because the recording for scenario 1 of the native speaker was much longer (59s) than the other recordings, I cut out some sentences, which were not

necessary for the scenario to be understandable. This resulted in the recording to have a duration of 38 seconds. Within the original recordings the non-native speaker used a mean number of 10 silent pauses, and the native speaker used a mean number of 9 silent pauses.

There were some factors within the recordings that could have affected the results, and were therefore manipulated. First of all, the number of pauses per minute between both speakers, and within the speakers between the two recordings, was not close to equal. The lowest number of pauses per minute was 21.29 and the highest was 26.89 pauses per minute. This could be expected, since L2 speakers in general use more pauses within their speech, than native speakers do (De Jong, 2018). Since this study tests whether the duration of pauses affects the perception of trustworthiness, and not the number of pauses, as already done in earlier research, I deleted a number of pauses within the non-native speech fragments and added a pause within a native speech fragment to level the pauses per minute for every fragment.

Besides the number of pauses, both speakers differed in their speech rate. The native speech was much faster (mean speech rate: 245 words per minute) than the non-native speech (mean speech rate: 160 words per minute). As mentioned in the literature review, speech rate is also one of the significant predictors of perceived fluency (Cucchirarini, Strik & Boves, 2000; De Jong, 2018; Kormos & Denes, 2004; Rositter, 2009). Since the duration of pauses should be the only fluency measure that could affect the perceived competency-based trustworthiness of the speakers in this experiment, this was manipulated. The speech rate of the native speaker was slowed down by factor 1.1 and the speech rate of the non-native speaker was sped up by factor 0.9 in PRAAT. After these manipulations and the manipulations of the empty pauses, a total number of 8 recordings were created. Table 1 and 2 show the total duration of the recordings in seconds, the number of pauses per minute, the number of pauses between and within grammatical units, the total duration of the empty pauses, the mean duration per pause, the speech rate, and the pitch of the voices in the recordings, for both fluency conditions after all manipulations were carried out.

Table 1. A table containing all factors of the four recordings in the NoPauses condition after all manipulations were carried out.

	Native Scenario 1	**Native Scenario 2**	**Non-native Scenario 1**	**Non-native Scenario 2**
Total length recording (sec)	37,5s	29,97s	28,33s	21,53s
Number of pauses (100-150ms)	8	6	6	4
Pauses within grammatical units	0	2	2	2
Pauses between grammatical units	8	4	4	2
Number of pauses per minute	12,8 pm	12,01 pm	12,71 pm	11,14 pm
Total duration of pauses (sec)	1,42s	1,12s	1,18s	0,79s
Mean duration per pause (sec)	0,18s	0,19s	0,2s	0,2s
Speech rate (words per minute)	240	250	222	234
Mean pitch	233,38Hz	246,43Hz	210,05Hz	196,86Hz

Table 2. A table containing all factors of the four recordings in the LongPauses condition after all manipulations were carried out.

	Native Scenario 1	**Native Scenario 2**	**Non-native Scenario 1**	**Non-native Scenario 2**
Total length recording (sec)	43,92s	33,51s	31,40s	23,80s
Number of pauses (750-1000 ms)	8	6	6	4
Pauses within grammatical units	0	2	2	2
Pauses between grammatical units	8	4	4	2
Number of pauses per minute	10,92 pm	10,74 pm	11,46 pm	10,08 pm
Total duration of pauses (sec)	7,87s	5,08s	4,84s	3,48s
Mean duration per pause (sec)	0,98s	0,85s	0,81s	0,87s
Speech rate (words per minute)	204	223	200	211
Mean pitch	234,85Hz	245,87Hz	209,97Hz	196,41Hz

3.2.2 Scenario/story

I created the scenarios for the recordings myself, but based them on those used by other researchers in earlier experiments (Bosker et al., 2014; Erickson et al., 1977; Elsbach & Elofson, 2000; Branum Ma-thesis, 2019). All scenarios were of the same type, argumentative, but in different circumstances and environments, implying different levels of expertise. The first scenario included a lecture on listening to music while studying, and the second scenario included a discussion with a partner about taking the bike rather than the car. None of the scenarios were created in such a way that they were likely to generate strong opinions related to politics or controversial issues. All scenarios were explained in easy-to-understand language, since the difference between this and hard-to-understand language is shown to have an impact on the perception of trustworthiness (Elsbach & Elofson, 2000).

The speakers received the scenarios including the arguments they had to use for the recordings. Both speakers recorded the scenarios once, which went directly well the first time. The speakers included each argument that was indicated in the scenario descriptions. The native speaker however added some additional information in scenario 1 that was not given in the scenario description. Since this also caused the recording of scenario 1 to be much longer than the other recordings, the sentences including additional information were cut out of the recording. This resulted in the recordings of both speakers to include approximately the same amount of arguments and information. The exact scenarios, as received by the speakers, can be found in appendix 1. The transcripts of the final recordings, used in the survey, can be found in appendix 2.

3.2.3 Fluency aspect

The manipulations of the empty pauses were based on one of the experimental designs used in Bosker et al. (2014b), which was already discussed extensively in the literature review section. The exact same durations were used, which means that empty pauses, with durations of more than 250ms, were manipulated. In the current experiment only two of the three conditions were created, namely the NoPauses and the LongPauses conditions. In the NoPauses condition all empty pauses of more than 250ms were manipulated to durations of 150ms or less. The empty pauses in the LongPauses condition were manipulated to durations between 750ms and 1000ms. As in the experiment of Bosker et al. (2014b) one empty pause was left as it was in the middle of the recordings, to make them sound more natural. The NoPauses and LongPauses conditions were chosen for this experiment, since the difference between the empty pauses in these two conditions are the

clearest and most distinct, and are therefore expected to have the largest chance of affecting the perception of competency-based trustworthiness. Every participant only took part in one survey, which means that every participant only was exposed to one of the two possible fluency conditions, as suggested by Bosker et al. (2014b).

The one empty pause in the middle of the recordings, which was left as it was, differed a lot between the recordings, some were about 350ms and others more than 1000ms therefore I made them all equal to around 350ms. This was done to make sure that the one pause in the middle would not affect the perceived competency-based trustworthiness of a speaker in a certain fluency condition.

As mentioned in the literature review, L2 speakers tend to pause more within constituents and within clauses and Analysis of Speech, than L1 speakers (De Jong, 2018). This was also the case in the recordings of the current study: the non-native speech included more pauses within these units. If I had only manipulated the pauses between constituents, clauses and AS, and not within these units, the recordings of the non-native speaker would have included a larger number of long pauses than the recordings of the native speaker. This could have affected the perceived fluency of the non-native speaker, which could in turn have affected the competency-based trustworthiness of this speaker. Therefore I chose to manipulate both pauses within and between these units, instead of only between, to make sure that the pauses within these units would not potentially affect the perceived fluency of the non-native speaker. All manipulations of the recordings were carried out in PRAAT (Boersma & Weenink, 2012), and can be found in table 1 and 2 above.

3.2.4 Perceived trustworthiness

To measure the speaker perception a 5-point likert-scale (totally disagree - totally agree) was created. The likert-scale items were based on the authoritativeness scale created by McCroskey (1966). The statements were however translated to Dutch, since the whole experiment was conducted in Dutch. The statements were created in such a way that they would measure the speaker perception and not the message perception of the recordings. I divided the statements into pairs, based on the same type of statements; every pair consisted of a positive formulated statement and a negative formulated statement. I made a selection of 10 of the positive statements from all pairs. Although it has been advocated for many years that adding negative items to a scale prevents response bias, it has more recently been shown that adding negative items to a scale affects the internal consistency of it, which in turn leads to ambiguity of results (Salzar, 2015; Roszkowski, 2010; Barnette, 2000). All participants

were presented with the same 10 statements during the survey. An example of a likert-scale item used in the survey is the following: '*Ik heb vertrouwen in deze spreker.'* (I have confidence in this speaker.). All likert-scale items that were used in the survey can be found in appendix 3.

After the survey, I contacted the participants that indicated their email address at the end of my survey. They received an email in which a link was included to the two recordings they had heard in the survey and five follow-up questions in Dutch. Most questions were related to the perception of the speaker, for example: 'On which characteristics of the speaker did you base your judgements?' and 'How would you describe this speaker?'. Other questions were more related to the different scenario descriptions, for example: 'Did the scenario descriptions affect the way you perceived the speaker?'. All five follow-up questions can be found in appendix 4.

3.3 Procedure

3.3.1 Quantitative research

The survey was created and hosted on the online Qualtrics survey platform. The participants were first presented with a consent form, incorporated in the survey, which explained the nature of the experiment. This also gave the participants the option to continue with the survey or exit. After the consent form, participants were presented with the first instructions that explained what was expected from them during the survey. This was followed by a test sound, to make sure the audio was working correctly. When the audio was working properly, the participants continued to the actual recordings. The 5-point likert-scales were presented at the same time as the recordings, to make sure that the statements did not only affect how participants listened to the second recording and rated the competency-based trustworthiness in the second scenario.

After the participants finished answering the likert-scales, they indicated some personal background information such as: age, gender, the language spoken at home when they grew up, and whether they had any hearing problems. At the end, they were asked whether I could contact them again for some follow-up questions to go into more detail about their decisions indicated in the survey. If they wanted to, they could enter their email address, if not, they could just exit the survey. The survey was conducted in Dutch and took the participants about 5 minutes.

3.3.2 Follow-up questions

The number of people who indicated their email address at the end of the survey was skewed across groups; in one of the groups, only two participants left their email address, while in another group 8 participants did. I contacted 8 of the 22 people, two per condition, who left their email address and sent them the email with the follow-up questions. I created a folder for each condition in Google Drive, to be able to send a link to the participants with the speech fragments, such that they could only listen to them, but not edit or download them. I send this link together with the five questions that can be found in appendix 4. Only 4 out of the 8 participants answered within two weeks, all four participants belonged to a different group.

3.4 Data analysis

3.4.1 Quantitative data

To analyse the likert-scale data I converted the likert-scale ratings to numbers from 1 to 5, in which strongly disagree was 1 and strongly agree was 5. I calculated the summability score (Goeman & De Jong, 2018) for both scenarios to see how high the likert-scale items were correlated with each other. If an item caused noise in the correlation, I left it out of the analysis. After that, I calculated the mean score per scenario for every participant. I performed a three way mixed ANOVA, since there are three independent variables in the current study: nativeness, fluency, and scenario. I performed post-hoc t-tests on specific independent variables if a significant effect was found in the three-way mixed ANOVA.

3.4.2 Follow-up questions data

After I received the emails, I saved the answers of the four participants in one file. I then compared the answers of the participants for all five questions. I looked for common factors indicated in the answers, such as nativeness of the speaker, the empty pauses, effect of scenario type, and perception of trustworthiness. If the same factors were present in certain answers I checked whether these participants belonged to the same group based on fluency or nativeness, to be able to draw some conclusions for a particular group.

4. Results

The results section is divided into quantitative results, provided by the 5-point likert-scales in the survey, and the results of the follow-up questions, provided by open questions

via email. The quantitative results include multiple statistical tests, and the results of the follow-up questions include examples and common factors found in the answers.

4.1 Competency-based trustworthiness perception

A 5-point likert-scale per scenario derived the competency-based trustworthiness perception of the speaker. A summability score (Goeman & De Jong, 2018) of the likert-scale items was calculated for each scenario. The summability scores turned out moderate for both scenario 1 (s = 0.439) and scenario 2 (s = 0.375). However likert-scale item 1 caused some noise in the summability scores of both scenarios. This item included the statement: 'I respect the opinion of the speaker'. The fact that this item caused some noise in the correlation can be explained, since respecting someone's opinion does not mean agreeing with that opinion and trusting the speaker on their opinion. Therefore, likert-scale item 1 was left out of the analysis. Without likert-scale item 1 the summability scores for scenario 1 (s = 0.481) and scenario 2 (s = 0.432) turned out higher. This indicated a suitably good level of internal cohesion, so all other 9 items were included in the mean scores for the analysis. The mean scores per participant, per scenario, as used in the analysis, can be found in appendix 5.

The descriptive statistics of the competency-based trustworthiness ratings of both speakers can be found in table 3. The table shows that both the native and the non-native speaker were rated as evenly competency-based trustworthy in the first (expert) scenario using no empty pauses (NoPauses). The native speaker was rated as more competency-based trustworthy than the non-native speaker in the first scenario using long empty pauses (LongPauses). The non-native speaker was rated as more competency-based trustworthy than the native speaker in the second (non-expert) scenario using both no empty pauses (NoPauses) and long empty pauses (LongPauses). The native speaker was rated as more competency based trustworthy in the first scenario than in the second scenario in both fluency conditions. The non-native speaker was rated as more competency-based trustworthy in the second scenario than in the first scenario in both fluency conditions.

Table 3. Means (SD in brackets) for each group for competency-based trustworthiness ratings divided by the two scenarios (N = 20 per group).

	Native		Non-native	
	NoPauses	**LongPauses**	**NoPauses**	**LongPauses**
Scenario 1	2.84 (0.51)	3.07 (0.72)	2.84 (0.71)	2.76 (0.57)
Scenario 2	2.40 (0.48)	2.51 (0.67)	2.91 (0.52)	2.94 (0.43)

To answer the research questions a three-way mixed ANOVA between subjects between nativeness and fluency and within subjects between scenarios was performed. The results indicated no significant main effect for nativeness between subjects F (1,76) = 1.792, p = 0.185, which means that the average rating for the non-native speaker group was not significantly higher (M = 2.86, SD = 0.56), than for the native speaker group (M = 2.71, SD = 0.65). The results also indicated no significant main effect for fluency between subjects F (1,76) = 0.414, p = 0.522, so that the average rating for the LongPauses condition group was not significantly higher (M = 2.82, SD = 0.63), than for the NoPauses condition group (M= 2.75, SD = 0.59). There was also no significant interaction between nativeness and fluency F (1,76) = 0.663, p = 0.418. The within subjects results of the ANOVA indicated a significant effect for scenarios F (1,76) = 9.462, p = 0.003, which means that the average rating for scenario 1 was significantly higher (M = 2.88, SD = 0.63), than for scenario 2 (M = 2.69, SD = 0.57). The interaction between scenario and nativeness was also significant F (1,76) = 26.187, p < .001, suggesting that scenario type affected the non-native group and the native group differently. The interaction between scenario and fluency within subjects was not significant F (1,76) = $6.640e^{-5}$, p = 0.994, which means that scenario type did not affect the NoPauses group and the LongPauses group any differently. The three-way interaction between scenario, nativeness, and fluency was also not significant F (1,76) = 0.901, p = 0.345, which means that scenario type did not affect the interaction between nativeness and fluency.

Figure 1. This figure shows the distribution of competency-based trustworthiness ratings for the native and non-native speaker in scenario 1 (expert). The bars show the 95% confidence intervals, the dots indicate outliers.

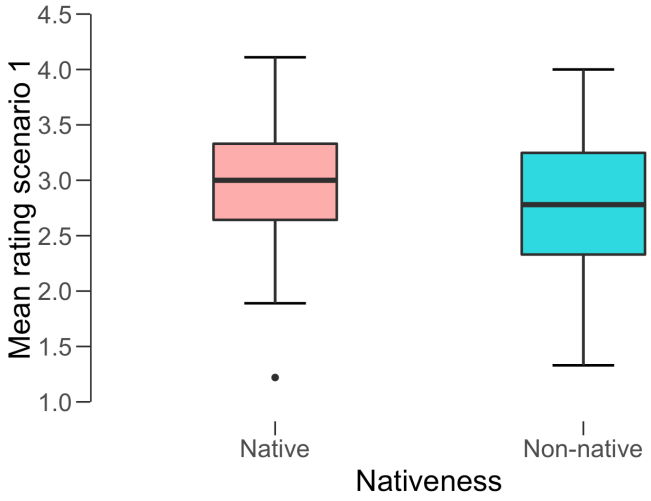
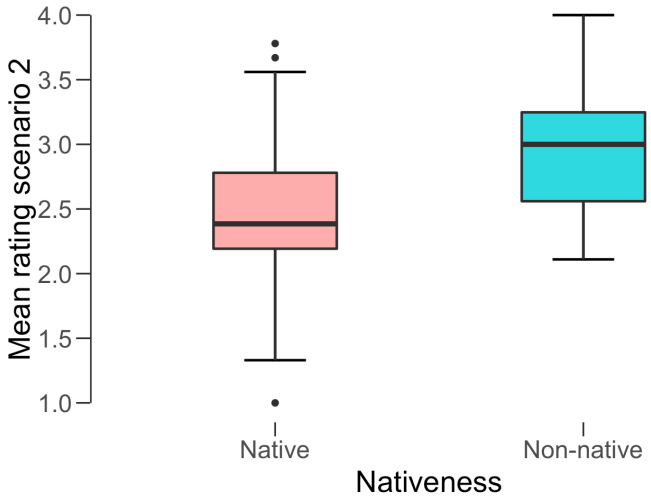


Figure 2. This figure shows the distribution of competency-based trustworthiness ratings for the native and non-native speaker in scenario 2 (non-expert). The bars show the 95% confidence intervals, the dots indicate outliers.



The effect for scenario and the interaction effect between scenario and nativeness are shown in figures 1 and 2. The figures show that in scenario 2 the native speaker was on average rated as less competency-based trustworthy than the non-native speaker, and as less competency-based trustworthy than both speakers in scenario 1. To investigate this interaction effect and the effect for scenario in more detail, a post-hoc analysis was carried out. For this, the data was split on scenarios, and two independent t-tests were carried out,

one for each scenario, to compare the native and the non-native group in both scenarios. For scenario 1 there was no significant difference between the native group (M = 2.96, SD = 0.63), and the non-native group (M = 2.80, SD = 0.64); t (78) = 1.124, p = 0.265. Levene's test indicated equal variances (F = 0.150, p = 0.699), so the degrees of freedom were unchanged. For scenario 2 the results indicated a significant difference between the native group (M = 2.45, SD = 0.58), and the non-native group (M = 2.93, SD = 0.47); t (78) = -3.979, p < .001. Levene's test indicated equal variances (F = 0.202, p = 0.655), so the degrees of freedom were unchanged. These results show that the native speaker was rated as significantly less competency-based trustworthy than the non-native speaker, only in the second scenario.

Figure 3. This figure shows the distribution of competency-based trustworthiness ratings for the native speaker in scenario 1 (expert) and scenario 2 (non-expert). The bars show the 95% confidence intervals, the dots indicate outliers.

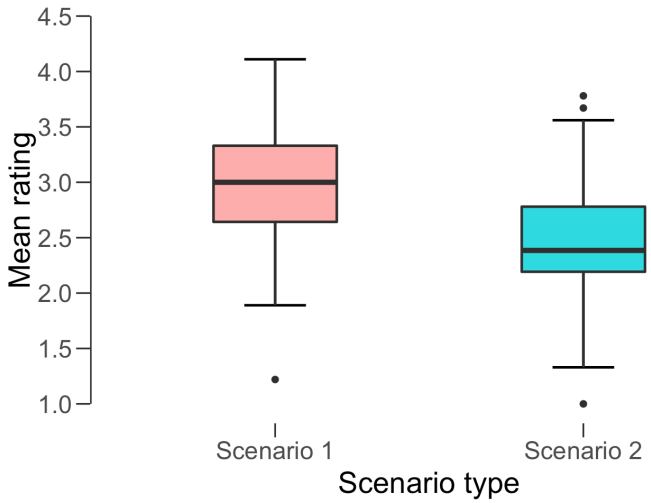
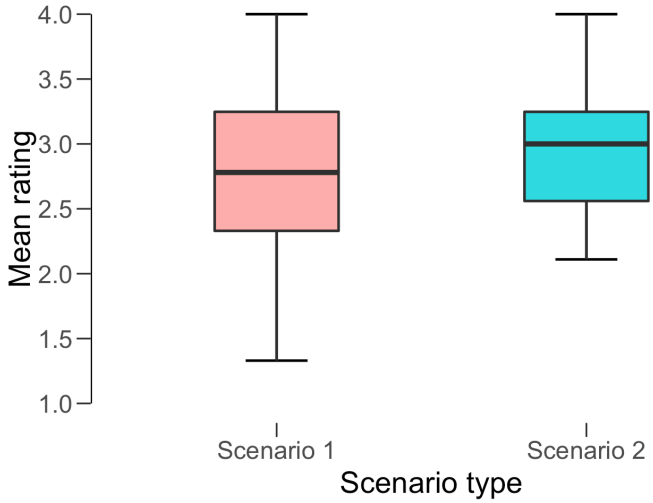


Figure 4. This figure shows the distribution of competency-based trustworthiness ratings for the non-native speaker in scenario 1 (expert) and scenario 2 (non-expert). The bars show the 95% confidence intervals.



Figures 3 and 4 show the interaction effect more clearly. The interaction shows that the native speaker was on average rated as more competency-based trustworthy in scenario 1 than in scenario 2. The non-native speaker was on average rated as more competency-based trustworthy in scenario 2 than in scenario 1. To investigate this effect in more detail, a second post-hoc analysis was carried out. For this, the data was split on nativeness, and two paired t-tests were carried out, one for each group, to compare the scenarios in both the native and non-native group. For the native group there was a significant difference between scenario 1 ($M = 2.96$, $SD = 0.63$), and scenario 2 ($M = 2.46$, $SD = 0.58$); $t(39) = 5.331$, $p < .001$. Shapiro Wilk's test indicated the data to be normally distributed ($W = 0.981$, $p = 0.726$), so the degrees of freedom were unchanged. For the non-native group the results indicated no significant difference between scenario 1 ($M = 2.80$, $SD = 0.64$), and scenario 2 ($M = 2.90$, $SD = 0.47$); $t(39) = -1.623$, $p = 0.113$. Shapiro Wilk's test indicated the data to be normally distributed ($W = 0.978$, $p = 0.630$), so the degrees of freedom were unchanged.

4.2 Follow-up questions

To get more insight into the underlying reasons for the ratings in the survey, the follow-up questions were analysed. No common factors were found between participants within one of the fluency groups, therefore only the analysis between participants based on the nativeness groups is discussed here. The first question of the follow-up questions asked on what characteristics of the speaker the participants based their judgements on the likert-

scales. Most of the participants indicated to have based it on the argumentation (message perception) and on how the argumentation was brought (speaker perception). One participant of the native group even mentioned to have paid attention to specific fluency measures; 'the articulation, speed of speech, stumbling over words, and speaking fluently' (translated from Dutch). For the second question, one participant in the native group indicated the speaker to be young, and spontaneous, but not an expert on the topic. The other participant in this group was more negative about the native speaker: 'She has an annoying voice and conveys her information in a negative way.' (translated from Dutch). The non-native speaker was indicated as smart by one participant, but especially characterized as not being Dutch by both participants in this group.

On question three both participants in the native group indicated the scenarios to have affected the expectations of the expertise of the speaker: 'The formality of the situation outlined gave me a verdict on the expertise, standing in front of a class versus discussing with a partner.' (translated from Dutch). However, one of the participants in this group indicated that it only affected her judgements of scenario 1 in a negative way: 'I expected a more professional text for the first scenario because it was a lecture for a school. But I think she gave arguments without substantiating them. So maybe I rated her more negatively due to higher expectations. In the second scenario, I don't think the description affected my rating as much because it was just a conversation with her boyfriend.' (translated from Dutch). This was expected based on the research of Smith et al. (1988), who indicated that listeners form an expectation based on the expertise of the speaker. However, both participants in the non-native group indicated the scenarios to not have affected their judgements of the speaker. On question four all participants indicated to not have noticed anything else about the speech fragments. None of them mentioned anything about the recordings being manipulated or about the empty pauses within the speech.

The last question asked whether the participants thought the speaker was trustworthy. Participants in both groups indicated the speaker to not be fully trustworthy, in some way. Two participants, one in the native and one in the non-native group, mentioned that the speakers did not use any references, which caused the speakers to be less trustworthy: 'The speaker mentions many arguments in the fragment, which radiates trustworthiness. However, I would like to know what source the argument is based on before I can say whether her arguments are correct.' (translated from Dutch). This same participant however also indicated; 'The fact that Dutch is not her native language does not affect the trustworthiness in my opinion.' (translated from Dutch). The other participant in the native group indicated

that the speaker would be trustworthy within an informal setting, but not as an expert on these topics. The other participant in the non-native group indicated that the speaker was not fully trustworthy, because ‘she says she thinks she knows’ (translated from Dutch). All full answers from the follow-up questions can be found in appendix 6.

5. Discussion

The experimental design in this study made use of one 5-point likert-scale, applied twice, for scenarios that differed in the level of implied expertise, to investigate the effect of empty pause manipulations, in native and non-native speech, on the perception of competency-based trustworthiness. No significant main effect for fluency was found between the NoPauses and the LongPauses group. This means that the empty pause manipulations in this study did not significantly affect the perception of competency-based trustworthiness of either the native or the non-native speaker. Based on Bosker et al. (2014b) this would not be expected, since they did find a main effect for fluency, such that the NoPauses condition was rated as more fluent than the ShortPauses and LongPauses conditions. However, they found their main effect for the perception of fluency, and not for the perception of competency-based trustworthiness. It could therefore be the case that either the NoPauses condition was not perceived as more fluent than the LongPauses condition, or the perception of fluency, based on the duration of empty pauses, does not directly affect the perception of competency-based trustworthiness. Either way the empty pause manipulations did not affect the competency-based trustworthiness in this experiment, such as other fluency measures did in earlier research (e.g., McGinnies & Ward, 1980; Erickson et al., 1978; Hosman & Stiltanen, 2011; Smith et al., 1988).

Further, no significant main effect was found for nativeness, which means that the native speaker was not perceived as more competency-based trustworthy than the non-native speaker overall. This result does not support the expectation based on the study of Barner-Rasmussen & Björkman (2007); non-native speakers are labelled as language-based out-group, which causes them to be perceived as less trustworthy. According to the results, this was not the case for the native and non-native speaker in the current study. The two participants in the non-native group both clearly labelled the speaker to be non-native, in the follow-up questions, however one also mentioned that the fact that the non-native speaker was not Dutch did not affect the trustworthiness of the speaker. The foreign accent of the non-native speaker indeed did not seem to affect the perception of competency-based

trustworthiness in a negative way. This means that there is a possibility that perceived accentedness does not affect competency-based trustworthiness of a speaker, which can be explained by Pinget et al. (2014); perceived fluency and perceived accentedness can be seen as separate constructs.

No interaction effect was found between nativeness and fluency. The empty-pause manipulations in the current study did not affect the ratings of either the native or the non-native speaker in a different way. This result does not support the expectations based on Branum (2019); the fluency manipulations did not affect the native speaker differently from the non-native speaker. It therefore seems that fluency manipulations do not affect competency-based trustworthiness in the same way as it affects perceived persuasiveness. Based on Bosker et al. (2014b), I expected the fluency manipulations to not affect the native speaker any differently from the non-native speaker. However, it is not only the case that the fluency manipulations did not affect the native speaker any differently from the non-native speaker, but it seems to not have affected either speaker at all. Overall, no effects between subjects were found in the current study, whereas effects would have been expected based on earlier research mentioned above.

Since the scenarios included different types of situations, scenario 1 implied the speaker to be an expert, and scenario 2 included an informal situation, the possible effects of both were investigated too. An overall effect for scenarios within subjects was found, which implies that the scenario descriptions did directly affect the way listeners perceived the speakers competency-based trustworthiness. This was expected based on Smith et al. (1988), who indicated that listeners form an expectation based on the expertise of the speaker, and rate speakers with higher expertise when these expectations are met. No interaction effect between fluency and scenario was found, which indicates that scenario type did not affect the fluency conditions any differently. This also means that the fluency manipulations in the current study did not affect the expectations based on expertise, as the disfluencies did in the study of Smith et al. (1988). A significant effect was found for the interaction between scenario and nativeness. This means that the scenarios did affect the native and the non-native group differently. No interaction effect between all three independent variables was found, which indicates that scenario type did not affect the interaction between nativeness and fluency.

Follow-up t-tests showed that the native speaker was perceived as significantly less competency-based trustworthy in scenario 2 (non-expert), than the non-native speaker. The other two follow-up t-tests could explain the effects more clearly; only the native speaker was

perceived as significantly more competency-based trustworthy in scenario 1 (expert) than in scenario 2 (non-expert). There was no significant effect found between the scenarios in the non-native group. This means that scenario type only affected the native group in the expected way based on Smith et al. (1988), but not the non-native group. An interesting factor is that in the follow-up questions only the participants in the native group indicated the scenario description to have affected their competency-based trustworthiness ratings, from which one indicated it to have affected her ratings in a negative way. The results however seem to indicate that higher expectations were made and met in scenario 1, while either lower expectations were made or higher expectations were not met in scenario 2. This raises the question whether the competency-based trustworthiness ratings for the native speaker in scenario 1 would have been higher if the expectations of all participants were met. On the other hand, the participants in the non-native group indicated the scenario descriptions to not have affected their perception of the competency-based trustworthiness of the speaker. This is reflected in the results since no significant effect was found between the two scenarios of the non-native speaker, which raises the question, whether the expectations based on expertise were adjusted when participants listened to the non-native speaker. Bosker et al. (2014a) and Hanulikova et al. (2012) already showed that we are able to adapt our predictive strategies when we know we are listening to a non-native speaker. Maybe the same effect applies to expectations based on expertise, such that expectations of expertise are adjusted as soon as we hear a non-native speaker, which causes us to create lower expectations. These lower expectations are then met, and they are rated as evenly competency-based trustworthy in both an expert and non-expert scenario.

Two last points for discussion are related to the fact that none of the participants indicated in the follow-up questions that they had noticed something about the recordings. First of all, they did not mention that the recordings were manipulated, although a lot of features were manipulated in the recordings. Not only the duration of empty pauses, but also the speech rate and the number of pauses were manipulated. However, also nothing was mentioned about the empty pauses in the speech fragments. Not even the one participant, who indicated multiple fluency measures to have affected her judgements, mentioned the duration of the pauses. This raises the question whether the participants did actually not notice the durations of the empty pauses at all, or whether they were noticed unconsciously but did not affect the perception of competency-based trustworthiness.

5.1 Limitations

Although this study shows some interesting results and raises some interesting questions for further research, there are also a number of limitations, which need to be taken into account when conducting research on this topic in the future. The finding that the duration of empty pauses did not affect competency-based trustworthiness asks for additional research. The fact that the participants did not indicate anything about the duration or the pauses themselves raises the question whether the participants did not notice the empty pauses at all, or whether the perception of this measure of disfluency did not directly affect the competency-based trustworthiness ratings. Further research could deal with this question by adding perceived fluency ratings to the experiment. By doing this, it can be investigated whether the durations of empty pauses are noticed or not and in that way also whether they affect the competency-based trustworthiness.

Some of the participants in the follow-up questions mentioned the argumentation of the speaker to have affected their judgements of competency-based trustworthiness. This relates to the message perception, while the 5-point likert-scale measured the speaker perception. Although the speaker perception is the aspect that was measured, it seems that the participants also took the message perception into account when rating the speakers. These two measures are difficult to keep as separate aspects of a speech fragment, when the participants are not instructed on what aspect of the speech fragment they have to pay attention to. The message perception could therefore have influenced the results in the current study.

Another limitation is that there were only two speakers in the experiment, one native (Dutch) speaker and one non-native (Slovak) speaker. Although a benefit is that an effect of the speaking style of both speakers would be consistent across listeners, it is difficult to draw conclusions and generalize these conclusions to the whole population based on only one native and one non-native speaker. Besides that, the non-native speaker only spoke in her L2, while she maybe could have had the same results if she spoke in her L1. This would mean that there are other factors, such as speaking style, which could have affected the ratings of competency-based trustworthiness of the non-native speaker. It could also be the case that the L1 of the non-native speaker affected how native speakers perceived her in her L2.

More limitations based on the speakers are related to the fact that both were female. The female speakers were chosen, since pitch affects the perception of trustworthiness (Belin et al., 2017). Although both speakers were in the scope of the normal distribution of female voices, their pitches differed around 30Hz. Maybe this difference did affect the competency-

based trustworthiness perception of the speakers. The native speaker was also indicated as having an annoying voice by one of the participants in the follow-up questions. Maybe this negative trait affected the way she perceived the competency-based trustworthiness of the speaker. Erickson et al. (1977) also showed that gender affected the perceived credibility of their speakers, such that listeners of the same sex as the speaker would rate the speaker as more credible than when the speaker was of the opposite sex. Although this was not investigated in the current study, gender could have affected the results, in such a way that male participants rated the speakers differently than female participants did. All the above-mentioned factors could be dealt with in future research by using a larger set of recordings from a speaker group with a variety of speakers of different ages, genders, and backgrounds, to make it more representative. This was however beyond the scope of the current study.

Another potential drawback is the fact that the recordings also included other disfluencies. The speakers had to create imperfect recordings, since the recordings needed to include empty pauses, to be able to manipulate these. However, this also caused the speakers to produce other disfluencies such as filled pauses, repetitions, and repairs. Although no effect has been found for fluency, these disfluencies have been shown to affect the perception of credibility, expertise and trustworthiness in earlier research (McGinnies & Ward, 1980; Erickson et al., 1978; Hosman & Stiltanen, 2011). Therefore, it could have affected the perception of competency-based trustworthiness of the speaker. In future research it would be better, to prevent or manipulate these disfluencies, to make sure they do not affect the results.

This study asks for an in depth representation of the communicative act of competency-based trustworthiness. It is therefore difficult to include the complex process of trustworthiness into a 5-point likert-scale. The likert-scale items are an abstract representation of the actual perception of credibility and competency-based trustworthiness. Besides that, the used likert-scale items created in 1966 by McCroskey, although shown to be reliable by later studies (Wanzenried & Powell, 1993; Elsbach & Elofson, 2000), can be expected to be somewhat out-dated. The communicative act of competency-based trustworthiness changes together with new generations and new technologies, such as social media platforms, to communicate. Maybe a new or an adjusted authoritative scale would be necessary to capture all these features, and be more representative for the new ways of communication.

A limitation I caused myself is the fact that the audio fragments in the survey were not forced to listen to, although the questions were forced to respond to. A participant could therefore have only listened a couple of seconds to the audio and already started to answer the likert-scale items. This means that some participants may not have listened through the whole

fragment and therefore also not heared all empty pauses within the speech fragments, which could have affected the results of the current study. For future research it is therefore important to make sure participants are forced to listen to the speech fragments.

A final limitation relates to the participation in my study. Due to time constraints only 80 people could participate in my survey, which resulted in 20 participants per group. This is however the lower limit for performing an ANOVA, so therefore it did not directly affect the results. Although 80 people participated in the survey, the number of people that left their email addresses was skewed across groups. Therefore only a total number of 8 participants were sent an email, from whom unfortunately only 4 answered the follow-up questions in the email. This made me unable to generalize the conclusions of the follow-up questions, since it was based on only one participant of each group.

6. Conclusion

This research has investigated a phenomenon, which impacts a wide range of people all over the world on a daily basis, through communication. The results did not directly answer the first research question, such that no effects of the fluency manipulations in either the native or non-native speech on the perception of competency-based trustworthiness were found. The native speaker was not perceived as more competency-based trustworthy, than the non-native speaker overall, as expected based on earlier research. However, the second research question can be answered partially; a significant effect was found for scenario type and for the interaction between scenario type and nativeness. The post-hoc analysis showed that the native speaker was perceived as significantly less competency-based trustworthy than the non-native speaker in the non-expert scenario. The second post-hoc analysis showed that the native speaker was perceived as more competency-based trustworthy in the expert scenario than in the non-expert scenario, while this same effect was not found for the non-native speaker. Although this research includes a number of limitations, as discussed in section 5, the results show that implied expertise does affect the perception of a native speaker, but not of a non-native speaker.

The fact that scenario type did affect the perceived competency-based trustworthiness of the native speaker, but not of the non-native speaker raises a number of questions that would be interesting for additional research. First of all, does this mean that we only create expectations of a native speaker? Or do we adjust our expectations based on expertise as soon as we hear a non-native speaker? If we adjust our expectations based on the expertise of a

non-native speaker, does this then mean that we expect less expertise from a non-native speaker? Or does it mean that we judge a non-native speaker on other aspects of expertise than a native speaker when giving ratings of competency-based trustworthiness? To answer these questions more extensive further research is needed on different measures of speech fluency, to be able to investigate how these aspects are processed in native and non-native speech, and how this relates to the perception of competency-based trustworthiness.

Two final suggestions for further research; the present study only used native listeners who listened to both native and non-native speech, and no distinction was made between male and female listeners. However, real world communication often involves both male and female speakers and listeners from a variety of backgrounds. This means that as well as the speaker can be non-native the listener can be non-native too. Rossiter (2009) found highly proficient non-native speakers to be able to rate perceived fluency on the same level as native listeners. However, native and non-native listeners may perceive competency-based trustworthiness differently, and so do male and female listeners. Therefore, research in which both female and male listeners from multiple language backgrounds are included could be fascinating for further research that is more representative for communication in the real world.

Overall, the results do not support the notion of native speakers to be more competency-based trustworthy than non-native speakers. Empty pause durations do not seem to affect the perception of competency-based trustworthiness either. In addition, this study has shown that implied expertise does affect the perception of competency-based trustworthiness differently for native and non-native speakers, such that the native speaker is affected by implied expertise in contrast to the non-native speaker. Although effects of fluency, as expected by earlier research, are not found, the significant effects for scenario type clearly show the value of further and more detailed research on this topic.

7. References

Author Removed At Request Of Original Publisher. (2016, 29 september). 11.4 Persuasive Strategies – Communication in the Real World. Pressbooks. https://open.lib.umn.edu/communication/chapter/11-4-persuasive-strategies/

Barner-Rasmussen, W., & Björkman, I. (2007). Language Fluency, Socialization and Inter-Unit Relationships in Chinese and Finnish Subsidiaries. Management and Organization Review, 3(1), 105–128. https://doi.org/10.1111/j.1740-8784.2007.00060.x

Barnette, J.J. 2000. Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement* 60, no. 3: 361–70.

Belin, P., Boehme, B., & McAleer, P. (2017). The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. *PLOS ONE*, *12*(10), e0185651. https://doi.org/10.1371/journal.pone.0185651

Bergmann, C., Sprenger, S. A., & Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychologica*, *161*, 25–35. https://doi.org/10.1016/j.actpsy.2015.07.015

Boehme, B., & Belin, P. (2014). Page 1 of 37 How trustworthy is your voice? The effects of voice manipulation on the perceived trustworthiness of novel speakers.

Boersma, P., & Weenink, D. (2012). *Praat: doing phonetics by computer* [computer program]. Retrieved from: http://www.praat.org/ (Version 6.0.36)

Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. Animal Behaviour, 82(1), 55–59. https://doi.org/10.1016/j.anbehav.2011.03.024

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & de Jong, N. H. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*(2), 159–175. https://doi.org/10.1177/0265532212455394

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014a). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, *75*, 104–116. https://doi.org/10.1016/j.jml.2014.05.004

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014b). The Perception of Fluency in Native and Nonnative Speech. *Language Learning*, *64*(3), 579–614. https://doi.org/10.1111/lang.12067

Branum, J. K. (2019). *Persuasion in native and non-native speech: the effect of speech rate on measures of belief change and perceived persuasiveness* (Thesis). Leiden University.

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 199–202). Aix-en-Provence, France: Laboratoire Parole et Langage.

Childers, D. G., & Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis. The Journal of the Acoustical Society of America, 90(4), 1841–1856. https://doi.org/10.1121/1.401664

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, *107*(2), 989–999. https://doi.org/10.1121/1.428279

De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. International Review of Applied Linguistics in Language Teaching. 54. 10.1515/iral-2016-9993.

De Jong, N. H. (2018). Fluency in Second Language Testing: Insights From Different Disciplines. Language Assessment Quarterly, 15(3), 237–254. https://doi.org/10.1080/15434303.2018.1477780

De Jong, N. H. & Bosker, H.R. (2013). Choosing a threshold for silent pauses to measure second language fluency.

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223–243. https://doi.org/10.1017/s0142716413000210

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second Language Fluency: Judgments on Different Tasks. *Language Learning*, *54*(4), 655–679. https://doi.org/10.1111/j.1467-9922.2004.00282.x

Elsbach, K. D., & Elofson, G. (2000). How the Packaging of Decision Explanations Affects Perceptions of Trustworthiness. *Academy of Management Journal*, *43*(1), 80–89. https://doi.org/10.5465/1556387

Erickson, B., Lind, E. A., Johnson, B. C., & O'Barr, W. M. (1978). Speech style and impression formation in a court setting: The effects of "powerful" and "powerless" speech.

Journal of Experimental Social Psychology, 14(3), 266–279. https://doi.org/10.1016/0022-1031(78)90015-x

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2008). The Role of Femininity and Averageness of Voice Pitch in Aesthetic Judgments of Women's Voices. Perception, 37(4), 615–623. https://doi.org/10.1068/p5514

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied linguistics*, *21*(3), 354-375.

Goeman, J. J., & De Jong, N. H. (2018). How well does the sum score summarize the test? Summability as a measure of internal consistency. *Educational Measurement: Issues and Practice*, *37*(2), 54-63. doi: https://doi.org/10.1111/emip.12181 x

Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*, *20*(1), S18-S29. doi: https://doi.org/10.1016/j.neuroimage.2003.09.013

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When One Person's Mistake Is Another's Standard Usage: The Effect of Foreign Accent on Syntactic Processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887. https://doi.org/10.1162/jocn_a_00103

Hardin, R. (2002). *Trust and Trustworthiness*. SAGE Publications. https://doi.org/10.7758/9781610442718

Hogg, M. A. (2020). Chapter 5 Social Identity Theory. *Contemporary Social Psychological Theories*, 112–138. https://doi.org/10.1515/9781503605626-007

Hosman, L. A., & Siltanen, S. A. (2011). Hedges, Tag Questions, Message Processing, and Persuasion. *Journal of Language and Social Psychology*, *30*(3), 341–349. https://doi.org/10.1177/0261927x11407169

Jones, K. (2012). Trustworthiness. *Ethics*, *123*(1), 61–85. https://doi.org/10.1086/667838

Kahng, J. (2014). Exploring Utterance and Cognitive Fluency of L1 and L2 English Speakers: Temporal Measures and Stimulated Recall. *Language Learning*, *64*(4), 809–854. https://doi.org/10.1111/lang.12084

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145–164. https://doi.org/10.1016/j.system.2004.01.001

Lennon, P. (1990). Investigating Fluency in EFL: A Quantitative Approach*. *Language Learning*, *40*(3), 387–417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x

Matzinger, T., Ritt, N., & Fitch, W. T. (2020). Non-native speaker pause patterns closely correspond to those of native speakers at different speech rates. *PLOS ONE*, *15*(4), e0230710. https://doi.org/10.1371/journal.pone.0230710

McCroskey, J. C. (1966). Scales for the measurement of ethos. *Speech Monographs*, *33*(1), 65–72. https://doi.org/10.1080/03637756609375482

McDougall, K., & Duckworth, M. (2017). Profiling fluency: An analysis of individual variation in disfluencies in adult males. *Speech Communication*, *95*, 16–27. https://doi.org/10.1016/j.specom.2017.10.001

McGinnies, E., & Ward, C. D. (1980). Better Liked than Right. Personality and Social Psychology Bulletin, 6(3), 467–472. https://doi.org/10.1177/014616728063023

Oleszkiewicz, A., Pisanski, K., Lachowicz-Tabaczek, K., & Sorokowska, A. (2016). Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. *Psychonomic Bulletin & Review*, *24*(3), 856–862. https://doi.org/10.3758/s13423-016-1146-y

Pinget, A. F., Bosker, H. R., Quené, H., & de Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, *31*(3), 349–365. https://doi.org/10.1177/0265532214526177

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, *14*(4), 423–441. https://doi.org/10.1080/01638539109544795

Rossiter, M. J. (2009). Perceptions of L2 Fluency by Native and Non-native Speakers of English. *Canadian Modern Language Review*, *65*(3), 395–412. https://doi.org/10.3138/cmlr.65.3.395

Roszkowski, M. J., & Soven, M. (2010). Shifting gears: consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, *35*(1), 113–130. https://doi.org/10.1080/02602930802618344

Salazar, M. S. (2015). The dilemma of combining positive and negative items in scales. *Psicothema*, *27*(2), 192–199. https://doi.org/10.7334/psicothema2014.266

Schulze, R., & Pishwa, H. (2015). The Exercise of Power in Communication: Devices, Reception and Reaction (2015de editie). Palgrave Macmillan. https://doi.org/10.1057/9781137478382

Segalowitz, N. (2010). Cognitive bases of second language fluency. Routledge.

Skehan, Peter & Foster, Pauline. (2012). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. Complexity, Accuracy and Fluency in Second Language Use, Learning and Teaching. 207-226.

Smith, V., Siltanen, S. A., & Hosman, L. A. (1998). The effects of powerful and powerless speech styles and speaker expertise on impression formation and attitude change. Communication Research Reports, 15(1), 27–35. https://doi.org/10.1080/08824099809362094

SurveyCircle – The Largest Community for Online Research. (z.d.). SurveyCircle. Geraadpleegd op 1 juni 2021, van https://www.surveycircle.com/en/

SurveySwap | Find Survey Participants Today. (z.d.). SurveySwap. Geraadpleegd op 1 juni 2021, van https://surveyswap.io/

Tsankova, E., Aubrey, A. J., Krumhuber, E., Möllering, G., Kappas, A., Marshall, D., & Rosin, P. L. (2013). Facial and Vocal Cues in Perceptions of Trustworthiness. *Computer Vision - ACCV 2012 Workshops*, 308–319. https://doi.org/10.1007/978-3-642-37484-5_26

Tavakoli, P. (2010). Pausing patterns: differences between L2 learners and native speakers. *ELT Journal*, *65*(1), 71–79. https://doi.org/10.1093/elt/ccq020

Wanzenried, J. W., & Powell, F. C. (1993). Source Credibility and Dimensional Stability: A Test of the Leathers Personal Credibility Scale Using Perceptions of Three Presidential Candidates. *Perceptual and Motor Skills*, *77*(2), 403–406. https://doi.org/10.2466/pms.1993.77.2.403

Williams, M. (2001). In Whom We Trust: Group Membership as an Affective Context for Trust Development. *The Academy of Management Review*, *26*(3), 377. https://doi.org/10.2307/259183

Yi H, Smiljanic R and Chandrasekaran B (2014) The neural processing of foreign-accented speech and its relationship to listener bias. *Front. Hum. Neurosci*. **8**:768. doi: 10.3389/fnhum.2014.00768

8. Appendices

8.1 Appendix 1

This appendix includes the scenario descriptions as received by the speakers, to create the recordings.

Scenario 1:

Je geeft een lezing op een middelbare school over het feit dat het luisteren van muziek geen goede invloed heeft op studeren. De reden hiervoor is dat aandacht bestaat uit twee delen: een onbewust deel en een bewust deel. Het onbewuste deel is altijd alert op achtergrond geluiden. Dit zorgt ervoor dat het onbewuste deel reageert op de muziek, waardoor je bewuste deel minder goed kan concentreren op de studiestof. Hierdoor neem je dus ook minder stof op en duurt het langer om de stof te leren. Je neemt dezelfde stof sneller op wanneer je in totale stilte studeert. Daarom kan je dus beter geen muziek luisteren tijdens het studeren, maar gewoon complete stilte hebben.

Probeer dit in maximaal 1 minuut te verwoorden alsof je voor een groep middelbare scholieren staat met de volgende argumenten erin:

- Aandacht bestaat uit twee delen
- Het onbewuste deel reageert op muziek
- Het bewuste deel kan minder goed concentreren op de stof
- Je neemt minder stof op
- Je doet er langer over

Je begint je verhaal met: “Het luisteren van muziek heeft geen goede invloed op studeren …”

Scenario 2:

Later kom je thuis en voer je een discussie met jouw partner over of jullie zullen fietsen of met de auto naar vrienden zullen gaan die avond. Je pleit voor fietsen, omdat dit gezonder is voor jezelf, doordat het bewegen zorgt voor minder ziektes, het helpt om te ontspannen, je er beter van kunt slapen en het helpt tegen overgewicht. Daarnaast is het beter voor het milieu omdat je met de auto wel uitstoot hebt en met de fiets niet. En als laatste, heb je met de fiets geen kans op file en met de auto wel.

Probeer dit in maximaal 1 minuut te verwoorden alsof je tegen jouw eigen partner praat met de volgende argument erin:

- Fietsen is gezonder

- Het zorg voor minder ziektes
- Het helpt om te ontspannen
- Je kunt er beter van slapen
- Het helpt tegen overgewicht
- Beter voor het milieu
- Geen kans op file

  Je begint je verhaal met: “Ik denk dat we beter met de fiets kunnen gaan …”

8.2 Appendix 2

This appendix includes the transcripts of the recordings as used in the survey, after the manipulations were carried out. These transcripts were the same for both fluency conditions; only the duration of the empty pauses differed.

Native speaker scenario 1

“het luisteren van muziek heeft geen goeie invloed op studeren ***empty pause*** het aandacht systeem bestaat uit twee delen ***empty pause*** één deel is het onbewuste en één deel het andere deel is het bewuste ***empty pause*** met het bewuste deel ben je aan het studeren onder begeleiding van een rustig piano muziekje waarvan je zelf denkt dat het niet afleidend is ***empty pause*** echter het onbewuste gedeelte is altijd alert op achtergrondgeluiden ***empty pause*** dit zorgt ervoor dat je niet met je volle aandacht in het bewuste gedeelte je kunt concentreren op de stof die je leest ***empty pause*** het resultaat hiervan is dat je minder stof opneemt terwijl je zo geconcentreerd aan het studeren lijkt te zijn ***empty pause*** en wanneer je in volle stilte studeert heb je hier geen last van en neem je dezelfde stof dus sneller op ***empty pause*** al met al het is is het dus niet zo handig om met muziek te studeren maar gewoon in complete stilte”

Native speaker scenario 2

“ik denk echt dat we beter met de fiets kunnen gaan hoor er zijn zoveel voordelen aan lekker fietsen in plaats van met de auto ***empty pause*** het is sowieso gezonder je hebt een beetje beweging je kan er van afvallen en het zorgt ervoor dat je minder kans hebt op ziektes ***empty pause*** het schijnt ontspannend te werken dat als je een stuk gefietst hebt dat je lekker fris en fruitig ***empty pause*** *uhm* aankomt en je kan beter slapen ’s avonds als je een beetje beweging hebt gehad ***empty pause*** sowieso fietsen is veel beter voor het milieu

dan in zo'n vieze stinkende file te gaan staan met je auto die je dus ook nog eens ontwijkt als je gaat fietsen want ja ***empty pause*** met de fiets heb je ook geen files"

Non-native speaker scenario 1

"het luisteren van muziek heeft geen goeie invloed op studeren want dan wordt je aandacht verdeeld ***empty pause*** het aandacht bestaat uit twee delen een onbewust deel en een bewust deel en het onbewuste deel is altijd alert op achtergrondgeluiden ***empty pause*** dat betekent dat wanneer je tijdens het studeren naar muziek luistert ***empty pause*** een deel van je aandacht reageert erop en de andere ***empty pause*** bewuste deel kan minder goed concentreren op de stof ***empty pause*** als je dit doet je neemt minder stof op en het duurt langer voordat je leert wat je moet leren ***empty pause*** daarom kan je dus beter geen muziek luisteren tijdens het studeren maar gewoon complete stilte hebben"

Non-native speaker scenario 2

"ik denk dat we beter met de fiets kunnen gaan ***empty pause*** het is niet alleen sneller want er is natuurlijk geen kans op file met de fiets maar ook beter voor het milieu het is ook gezonder want met de lichamelijke beweging die je tijdens ***empty pause*** tijdens het fietsen maakt *uh* zorg je voor minder ziektes en het helpt ook tegen overgewicht ***empty pause*** en natuurlijk na de hele avond zitten als je 30 minuten gaat fietsen ***empty pause*** je hele lichaam ontspant en je kunt er beter van slapen"

8.3 Appendix 3

The statements as used in the surveys in Dutch, English translations in brackets:

1. Ik respecteer de opinie van deze spreker over dit onderwerp. (I respect this speaker's opinion on the topic.)
2. Deze spreker is goed geïnformeerd over het onderwerp. (The speaker is well-informed on this subject.)
3. Ik beschouw deze spreker als expert op dit gebied. (I would consider this speaker to be an expert on the topic.)
4. Deze spreker is aan autoriteit op dit gebied. (This speaker is an authority on the topic.)
5. Deze spreker heeft veel ervaring met dit onderwerp. (This speaker has had substantial experience with this subject.)
6. Deze spreker heeft veel kennis van factoren die bij dit onderwerp een rol spelen. (This speaker has considerable knowledge of the factors involved with this subject.)

7. Er zijn maar weinig mensen die zo gekwalificeerd zijn om over dit onderwerp te spreken als deze spreker. (Few people are as qualified to speak on this topic as this speaker.)
8. Deze spreker heeft een hoge status in onze samenleving. (This speaker has high status in our society.)
9. De spreker is een betrouwbare bron van informatie over dit onderwerp. (This speaker is a reliable source of information on the topic.)
10. Ik heb vertrouwen in deze spreker. (I have faith in this speaker.)

8.4 Appendix 4

The follow-up questions that were sent via email, English translations in brackets.

1. 'Op welke eigenschappen van de spreker heb je jouw beoordeling gebaseerd?' (On which characteristics of the speaker did you base your judgments?)
2. 'Hoe zou je de spreker omschrijven?' (How would you describe the speaker?)
3. 'Hadden de scenario omschrijvingen veel invloed op hoe je de spreker hebt beoordeeld? Zo ja, waarom?' (Did de scenario descriptions affect how je judged the speaker? If yes, why?)
4. 'Is je nog iets anders opgevallen aan de fragmenten? Zo ja, wat?' (Did anything about the speech fragments stood out to you? If yes, what?)
5. 'Vertrouw je deze spreker? Leg uit waarom wel of waarom niet. ('Do you trust this speaker? Explain why you do or don't.)

8.5 Appendix 5

This appendix includes the mean rating data from the survey per participant per scenario used for the analysis.

Table 4. This table includes the mean ratings, from the 9 likert-scale items used in the analysis, for scenario 1 and scenario 2 per participant. The mean rating scores were calculated from the ratings converted to numbers (strongly disagree = 1, disagree = 2, neutral = 3, agree = 4, strongly agree = 5). Nativeness and fluency indicate which groups the participants were in.

Nativeness	Fluency	Scenario 1	Scenario 2
Native	**NoPauses**	2,67	2,22
		3,78	2,33
		2,00	2,11

	3,67	2,22
	2,22	2,22
	2,67	1,67
	2,56	2,89
	3,22	2,00
	3,00	2,78
	1,89	2,67
	3,00	1,78
	3,33	3,00
	2,56	2,22
	2,89	2,00
	2,78	3,11
	2,44	2,00
	3,44	3,56
	2,78	2,22
	2,67	2,44
	3,33	2,56
LongPauses	3,22	2,56
	4,11	3,78
	2,44	2,22
	3,33	2,78
	3,33	3,67
	2,89	2,78
	3,89	2,22
	3,11	2,33
	2,89	2,78
	2,67	2,00
	3,22	2,22
	3,11	2,44
	2,00	1,00
	3,11	2,44
	3,44	3,22
	4,11	2,67

		1,22	1,33
		3,56	2,11
		2,11	2,56
		3,67	3,11
Non-native	**NoPauses**	2,67	2,56
		3,67	3,22
		2,78	3,11
		1,33	2,22
		3,00	3,00
		3,11	3,56
		3,44	3,33
		3,00	3,11
		2,44	2,56
		2,44	3,00
		2,11	2,44
		2,67	2,11
		2,44	2,89
		3,22	3,22
		4,00	3,33
		2,78	2,56
		1,78	2,22
		4,00	4,00
		3,67	3,44
		2,22	2,22
	LongPauses	2,78	2,67
		2,00	2,56
		3,33	3,33
		2,44	2,56
		3,11	3,22
		3,00	3,11
		2,44	3,56
		3,33	3,22
		2,00	3,33

	3,11	3,33
	3,33	2,78
	3,11	2,67
	2,22	3,11
	2,67	3,22
	2,00	2,33
	2,33	2,44
	2,22	2,89
	2,33	2,33
	3,89	3,78
	3,56	2,44

8.6 Appendix 6

This appendix includes the raw data from the follow-up questions.

Participant in the NoPauses Native group:

1. Op basis van hoe ze de informatie bracht en welke argumenten ze hiervoor gebruikte. Verder weet ik ook niet haar achtergrondinformatie (hoe oud ze is, opleiding, werk ervaring).
2. Vervelend. Ze heeft een vervelende stem en brengt haar informatie op een negatieve manier.
3. Ja, ik verwachte bij het eerste scenario een professionelere tekst omdat het een lezing voor een school was. Maar naar mijn idee gaf ze argumenten zonder deze te onderbouwen. Dus misschien dat ik haar negatiever beoordeelde door hogere verwachtingen. Bij het tweede scenario had de omschrijving denk ik niet zoveel invloed op mijn beoordeling, omdat het gewoon een gesprek met haar vriend was.
4. Nee.
5. Nee want ze gebruikt geen bronnen of wetenschappelijk onderzoek om haar argumenten te ondersteunen. Ze geeft eigenlijk een soort van haar eigen mening, dus dat vind ik niet betrouwbaar.

Participant in the LongPauses Native group:

1. De articulatie, snelheid van spreken, struikelen over woorden of vloeiend sprekend.

2. Jong, spontaan, enthousiast, maar wellicht geen expert op de besproken gebieden.
3. De formaliteit van de situatie die werd geschetst gaf me een vooroordeel over de expertise, voor een klas staan vs met een partner bespreken.
4. Niet echt.
5. Redelijk. Ik zou deze persoon als betrouwbaar zien in gesprek als vrienden maar misschien niet als expert in de besproken onderwerpen.

Participant in the NoPauses Non-native group:

1. Haar argumentatie
2. Een vrouw waarvan Nederlands niet haar moedertaal is.
3. Naar mijn idee hadden ze geen invloed.
4. Niet per se.
5. De spreker noemt veel argumenten in het fragment, wat betrouwbaarheid uitstraalt. Ik zou echter graag willen weten op welke bron de argumentatie gebaseerd is voordat ik kan zeggen of haar argumenten ook kloppen. Het feit dat Nederlands niet haar moedertaal heeft geen invloed op de betrouwbaarheid naar mijn mening.

Participant in the LongPauses Non-native group:

1. Ze wist het niet met zekerheid maar dacht dat het zo is
2. wel slim en iemand die niet zomaar iets aanneemt
3. nee
4. dat ze niet Nederland is
5. niet helemaal omdat ze zegt dat ze denkt het te weten