Summer 7-17-2015

Interviewer Voice Characteristics and Data Quality

Nuttirudee Charoenruk
University of Nebraska-Lincoln, ncharoenruk2@unl.edu

INTERVIEWER VOICE CHARACTERISTICS AND DATA QUALITY

by

Nuttirudee Charoenruk

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Survey Research & Methodology

Under the Supervision of Professor Kristen M. Olson

Lincoln, Nebraska

July, 2015

INTERVIEWER VOICE CHARACTERISTICS AND DATA QUALITY

Nuttirudee Charoenruk, Ph.D.

University of Nebraska, 2015

Adviser: Kristen M. Olson

As an aural mode, interviewer voices play an important part in telephone surveys. Telephone interviewers are typically instructed to read questions with a proper phrasing and inflection and to read questions at a speech rate of 2 words per second (wps). However, there is no study that examines whether these interviewer voices affect data quality. In this dissertation, I examine how interviewer voice characteristics are associated with data quality in socially desirable, undesirable, and complex questions.

Data for this study come from the Work and Leisure Today Survey (NSF SES-1132015). I examined the first turn that interviewers read a survey question (n=4,689). Pitch, intonation, speech rate, and disfluencies are both objectively measured by the Praat program and subjectively evaluated by coders. In addition, coders evaluated five interviewer personality traits (expertise, trustworthiness, reliability, confidence, and easiness to understand) from interviewer voices. I examined four sets of data quality indicators including problematic respondent behaviors, item nonresponse, the directional hypothesis of "more/less is better," and rounding.

Analyses showed both objective and subjective voice characteristics affect data quality; however, the effects are inconsistent across data quality indicators.

Interviewers obtain better data quality when they read questions with moderate intonation and disfluencies. The voice characteristic with the largest effect on data quality is speech rate. Interviewers obtain better data quality when they read neutral questions with 2 wps, but read socially undesirable questions more quickly. Results suggest that interviewers should be trained to read questions with moderate intonation and disfluencies. In addition, to maximize data quality, interviewers should read neutral questions with the recommended speech rate of 2 wps, but read socially undesirable questions more quickly.

I also found that listeners can perceive interviewers' personality traits (credibility and easiness to understand) from interviewers' voices, and these personality traits tend to affect data quality. Credibility affects data quality in sensitive questions while easiness to understand affects data quality in complex questions. In addition, I found credibility mediates the effect of speech rate on respondents interrupting questions with answers. Moreover, easiness to understand mediates the effects of intonation and speech rate on item nonresponse rates.

To my father, Adm. Chainarong Charoenruk, my mother, Pol. Col. Wannee Charoenruk, and my brother, Lt. JG. Note Charoenruk

ACKNOWLEDGEMENTS

I must thank numerous people. First, I would like to gratefully acknowledge the guidance, support, and encouragement of my dissertation chair, Kristen Olson. Kristen is a wonderful advisor. She gave me tremendous advice and support. Kristen always believes in me and the importance of my dissertation. Thank you so much, Kristen. I appreciate for everything that you have done to me not only for my dissertation but also for all researches that I had done with you.

Additionally, I am grateful for advice and mentoring from Jolene Smyth, one of my dissertation committee members. Jolene always provides me feedback and advice in writing my dissertation as well as papers, research proposals, and other work with her. I also would like to recognize my other committee members Robert Belli and Christina Falci who provided suggestions and helpful criticism for improving the dissertation.

This work would not have been possible without the funding from the National Science Foundation for dissertation grant (NSF SES-1356985). Chulalongkorn University provided a support of my Ph.D. in Survey Research and Methodology Program (SRAM), University of Nebraska-Lincoln. Part of data in this study was funded by the National Science Foundation (NSF SES-1132015) awarded to SRAM faculty at the University of Nebraska-Lincoln.

I am indebted to my coders, Jacob Cox, Luke Biggs, Dina Morales, Jennifer Figueroa, Tom Blackwell, and Alex Goodman. Thank you for your hard work on coding voice files in this dissertation.

I sincerely thank Barb Rolfes and Renae Reis. They provide so much assistance for SRAM graduate students. My SRAM friends also deserve acknowledgement, especially Mathew Stange. We were working and studying for

comprehensive exam together. We knew that graduate life was tough and we always encourage each other. I appreciate all of his support and advice. In addition, I appreciate my other SRAM friends: Rebecca Powell, Jinyoung Lee, Ana Lucia Cordova Cazar, Beth Cochran, Quan Zhou, and Kay Ricci. Thank you for making my Ph. D. life enjoyable.

Thanks to my friends at Thailand for amazing encouragement, especially Kanyawadee Raksrisak, Nantaporn Techawongchimplee, Jiraphawee Wisitwaithayakun, Thanthip Pruangchana, and Priyapha Suphamahitorn.

Finally, I would like to make a special acknowledgement to my parents, Chainarong and Wannee, and my brother, Note, for unconditional love, support, understanding, and believing in me. I love you.

Thank you all!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Telephone surveys frequently contain socially desirable, socially undesirable, and complex questions that tend to produce problems for respondents (Tourangeau & Yan 2007; Fowler 1992). Previous research has examined how interviewers in telephone surveys affect data quality in these types of questions (Groves, et al. 2009; Tourangeau & Yan 2007), primarily through examining interviewer behaviors (Dykema, et al. 1997) and interviewer's demographic characteristics (Krysan & Couper 2003; Kane & Macaulay 1993; Grove & Fultz 1985). Questions remain, however, about whether an interviewer's voice characteristics also affect data quality. Interviewer voice characteristics may influence data quality by affecting the respondents' perception of an interviewer's personality traits such as credibility (Miller, et al. 1976), and thus can affect their responses (Blair 1977; Barath & Cannell 1976). In this dissertation, I examine whether interviewer voice characteristics affect data quality in socially desirable, socially undesirable, and complex questions.

This dissertation contains three main objectives (Figure 1.1). First, I evaluate whether a listener's subjective perceptions of an interviewer's voice characteristics (rated speech rate, pitch, intonation, and disfluency) and their assessment of five interviewer personality traits[1] (confidence, easiness to understand, reliability, trustworthiness, and expertise) are associated with the interviewer's objective acoustic voice characteristics including speech rate, pitch, intonation, and disfluency. The second objective is to examine whether objective acoustic voice characteristics of telephone survey

[1] Both paralinguistic and survey research study use various terms to define personality traits. These include personal characteristics, personality assessment, and personal attributions. For the purpose of this dissertation, I will use the term "personality traits."

interviewers are associated with data quality in socially desirable, socially undesirable, and complex questions. In the third objective, I investigate how subjective voice characteristics and perceptions of interviewer personality traits affect data quality and whether subjective perceptions of an interviewer's personality traits mediate the relationship between objective acoustic voice characteristics and data quality.

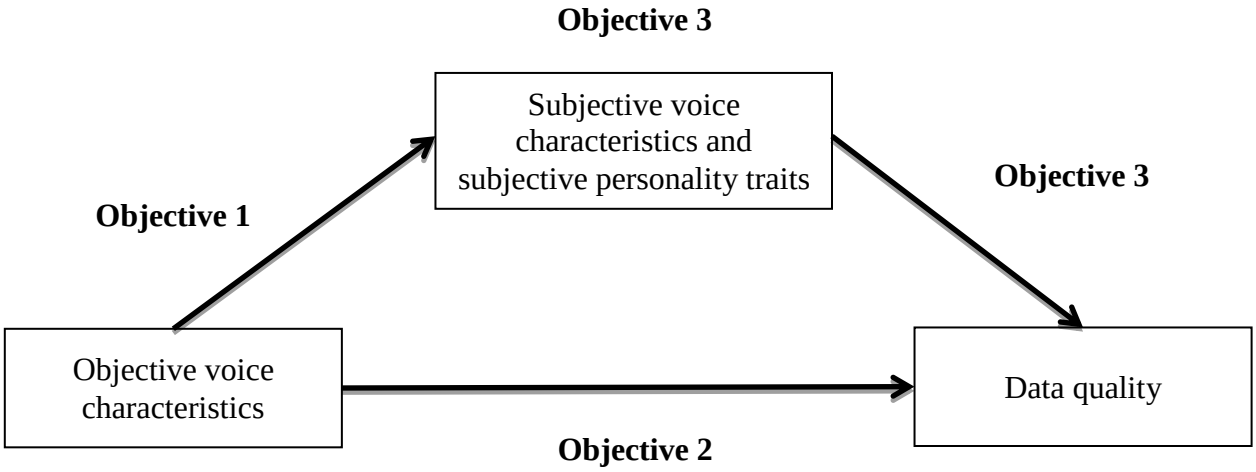


Figure 1.1 Model for the three objectives of this dissertation

It is important to understand how interviewer voices affect data quality in telephone interviews because these voices are the primary means of communication to respondents, as this mode lacks the ability to use show cards or self-administered questionnaires. Interviewer voices have been shown to play an important role in recruitment of sampled persons (e.g., Benki, et al. 2011; Van der Vaart, et al. 2005; Oksenberg, et al. 1986), but the effects on how respondents answer questions are virtually unexplored. If voice characteristics of interviewers affect data quality, we will be able to select or train interviewers to modify some of their vocal characteristics with the goal of maximizing data quality. Moreover, results from this dissertation will be useful for selecting interviewers based on voice characteristics for audio computer-assisted self-

interviewing (ACASI), telephone audio-CASI (T-ACASI), and interactive voice response (IVR) systems with the goal of minimizing measurement error.

1.1 Background and significance

1.1.1 Data quality in telephone surveys: interviewers, questions, respondents, and indicators

Telephone surveys are an important method for collecting data about populations and have been used by researchers in sociology, economics, political science, and public health (Groves, et al. 2009). For example, each month, the Survey of Consumers measures changes in U.S. consumer attitudes and expectations. The Behavioral Risk Factor Surveillance System interviews over 500,000 U.S. residents via telephone about their health-related risk behaviors and chronic health conditions. Although there has been growth in the use of other modes such as mail and web (e.g., Dillman, et al. 2014), telephone surveys remain frequently used either as the primary recruitment and data collection mode or as a lower cost mode for follow-up rounds of longitudinal surveys to collect timely data from a representative sample (Steeh 2008). Thus, it is important to understand factors related to data quality in this mode. Interviewers, questions, and respondents in telephone surveys can affect survey responses, which can in turn impact data quality (Biemer & Lyberg 2003; Groves 1989).

Telephone interviewers and data quality

Data quality consists of several dimensions such as accuracy, timeliness, richness of detail, and accessibility (Biemer & Lyberg 2003). In this dissertation, I focus only on the data accuracy dimension. In particular, I focus on measurement error, which is one component of data accuracy (Groves 1989). Measurement error, i.e. the difference

between the true value of survey estimates and estimates from survey responses, can arise from interviewers (Groves, et al. 2009; Groves 1989). For example, interviewers may falsify data or deviate from their standardized behaviors, which may in turn influence survey responses (Biemer & Lyberg 2003; Fowler & Mangione 1990).

Telephone interviewers can increase the variance of survey estimates and lead to systematic biases (Biemer & Lyberg 2003; Kane & Macaulay 1993; Groves 1989; Groves & Magilavy 1986). Interviewers act as clustering agents, increasing the variance of a mean by a factor of $1+(b-1)\rho$, where b is the average interviewer workload and ρ is the intracluster correlation coefficient due to interviewers. In a typical telephone survey, workloads of b=30 and ρ=0.03 inflate the variance of the mean by 1.87, substantially widening confidence intervals. Demographic characteristics of interviewers such as gender or race and interviewer behaviors such as probing have been shown to affect data quality (e.g., Schaeffer & Dykema 2011; Groves & Fultz 1985). Yet few investigations have examined characteristics of an interviewer's voice as a source of measurement error.

Questions and data quality

Although there are many types of questions asked in social surveys, socially desirable, undesirable and complex questions are consistently prone to measurement errors and problems for interviewers and respondents. Additionally, there are changes in a speaker's vocal patterns for these types of questions (Bachorowski 1999). Thus, these three question types are good candidates for evaluating the association between vocal characteristics and data quality.

Socially desirable and socially undesirable questions are often asked in telephone surveys and are particularly prone to measurement error (Groves, et al. 2009;

Tourangeau, et al. 2000). It is well established that respondents systematically misreport sensitive behaviors, especially in interviewer-administered surveys (Tourangeau & Yan 2009) and that demographic characteristics (e.g., gender and race) of interviewers affect reports to sensitive questions (e.g., Axinn 1991). In general, respondents tend to edit or censor a truthful response to be more in alignment with social norms (Tourangeau, et al. 2000). That is, respondents are more likely to overreport desirable behaviors such as voting and to underreport undesirable behaviors such as illicit drug use or heavy drinking, resulting in measurement error (Kreuter, et al. 2008). In addition to measurement error, nonignorable item missingness may occur for surveys with sensitive questions (Tourangeau & Yan 2007) because people are less likely to respond to socially undesirable items (Tourangeau, et al. 2010) and more likely to respond to socially desirable items (Sakshaug, et al. 2010).

Complex questions are also often asked in telephone surveys. Although there are many different types of questions, complex questions consistently have more problematic interviewer and respondent behaviors during an interview and these behaviors are often associated with decreases in data quality (Fowler 2011; Schaeffer & Dykema 2011; Schnell & Kreuter 2005). Features of complex questions include long questions, syntactical complexity, instructions, introductions, and ambiguous words, or questions that ask for retrospective reports, frequency or other quantitative reports (Holbrook, et al. 2006; Knauper, et al. 1997; Fowler 1992). For example, questions about a respondent's income are complex because respondents may not retrieve all the relevant information from memory or may find it difficult to add all of their income resources together (Jans 2010). Previous research has found that difficult or complex questions create problems

for the cognitive response process and encourage respondents to use a satisficing response strategy (Krosnick 1991) where they skip or truncate the cognitive response process and provide either incomplete or biased reports, or "don't know" answers. This kind of breakdown of the cognitive response process leads to a decrease in accuracy and completeness of reports (Knauper, et al. 1997).

Respondents and data quality

Answering telephone survey questions require respondents to have memory and verbal skills (Dillman, et al. 2014). Respondents have to consider the question and response options while they come up with their answers. To answer survey questions, respondents go through four cognitive response process steps: 1) comprehend questions, 2) retrieve relevant information, 3) make a judgment, and 4) map answers with response options (Tourangeau, et al. 2000). Respondents who have lower cognitive ability are more likely to experience difficulty holding the questions and response options in working memory, and thus have trouble in processing the four components of response process (Knauper, et al. 1997). As such, data quality reduces as respondent's cognitive ability declines. Previous research used respondent's age and education as proxy variables of cognitive ability (Knauper 1999; Narayan & Krosnick 1996; Groves 1989). Thus, this study controls for respondent age and education.

Respondent's age. Older people have less ability to store and process information in working memory than younger people (Knauper 1999). Older respondents tend to experience failures when retrieving information from their memory (Groves 1989). Older respondents also have greater mapping difficulties, provides higher rates of inadequate responses, and are more likely to produce response order effects than younger

respondents (Holbrook, et al. 2006; Knauper 1999; Belli, et al. 1999). Moreover, instead of being task-oriented, older respondents are more likely to build rapport with interviewers than younger respondents (Groves 1989), although there are mixed results for the association between interviewer-respondent rapport and data quality (Bilgen & Belli 2010; Dijkstra 1987).

Respondent's education. Respondents with lower levels of education are more likely to have comprehension problems and more likely to be influenced by irrelevant cues (Narayan & Krosnick 1996; Groves 1989; Schuman & Presser, 1981). Thus, data quality is lower for less educated respondents. For example, Narayan and Krosnick (1996) found higher response effects such as acquiescence and no-opinion effect in respondents with less education compared to highly educated respondents.

Data quality indicators

There are many ways to evaluate data quality. Ideally, a "gold standard" is available on the frame for all questions in the survey. However, in random digit dial telephone surveys, gold standard data do not usually exist because of a lack of frame data. Instead, many other data quality indicators are used.

One measure of data quality comes from the interaction between the interviewer and respondent during the interview. Respondent behaviors can manifest indicating that respondents had cognitive problems when answering a survey question, thus potentially decreasing data quality (Schaeffer & Dykema 2011, Fowler 2011, Dykema, et al. 1997, Fowler & Cannell 1996, Fowler & Mangione 1990). For example, when a question contains an unclear term, respondents may request clarification, express uncertainty about a question, or give an answer that does not meet the question's objective (Fowler 2011,

Fowler & Cannell 1996, Fowler 1992). Moreover, respondents who give a qualified answer indicate uncertainty about their final answer (Dykema, et al. 1997). In addition, respondents who interrupt questions with answers will not hear all information that interviewers would like to ask, thus, they may give inaccurate responses or give an answer that does not meet the question's objective (Fowler & Cannell 1996, Fowler 1992, Fowler & Mangione 1990).

Item nonresponse and rounding (e.g., reporting units in multiples of 5 and 10) are also used to evaluate data quality (Tourangeau, et al. 2000). Moreover, two assumptions are normally made in socially desirable and undesirable items- lower reports of socially desirable behaviors and higher reports of socially undesirable behaviors are more accurate (Kreuter, et al. 2008). These two assumptions arise because social desirability concerns lead respondents to overreport socially desirable behaviors and underreport socially undesirable behaviors.

In this dissertation, I use the item nonresponse rate, rounding, the directional hypotheses of "more is better" and "less is better," and respondent behaviors (interrupting questions with an answer, expressing uncertainty about a question, requesting clarification, giving qualified answers, and giving a response that does not meet the question's objective) as measurement error outcomes.

1.1.2 Interviewer voice characteristics

Telephone surveys are an aural, rather than a visual, mode (Conrad, et al. 2008; Groves 1990). Respondents hear what an interviewer asks and the interviewer hears what respondents answer (Dillman, et al. 1996). As such, the interviewer's voice is an important part of the mode. A speaker's voice can convey much more information than

simply the meaning of words or sentences themselves (Groves, et al. 2008). A speaker's voice characteristics can provide information about a speaker's personality traits such as credibility and confidence (Broome 2012; Van der Vaart, et al. 2005; Smith & Shaffer 1995; Oksenberg & Cannell 1988; Apple, et al. 1979; Miller, et al. 1976; Pearce & Conking 1971) and speaker's demographic characteristics such as social class, age, and sex (Groves, et al. 2008; Oksenberg, et al. 1986; Smith 1979; Harms 1961).

Measures of interviewer voice characteristics

Interviewer voice characteristics have been previously measured either subjectively or objectively. For example, Oksenberg et al. (1986) used raters to evaluate an interviewer's voice characteristics such as rate of speaking, intonation, loudness, and flow of words. Previous research has also used raters to evaluate an interviewer's personality traits such as being confident, professional, and pleasant to listen to (Broome 2012). In addition to being subjectively rated by judges, voice characteristics can also be objectively measured using computer software such as Praat. For example, Benki et al. (2011) used Praat to obtain the pitch of telephone interviewer voices in survey introductions to study unit nonresponse. In this dissertation, interviewers' voice characteristics are measured both subjectively and objectively.

Four voice characteristics, including rate of speech, pitch, intonation, and disfluency, are measured in this dissertation. First, *rate of speech* or verbal rate is the number of words per unit time for the period of speech (Webb 1969). The average adult's speech rate in English is between 2.5 and 3.2 words per second (wps) (Tauroza & Allizon 1990). Second, *pitch* is the fundamental frequency, i.e. the rate of laryngeal vibration (Broome 2012; Benki, et al. 2011). On average, voice pitch for men is 120 Hertz (Hz)

and for women is 210 Hz (Traunmuller & Eriksson 1993). Third, *intonation* is variation in pitch (Oksenberg & Cannell 1988). It refers to a pattern of rises and falls in pitch and the patterns of stress in language (Kent & Read 2002). Finally, *disfluency* is the parts of speech that are not words such as fillers (like ums and uhs) and pauses (Conrad, et al. 2008). Disfluences make up about 6% of speech (Bortfeld, et al. 2001).

In this dissertation, I hypothesize that an interviewer's objective voice characteristics such as pitch and intonation affect a listener's subjective rating of these voice characteristics (e.g., ratings of pitch and intonation) and their subjective perception of other interviewer's personality traits (e.g., easiness to understand; objective 1). Moreover, I hypothesize that objective voice characteristics directly affect data quality (objective 2) and that subjective perceptions of an interviewer's personality traits mediate the relationship between objective voice characteristics and data quality (objective 3). Because I hypothesize that objective acoustic voice characteristics may affect data quality to the extent that they are perceived by outside listeners (objective 3 in this study), objective 1 aims to establish how these objective characteristics are perceived by listeners.

1.1.3 Objective 1: Subjective and objective voice characteristics

Previous research has studied the association between subjective and objective voice characteristics. Objective measurements of pitch, speech rate, and intonation are strongly correlated with their subjectively rated counterparts (Van der Vaart, et al. 2005; Oksenberg & Cannell 1988), and objective and subjective evaluations of the number of pauses are moderately correlated (Oksenberg & Cannell 1988). I expect to observe similar results as previous research. That is, I expect that *listeners will perceive high*

(low) pitched voices as having a high (low) pitch, that voices with a faster (slow) speech rate will be rated as speaking quickly (slowly), that voices with a high variation in pitched voice (low) will be rated as speaking with varied (flat) intonation, and that voices with high (low) disfluency will be perceived as speaking with high (low) disfluency.

More importantly, I hypothesize that objectively measured voice characteristics may affect a listener's perception of an interviewer's personality traits (confidence, easiness to understand, reliability, trustworthiness, and expertise). Previous research has found a relationship between objective voice characteristics and subjective evaluations of that speaker's personality traits.

Pitch. High-pitched voices are perceived as less truthful, less reliable, less trustworthy, less easy to understand, and more confident than lower-pitched voices (Tigue, et al. 2012; Dey, et al 2006; Van der Vaart, et al. 2005; Apple, et al. 1979; Scherer, et al. 1973). In addition, previous research found associations between pitch and perceptions of professionalism but the direction of the association is mixed (Broome 2012; Van der Vaart, et al. 2005). Therefore, I expect *listeners to perceive interviewers who read a question with higher pitched voices as less reliable, less trustworthy, less easy to understand, and more confident than those who read a question with lower pitched voices*. In addition, because of mixed findings on the direction of the association between pitch and a listener's perception of expertise, I expect that *the listeners can judge whether an interviewer is an expert based on voice pitch, but that the direction cannot be hypothesized*.

Intonation. Voices with varied intonation are perceived as less reliable, less trustworthy, more confident, easier to understand, and more professional than voices with

less variation (Natsumi 2013, Van der Vaart, et al. 2005; Oksenberg, et al. 1986; Brooke & Ng 1986; Scherer, et al. 1973). As such, I expect *listeners to perceive voice with varied intonation as less reliable, less trustworthy, more confident, easier to understand, and more expert than voice with flat intonation*.

Rate of speaking. Rapid speech is perceived as more credible, more trustworthy, more professional, and expressing more confidence than slow speech (Broome, 2012; Smith & Shaffer 1995; Oksenberg & Cannell 1988; Apple, et al. 1979; Miller, et al. 1976; Scherer, et al. 1973; Pearce & Conking 1971), but also reduces comprehension and may be hard to understand (Miller, et al. 1976). As such, I expect *listeners to perceive interviewers who read questions at a faster rate of speech as more trustworthy, more reliable, more expert, and more confident, but less understandable than those who speak more slowly*.

Disfluencies. Highly disfluent speakers are judged as less confident, less credible, less trustworthy, and less expert than more fluent speakers (Conrad, et al. 2008; Castro & de Moraes 2008; Ehlen 2007; Bortfeld, et al. 2001; Ketrow 1990; Oksenberg & Cannell 1988; Apple, et al. 1979; Miller et al. 1976; Miller & Hewgill 1964). In addition, perceived easiness to understand has been found to be negatively associated with the use of fillers (Pytko & Reese 2013). Thus, I expect *listeners to perceive interviewers who read a question with more disfluencies as less confident, less trustworthy, less reliable, less expert, and less easy to understand than those who read questions with fewer disfluencies*.

As mentioned earlier, examining how respondents perceive interviewer personality traits from interviewer voice is important because perception of an

interviewer's personality traits may mediate the effects of objective interviewer voice characteristics on data quality (objective 3 in this dissertation). Moreover, for socially desirable and undesirable questions, it has been assumed that a lack of visual presence of an interviewer in a telephone survey increases anonymity for respondents (Jans 2010). However, if respondents can perceive an interviewer's personality traits based on their voices, telephone interviews may not provide as much anonymity as expected for socially desirable and socially undesirable questions (Jans 2010).

1.1.4 Objective 2: Objective voice characteristics and data quality

Previous research has examined how interviewer behaviors in telephone interviews affect data quality (e.g., Fowler 2011; Dykema, et al. 1997). However, there is surprisingly little relevant research on the relationship between voice characteristics and data quality. Objective measures of interviewer voice characteristics have been used to examine the effect of voice characteristics on unit nonresponse (e.g., Benki, et al. 2011; Van der Vaart, et al. 2005). But, to my knowledge, only one study (Jans 2010) has examined the effect of objectively measured *respondent* voice characteristics on data quality, finding weak associations, and no studies have examined the relationship between objectively measured *interviewer* voice characteristics and data quality. Hypotheses for examining the effects of objective interviewer voice on data quality are discussed below.

Pitch. A high level of emotional arousal (e.g. fear and anxiety) is associated with an increase in voice pitch (Bachorowski 1999). Thus, interviewers who ask questions with higher pitched voices (accounting for gender) may be perceived as asking questions that are more sensitive than interviewers who ask questions with lower pitched voices. As

such, I expect *more socially desirable reports among interviewers asking socially desirable and undesirable questions with higher pitched voices, compared to those asking questions with lower pitched voices.* In addition, respondents tend to interrupt questions with answers, give a qualified answer, express uncertainty about a question, and give an answer that does not meet the question's objective in sensitive questions (Jans 2010; De la Puente & McKay 1995). As such, because questions asked with higher pitched voices are perceived as being more sensitive, I expect *a higher proportion of problematic respondent behaviors among interviewers asking socially desirable and undesirable questions with higher pitched voices, compared to those asking the questions with lower pitched voices*. On the other hand, interviewers who ask questions with higher pitched voices may be perceived as more attractive and are more likely to receive responses from respondents (Ketrow 1990, Oksenberg, et al. 1986). As such, I expect *lower item nonresponse rates among interviewers asking questions with higher pitched voices, compared to those asking questions with lower pitched voices.*

Intonation. Listeners tend to perceive voices with lower pitch variability as less credible (Addington 1968). Respondents are more likely to provide better data quality to interviewers whom they perceive as more credible (Groves 1990). As such, I expect that *interviewers who read questions with higher pitch variation (more intonation) will obtain better data quality (lower item nonresponse, rounding, and responses prone to socially desirable bias) than those who read questions with lower pitch variation*. In addition, respondents are more likely to give an answer that does not meet the question's objective and give qualified answers in sensitive questions (Jans 2010; De la Puente & McKay 1995). Respondents are more likely to trust interviewers and are more willing to give

better quality answers to more credible interviewers (Groves 1990; Ohanian 1990; Hovland, et al. 1953). As such, I expect that *interviewers who are perceived as more credible (i.e. those who read a question with higher intonation) will obtain fewer problematic respondent behaviors compared to those who are perceived as less credible (i.e. those who read a question with lower intonation).*

Rate of speaking. I have two competing hypotheses. Previous paralinguistic studies found that rapid speech is perceived as more credible and more persuasive than slow speech (Smith & Shaffer 1995; Oksenberg & Cannell 1988; Apple, et al. 1979; Miller, et al. 1976; Pearce & Conking 1971). Thus, respondents may perceive interviewers who read questions more quickly as more credible than those who ask questions more slowly. As such, from paralinguistic research, I expect *interviewers who read questions with a faster pace to be more likely to obtain better data quality (lower item nonresponse, rounding, responses prone to socially desirable bias, and proportion of problematic respondent behaviors) compared to those who read questions with a slower pace.*

On the other hand, survey practice (e.g., Fowler & Mangione 1990; Cannell, et al. 1981) suggests that interviews should be conducted at a slow pace. Fast interviews lead respondents to perceive that quick answers are acceptable, and not to take the time to give accurate and complete answers, thus, decreasing data quality (Fowler & Mangione 1990; Fowler 1966). As such, from survey research, I expect *interviewers who ask questions with faster rates of speaking will obtain lower data quality (higher item nonresponse rates, rounding, and socially (un)desirable answers) than those who ask questions at a slower pace.* Moreover, respondents may not understand a question asked more quickly,

and thus request clarification about a question, express uncertainty about a question, and give an answer that does not meet question objective (Ongena & Dijkstra 2007; Fowler & Cannell 1996). As such, I expect that *interviewers who ask questions more quickly will obtain higher rates of problematic respondent behaviors than those who ask questions more slowly.*

Disfluencies. Similar to speech rate, I have two competing hypotheses for disfluencies. Paralinguistic research (e.g., Ketrow 1990; Miller, et al. 1976) shows that listeners tend to perceive speech with higher rates of disfluencies as less credible compared to speech with lower rates of disfluencies, thus leading to lower perceptions of credibility for interviewers with higher levels of disfluencies. As such, I expect *higher item nonresponse rates, rounding, socially (un)desirable answers, and proportions of problematic respondent behaviors among questions read with more disfluencies than those read with fewer disfluencies.*

Survey research has shown that disfluencies tend to affect respondents at the comprehension stage. Disfluencies in speech have a "disfluency advantage" allowing respondents to have more time to think about their responses, and thus increase data quality (Brennan & Schober 2001; Bradburn, et al. 1987). Moreover, disfluency rates increase when discussing an unfamiliar domain and talking in long sentences (Bortfeld, et al. 2001), indicating that speakers have difficulties deciding what to say and how to say it (Kidd, et al. 2011; Clark 2002). These fillers can alert listeners that upcoming speech will be complex so that they should pay attention to what speakers will say (Clark & Tree 2002). As such, the competing hypothesis is that I expect *lower item nonresponse rates, rounding, and socially (un)desirable answers among questions read with more*

disfluencies than those read with fewer disfluencies. In addition, because of the disfluency advantage that allows respondents to have more time to interpret questions and retrieve information, respondents are less likely to give an answer that does not meet the question's objectives or give a qualified answer (Bradburn, et al. 1987). As such, I expect *fewer problematic respondent behaviors among questions read with more disfluencies than those read with fewer disfluencies.*

Understanding the effects of voice characteristics on data quality is important because many survey organizations recommend that interviewers read questions at a pace of two words per second and with proper phrasing and inflection (Guenzel, et al. 1983). However, there is no empirical research that examines whether these voice characteristics have an effect on data quality. The recommended speech rate of 2 wps is slower than the rate of speech in ordinary conversation, and this may affect data quality.

1.1.5 Objective 3: Subjective voice characteristics and data quality

Little research has examined the effects of subjectively rated voice characteristics and subjectively evaluated interviewer personality traits on data quality. Previous research on the effects of subjectively rated voice characteristics on unit nonresponse finds that interviewers who are rated as having higher pitched voices, faster rates of speaking, greater loudness, falling intonation, and clearer and more distinct pronunciation have higher response rates (Groves, et al. 2008; Oksenberg & Cannell 1988; Oksenberg, et al. 1986). These interviewers are also judged as being more pleasant, cheerful, friendly, enthusiastic, interested, intelligent, educated, professional, and confident (Oksenberg & Cannell 1988; Oksenberg, et al. 1986).

I expect that subjectively measured voice characteristics will affect data quality in ways consistent with the above discussion for objectively measured voice characteristics, but that the effects may be stronger (Van der Vaart, et al. 2005). In addition, I hypothesize that respondent's perceptions of an interviewer's personality traits affect data quality. Respondents are more likely to provide better data quality to interviewers perceived as more credible (Groves 1990). Previous research found reliability, trustworthiness, expertise, and confidence all contribute to an underlying "credibility" construct (Sah, et al. 2013; Ohanian 1990; Hovland, et al. 1953). As such, I expected that *interviewers who are perceived as being more confident, more reliable, more trustworthy, and having more expertise (i.e. being more credible) will receive better data quality (less rounded answers, fewer answers prone to socially desirable bias, fewer item nonresponse rates, and fewer problematic respondent behaviors) than those who are perceived as being less confident, less reliable, less trustworthy, and having less expertise (i.e. being less credible).*

In addition to credibility, respondents are more likely to give better quality answers when they more easily comprehend questions read by interviewers (Japec 2008). Easiness to understand may play an especially important role for complex questions because understandability can affect a listener's comprehension of these questions (Miller et al. 1976). As such, *I expect higher data quality among interviewers whose voices are perceived as being easier to understand than those whose voices are perceived as being less easy to understand*.

This examination is important because the subjective perceptions of voice characteristics may mediate the effect of objective voice characteristics on data quality.

For example, Barath and Cannell (1976) found that interviewers whose voices were rated as rising at the end of questions received higher rates of acquiescent reports of sensitive health conditions than those whose voices were rated as falling. Blair (1977), looking at nonsensitive questions, found the opposite. One reason that Barath and Cannell (1976) and Blair (1977) found opposite effects of interviewer voice rising may be due to perceptions of reliability or trustworthiness of the interviewer. As such, in this dissertation, I expect that *interviewers' personality traits will mediate the effect of interviewer voice characteristics on data quality*.

Paralinguistic research usually examines voice characteristics by the speaker's gender since voice characteristics differ for men and women (Bortfeld, et al. 2001). Females have higher pitched voices, greater variability of pitch, somewhat slower speech rates, and use fewer fillers relative to males (Yuan, et al. 2006; Bortfeld, et al. 2001; Kent & Read 2002). Interviewers are more likely to be perceived positively if they follow their expected voice pattern, and thus obtain better quality answers (Benki, et al. 2011; Rubin 1992). For example, lower pitched voices are judged as more attractive in male speakers while higher pitched voices are judged as more attractive in female speakers (Benki, et al. 2011; Oksenberg, et al. 1986). Listeners are more persuaded by and compliant to speakers whose voices are perceived as more attractive (Ketrow 1990). As such, I expect that female interviewers with high pitched voices will have lower item nonresponse rates relative to those with low pitched voices, but the opposite direction is expected for male interviewers.

Besides gender, interviewer experience may affect the rate of speaking. More experienced interviewers tend to conduct interviews faster than inexperienced

interviewers (Olson & Bilgen 2011). Rapid speech may be hard to understand (Miller, et al. 1976). Thus, I expect that more experienced interviewers will have higher item nonresponse rates than inexperienced interviewers. As such, in this study, I also examine interaction effects between interviewer voice characteristics and interviewer sex and experience. In addition, as mentioned earlier, a speaker's vocal characteristics are more likely to change in socially desirable, undesirable, and complex questions. Thus, I also examine whether question characteristic moderate the effects of interviewer voice characteristics on respondent's perceptions of an interviewer's personality traits and data quality.

1.2 Research design and methods

Data

Data for this study come from the Work and Leisure Today Survey conducted in July and August 2013 by AbtSRBI (NSF SES-1132015). It is a landline RDD CATI survey with 450 completed interviews collected by 20 interviewers. Interviewers completed an average of 22.5 interviews. To increase the stability of the analyses, interviewers who conducted fewer than 10 interviews are eliminated from the study (Olson & Peytchev 2007). As such, I analyze 432 interviews conducted by 19 interviewers (9 female and 10 male interviewers) in this dissertation. The interviews were transcribed and behavior coded as part of an ongoing NSF grant (NSF SES-1132015).

Questions

The questionnaire has 54 questions. For this dissertation, I consider 24 candidate questions including socially desirable, socially undesirable, complex, and neutral (not complex and not socially desirable/undesirable) questions. I have selected twelve (three

of each of the four categories) of these 24 questions based on the criterion that questions have to contain both item nonresponse and item response and that there is sufficient variability in responses. The wording of the selected twelve questions appears in Appendix A. The topics of these questions include employment status and volunteer work, activities for leisure such as using internet and exercise, and substance use such as drinking alcohol.

Measures of voice characteristics and interviewer's personality traits

In this dissertation, I examine the first turn that an interviewer read a survey question. I measure voice characteristics using objective acoustic measures from a computer program and subjective rating measures from judges' ratings of the voice characteristics. To obtain the objective acoustic measures, I use the Praat computer software program to measure an interviewer's pitch, intonation, speech rate, and disfluency (http://www.fon.hum.uva.nl/praat/) from 4,689 voice files. Pitch is measured by the mean pitch over the question reading. Intonation is measured by the standard deviation of pitch over the question reading. Speech rate is the number of words per second (from the beginning of the question to the end of the question). Lastly, disfluencies is measured by number of fillers. Number of words and number of fillers are coded from interview transcripts.

Besides objective measures, the 4,689 voice files are subjectively evaluated for measures of interviewer voice characteristics and interviewers' personality traits. For the subjective measures of voice characteristics, raters are asked to evaluate voice characteristics on seven point scales (Groves, et al. 2008; Oksenberg, et al., 1986). Specifically, raters evaluate the interviewer's average pitch on a range from low (1) to

high (7), the interviewer's intonation (pitch variation) from small range (1) to large range (7), the interviewer's speech rate from slow (1) to fast (7), and the interviewer's disfluency from low (1) to high (7). Five interviewer personality traits - including confidence, easiness to understand, reliability, trustworthiness, and expertise - also are rated from low (1) to high (7). First, confidence is the extent to which the interviewer is self-assured and conducts the interview with poise. Second, ease of understanding is the extent to which the interviewer's voice is easy to understand. Third, reliability is the extent to which interviewers say something that can be believed. Fourth, trustworthiness is the degree of confidence in an interviewer to ask a valid survey question and to keep respondents' answers confidential. Fifth, expertise is the extent to which an interviewer is good at her/his job in asking a survey question. Raters also determine whether they think that the interviewer is male or female from the interviewer's voice (See Appendix B for coder instruction).

Measures of data quality

I first evaluate item nonresponse rates for all questions. Second, I use the directional hypothesis of "more is better" for socially undesirable questions and "less is better" for socially desirable questions. Third, rounding– answering using prototypical responses – is used for complex questions and neutral questions.

In addition to these indicators of data quality, I use five respondent behaviors that previous research has found to be associated with data quality as additional dependent variables (Schaeffer & Dykema 2011; Dykema, et al. 1997). These include that the respondent 1) interrupts questions with an answer, 2) expresses uncertainty about a question, an answer or has difficulty answering, 3) requests clarification, 4) gives a

qualified answer, and 5) gives a response that does not meet the question's objective. The behavior coding was done at a turn level within a question-answer-feedback sequence (See Kirchner & Olson (2014) for more detail about behavior coding process). These five respondent behaviors are evaluated for all questions. A summary of data quality used in this study appears in Appendix C.

Data quality analysis

Each respondent is assigned to be interviewed by one interviewer, and interviewers obtain responses from multiple respondents. Thus, multi-level modeling of nested data is used for analysis (O'Muircheartaigh & Campanelli 1998; 1999). For each model, respondent age and education are control variables. These analyses are described in more detail in each subsequent chapter.

1.3 Outline of dissertation

The next three chapters correspond to each of the dissertation's objectives. In chapter 2, I examine the associations between objectively and subjectively measured voice characteristics (Objective 1). In Chapter 3, I examine the association between objective measures of interviewer voice characteristics and data quality (Objective 2). In Chapter 4, I examine whether the subjective measures of interviewer personality traits mediate the relationship between objective measures of interviewer voice characteristics and data quality (Objective 3). In Chapter 5, I summarize and discuss the findings from three chapters, provide significant implications from this dissertation, identify limitations in this research, and outline future research.

CHAPTER2: SUBJECTIVE AND OBJECTIVE VOICE CHARACTERISTICS

Introduction

As an aural mode, an interviewer's voice is an important part of telephone surveys. A speaker's voice can convey information about a speaker's personality traits (e.g., trustworthiness and confidence), demographic characteristics (e.g., gender and age), and emotional state (e.g., anger and happiness) (Broome 2012; Van der Vaart, et al. 2005; Kent & Read 2002; Smith & Shaffer 1995; Oksenberg & Cannell 1988; Apple, et al. 1979; Miller, et al. 1976; Pearce & Conking 1971). Consequently, respondents may be able to perceive personality traits of interviewers from their voices, and these perceptions may affect data quality.

In this dissertation, I hypothesize that objective voice characteristics that can be obtained from computer software may also be perceived by respondents and to the extent that they are perceived, may affect data quality (Chapter 4 in this dissertation). Thus, in this chapter, I examine whether listeners can actually perceive objective voice characteristics (Figure 2.1). Specifically, I evaluated how coders' perceptions of interviewers' voice characteristics (rated pitch, rated intonation, rated speech rate, and rated disfluencies) and their assessment of five personal traits of interviewers (expertise, confidence, reliability, trustworthiness, and easiness to understand) are associated with objective measures of interviewers' voices.

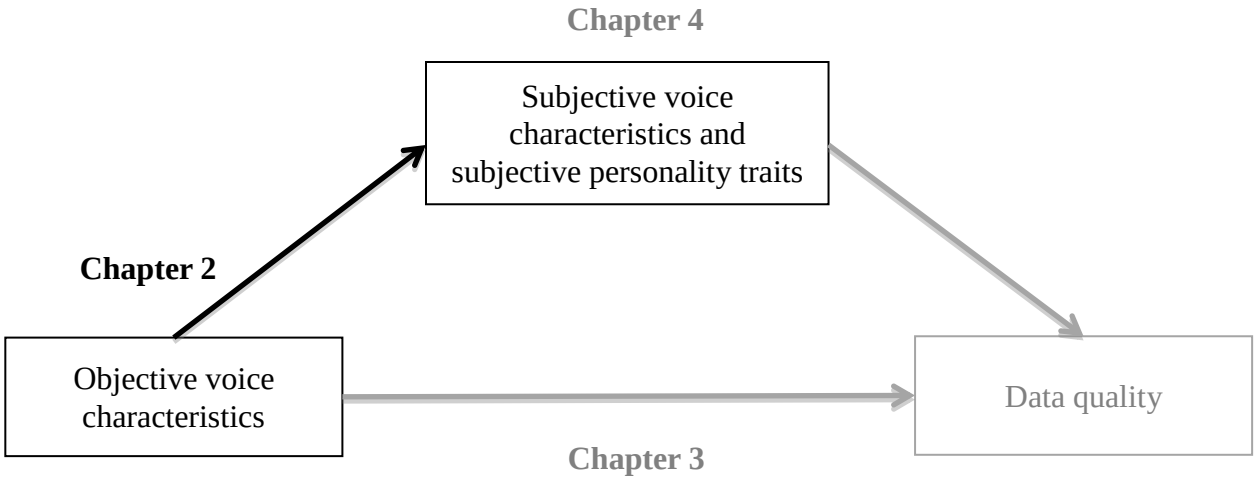


Figure 2.1 Conceptual model for the relationship between interviewer voice characteristics and data quality

2.1 Literature review

A speaker's voice can convey much more information than simply the meaning of words or sentences themselves (Groves, et al. 2008). For example, listeners tend to perceive persons who speak with high-pitched voices as less truthful and more nervous than people whose voices are low-pitched (Apple, et al. 1979). Because telephone interviewers have only an aural channel of communication, respondents may perceive that interviewers have certain personality traits from their voices as they ask survey questions. In this research, I examine whether listeners perceive personality traits of interviewers (expertise, confidence, reliability, trustworthiness, and easiness to understand) from interviewer voice characteristics, including pitch, intonation, rate of speaking, and disfluencies.

Pitch is the fundamental frequency, i.e., a rate of vocal cord vibration that looks like a waveform (Broome 2012; Benki, et al. 2011; Groves, et al. 2008; Johnson 2003). The frequency, in hertz (Hz), refers to the number of repeated waves of sound per second (Johnson 2003). For example, 100 Hz means that the waveform has 100 waves per

second. Range of pitch differs by gender — women's voices have a higher range of pitch than men's voices (Kent & Read 2002). The typical pitch is 120 Hz for men and 210 Hz for women (Traunmuller & Eriksson 1993).

Intonation is variation in pitch (Oksenberg & Cannell 1988). It refers to a pattern of rising and falling pitch and to the patterns of stress in a language (Kent & Read 2002). Intonation can convey the meaning of sentences (Mackey 1987). Although speakers' voices tend to drop at the end of declarative statements, they tend to rise in asking questions (Blair 1977; Barath & Cannell 1976). Proper phrasing and inflection are recommended when asking telephone interview questions (Cannell, et al. 1981).

Rate of speech or verbal rate is the number of words per unit of time over the duration of speech (Webb 1969). Males speak, on average, slightly faster than females (Yuan, et al. 2006). In ordinary conversation, adults speak at a rate of speech between 2.5 and 3.2 words per second (wps) (Tauroza & Allizon 1990). However, in telephone surveys, interviewers are typically instructed to ask questions at an average pace of two wps (Guenzel, et al. 1983; Cannell, et al. 1981). This recommended speech rate is slower than the rate of speech in ordinary conversation, and this may affect listeners' perceptions.

Disfluency is the parts of speech that are not words, such as fillers (like ums and uhs) and pauses (Conrad, et al. 2008). Disfluencies typically make up about 6% of speech (Bortfeld, et al. 2001). Disfluencies can be a cue of comprehension difficulty (Conrad, et al. 2008; Schober & Bloom 2004). Pauses in answering questions can indicate that respondents are having trouble answering them (Schober & Bloom 2004; Brennan & William, 1995). Similarly, disfluencies may also imply that interviewers are having

difficulty asking questions, especially when saying unfamiliar words or words they use infrequently (Kidd, et al. 2011).

In this study, I extracted interviewer voice characteristics from interviewers' readings of 12 survey questions. If the question was read twice, only the first reading was used. Interviewers' voice characteristics — including pitch, intonation, speech rate, and disfluencies — can be measured objectively by computer software programs and subjectively by raters.

2.1.1 Objective measures and subjective ratings of interviewers' voice characteristics

Objective voice characteristics. Voice characteristics can be objectively measured by computer software programs such as Praat and Sequence Viewer. For example, Groves et al. (2008) and Benki et al. (2011) used Praat to ascertain the pitch of interviewers' voices in survey introductions to study unit nonresponse. In my study, Praat was used to extract information about interviewer voice characteristics (described below).

Subjective voice characteristics. Previous research has used raters to evaluate interviewers' voice characteristics, such as speech rate, intonation, loudness, and flow of words (Broome 2012; Oksenberg, et al. 1986). In this study, I recruited six undergraduate students to be coders (details in the Methods section below). They rated pitch, intonation, speech rate, and disfluencies on a seven-point scale based on definitions listed in Table 2.1.

Table 2.1 Definitions of subjective ratings of interviewers' voice characteristics

Measures	Definition
Pitch	The degree of perception of an interviewer's voice from low-pitched (1) to high-pitched (7)
Intonation	The degree of variation in voice pitch from flat (1) to varied (7)

Speech rate	The speed with which an interviewer reads survey questions from slow (1) to fast (7)
Disfluencies	The degree to which an interviewer has a part of speech that are not words, such as stutters, from low (1) to high (7)

Previous research found strong correlations between objective measurements and subjective ratings of pitch, speech rate, and intonation (Van der Vaart, et al. 2005; Oksenberg & Cannell 1988), and moderate correlations between objective and subjective evaluations of the number of pauses (Oksenberg & Cannell 1988). Consistent with previous research, I hypothesize that the objective measures and subjective ratings of interviewers' voice characteristics are correlated. Specifically, I expect the following: 1) that raters will perceive high (low) pitched voices as having a high (low) pitch; 2) that voices with a faster (slower) speech rate will be perceived as speaking quickly (slowly); 3) that raters will be able to perceive high (low) variations in pitch as having a varied (flat) intonation; and 4) that raters will be able to perceive interviewers' speech with a high (low) number of fillers as having high (low) disfluencies.

2.1.2 Objective measures of interviewers' voice characteristics and subjective ratings of interviewers' personality traits

Certain personality traits of a speaker can be detected through his or her voice (Ketrow 1990; Apple, et al. 1979). As such, respondents may perceive interviewers' personal traits from their voices. In this study, I hypothesize that objective voice characteristics may affect coders' perceptions of five interviewer personality traits, including expertise, confidence, reliability, trustworthiness, and easiness to understand. Coders in this study rated five interviewer personality traits on a seven-point scale, i.e.

from low (1) to high (7), based on definitions of interviewer personality traits as listed in Table 2.2.

Table 2.2 Definitions of interviewer personality traits

Measures	Definition
Confidence	The extent to which the interviewer is self-assured and poised in conducting the interview
Easiness to understand	The extent to which the interviewer's voice is easy to understand
Reliability	The extent to which an interviewer says something that can be believed
Trustworthiness	The degree of confidence that an interviewer will ask valid survey questions and keep respondents' answers confidential
Expertise	The extent to which an interviewer is good at his or her job in asking survey questions

Previous paralinguistic research has found relationships between objective voice characteristics and subjective evaluations of that speaker's personal traits. I expect that results found in this study will be consistent with this previous research.

Pitch. High-pitched voices are perceived as less truthful, less reliable, less trustworthy, less easy to understand, and more confident than lower-pitched voices (Tigue, et al. 2012; Dey, et al 2006; Van der Vaart, et al. 2005; Apple, et al. 1979; Scherer, et al. 1973). In addition, previous research found associations between pitch and perceptions of professionalism; however, the direction of the association is mixed — Broome (2012) and Van der Vaart, et al. (2005) found a negative association, but Oksenberg, et al. (1986) found a positive association. Therefore, I expect coders to perceive interviewers who read a question with higher pitched voices as less reliable, less trustworthy, less easy to understand, and more confident than those who read a question with lower pitched voices. In addition, because of mixed findings on the direction of the association between pitch and a listener's perception of expertise, I expect that the coders

can judge whether an interviewer is an expert based on voice pitch, but that the direction cannot be hypothesized.

Intonation. Voices with broad variations in pitch are perceived as less reliable, less trustworthy, more confident, easier to understand, and more professional than voices with less variation (Natsumi 2013, Van der Vaart, et al. 2005; Oksenberg, et al. 1986; Brooke & Ng 1986; Scherer, et al. 1973). As such, I expect coders to perceive interviewers who read questions with high intonation (more variable pitch) as less reliable, less trustworthy, more confident, easier to understand, and more expert than those reading questions with low intonation (less variable pitch).

Speech rate. Rapid speech is perceived as more credible, more trustworthy, more professional, and expressing more confidence than slow speech (Broome, 2012; Smith & Shaffer 1995; Oksenberg & Cannell 1988; Apple, et al. 1979; Miller, et al. 1976; Scherer, et al. 1973; Pearce & Conking 1971). However, speaking quickly also reduces comprehension and may be harder to understand (Miller, et al. 1976). Based on this previous research, I expect coders to perceive interviewers who read questions at a faster rate of speech as more trustworthy, more reliable, more expert, and more confident, but less understandable than those who speak more slowly.

Disfluencies. Disfluencies, such as longer pauses and use of more fillers, can indicate a speaker's lack of confidence and that the speaker is encountering difficulties (Conrad, et al. 2008; Ehlen 2007; Bortfeld, et al. 2001). Highly disfluent speakers also are judged as less credible, less trustworthy, and less expert than more fluent speakers (Castro & de Moraes 2008; Ketrow 1990; Oksenberg & Cannell 1988; Apple, et al. 1979; Miller et al. 1976; Miller & Hewgill 1964). In addition, perceived easiness to understand

has been found to be negatively associated with the use of fillers (Pytko & Reese 2013). Thus, I expect coders to perceive interviewers who read a question with more disfluencies as less confident, less trustworthy, less reliable, less expert, and less easy to understand than those who read questions with fewer disfluencies.

The effects of objective voice characteristics on subjective ratings of interviewers' personality traits may not be linear. For example, it may be easier to understand a speaker talking at a normal speech rate, i.e. between 2.5 and 3.2 wps, than someone speaking at either slower or faster speech rates. Because of this possibility, I also examine the nonlinear relationships between interviewers' voice characteristics and interviewers' personality traits.

In addition, this study accounts for interviewer's demographic characteristics and question characteristics that may affect listeners' perceptions of interviewers' personal traits.

Interviewer's demographic characteristics. Voice characteristics may differ by an interviewer's gender and experience, and these interviewer characteristics may moderate relationships between interviewers' voice characteristics and the subjective ratings of their personality traits. Males speak slightly faster than females (Yuan, et al. 2006). In addition, interview length decreases as an interviewer has more interviewing experience (Olson & Peytchev 2007). Slower speech has a comprehension advantage over fast rates of speech (Hayati 2010). Consequently, female and inexperienced interviewers may be rated as easier to understand than male and experienced interviewers.

Question characteristics. Socially desirable, undesirable, and complex questions are good candidates for use in evaluating the association between vocal characteristics and subjective ratings of interviewer personality traits. This is because a speaker's vocal patterns are more likely to change with these types of questions (Bachorowski 1999). For example, because a high level of emotional arousal (e.g., fear and anxiety) is associated with an increase in voice pitch (Bachorowski 1999), interviewers may read socially undesirable questions with higher pitched voices than they use with other questions, and this may affect listeners' perceptions of personal traits such as trustworthiness.

2.2. Data and methods

2.2.1 Data

Data for this study came from the Work and Leisure Today Survey conducted in July and August 2013 by AbtSRBI (NSF SES-1132015). It is a landline RDD CATI survey administered by 22 interviewers with 450 completed interviews. To increase the stability of the analyses, interviewers who conducted fewer than 10 interviews were eliminated from the study (Olson & Peytchev 2007). These deletions left 432 interviews for my analysis. These interviews were conducted by 19 interviewers (9 female and 10 male). Overall, 4,689 voice files containing the first turn of an interviewer reading a question were examined for this study.

2.2.2 Questions

The questionnaire has 54 questions. For this study, I considered 24 candidate questions representative of the four categories — socially desirable, socially undesirable, complex, and neutral (not complex and not socially desirable/undesirable). I selected 12 (three from each of the above four categories) from these 24 questions. My selections

were based on the criteria that questions have to contain both item nonresponse and item responses and sufficient variability in responses. The wording of the 12 questions selected appears in Appendix A. The topics of these questions cover employment status and volunteer work, leisure activities such as exercise and Internet use, and substance use such as drinking alcohol.

2.2.3 Measures of voice characteristics and personality traits of interviewers

Objective voice characteristics. I used the Praat computer program (www.fon.hum.uva.nl/praat/) to do a speech analysis. Figure 2.2 shows an example of the waveform (upper panel) and the spectrogram (lower panel) of speech that the Praat program produces.

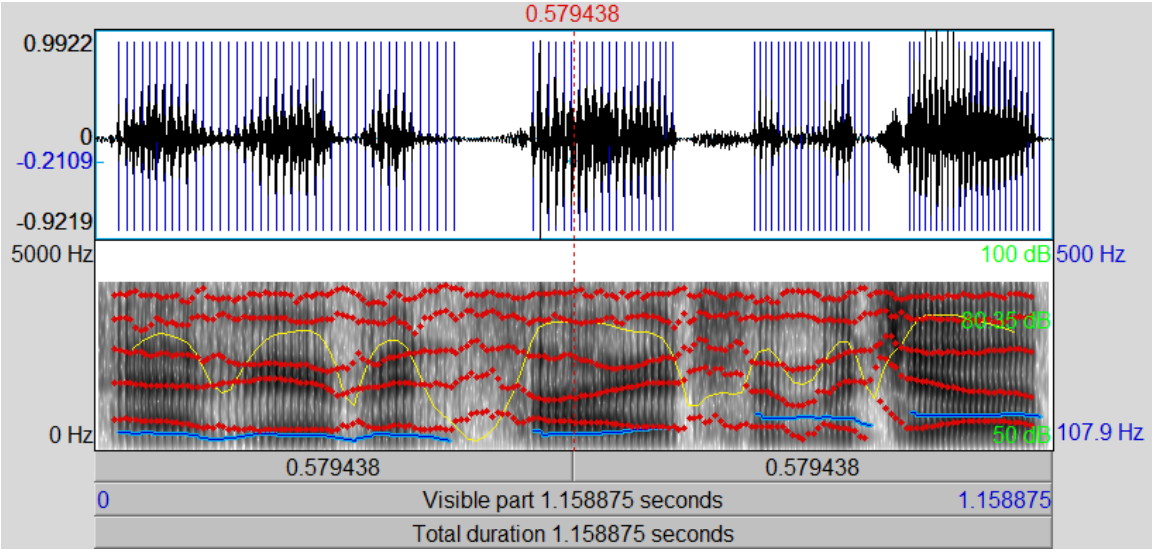


Figure 2.2 An example of a speech waveform (upper panel) and a speech spectrogram (lower panel)

The waveform of a sound (the upper panel in Figure 2.2) represents sound pressure as a function of time (Boersma & Weenink 2014). Sound pressure is the variation in air pressure that a human is able to perceive as sound. The greater the variation, the louder the sound that is heard. The waveform shows silences and also the

duration of utterances (Boersma & Weenink 2014). The spectrogram (the lower panel in Figure 2.2), which is associated with the waveform, displays visually the degree of sound pressure at various frequencies (in Hz; on the vertical axis) along time (on the horizontal axis) (Boersma & Weenink 2014). Sound that is heard normally contains a combination of many frequencies that add up to a single complex wave of sound (i.e., a speaker's voice). The spectrum shows the different sound pressures at specific frequencies of voices and time. The darker parts of the spectrogram (lower panel) mean that the sound has a higher sound pressure (volume) for that frequency and time. In addition, information about pitch can be visually presented in the spectrogram as a series of blue dots (Styler 2014).

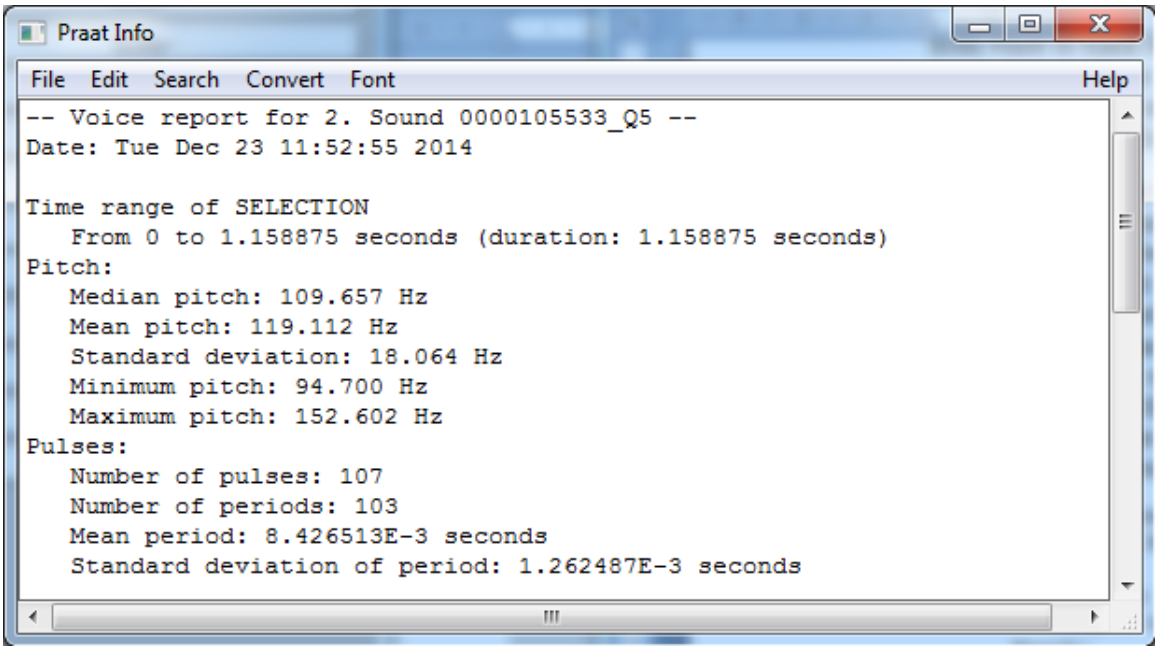


Figure 2.3 An example of a speech analysis summary produced by Praat

In addition to the waveform and the spectrogram of a sound, Praat summarizes its speech analysis in a text file format, as shown in Figure 2.3. I used the mean of pitch over the period that an interviewer first read a survey question as the measurement of pitch

and the standard deviation of pitch over the same period as a measurement of intonation. In addition, the total duration time that interviewers read a survey question was used to compute the speech rate, which is defined as the number of words per second (wps). Moreover, I used the number of fillers — including Um, Uh, Ah, Mm, Hm, and Oh — as the measure of disfluencies. The number of words and fillers extracted from the interview transcript was read into the Sequence Viewer program.[2]

Subjective voice characteristics and ratings of interviewers' personality traits. Six undergraduate students (4 males and 2 females) were recruited as coders based on a requirement that that they did not have any prior survey interviewing experience. This requirement allowed the coders to better represent uninformed telephone respondents rather than a trained interviewer. The coders listened to the voice file containing the first turn in which an interviewer reads a survey question and rated four interviewer voice characteristics and five interviewer personality traits on seven-point scales (see Appendix B for the coders' instructions). The coders evaluated four interviewer voice characteristics, including pitch from low (1) to high (7), intonation from flat (1) to varied (7), rate of speech from slow (1) to fast (7), and disfluencies from low (1) to high (7). In addition, the coders evaluated five interviewer personality traits — confidence, easiness to understand, reliability, trustworthiness, and expertise — from low (1) to high (7). Moreover, coders were asked to identify whether an interviewer was male or female. The coders were trained before they conducted the actual coding. Because the subjective ratings in this study were coded by multiple coders, I used the average of the ratings

[2] Sequence Viewer is a program used to analyze sequential data such as interviewer-respondent interactions in which each question can indicate the beginning of a new sequence (Dijkstra 2009).

interviewers' voice characteristics and interviewers' personality traits across the coders in the analyses.

Assessing coder reliability. The reliability of each of the subjective ratings was measured for continuous outcomes by an Intraclass Correlation Coefficient (ICC). The ICC was calculated by using the mean of ratings as a unit of analysis, i.e., ICC (3,k), which is the ICC for a fixed set of k raters (Shrout & Fleiss 1979).

I conducted a pilot study to test the instructions and gain an initial measurement of the reliability of the subjective voice measures. Four survey research and methodology (SRAM) graduate students (1 male and 3 females) were selected to be coders in the pilot study. I randomly selected 15 voice files in each of the four types of questions (for a total of 60 voice files) to be evaluated by all four pilot coders. Results from the pilot study revealed that all of the subjective measures had adequate reliability ($0.42<ICC(3,k)<0.88$). After debriefing the coders in the pilot study, the coding protocol was not modified because, as I had expected, they all understood the instructions consistently. The reliability of the undergraduate coders' ratings is provided below.

2.2.4 Analyses

I conducted two main analyses in this study. First, I evaluated the associations between objective and subjective measures of the interviewers' voice characteristics – that is, pitch, intonation, speech rate, and disfluencies — by examining the Pearson's correlations between these two measures.

Next, I examined the relationship between the objective measures of interviewers' voice characteristics and the subjective ratings of interviewers' personality traits. The analyses were done for all 12 questions in one model for each interviewer's personality

trait. A two-level multilevel model was estimated to account for variation due to questions (Level 1) and interviewers (Level 2) (O'Muircheartaigh & Campanelli 1998; 1999). I also tested for a significant interviewer variance effect (var(U_{0j})) through a base model using a mixture of chi-square distributions (Rabe-Hesketh & Skrondal 2012). In general, the model estimated in this chapter is:

$$Rating_{ij} = \gamma_{00} + \gamma_{01} IChar_j + \sum_{a=1}^{A} \gamma_{a0} VoiceChar_{aj} + \sum_{a=1}^{A} \gamma_{a1} IChar_j \times VoiceChar_{aj} + \sum_{b=A+1}^{B} \gamma_{b0} QChar_{bj} + \sum_{c=B+1}^{C} \gamma_{c0} VoiceChar_j \times QChar_j + U_{0j} + e_{ij}$$

where $Rating_{ij}$ = rated confidence, rated expertise, rated reliability, rated trustworthiness, and rated easiness of understand,

$VoiceChar_{ij}$ = pitch (centered at 165 Hz), intonation (centered at 40 Hz), rate of speaking (centered at 3.5 wps), and number of fillers,

$QChar_{ij}$= neutral, complex, socially undesirable, and socially desirable questions,

$IChar_j$ = interviewer's gender and experience,

U_{oj} = random interviewer effect, and

e_{ij} = residuals.

Interviewers' demographic characteristics are dichotomous variables in which male interviewers and interviewers whose experience is higher than 1 year are the reference groups. Neutral questions are the reference group for types of questions. As mentioned earlier, nonlinear relationships between objective voice characteristics and subjective ratings of interviewers' personality traits may exist. Because of this possibility, I also examined the squared terms of pitch, intonation, speech rate, and number of fillers.

2.3 Results

2.3.1 Descriptive statistics for the objective and subjective measures of interviewers' voice characteristics and subjective ratings of interviewers' personality traits

Table 2.3 presents descriptive statistics on the objective measures of interviewers' voice characteristics, subjective ratings of interviewers' voice characteristics, and subjective ratings of interviewers' personality traits. The average objective measures of interviewers' pitch, intonation, speech rate, and number of fillers is 168 Hz, 42 Hz, 3.5 wps, and 0.2 fillers per voice file, respectively.

Table 2.3 Descriptive statistics on objective measures of interviewers' voice characteristics, subjective measures of interviewers' voice characteristics, and subjective measures of interviewers' personality traits

Measures	Mean	S.D.	Min.	Max.
Objective measures of interviewers' voice characteristics				
Pitch	167.63	40.04	89.38	315.11
Intonation	41.72	19.91	3.87	160.58
Speech rate	3.49	0.95	0.55	6.88
Disfluency	0.2	0.48	0	4
Subjective measures of interviewers' voice characteristics				
Pitch (1=low, 7=high)	3.49	0.85	1.6	6
Intonation (1=flat, 7=varied)	3.04	0.62	1.4	5.20
Speech rate (1=slow, 7=fast)	3.94	0.76	1.75	6.71
Disfluency (1=low, 7=high)	1.87	0.65	1	5.50
Subjective measures of interviewers' personality traits				
Confidence (1=low, 7=high)	5.76	0.53	2.67	7
Easiness to understand (1=low, 7=high)	5.90	0.49	3.00	7
Reliability (1=low, 7=high)	5.89	0.43	3.40	7
Trustworthiness (1=low, 7=high)	5.74	0.50	4.00	7
Expertise (1=low, 7=high)	5.87	0.50	3.40	7
Female interviewer (1=female, 0=male)	0.50	0.48	0	1

Six coders rated interviewers' voice characteristics and interviewers' personality traits on a seven-point scale. The average ratings of interviewers' pitch (1=low, 7=high), intonation (1=flat, 7=varied), speech rate (1=slow, 7=fast), and disfluencies (1=low, 7=high) are 3.49, 3.04, 3.94, and 1.87, respectively. In addition, the average ratings of

interviewers' confidence, easiness to understand, reliability, trustworthiness, and expertise where 1=low and 7=high for each trait are 5.76, 5.90, 5.89, 5.74, and 5.87, respectively. The coders also coded whether an interviewer was a male or female based on the interviewer's voice. On average, the coders identified 50% of interviewers' voices as male. The coders' gender identification matched the actual gender of interviewers 96.67% of the time.

Table 2.4 ICC (3,k) and assessment of reliability based on Munro's criteria for subjective ratings of interviewers' voice characteristics and personality traits

	ICC(3,k)	Munro's criteria
Interviewers' voice characteristics		
Pitch	0.637	moderate reliability
Intonation	0.617	moderate reliability
Speech rate	0.820	high reliability
Disfluency	0.636	moderate reliability
Interviewers' personality traits		
Confidence	0.599	moderate reliability
Easiness to understand	0.456	low reliability
Reliability	0.431	low reliability
Trustworthiness	0.468	low reliability
Expertise	0.629	moderate reliability
Interviewer's gender	0.987	high reliability

ICC (3, k) was used to measure the reliability of the subjective ratings of interviewers' voice characteristics and interviewers' personality traits. Munro (2005) characterized ICC values of 0.26-0.49 as low, 0.50-0.69 as moderate, 0.70-0.89 as high, and 0.9-1 as very high. Values of ICC (3, k) for subjective ratings of interviewers' voice characteristics and interviewers' personality traits are shown in Table 2.4. Using Munro's criteria of reliability, there were high reliability in speech rate (ICC=0.820) and interviewer gender (ICC=0.987), and moderate reliability in pitch (ICC=0.637), intonation (ICC=0.617), disfluencies (ICC=0.636), confidence (ICC=0.599), and

expertise (ICC=0.629). However, low reliability was found in easiness to understand (ICC=0.456), reliability (ICC=0.431), and trustworthiness (ICC=0.468).

2.3.2 Associations between objective and subjective interviewers' voice characteristics

Table 2.5 shows the Pearson's correlations between the objective and subjective interviewers' voice characteristics. Consistent with the hypotheses, the objective and subjective measures of pitch, intonation, speech rate, and disfluencies are positively correlated: pitch (r=0.79); intonation (r=0.49); speech rate (r=0.52); and disfluencies (r=0.36) (See the diagonal correlations in Table 2.5). These positive moderate-to-high correlations imply that listeners can perceive interviewers' voice characteristics as being quite similar to those determined by the computer program.

Table 2.5. Pearson's correlations between the objective and subjective interviewers' voice characteristics

Objective measures	Subjective measures			
	Rated pitch	Rated intonation	Rated speech rate	Rated disfluencies
Pitch	0.79 **	0.42 **	-0.03	-0.10 **
Intonation	0.46 **	0.49 **	-0.07 **	-0.04 *
Speech rate	-0.07 **	-0.07 **	0.52 **	-0.26 **
Disfluencies	-0.10 **	0.11 **	0.09 **	0.36 **

Note **p<0.01, *p<0.05

The off-diagonal correlations indicate the positive moderate correlation between the subjective measure of pitch and the objective measure of intonation (r=0.46) and between the subjective measure of intonation and the objective measure of pitch (r=0.42). This may be because intonation is defined as the degree of variation in voice pitch. As such, these two voice characteristics are moderately correlated. Unlike these two off-diagonal correlations, the other off-diagonal correlations are quite low, implying that coders cannot perceive interviewers' voice characteristics from other objective measures

of voice characteristics. For example, the coders could not perceive from an interviewer's pitch whether he or she read quickly or slowly (r=-0.07).

2.3.3 Interviewer voice characteristics and interviewer personality traits

Table 2.6 shows results from the regression in which interviewer voice characteristics were predicting perceptions of interviewers' personality traits. Variance across interviewers in all five rated interviewer personality traits are significant ($p<0.01$). As rated by the coders, interviewers vary in their confidence - 35.4 percent of the variation is due to interviewers (χ^2=1751.64, $p<0.01$); in their easiness to understand (ICC due to interviewers=27.5%; χ^2=1192.56, $p<0.01$), in their reliability (ICC due to interviewers=24.5%; χ^2=1035.48, $p<0.01$), in their trustworthiness (ICC due to interviewers=33.7%; χ^2=1508.07, $p<0.01$), and in their expertise (ICC due to interviewers=37.7%; χ^2=1945.16, $p<0.01$).

Overall, listeners could perceive interviewers' personality traits from their voices, i.e. the objective measures of vocal characteristics predicted the subjective ratings of personality traits of interviewers. The effects of objective measure of pitch on interviewers' personality traits do not vary by interviewer's demographic characteristics and question types. However, some effects of intonation, speech rate, and fillers on rated interviewers' personality traits are moderated by interviewer sex and question types.

Pitch. As shown in Table 2.6, objectively measured pitch has a curvilinear relationship with rated confidence, reliability, trustworthiness, and expertise. Interviewers who read a question with higher pitched voices are rated as being more confident, more reliable, more trustworthy, and having more expertise, but with a decelerating rate (Figure 2.4). The findings for reliability and trustworthiness are opposite the

hypothesized direction, but the findings for confidence and expertise are consistent with the hypothesized direction.

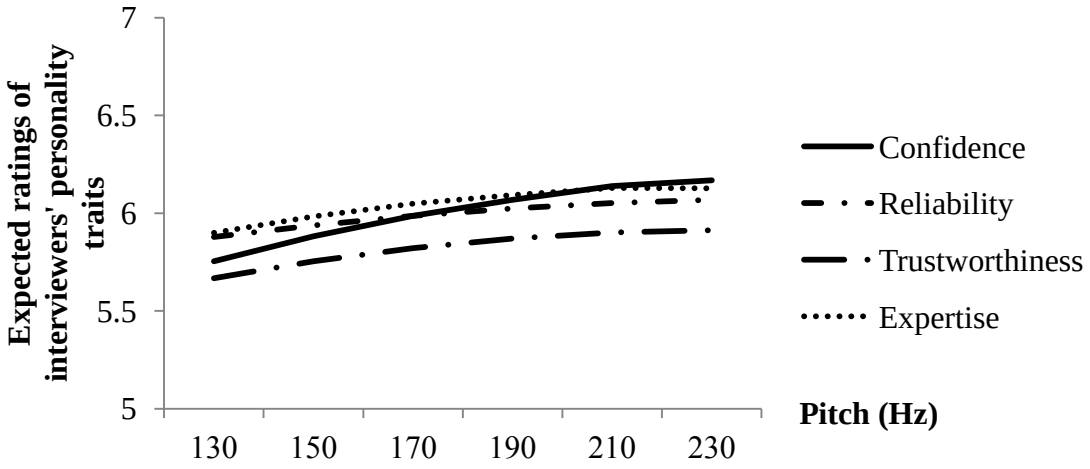


Figure 2.4 Expected ratings of interviewers' personality traits by pitch

Intonation. As shown in Table 2.6, intonation squared is significantly and negatively associated with all five expected ratings of interviewer's personality traits—confidence (coefficient=-0.00005, $p<0.01$); easiness to understand (coefficient=-0.00001, $p<0.01$); reliability (coefficient=-0.00001, $p<0.01$); trustworthiness (coefficient=-0.00004, $p<0.01$); and expertise (coefficient=-0.00004, $p<0.01$). Figure 2.5 presents an inverse U-shape association between expected ratings of interviewers' personality traits and intonation. The expected ratings of interviewer's personality traits are higher when interviewers read questions with moderate intonation (around 40-60 Hz), compared to either lower or higher intonation. In addition, as Figure 2.5 shows, the negative effects of intonation on perceptions of interviewer's personality traits are more apparent when interviewers read questions with an intonation higher than 80 Hz, an intonation level observed among only 5% of voice files in this study. This implies that when interviewers read questions with an intonation higher than its typical range, perceptions of these interviewer personality traits dramatically decline.

Table 2.6 Hierarchical logistic model predicting subjective ratings of interviewer personality traits by objective voice characteristics

	Confidence		Easiness to understand		Reliability		Trustworthiness		Expertise	
	coefficient (SE)		coefficient (SE)		coefficient (SE)		coefficient (SE)		coefficient (SE)	
Intercept	6.027(0.09)	**	6.070(0.08)	**	5.991(0.07)	**	5.778(0.087)	**	6.049(0.10)	**
Pitch	0.005(0.0004)	**	0.0006(0.0004)		0.002(0.0004)	**	0.003(0.0004)	**	0.003(0.0004)	**
Pitch^2	-0.00003(0.000004)	**			-0.00001(0.000004)	**	-0.00002(0.0000004)	**	-0.00002(0.000004)	**
Intonation	0.0004(0.0006)		0.001(0.0007)	*	0.0007(0.0005)		0.0005(0.0005)		0.0003(0.0006)	
Intonation^2	-0.00005(0.000012)	**	-0.0001(0.00001)	**	-0.0001(0.00001)	**	-0.00004(0.00001)	**	-0.00004(0.00001)	**
Speech rate	0.135(0.018)	**	-0.032(0.02)		0.108(0.006)	**	0.062(0.007)	**	0.107(0.009)	**
Speech rate^2	-0.017(0.006)	**	-0.051(0.005)	**	-0.029(0.004)	**	-0.020(0.004)	**	-0.033(0.004)	**
Filler	-0.118(0.01)	**	-0.046(0.02)	*	-0.075(0.01)	**	-0.0009(0.016)		-0.123(0.01)	**
Iwer experience <1	-0.129(0.15)		0.103(0.12)		-0.029(0.10)		0.058(0.13)		-0.031(0.15)	
Female Iwer	-0.276(0.13)		-0.034(0.11)		-0.039(0.09)		0.093(0.12)		-0.085(0.14)	
Desirable Qs	0.022(0.02)		-0.051(0.02)	**	0.035(0.01)	*	0.002(0.01)		0.0003(0.02)	
Complex Qs	-0.010(0.02)		-0.189(0.02)	**	0.005(0.02)		-0.015(0.02)		-0.024(0.02)	
Undesirable Qs	-0.090(0.02)	**	-0.128(0.02)	**	-0.065(0.02)	**	-0.118(0.02)	**	-0.119(0.02)	**
Voice * Iwer chars										
Intonation*female Iwer	0.005(0.0008)	**			0.002(0.0008)	*	0.003(0.0008)	**	0.003(0.001)	**
Speech rate*female Iwer			0.047(0.01)	**					0.037(0.013)	*
Filler*female Iwer							-0.069(0.02)	**		
Voice * question chars										
Intonation*desirable Qs			-0.0001(0.0008)							
Intonation*complex Qs			0.003(0.0008)	**						
Intonation*undesirable Qs			0.0002(0.0008)							

Note **p<0.01, *p<0.05

Table 2.6 Continued.

	Confidence		Easiness to understand		Reliability		Trustworthiness		Expertise	
	coefficient (SE)		**coefficient (SE)**		**coefficient (SE)**		**coefficient (SE)**		**coefficient (SE)**	
Voice * question chars										
Speech rate*desirable Qs	0.054(0.03)	*	-0.009(0.02)							
Speech rate*complex Qs	-0.068(0.02)	**	0.279(0.02)	**						
Speech rate*undesirable Qs	-0.054(0.02)	*	0.085(0.02)	**						
Filler*desirable Qs			0.004(0.03)							
Filler*complex Qs			-0.033(0.04)							
Filler*undesirable Qs			-0.136(0.04)	**						
Variance components										
Random interviewer	0.080(0.03)	**	0.05(0.02)	**	0.04(0.01)	**	0.064(0.02)	**	0.083(0.03)	**
Residual	0.146(0.003)	**	0.132(0.003)	**	0.123(0.003)	**	0.126(0.003)	**	0.137(0.002)	**
Model fit										
AIC	4416.6		3952.7		3637.8		3717.6		4122.9	
BIC	4434.5		3975.4		3652.9		3733.7		4139	
N	4689		4689		4689		4689		4689	

Note **p<0.01, *p<0.05

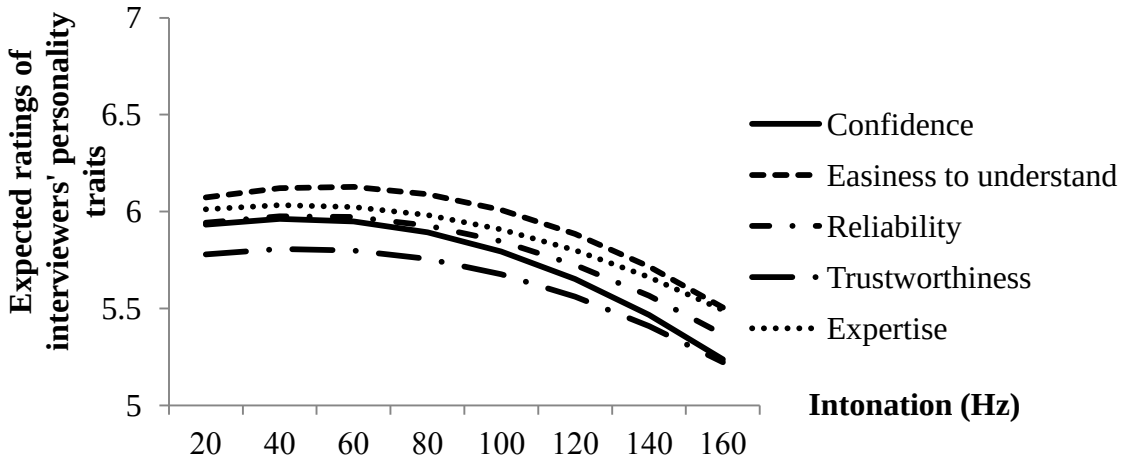


Figure 2.5 Expected ratings of interviewers' personality traits by intonation

Speech rate. A curvilinear relationship between objectively measured speech rate and rated interviewer personality traits was found (Table 2.6). Figure 2.6 presents expected ratings of interviewers' personality traits across speech rate levels. As expected, listeners rated a fast speech rate as more confident, more reliable, more trustworthy, and more expert compared to a slow speech rate. In contrast, a non-monotonic relationship between rated easiness to understand and speech rate was found. A speech rate between 3 to 4 words per second (wps) was rated as easier to understand than a speech rate of 2 wps or a speech rate of 5 wps.

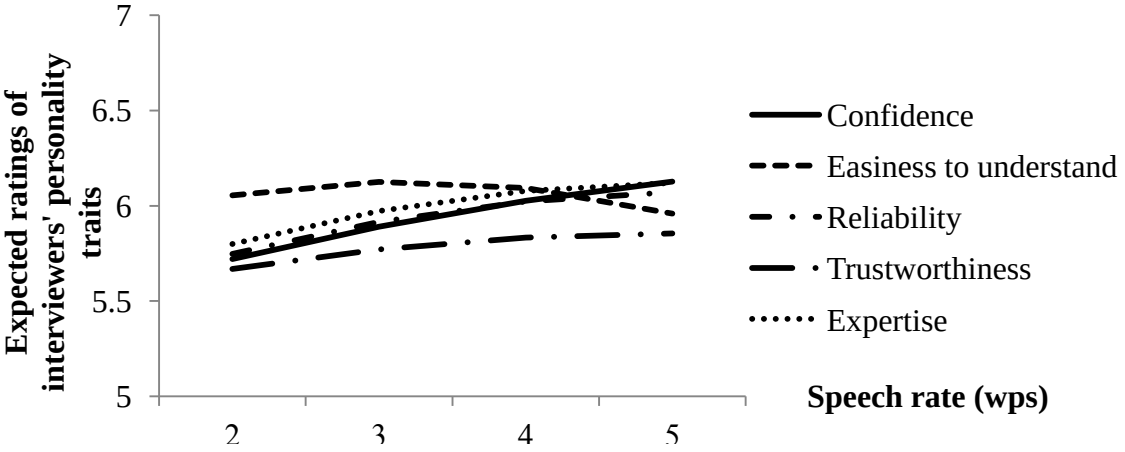


Figure 2.6 Expected ratings of interviewers' personality traits by speech rate

Fillers. As expected, fillers negatively and significantly affect rated confidence (coefficient=-0.118, p<0.01), easiness to understand (coefficient=-0.046, p=0.02), reliability (coefficient=-0.075, p<0.01), and expertise (coefficient=-0.123, p<0.01) (Table 2.7). Interviewers who asked questions with more fillers were rated as being less confident, less easy to understand, less reliable, and less expert than those who asked questions with more fillers, as displayed in Figure 2.7.

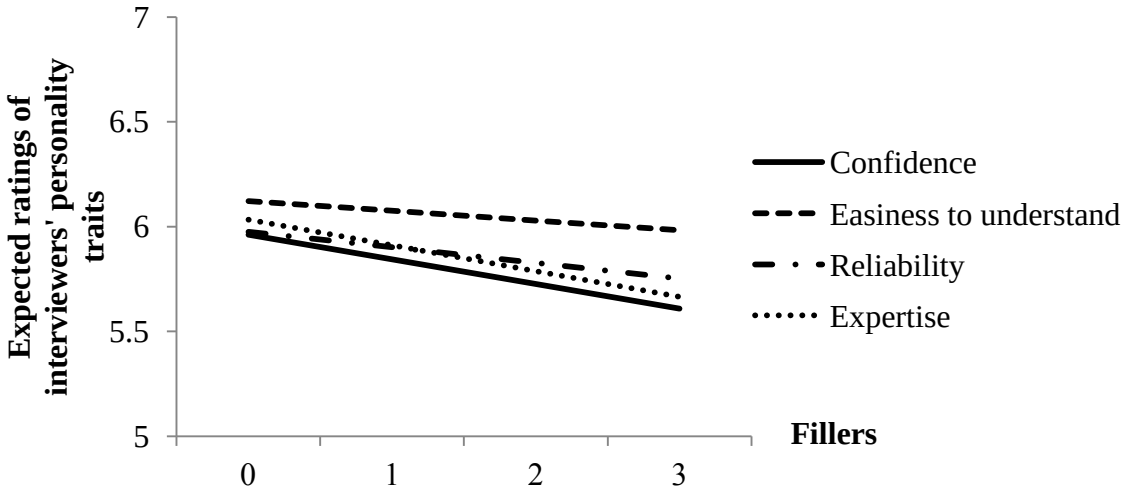
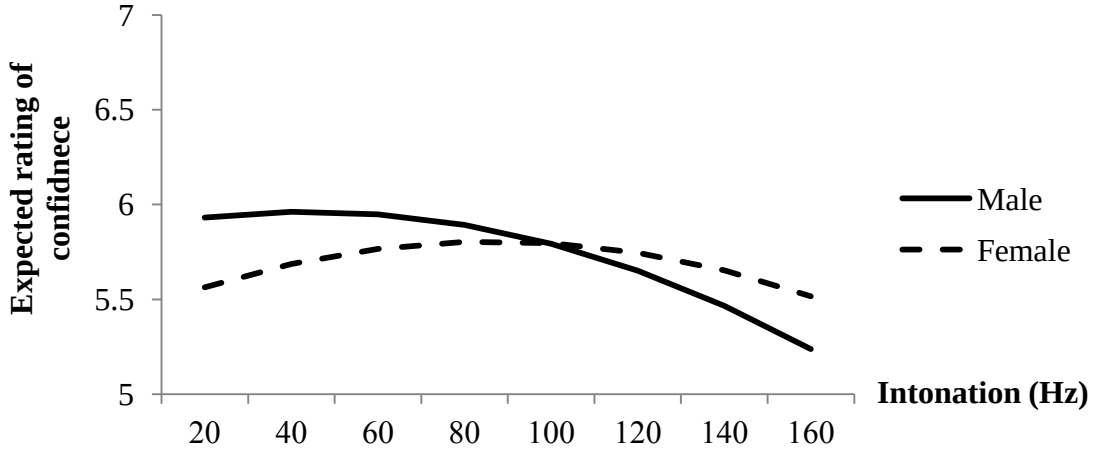


Figure 2.7 Expected ratings of interviewers' personality traits by fillers

Interviewer voices and interviewer demographic characteristics

I examined whether interviewer sex and experience were associated with rated interviewer personality traits, and whether these two characteristics moderate the effects of objective measures of interviewer voices on rated personality traits. Rated interviewer personality traits are only associated with interviewer sex, not interviewer experience. Female interviewers were rated as less confident in asking survey questions than male interviewers; however, the effect is just marginally significant (coefficient=-0.276, p= 0.05). In addition, I found that interviewer sex moderates some effects of intonation, speech rate, and filler on rated personality traits of interviewers.

Intonation and interviewer sex. Interviewer sex moderates the effects of intonation on rated confidence (coefficient=0.005, $p<0.01$), rated reliability (coefficient=0.002, $p=0.02$), rated trustworthiness (coefficient=0.003, $p<0.01$), and rated expertise (coefficient=0.003, $p<0.01$). As shown in Figure 2.8, for male voices, the highest expected ratings of confidence, reliability, trustworthiness, and expertise is at an intonation of 40 Hz. However, for female voices, the highest expected ratings of confidence, reliability, trustworthiness, and expertise is at an intonation ranging between 60-80 Hz. Consistent with gender stereotypes in paralinguistic research (Kent & Read 2002), this study found male interviewers ask questions with flatter intonation than female interviewers (mean intonation $_{male}$=34.82 Hz, mean intonation $_{female}$=48.89 Hz; t=-25.84, $p<0.01$). As Figure 2.8 shows, male interviewers were rated as being more confident, more reliable, and having more expertise than female interviews at low intonation; however, the opposite direction was found at high intonation. This implies that negative impressions (e.g., lower ratings on each personality trait) occur for male voices with high intonation and for female voices with low intonation.



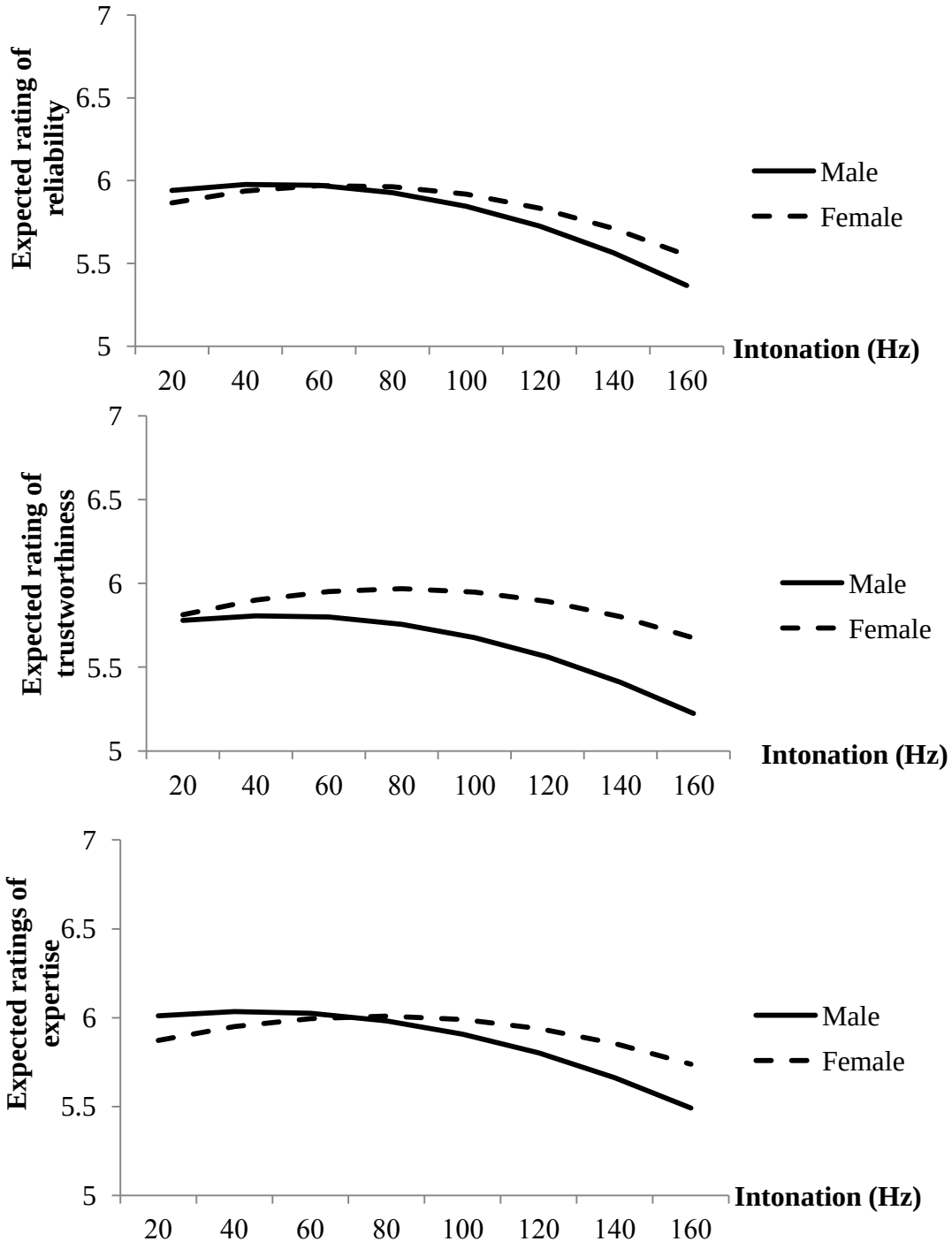


Figure 2.8 Expected ratings of interviewers' personality traits by intonation and interviewer sex

Speech rate and interviewer sex. Interviewer sex moderates the effects of speech rate on rated easiness to understand (coefficient=0.047, $p<0.01$) and expertise

(coefficient=0.037, p=0.04). Figure 2.9 shows the effects of speech rate on rated easiness to understand and expertise by gender. However, the effects are modest. As speech rate increases, male interviewers were rated as slightly more expert and easier to understand than female interviewers. However, at a speech rate of 4 and 5 wps, male voices were rated as slightly less easy to understand than female voices.

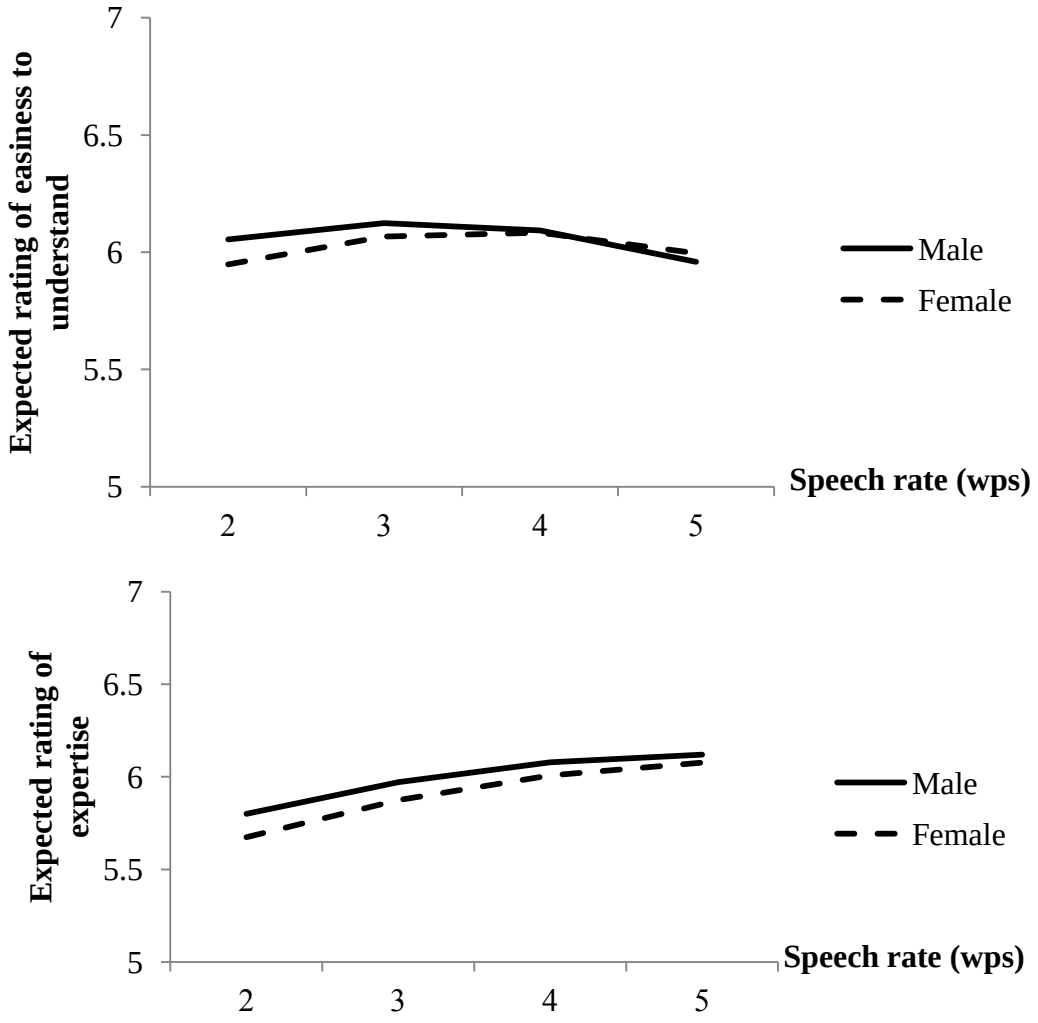


Figure 2.9 Expected ratings of easiness to understand (upper panel) and expertise (lower panel) by speech rate and interviewer sex

Fillers and interviewer sex. As shown in Table 2.6, the negative effect of fillers on rated trustworthiness (coefficient=-0.069, p<0.01) was significant only for female

interviewers. Female interviewers who read questions with fewer fillers were rated as more trustworthy than those who read questions with more fillers.

Interviewer voices and question characteristics

In addition to interviewer sex, ratings of interviewers' personality traits vary by question type (Table 2.6). Interviewers reading socially undesirable questions were judged as being less confident (coefficient=-0.090, $p<0.01$), less easy to understand (coefficient=-0.128, $p<0.01$), less reliable (coefficient=-0.065, $p<0.01$), less trustworthy (coefficient=-0.118, $p<0.01$), and less expert (coefficient=-0.119, $p<0.01$) than interviewers reading neutral questions. In addition, compared to neutral questions, interviewers were rated as being less easy to understand for complex questions (coefficient=-0.189, $p<0.01$) and for socially desirable questions (coefficient=-0.051, $p<0.01$). Moreover, interviewers reading socially desirable questions were rated as more reliable than those reading neutral questions (coefficient=0.035, $p=0.02$). Additionally, question types moderate the effects of intonation, speech rate, and fillers on rated interviewers' personality traits, especially rated confidence and easiness to understand.

Intonation and question types. As shown in Table 2.6, the effect of intonation on rated easiness to understand differs between complex and neutral questions (coefficient=0.003, $p<0.01$). While the highest rating of easiness to understand was at intonation levels around 60 Hz for neutral questions, the highest rating of easiness to understand was at intonation levels around 80 Hz for complex questions (Figure 2.10). At intonation levels less than 100 Hz, listeners rated neutral question as easier to understand than complex questions. However, beyond intonation levels of 100 Hz, listeners rated complex questions as easier to understand than neutral questions– the effect is modest.

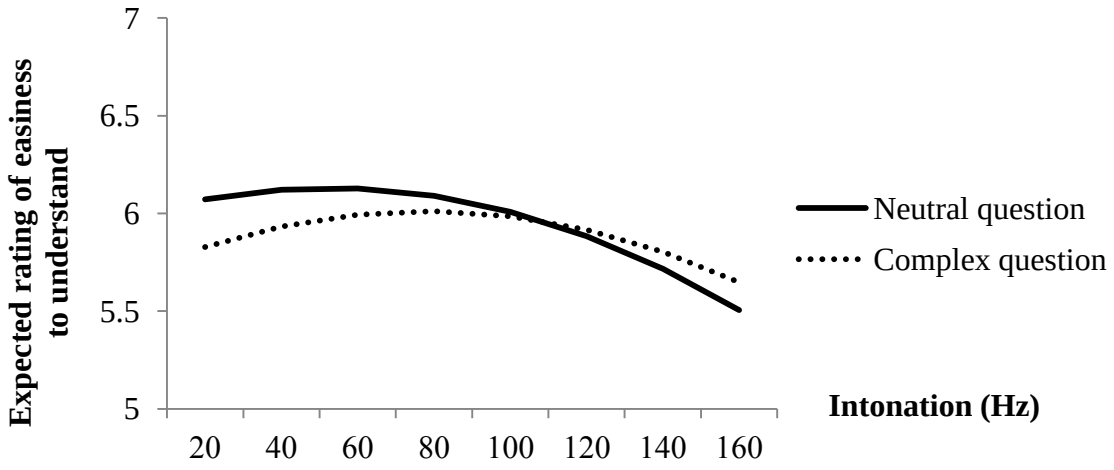


Figure 2.10 Expected ratings of easiness to understand in neutral and complex questions by intonation

Speech rate and question types. The effects of speech rate on rated confidence and easiness to understand differs by question type (Table 2.6). The direction of the effects of speech rate on rated confidence is the same across question types, although the strength of the relationship differs (Figure 2.11 upper panel). As speech rate increases, rated confidence increases with decelerating rate. The effect was strongest in socially desirable questions compared to other question types. In contrast, the *direction* of the effects of speech rate on rated easiness to understand significantly *differs* by question types. An inverse U-shaped relationship was found for neutral, desirable, and undesirable questions (Figure 2.11 lower panel). For complex questions, however, a positive association between speech rate and rated easiness to understand was found (coefficient=0.279, $p<0.01$; Table 2.6); that is, complex questions asked more quickly are rated as easier to understand (Figure 2.11 lower panel). Moreover, the expected rating of easiness to understand is highest at speech rate of 3 wps for neutral and socially desirable questions and at speech rate of 4 wps for socially undesirable questions.

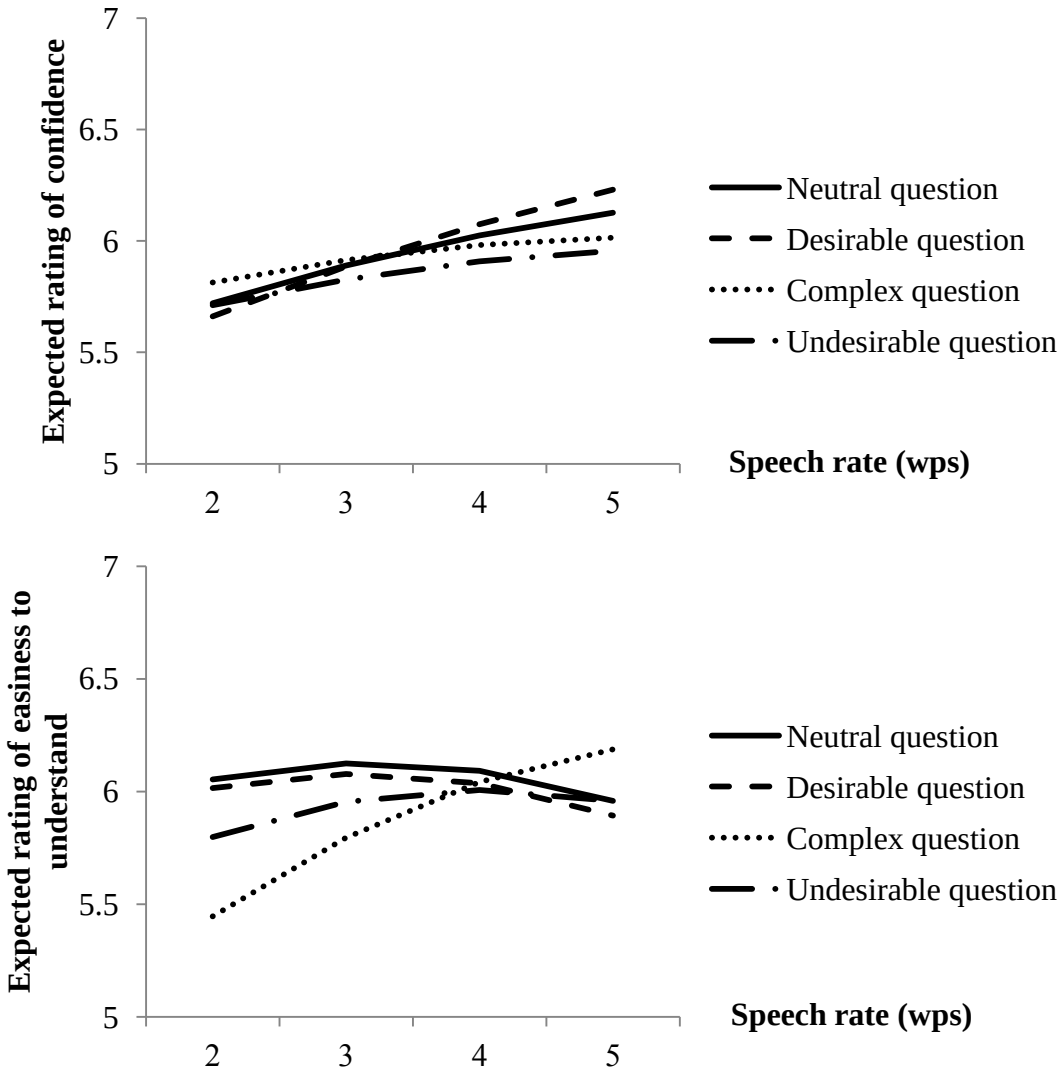


Figure 2.11 Expected ratings of confidence (upper panel) and easiness to understand (lower panel) by speech rate and question type

Fillers and question types. The negative effect of fillers on rated easiness to understand was stronger in socially undesirable questions than in neutral questions (coefficient=-0.136, $p<0.01$; Table 2.6). As shown in Figure 2.12, survey questions read with more fillers are rated as less easy to understand in socially undesirable questions than in neutral questions.

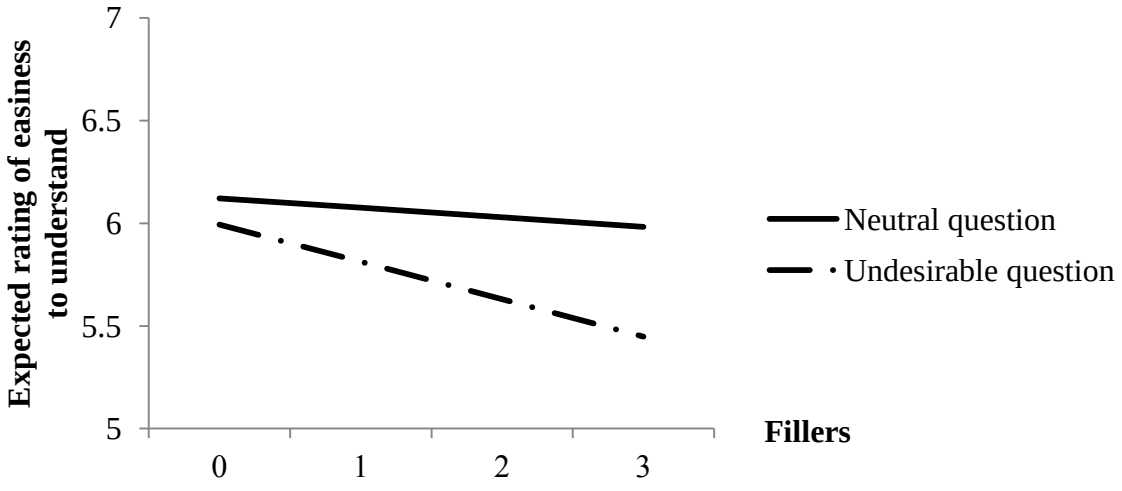


Figure 2.12 Expected ratings of easiness to understand in neutral and socially undesirable questions by fillers

2.4 Conclusion and discussion

Correlations between the subjective and objective measures of pitch, intonation, speech rate, and disfluency were examined. Consistent with Van der Vaart, et al. (2005) and Oksenberg & Cannell (1988), listeners perceived high (low) pitched voices as having a high (low) pitch, perceived voices with a faster (slow) speech rate as speaking quickly (slowly), perceived high (low) variation in pitch voices as having a varied (flat) intonation, and perceived interviewer's speech with high (low) number of fillers as having high (low) disfluencies. The findings show that listeners can perceive interviewer voice characteristics as measured by a computer program.

More importantly, I examined how objective voice characteristics affect a listener's perception of interviewers' personality traits (expertise, confidence, reliability, trustworthiness, and easiness to understand). I found listeners perceive personality traits of interviewers from interviewer voice characteristics. However, some effects of intonation, speech rate, and disfluencies on rated interviewer personality traits were moderated by interviewer sex and question types.

Pitch. The effect of the objective measure of pitch on rated interviewer personality traits did not vary by interviewer's demographic and question characteristics. Listeners rated interviewers reading questions with higher pitched voices as more reliable, more confident, more expert, and more trustworthy than those reading questions with lower pitched voices. Results for reliability and trustworthiness are opposite my hypothesis. Low pitched voices are associated with undesirable traits such as dishonesty and dominance, whereas high pitched voices are associated with positive personality traits such as friendliness (Boehme 2014). Listeners may consider these positive and negative traits to be parts of trustworthiness and reliability versus untrustworthiness and unreliability respectively, and thus perceive higher pitched voices as more trustworthy and more reliable (Boehme 2014). Results in this study suggest that survey organizations should recommend that interviewers read questions with higher pitched voices.

Intonation. Overall, intonation has an inverse U-shaped association with expected ratings of interviewer personality traits, although this effect varies across interviewer sex and question types. Male interviewers who read questions with less intonation (variation in voice pitch around 40 Hz) and female interviewers who read questions with moderate intonation (around 60-80 Hz) are perceived as more confident, more reliable, more trustworthy, and more expert compared to either too high or too low intonation. The negative association between intonation and perceptions of interviewer personality traits are more apparent when an interviewer reads a question with an intonation higher than 80 Hz, an intonation observed among only 5% of voice files in this study. Listeners may think it strange when interviewers read questions with too much intonation. However, within a typical range of intonation (intonation around 20-80 Hz, accounting for 95% of

voice files), there are positive associations between intonation and perceptions of interviewer's personality traits. As with pitch, results for trustworthiness and reliability are opposite my hypothesis. Similar to pitch, lower intonation is associated with undesirable personality traits such as dominance and dishonesty while higher intonation is related to positive personality traits such as friendliness. Listeners may include these positive and negative personality traits when judging whether voices are trustworthy and reliable (Boehme 2014).

Consistent with gender stereotypes of voices that females have higher intonation than males, perceptions of interviewer personality traits for male interviewers were more positive at *lower* intonation levels than for female interviewers. In contrast, perceptions of interviewer's personality traits for female interviewer were more positive at *higher* intonation levels than for male interviewers. In addition to interviewer sex, the effect of intonation on rated easiness to understand varies by question types. To be perceived as easier to understand, interviewers should read complex questions with more intonation (around 80 Hz) than neutral questions (around 60 Hz).

Speech rate. Overall, listeners rated interviewers' voices with a speech rate faster than 2 wps as more confident, more reliable, more trustworthy, and more expert than those with a speech rate of 2 wps. The effect of speech rate on rated confidence was strongest for socially desirable questions compared to other question types. In addition, as speech rate increases, male interviewers were rated as more expert than female interviewers.

A non-monotonic effect of speech rate on rated easiness to understand was found, with variation across question types. Overall, the highest rating of easiness to understand

was found when interviewers read questions with a speech rate of 3 words per second. For complex questions, a speech rate faster than 3 words per second was perceived as easier to understand. It may be hard for listeners to hold all of the relevant information about complex questions in working memory when interviewer read question slowly. In addition, for socially undesirable questions, the highest rating of easiness to understand is at a speech rate of 4 words per second. It may be awkward to listen to someone asking sensitive questions with a slow pace. Thus, listeners may find it more pleasant (or at least less awkward) to listen to a sensitive question quickly, and thus perceive a faster speech rate as easier to understand.

Results for the effects of speech rate on rated interviewer personality traits imply that when interviewers follow the survey instructions to read question at the recommended pace of 2 wps, respondents may perceive interviewer personality traits as having low confidence, low reliability, low trustworthiness, low expertise, and low easiness to understand.

Fillers. Listeners rated interviewers who read questions with fewer fillers as more confident, more reliable, more expert, and easier to understand than those who read questions with more fillers. In addition, the negative effect of fillers on rated trustworthiness was significant only for female interviews, and the negative effects of fillers on rated easiness to understand was stronger for socially undesirable questions compared to neutral questions.

Interviewer's demographic and question characteristics. This study only found a direct effect of interviewer sex on rated confidence. Female voices were rated as less confident than male voices; however, the effect is just marginally significant.

Additionally, question types affect rated interviewers' personality traits, especially for socially undesirable questions. Interviewers reading socially undesirable questions were rated as being less confident, less easy to understand, less reliable, less trustworthy, and less expert compared to those reading neutral questions.

Complex and socially desirable questions were rated as less easy to understand than neutral questions. In this study, complex questions include questions that contain ambiguous words such as "Kaninhop" (a Swedish sport involving rabbits jumping over obstacles) in question 13E and amount of minutes spent on a computer on a "typical" day in question 19. It may be hard for respondents to interpret words such as "Kaninhop" and "typical." As such, listeners may perceive complex questions as less easy to understand than other types of questions. In addition, some socially desirable questions in this study include long introductions (e.g., Q8 has a long definition of volunteer activities in the introduction; See Appendix A for question wording). Thus, listeners may perceive this kind of question as hard to understand.

As expected, listeners can perceive interviewers' personality traits from interviewer voice characteristics. As mentioned earlier, the perception of interviewer personality traits may affect data quality. For example, respondents may perceive interviewers who ask questions with higher pitch variation as more reliable, and this may decrease underreports of socially undesirable behaviors. In chapter 4, I examine whether perceived interviewer personality traits mediate the relationship between objective voice characteristics and data quality.

2.5 Limitations

Listeners at different age cohorts judge voices based on different criteria - young people use different standards for those who are from their generation from older people (Ketrow 1990). As such, the results in this study may be affected by the coder's age because coders in this study are more likely to be young. Unfortunately, because the number of coders (6 coders) in this study is quite small, variability of coders is not large enough to be included as a separate level in the multilevel analysis. In addition, a sample of voice files was analyzed in this study. More voice files analyzed will increase statistical power in this study.

CHAPTER 3: OBJECTIVE VOICE CHARACTERISTICS AND DATA QUALITY

Introduction

Many national surveys use telephone interviewing as the primary mode of data collection (Groves, et al. 2009). Telephone interviews are an aural mode, and speech can provide information about the speakers such as their gender and emotional state (Kent & Read 2002). As such, the voice of interviewers can convey more information than simply the meaning of words or sentences themselves (Groves, et al. 2008), and thus may affect survey responses. That is, telephone interviewers' voices might reveal information about interviewers to respondents. However, there is little relevant research on the relationship between interviewer voice characteristics and data quality.

Interviewer training generally focuses on intonation and rate of speech. Interviewers are advised to ask questions with proper phrasing and inflection and to speak at an average rate of two words per second (Guenzel, et al. 1983; Cannell, et al. 1981). Even though these vocal characteristics appear in training protocols, there is no empirical research that examines whether these voice characteristics have an effect on data quality. In this chapter, I examine whether objective measures of interviewer voices predict data quality. In particular, this study examines four interviewer voice characteristics—pitch, intonation, rate of speech, and disfluencies—and how these affect data quality as measured through respondent behaviors, item nonresponse, rounding, and reports on sensitive items.

3.1 Literature review

Telephone interviewers can increase the variance of and lead to systematic biases in survey estimates (West & Olson 2010; Biemer & Lyberg 2003; Kane & Macaulay

1993; Groves 1989; Groves & Magilavy 1986; Kish 1962). The decisions that interviewers make about how to ask questions, probe for clarification, and provide feedback can affect the variance of estimates (Biemer & Lyberg 2003). Even when interviewers read questions as worded, however, interviewer variance can occur if interviewers vary in their voice characteristics when asking questions (Groves 1989). Similarly, interviewers can be a source of bias through an interviewer's experience and traits (Groves, et al. 2009; Biemer & Lyberg 2003). For example, more experienced interviewers may be more careless in interviewing than less experienced interviewers, resulting in measurement error (Olson & Bilgen 2011; Groves, et al. 2009).

Survey interviewing contains two main actors: interviewers and respondents (Japec 2008). Figure 3.1 presents a conceptual model for how interviewer voice characteristics affect data quality. The model applies for an interviewer reading a survey question (the upper part of the model) and the respondents' response process (the lower part of the model).

At the top of Figure 3.1, the question/answering process starts by interviewers understanding the survey questions that they are asking. If each interviewer understands a survey question differently, interviewer variance may increase (Japec 2008; Campanelli, et al. 1991). Then, standardized interviewing procedures require interviewers to read questions as written, and to use standardized methods of clarification and feedback (Fowler & Mangione 1990). Unexplored, but long hypothesized, is that interviewers are a source of measurement error if they differ in voice characteristics while they read survey questions exactly as worded (Groves 1989).

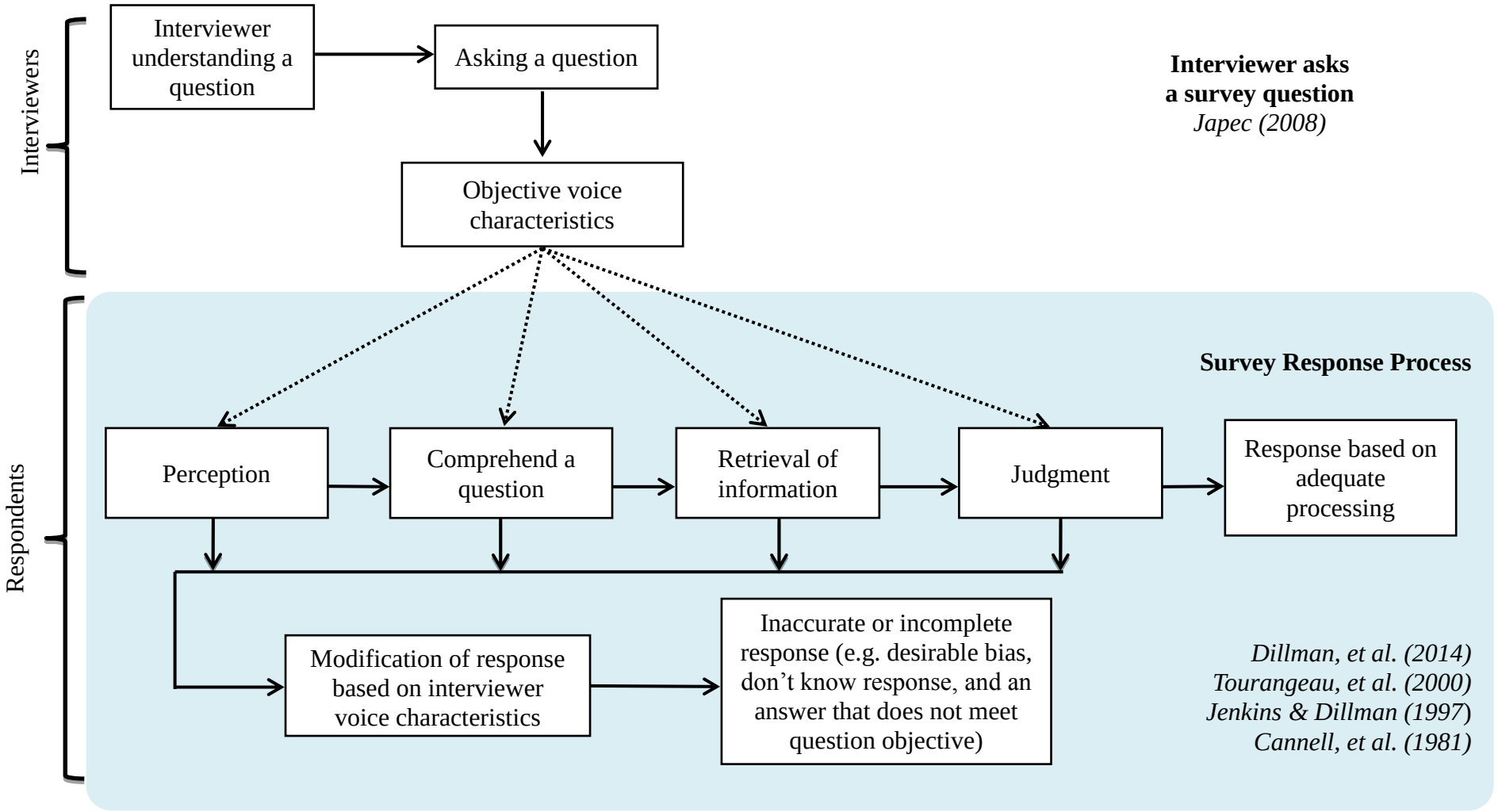


Figure 3.1 The effect of interviewer voice characteristics on data quality

As shown in the lower part of Figure 3.1, to answer survey questions, respondents 1) comprehend questions, 2) retrieve relevant information, 3) make a judgment, and 4) map answers to response options (Tourangeau, et al. 2000). Generally, respondents can take two tracks to engage in these four components. On one track, respondents optimize by thoughtfully following the four elements of the survey response process. On another track, respondents satisfice by modifying their answers and giving inaccurate or incomplete responses (Tourangeau, et al. 2000; Krosnick 1991; Cannell, et al. 1981).

Although Tourangeau and his colleagues (2000) defined comprehending survey questions as the first stage of survey response process, Jenkins and Dillman (1997) argued that there is a perception stage that occurs before the comprehension stage in self-administered surveys. Respondents use preattentive processing to perceive information such as the layout in a questionnaire to detect where they should start on the questionnaire before they comprehend survey questions (Jenkins & Dillman 1997). Thus, perception is a stage in which respondents use their previous knowledge to interpret and respond to survey stimuli.

In telephone surveys, respondents perceive interviewer voice characteristics before they attend to the content of the survey question (Johnson 2003; Kent & Read 2002). That is, interviewer voice characteristics such as pitch, intonation, speech rate and disfluencies are likely to affect respondents' processing of a survey question at the perception stage. These characteristics are also likely to affect respondents' ability to comprehend a question, retrieve relevant information, make judgments, and report answers (Dillman, et al. 2014). I now examine these expectations in more detail.

Interviewer vocal characteristics and data quality

I look at four different interviewer vocal characteristics – pitch, intonation, speech rate, and disfluencies. ***Pitch*** is the fundamental frequency, i.e. the rate of vocal cord vibration (Broome 2012; Benki, et al. 2011; Groves, et al. 2008; Johnson 2003). On average, voice pitch for men is 120 Hertz (Hz) and for women is 210 Hz (Traunmuller & Eriksson 1993). ***Intonation*** is variation in pitch (Oksenberg & Cannell 1988). It refers to a pattern of rises and falls in pitch and the patterns of stress in language (Kent & Read 2002). ***Rate of speech*** is the number of words per unit of time for the speech (Webb 1969). The average adult's speech rate in English is between 2.5 and 3.2 words per second (wps) (Tauroza & Allizon 1990). Males have a slightly faster speaking rate than females (Yuan, et al. 2006). ***Disfluency*** is the parts of speech that are not words but are fillers (e.g., ums and uhs) and pauses (Conrad, et al. 2008). Disfluencies make up about 6% of speech (Bortfeld, et al. 2001).

The association between interviewer voice characteristics and data quality likely varies over types of questions. Socially desirable, undesirable, and complex questions are often asked in telephone surveys and are consistently prone to measurement errors and problems for interviewers and respondents (Tourangeau, et al. 2000; Groves, et al. 2009). Although measurement error is defined as the difference between the true value of a measure and estimates from survey responses (Olson 2006; Biemer & Lyberg 2003; Groves 1989), true values are difficult to obtain. Thus, alternative measures are used for evaluating measurement error.

One measure of data quality comes from the interaction between the interviewer and respondent during the interview. Respondent behaviors can manifest indicating that

respondents had cognitive problems when answering a survey question, thus potentially decreasing data quality (Schaeffer & Dykema 2011, Fowler 2011, Dykema, et al. 1997, Fowler & Cannell 1996, Fowler & Mangione 1990). For example, when a question contains an unclear term, respondents may request clarification, express uncertainty about a question, or give an answer that does not meet question's objective (Fowler 2011, Fowler & Cannell 1996, Fowler 1992). Moreover, respondents who give a qualified answer indicate uncertainty about their final answer (Dykema, et al. 1997). In addition, respondents who interrupt questions with answers will not hear all information that interviewers would like to ask, thus, they may give inaccurate responses or give an answer that does not meet question objectives (Fowler & Cannell 1996, Fowler 1992, Fowler & Mangione 1990).

Item nonresponse, rounding (i.e., reporting units in multiples of 5 and 10), and responses prone to socially desirable bias also have been used as more direct data quality indicators (Tourangeau, et al. 2000). Respondents overreport socially desirable behaviors, such as voting, and underreport socially undesirable behaviors, such as illicit drug use (Kreuter, et al. 2008; Tourangeau & Yan 2007). As such, higher reports of socially undesirable and lower reports of socially desirable behaviors are assumed to be more accurate, i.e. the hypothesis of "more/less is better" (Kreuter, et al. 2008; Tourangeau, et al. 2000). Nonignorable item missingness also may occur for surveys with sensitive questions (Tourangeau & Yan 2007) because people are less likely to respond to socially undesirable items and more likely to respond to socially desirable items (Sakshaug, et al. 2010). Interviewer fixed characteristics (e.g., gender and race) have been found to affect reports to sensitive questions (e.g., Dykema, et al. 2012; Axinn 1991), but the effects of

interviewer vocal characteristics on quality of reports to these questions has not been examined.

Complex questions can create problems for the cognitive response process in which respondents use a satisficing response strategy; they skip or truncate the cognitive response process and provide either incomplete or biased reports, or "don't know" answers (Knauper, et al. 1997; Krosnick 1991). This satisficing results in problematic interviewer and respondent behaviors during interviews (e.g. respondents request for clarification and provide inadequate answers) (Fowler 2011; Schaeffer & Dykema 2011; Schnell & Kreuter 2005; Fowler & Cannell 1996), as well as item nonresponse and rounding (e.g., reporting units in multiples of 5 and 10) (Tourangeau, et al. 2000; Dykema, et al. 1997; Fowler & Cannell 1996).

Socially desirable, undesirable, and complex questions are good candidates for evaluating the association between voice characteristics and data quality because a speaker's vocal pattern changes for these types of questions (Bachorowski 1999). For example, disfluencies increase when discussing an unfamiliar domain and talking in long sentences (Bortfeld, et al. 2001). Because complex questions frequently contain unfamiliar words or have long sentences, interviewers may read complex questions with higher disfluency rates than other types of questions.

Pitch. I expect that pitch will affect respondents at the perception stage, and then affect the mapping stage. A high level of emotional arousal (e.g. fear and anxiety) is associated with an increase in voice pitch (Bachorowski 1999). Thus, interviewers who ask questions with higher pitched voices (accounting for gender) may be perceived as asking questions that are more sensitive than interviewers who ask questions with lower

pitched voices. As such, I expect *more socially desirable reports among interviewers asking socially desirable and undesirable questions with higher pitched voices, compared to those asking the questions with lower pitched voices.* In addition, respondents tend to interrupt questions with answer, give a qualified answer, express uncertainty about a question, and give an answer that does not meet question objective in sensitive questions (Jans 2010; De la Puente & McKay 1995). As such, because asking questions with higher pitched voices are perceived as asking more sensitive questions, I expect *a higher proportion of problematic respondent behaviors among interviewers asking socially desirable and undesirable questions with higher pitched voices, compared to those asking the questions with lower pitched voices*. On the other hand, interviewers who ask questions with higher pitched voices may be perceived as more attractive and are more likely to receive responses from respondents (Ketrow 1990, Oksenberg, et al. 1986). As such, I expect *lower item nonresponse rates among interviewers asking questions with higher pitched voices, compared to those asking questions with lower pitched voices.*

Intonation. Similar to pitch, I expect that intonation will affect respondents at the perception stage, and then affect the mapping stage. Listeners tend to perceive voices with lower pitch variability as less credible (Addington 1968). Respondents are more likely to provide better data quality to interviewers whom they perceive as more credible (Groves 1990). As such, I expect that *interviewers who read questions with higher pitch variation (more intonation) will obtain better data quality (lower item nonresponse, rounding, and responses prone to socially desirable bias) than those who read questions with lower pitch variation*. In addition, respondents are more likely to give an answer that does not meet the question's objective and give qualified answers in sensitive questions

(Jans 2010; De la Puente & McKay 1995). Because respondents are more likely to trust interviewers and are more willing to give better data quality answer to interviewers whom they are perceived as more credible (Groves 1990; Ohanian 1990; Hovland, et al. 1953), I expect that *interviewers whom respondents perceived as more credible (i.e. those who read a question with higher intonation) will obtain lower proportions of problematic respondent behaviors compared to those whom they perceived as less credible (i.e. those who read a question with lower intonation).*

Nonlinear associations between intonation and the data quality indicators may occur. Asking question with too much intonation may result in negative perceptions of the interviewer. This effect may be greater for male interviewers. Males tend to have lower variability in their voice pitch than females (Benki, et al. 2001; Rubin, 1992), and as such, males with high levels of variability in their voice pitch may be perceived as less credible.

Rate of Speaking. Speech rate can affect respondents either at the perception stage or at the other stages. I have two competing hypotheses. Previous paralinguistic studies found that rapid speech is perceived as more credible and more persuasive than slow speech (Smith & Shaffer 1995; Oksenberg & Cannell 1988; Apple, et al. 1979; Miller, et al. 1976; Pearce & Conking 1971). Thus, respondents may perceive interviewers who read questions more quickly as more credible than those who ask questions more slowly. As such, from paralinguistic research, I expect *interviewers who read questions with a faster pace to be more likely to obtain better data quality (lower item nonresponse, rounding, responses prone to socially desirable bias, and proportion of problematic respondent behaviors) compared to those who read questions with a*

slower pace. However, a nonlinear relationship between speech rate and the data quality indicators may occur. Listeners perceive extremely fast speech as less attractive than moderately fast speech (Street, et al. 1983). Thus, I expect that *interviewers who read questions extremely quickly will receive higher item nonresponse rates than those who read questions with a moderate pace.*

On the other hand, survey practice (e.g., Fowler & Mangione 1990; Cannell, et al. 1981) suggests that interviews should be conducted at a slow pace. Fast interviews lead respondents to perceive that quick answers are acceptable, and to not take the time to give accurate and complete answers, thus, decreasing data quality (Fowler & Mangione 1990; Fowler 1966). As such, from survey research, I expect *interviewers who ask questions with faster rates of speaking will obtain lower data quality (higher item nonresponse rates, rounding, and socially (un)desirable answers) than those who ask questions at a slower pace.* Moreover, respondents may not understand a question asked more quickly, and thus request clarification about a question, express uncertainty about a question, and give an answer that does not meet question objective (Ongena & Dijkstra 2007; Fowler & Cannell 1996). As such, I expect that *interviewers who ask questions more quickly will obtain higher proportion of problematic respondent behaviors than those who ask questions more slowly.*

Disfluencies. Similar to speech rate, I have two competing hypotheses for disfluencies. Paralinguistic research (e.g., Ketrow 1990; Miller, et al. 1976) shows that listeners tend to perceive speech with higher rates of disfluencies as less credible compared to speech with lower rates of disfluencies, thus leading to lower perceptions of credibility for interviewers with higher levels of disfluencies. As such, I expect *higher*

item nonresponse rates, rounding, socially (un)desirable answers, and proportions of problematic respondent behaviors among questions read with more disfluencies than those read with fewer disfluencies.

Survey research has shown that disfluencies tend to affect respondents at the comprehension stage. Disfluencies in speech have a "disfluency advantage" allowing respondents to have more time to think about their responses, and thus increase data quality (Brennan & Schober 2001; Bradburn, et al. 1987). Moreover, disfluency rates increase when discussing an unfamiliar domain and talking in long sentences (Bortfeld, et al. 2001), indicating that speakers have difficulties deciding what to say and how to say it (Kidd, et al. 2011; Clark 2002). These fillers can alert listeners that upcoming speech will be complex so that they should pay attention to what speakers will say (Clark & Tree 2002). As such, I expect *lower item nonresponse rates, rounding, and socially (un)desirable answers among questions read with more disfluencies than those read with fewer disfluencies.* In addition, because of the disfluency advantage that allows respondents to have more time to interpret questions and retrieve information, respondents are less likely to give an answer that does not meet the question's objectives or give a qualified answer (Bradburn, et al. 1987). As such, I expect *lower proportions of problematic respondent behaviors among questions read with more disfluencies than those read with fewer disfluencies.*

A nonlinear association between disfluencies and the data quality indicators may occur. Conrad, et al. (2013) found fillers have an inverse-U shaped association with the likelihood of survey participation. Respondents may have negative perceptions to interviewer's voices that either have many disfluencies or have perfect fluency, i.e. no

disfluencies. That is, respondents perceive interviewer speech without fillers, i.e. perfectly fluent, as overly scripted and perceive interviewer speech with lots of fillers as being too relaxed (Benki, et al. 2011). As such, interviewers who read questions with a moderate disfluency rate may receive better data quality than those who read question with either too many disfluencies or no disfluencies.

Interviewer sex and experience may moderate the relationship between interviewer voice characteristics and data quality. Females have higher pitched voices, greater variability of pitch, somewhat slower speech rate, and use fewer fillers relative to males (Yuan, et al. 2006; Bortfeld et al. 2001; Kent & Read 2002). Negative stereotypes can occur if males/females deviate from their expected voice pattern, and this deviation may affect survey responses (Benki, et al. 2001; Rubin 1992). As such, I expect that *interviewers whose voice characteristics deviate from their gender stereotype will receive lower data quality compared to those whose voice characteristics follow their gender stereotype.* Interviewer experience may also moderate the effect of speech rate and disfluencies on data quality. For example, a disfluency advantage may be greater for inexperienced interviewers than experienced interviewers (leading to longer interviews for inexperienced interviewers, Olson & Peytchev 2007). As such, *inexperienced interviewers may obtain better data quality than experienced interviewers*. As such, I examine interaction effects between interviewer voice characteristics and interviewer sex and experience.

Respondents with lower cognitive ability tend to experience greater difficulties holding the questions and response options in working memory, and thus data quality is lessened as respondent cognitive ability declines (Knauper, et al. 1997). Commonly used

proxies for cognitive ability are respondent age and education—data quality tends to be lower for older and less educated respondents compared to younger and more educated respondents (Knauper 1999; Narayan & Krosnick 1996; Groves 1989). Thus, age and education are included as control variables.

3.2. Data and methods

3.2.1 Data

Data for this study come from the Work and Leisure Today Survey conducted in July and August 2013 by AbtSRBI. It is a landline RDD CATI survey administered by 22 interviewers with 450 completed interviews. To increase the stability of the analyses, interviewers who conducted fewer than 10 interviews are eliminated from the study (Olson & Peytchev 2007). As such, I analyze 432 interviews conducted by 19 interviewers (9 female and 10 male interviewers). Each of the interviews was transcribed and behavior coded as part of an ongoing NSF grant (NSF SES-1132015).

The questionnaire has 54 questions. I considered 24 candidate questions including socially desirable, socially undesirable, complex, and neutral (not complex and not socially desirable/undesirable) questions. Neutral questions serve as a comparison for both socially (un)desirable and complex questions. I selected twelve questions of these 24 questions (Q8, Q13A, and Q21F for socially desirable questions; Q5, Q21C, and Q21D for socially undesirable questions; Q13E, Q19, and Q20 for complex questions; Q2, Q13D, and Q21A for neutral questions) based on the criteria that questions contain both item nonresponse and item response and sufficient variability in responses. The wording of the twelve questions appears in Appendix A. The topics of these questions include employment status and volunteer work, activities for leisure such as using internet and

exercise, and substance use such as drinking alcohol. Due to the selection of 12 questions from 24 candidate questions and omitting voice files that had bad quality (n=42) or contained respondents interrupting interviewers when they read the question (n=40), a sample of 4,689 voice files of interviewers reading individual questions were analyzed in this study.

3.2.2 Measures of voice characteristics

All of the measures of interviewer voice characteristics come from the interviewer's reading of a survey question. If the question was read twice, only the first reading was used. Praat (www.fon.hum.uva.nl/praat/), a speech analysis program, was used to extract voice information about pitch, intonation, and total duration time of the first turn that an interviewer read the survey question. Pitch is extracted by using the automated pitch measurement technique (Boersma & Weenink 2014). In this study, the mean pitch over the question reading is used as the measure of pitch and the standard deviation of the pitch over that question reading is used as the measure of intonation. In addition, total duration of time is used to calculate the rate of speaking defined as the number of words in a turn per second. Number of words and fillers are extracted from interview transcripts read into the Sequence Viewer program.[3] Number of fillers including Um, Uh, Ah, Mm, Hm, and Oh are used as the measure of disfluencies.

3.2.3 Data quality analysis

I extracted four objective measures of interviewer voice characteristics including pitch, intonation, rate of speech, and number of fillers from the sampled 4,689 voice files.

[3] Sequence Viewer is a program used to observe and analyze sequential data such as interviewer-respondent interactions in which each question can indicate the beginning of a new sequence (Dijkstra 2009).

Each respondent was interviewed by one interviewer, and interviewers obtain responses from multiple respondents. Thus, multi-level modeling is used for analysis (O'Muircheartaigh & Campanelli 1998; 1999). I examined four data quality indicators: 1) respondent behaviors associated with data quality, 2) item nonresponse, 3) rounding, and 4) the directional "more/less is better" hypotheses- higher reports of the socially undesirable behaviors and lower reports of socially desirable behaviors indicate better data quality. Respondent behaviors and item nonresponse were analyzed for all 12 questions in one model each. Rounding was used as a data quality indicator only for a subset of the neutral and complex questions and the directional hypotheses was used as a data quality indicator only for socially (un)desirable questions. As such, rounding and the directional hypotheses were analyzed for each question separately (See Table C.1 in Appendix C for the summary of data quality indicators analyzed for each question and Table C.2 for the descriptive statistics of data quality indicators). For each of the data quality indicators, I also tested for a significant interviewer variance effect through a base model using a mixture of chi-square distributions (Rabe-Hesketh & Skrondal 2012).

Dichotomous variables of interviewer sex and experience and respondent age and education were included in the models—interviewer experience: one year of experience at the survey organization or less (25%) and more than one year of experience (75%); respondent age: respondents whose age is 60 years old or less (47%) versus greater than 60 years old (53%); and respondent education: respondents whose education is high school or less (30%) versus higher than high school (70%).

Respondent behavior and item nonresponse analyses

There are six dependent variables available for all of the question readings for each respondent. The first set of dependent variables are a set of five respondent behaviors previously found to be associated with data quality (Schaeffer & Dykema 2011; Dykema, et al. 1997): the respondent 1) interrupts questions with an answer, 2) expresses uncertainty about a question, 3) requests clarification, 4) gives qualified answers, and 5) gives a response that does not meet the question's objective (See Kirchner & Olson (2014) for more detail about behavior coding process). Each of the behaviors is coded for whether or not it occurred at least once during the interviewer-respondent interaction for each question. Additionally, item nonresponse was examined by using a dichotomous variable where 1 indicates the respondent did not answer the item and 0 indicates they did answer.

For these six dependent variables, there are three sources of variation - questions, respondents, and interviewers. Thus, a three-level multilevel model was estimated. A logit link with a binary distribution was used to estimate the following model.

$$logit(y_{tij}) = \gamma_{000} + \gamma_{001}IChar_j + \gamma_{010}RChar_{ij} + \sum_{a=1}^{A}\gamma_{a00}VoiceChar_{aij}$$
$$+ \sum_{a=1}^{A}\gamma_{a01}IChar_j \times VoiceChar_{aij} + \sum_{b=A+1}^{B}\gamma_{b00}QChar_{bij}$$
$$+ \sum_{c=B+1}^{C}\gamma_{c00}VoiceChar_{ij} \times QChar_{ij} + V_{00j} + U_{0ij} + e_{tij}$$

where y_{tij} = respondent behaviors and item nonresponse,

$VoiceChar_{ij}$ = pitch (centered at 165 Hz), intonation (centered at 40 Hz), rate of speaking (centered at 3.5 wps), and number of fillers,

$QChar_{ij}$= neutral (reference group), complex, socially undesirable, and socially desirable questions,

$IChar_j$ = interviewer's sex and experience,

$RChar_j$ = respondent's age and education,

V_{ooj} = random interviewer effect,

U_{0ij} = random respondent effect, and

e_{tij} = residual.

Rounding and directional hypotheses

In addition to respondent behaviors and item nonresponse, I examined two other data quality indicators. First, rounding is measured using a dichotomous variable indicating whether the respondent gave a prototypical answer (e.g., 5, 10, 15 or multiples of 60 minutes for question 19 asking about number of minutes that respondents spend on computer on a typical day). Second, the directional "more/less is better" hypotheses are tested using the responses to the survey questions (e.g., engage in socially undesirable behaviors at least one time or engage in socially desirable behaviors few times); the "yes" response is predicted for the socially undesirable questions (e.g., respondents answered "yes" they have ever been fired from a job) and the "no" response is predicted for the socially desirable questions (e.g., respondents answered "no" they have not done any volunteer activities in the last 12 months) (See Table C.1 in Appendix C for more detail).

Each item was modeled separately. Two-level hierarchical models predicting data quality were estimated to account for variation due to respondents and interviewers (O'Muircheartaigh & Campanelli 1999). Because the data quality indicators for the directional hypotheses and rounding are dichotomous variables, a logit link with a binary distribution was used to estimate the models. In general, using the terms defined above, the model is:

$$logit(data\ quality\ indicator_{ij}) = \gamma_{00} + \gamma_{01} Ichar_j + \sum_{a=1}^{A} \gamma_{a0} VoiceChar_{ij}$$

$$+ \sum_{a=1}^{A} \gamma_{a1} VoiceChar_{ij} \times Ichar_j + \sum_{b=A+1}^{B} \gamma_{b0} RChar_{ij} + V_{0j} + e_{ij}$$

As mention earlier, because interviewer voice characteristics may have a nonlinear relationship with the data quality indicators, I also examined the squared terms of pitch, intonation, speech rate, and number of fillers.

3.3 Results

3.3.1 Descriptive statistics for pitch, intonation, rate of speaking, and fillers

The average pitch, intonation, rate of speech, and number of fillers is around 168 Hz, 42 Hz, 3.5 wps, and 0.2 fillers per voice file respectively. Table 3.1 presents descriptive statistics on pitch, intonation, rate of speech, and number of fillers by interviewer sex and interviewer experience. The differences in interviewer voice characteristics by interviewer sex and interviewer experience were all significant. Male interviewers ask questions with lower pitched voices ($t=-76.67$, $p<0.01$), with flatter intonation ($t=-25.84$, $p<0.01$), with a faster rate of speaking ($t=7.02$, $p<0.01$), and use more fillers ($t=10.46$, $p<0.01$) than female interviewers. Interviewers whose experience is 1 year or less read questions with higher pitched voices ($t=2.66$, $p<0.01$), with flatter intonation ($t=-5.15$, $p<0.01$), with a slower rate of speaking ($t=-4.03$, $p<0.01$), and use fewer fillers ($t=-13.20$, $p<0.01$) than those whose experience is greater than 1 year.

Table 3.2 shows Pearson product-moment correlations between the objective measures of interviewer voice characteristics including pitch, intonation, speech rate, and disfluencies. The correlation between intonation and pitch is quite high ($r=0.56$, $p<0.01$). This may be because intonation is defined as the variability of pitch. The correlations between speech rate and pitch, disfluencies and pitch, and speech rate and intonation are

Table 3.1 Descriptive Statistics on Pitch, Intonation, Rate of Speech, and Number of Fillers by Interviewer Sex and Interviewer Experience

	Interviewer Sex						**Interviewer Experience**					
	Male			**Female**			**Experience 1 Year or Less**			**Experience Greater than 1 Year**		
Objective voice characteristics	**Mean *(S.D.)***	**Min**	**Max**	**Mean *(S.D.)***	**Min**	**Max**	**Mean *(S.D.)***	**Min**	**Max**	**Mean *(S.D.)***	**Min**	**Max**
Pitch (Hz)	138.23 *(21.94)*	89.38	250.04	198.11 *(30.67)*	126.98	315.11	170.64 *(46.5)*	89.38	315.11	166.64 *(37.6)*	90.08	312.82
Intonation (Hz)	34.82 *(17.95)*	3.86	160.58	48.89 *(19.29)*	5.86	118.72	39.24 *(18.5)*	3.87	160.58	42.54 *(20.3)*	5.86	137.58
Rate of speech (wps)	3.59 *(0.97)*	0.55	6.89	3.40 *(0.91)*	0.62	6.72	3.40 *(0.96)*	0.62	6.48	3.53 *(0.94)*	0.54	6.89
Filler	0.27 *(0.53)*	0	4	0.13 *(0.42)*	0	4	0.07 *(0.31)*	0	4	0.24 *(0.52)*	0	4

Table 3.2 Pearson's correlations between the objective interviewers' voice characteristics

	Objective measures of interviewer voice characteristics							
Objective measures of interviewer voice characteristics	**Pitch**		**Intonation**		**Speech rate**		**Disfluencies**	
Pitch	1.00		0.56	**	-0.09	**	-0.08	**
Intonation	0.56	**	1.00		-0.07	**	0.03	
Speech rate	-0.09	**	-0.07	**	1.00		-0.01	
Disfluencies	-0.08	**	0.03		-0.01		1.00	

Note **p<0.01, *p<0.05

quite low ($|r|<0.1$, $p<0.01$) indicating a weak correlation. That is, they are not likely to vary together, e.g. an increase in voice pitch is weakly associated with a decrease in speech rate ($r=-0.09$, $p<0.01$). In addition, there is no association between intonation and disfluencies and speech rate and disfluencies.

3.3.2 Respondent behaviors

I examined five respondent behaviors associated with data quality including interrupting questions with answers (10% of voice files), expressing uncertainty about a question (11% of voice files), requesting clarification about a question (20% of voice files), giving a qualified answer (14% of voice files), and giving a response that does not meet the question's objective (22% of voice files). Table 3.3 shows results from the hierarchical logistic regression models of interviewer voice characteristics predicting these five problematic respondent behaviors. Interviewer and respondent random effects were significant only for interrupting questions with answers and giving a response that does not meet the question's objectives ($p<0.01$) (See Table C.3 in Appendix C).

Vocal Characteristics and Question Characteristics. Overall, interviewer voice characteristics inconsistently affect respondent behaviors. Pitch affects only one respondent behavior- the respondent giving a response that does not meet the question's objective (coefficient= 0.007, $p<0.01$). As expected, respondents are more likely to give a response that does not meet the question objective to interviewers who read a question with higher pitched voices compared to those who read a question with lower pitched voices. There is no significant association between pitch and interviewer and question characteristics.

Table 3.3 Hierarchical logistic model predicting respondent behaviors by objective voice characteristics

	Interrupt questions with answers		Express uncertainty about a question		Request clarification about a question		Give a qualified answer		Give a response that does not meet the question's objective	
	coefficient (SE)		coefficient (SE)		coefficient (SE)		coefficient (SE)		coefficient (SE)	
Main effects										
Intercept	-2.053(0.23)	**	-3.174(0.28)	**	-2.058(0.18)	**	-1.669(0.15)	**	-0.846(0.14)	**
Pitch	0.002(0.003)		-0.001(0.003)		-0.001(0.003)		0.001(0.002)		0.007(0.002)	**
Intonation	-0.001(0.004)		-0.0001(0.007)		0.001(0.005)		-0.002(0.003)		-0.005(0.003)	
Intonation^2			0.0005(0.0001)	**	0.0004(0.0001)	**				
Speech rate	0.513(0.17)	**	0.379(0.21)		0.205(0.14)		0.308(0.13)	*	0.182(0.12)	
Speech rate^2							-0.141(0.05)	**		
Fillers	-0.076(0.11)		-0.04(0.14)		-0.084(0.10)		0.110(0.12)		-0.295(0.16)	*
Interviewer's experience < 1 year	0.153(0.28)		-0.203(0.35)		-0.099(0.20)		0.072(0.14)		0.139(0.17)	
Female interviewer	-0.104(0.30)		0.270(0.36)		0.276(0.23)		0.211(0.18)		-0.504(0.19)	*
Desirable question	0.503(0.14)	**	-0.099(0.20)		0.218(0.12)		-0.650(0.13)	**	-0.069(0.11)	
Complex question	0.779(0.15)	**	1.822(0.17)	**	1.597(0.12)	**	0.455(0.12)	**	0. 935(0.11)	**
Undesirable question	-0.723(0.19)	**	-0.864(0.26)	**	-1.096(0.17)	**	-1.185(0.16)	**	-0.867(0.13)	**
R whose education is high school or less	-0.068(0.14)		0.156(0.14)		0.276(0.23)		0.023(0.13)		0.341(0.11)	**
R whose age is 60 or less	-0.684(0.13)	**	-0.243(0.13)		-0.012(0.10)		0.093(0.12)		-0.652(0.10)	**
Interaction btw voice and Q char										
Intonation*desirable question			-0.002(0.01)		-0.013(0.01)	*				
Intonation*complex question			-0.021(0.01)	**	-0.020(0.01)	**				
Intonation*undesirable question			-0.023(0.01)	*	-0.023(0.01)	**				

Note **p<0.01, *p<0.05; n = 4,689

Table 3.3 Continued.

	Interrupt questions with answers		Express uncertainty about a question		Request clarification about a question		Give a qualified answer		Give a response that does not meet the question's objective	
	coefficient (SE)		**coefficient (SE)**		**coefficient (SE)**		**coefficient (SE)**		**coefficient (SE)**	
Speech rate*desirable question	-0.223(0.22)		-0.379(0.30)		-0.2880.19)		0.153(0.20)		0.096(0.16)	
Speech rate*complex question	-0.388(0.19)	*	-1.576(0.22)	**	-1.08(0.15)	**	-0.155(0.15)		0.203(0.13)	
Speech rate*undesirable question	-0.879(0.23)	**	-0.982(0.29)	**	-0.691(0.19)	**	-0.508(0.19)	**	-0.540(0.15)	**
Interaction btw voice and Iwer char										
Speech rate*inwer exp < 1 yr					0.312(0.11)	**				
Fillers*inwer exp < 1 yr					0.514(0.26)	*				
Fillers*female interviewer							-0.611(0.23)	**	0.470(0.18)	**
Variance components										
2-level variance (respondents)	0.423(0.10)	**	0.175(0.11)		0.144(0.06)		0.589(0.09)		0.392(0.07)	**
3-level variance (interviewers)	0.212(0.11)	**	0.342(0.16)		0.096(0.05)		0		0.049(0.04)	**
Residual variance	3.29		3.29		3.29		3.29		3.29	
Model fit										
Generalized Chi-square	3737.2		4206.52		4346.29		3742.15		4033.17	

Note **p<0.01, *p<0.05; n = 4,689; the variance components from the base model is presented in Table C.3. in Appendix C.

As shown in Table 3.3, intonation affects only some respondent behaviors. Intonation has a U-shaped association with respondents expressing uncertainty about a question (coefficient for intonation=-0.0001, p=0.99; coefficient for intonation^2=0.0005, p<0.01) and requesting clarification about a question (coefficient for intonation=0.001, p=0.85; coefficient for intonation^2=0.0004, p<0.01). There is no association between intonation and interviewer characteristics.

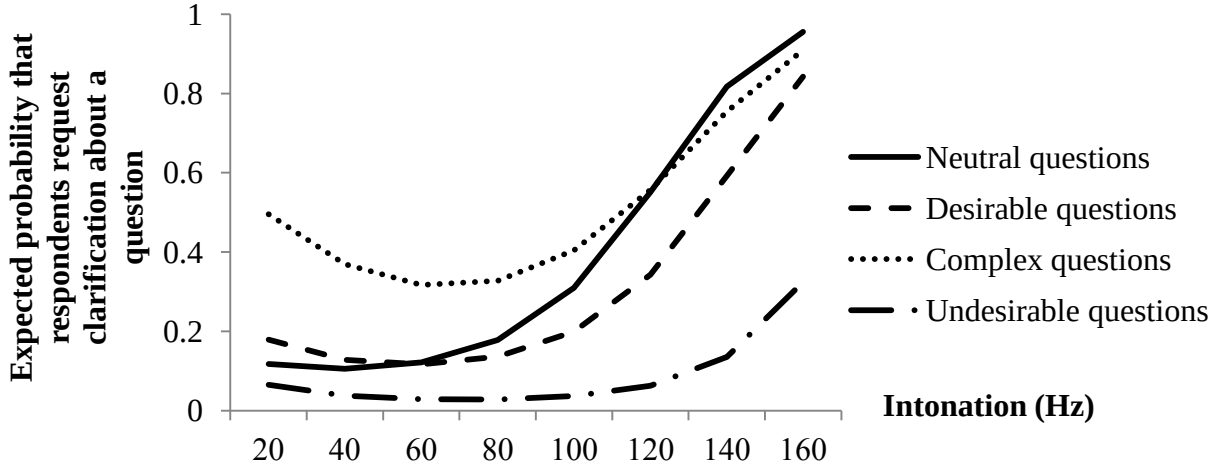


Figure 3.2 Expected probability that respondents request clarification about a question by intonation

Question characteristics moderate the effect of intonation on respondents requesting clarification and expressing uncertainty about a question. For example, Figure 3.2 shows the U-shaped association between intonation and expected probability that respondents **request clarification about a question**. Respondents are less likely to request clarification about a question when interviewers read a question with moderate intonation (intonation between 40 and 80 Hz). The strength of this association varies across question types, with the strongest association between intonation and clarification requests being for neutral questions. The direction of the effects of intonation on respondents **expressing uncertainty about a question** was the same for all question

types, but the effects were strongest in neutral questions. The same U-shaped relationship holds for respondents expressing uncertainty about a question as for clarification requests. That is, respondents are less likely to express uncertainty about a question when interviewers read a question with moderate intonation (intonation between 40 and 80 Hz).

For speech rate, as an interviewer reads a survey question faster, the proportions of respondent **interrupting questions with answers** and **giving a qualified answer** significantly increase, although this effect varies across question type.

While speech rate is *positively associated* with respondent behaviors in **neutral** questions (i.e., consistent with the hypothesis from survey practice), it is *negatively associated* with respondent behaviors in **socially undesirable** questions (i.e., consistent with the hypothesis from paralinguistic research). Respondents are more likely to engage in any of these five behaviors when interviewers read **socially undesirable questions** with a speech rate of 2 words per second (wps) compared to a faster pace. For example, as interviewers' speech rate increases, the expected probability of respondents giving a qualified answer across speech rate decreases for socially undesirable questions, but increases for all other question types (Figure 3.3).

In **complex questions**, the effects of speech rate on respondent behaviors are mixed. Figure 3.4 shows the effects of speech rate on respondents expressing uncertainty about a question, requesting clarification, and interrupting questions with answers in complex questions. Speech rate is negatively associated with respondents expressing uncertainty about a question and requesting clarification about a question. Respondents are *more likely* to express uncertainty about a question and request clarification about a question when interviewers read complex questions more slowly. In contrast, respondents

are *less likely* to interrupt complex questions with answers when interviewers read complex questions more slowly.

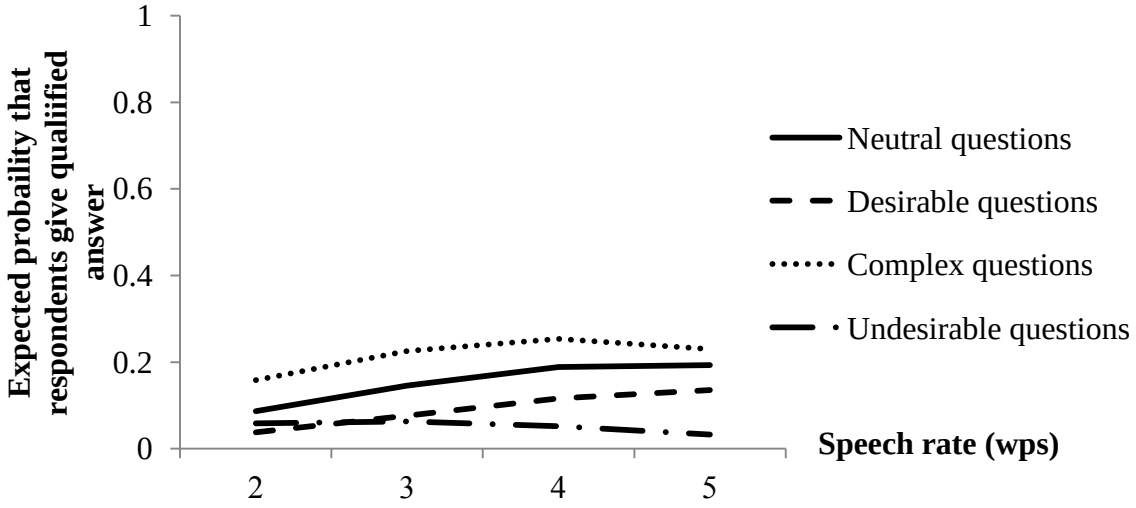


Figure 3.3 Expected probability that respondents give a qualified answer by speech rate and question type

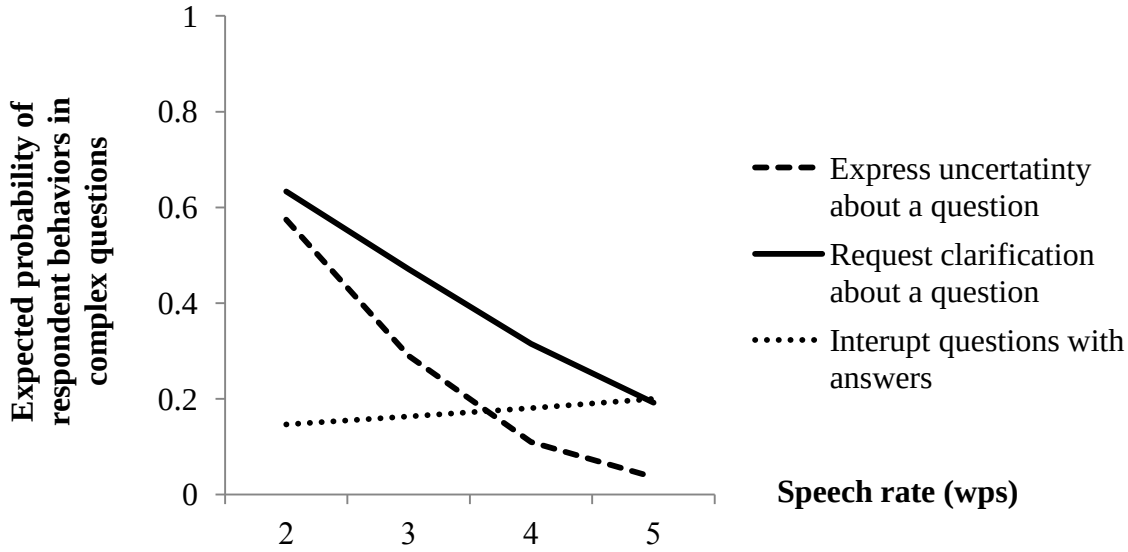


Figure 3.4 Expected probabilities that respondents express uncertainty about a question, request clarification, and interrupt questions with answers in complex questions by speech rate

Question Characteristics. Compared to neutral questions, all five respondent behaviors were more prevalent in complex questions but less prevalent in socially undesirable questions ($p<0.01$, Table 3.3).

Interviewer Characteristics. In addition to question characteristics, interviewer characteristics moderate the effect of voice characteristics on some respondent behaviors. Interviewer experience moderates the effect of **speech rate** and **fillers** on respondents **requesting clarification about a question**, with the effects found only for inexperienced interviewers. Respondents are more likely to request clarification about a question from inexperienced interviewers who read questions with a faster pace (coefficient=0.312, $p<0.01$) and use more fillers (coefficient=0.514, $p=0.04$) compared to those who read questions slower and use fewer fillers.

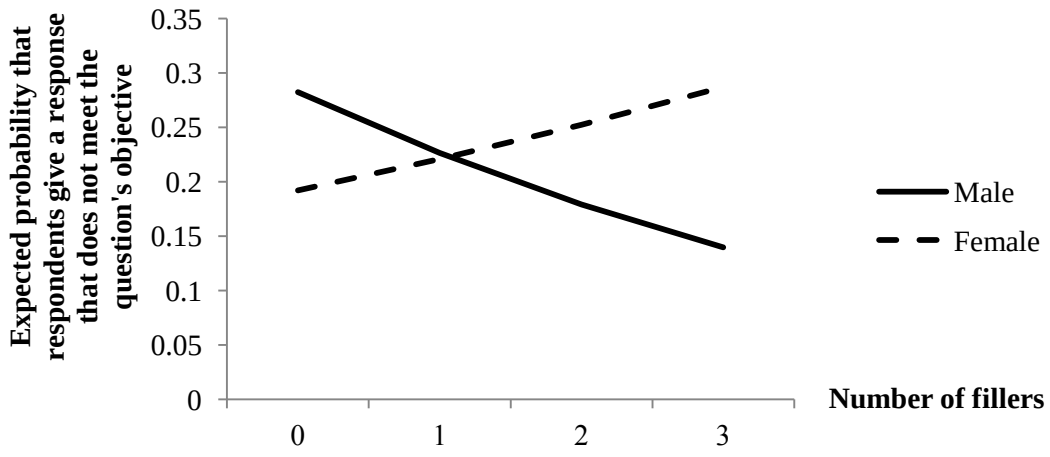


Figure 3.5 Expected probability that respondents give a response that does not meet the question's objective by fillers

Regarding interviewer sex, the effect of **fillers** on respondents **giving a qualified answer** was only found for female interviewers. Respondents are less likely to give a qualified answer to female interviewers who use more fillers compared to those who use fewer fillers (coefficient=-0.611, $p<0.01$). Moreover, the effect of **fillers** on respondents giving a response that **does not meet the question's objective** varied by interviewer's sex. As shown in Figure 3.5, respondents are less likely to give a response that does not meet the question's objective (coefficient=-0.295, $p=0.01$) to male interviewers when

they use more fillers compared to when they use fewer fillers. The opposite effect was found for female interviewers (coefficient=0.470, p<0.01).

Respondent Characteristics. Respondent characteristics affect some respondent behaviors. Respondents whose education is high school or less are more likely to give a **response that does not meet the question's objective** compared to those whose education is greater than high school (coefficient=0.341, p<0.01). In addition, **interrupting questions with answers and giving a response that does not meet the question's objective** are less prevalent among respondents whose age is 60 or less compared to those whose age is greater than 60 years old (p<0.01, Table 3.3).

3.3.3 Item nonresponse

The overall item nonresponse rate across these 12 questions is 4.6 percent. Overall, 1.7 percent of the variance on the item nonresponse rate was due to differences between respondents (χ^2=0.24, p=0.31) and 11 percent of the variation in the item nonresponse rate was due to differences between interviewers (χ^2=5.59, p<0.01) (See Table C.4 in Appendix C). Table 3.4 shows the results of analyses that examined the extent to which interviewer voice characteristics predict item nonresponse.

Table 3.4 Hierarchical logistic model predicting item nonresponse by objective voice characteristics

	coefficient (SE)	
Main effects		
Intercept	-3.271(0.33)	**
Pitch	-0.001(0.004)	
Intonation	-0.013(0.006)	*
Intonation^2	0.0003(0.0001)	*
Speech rate	-0.433(0.10)	**
Fillers	0.032(0.18)	
Interviewer's experience < 1 year	0.376(0.43)	
Female interviewer	-0.402(0.45)	
Desirable question	-1.290(0.29)	**

Complex question	0.399(0.19)	*
Undesirable question	0.164(0.22)	
R whose education is high school or less	0.091(0.169)	
R whose age is 60 or less	-0.089(0.16)	
Interaction btw voice and Iwer char		
Speech rate*inwer exp < 1 yr	-0.313(0.15)	*
Variance components		
2-level variance (respondents)	0.203(0.15)	
3-level variance (interviewers)	0.510(0.24)	**
Residual variance	3.29	
Model fit		
Generalized Chi-square	3827.22	

Note **p<0.01, *p<0.05; n = 4,689; the variance components from the base model is presented in Table C.4. in Appendix C.

In contrast to the hypotheses, interviewer pitch and fillers do not affect the item nonresponse rate. As shown in Table 3.4, however, interviewer intonation and speech rate *are* associated with item nonresponse. A curvilinear relationship between intonation and item nonresponse was found, displayed in Figure 3.6 (coefficient= -0.013, p=0.02 for intonation; coefficient= 0.0002, p=0.03 for intonation^2). The expected item nonresponse rates are lowest when interviewers read questions with moderate intonation (around 40-80 Hz). There is no association between intonation and interviewer and question characteristics.

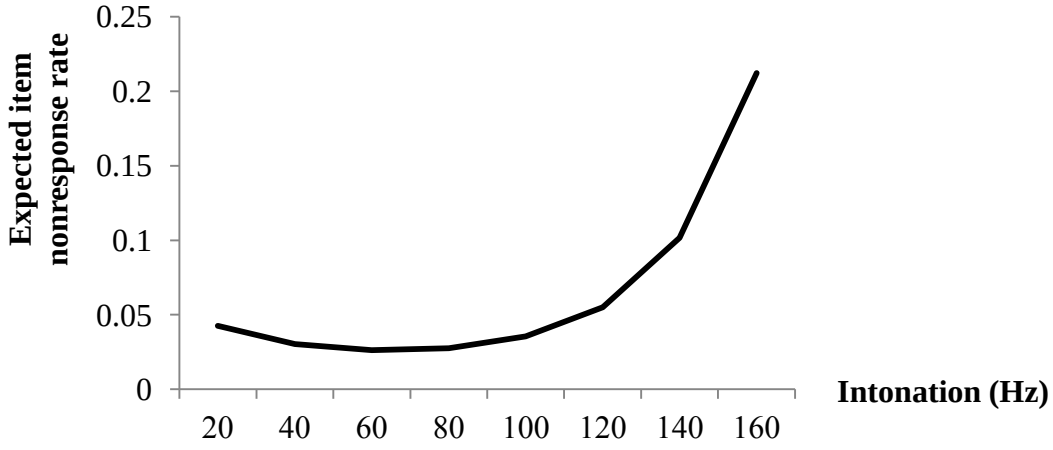


Figure 3.6. Expected item nonresponse rate by intonation

As shown in Table 3.4, speech rate negatively affects item nonresponse rates, but the effect is moderated by interviewer experience. As shown in Figure 3.7, consistent with the hypothesis from paralinguistic research, the item nonresponse rate decreases for all interviewers as they speak more quickly. Experienced interviewers who read questions with a faster pace have lower item nonresponse rates than those who read questions with a slower pace (coefficient= -0.433, $p<0.01$), but the effect is stronger for inexperienced interviewers as expected (Figure 3.7). There is no association between speech rate and question characteristics.

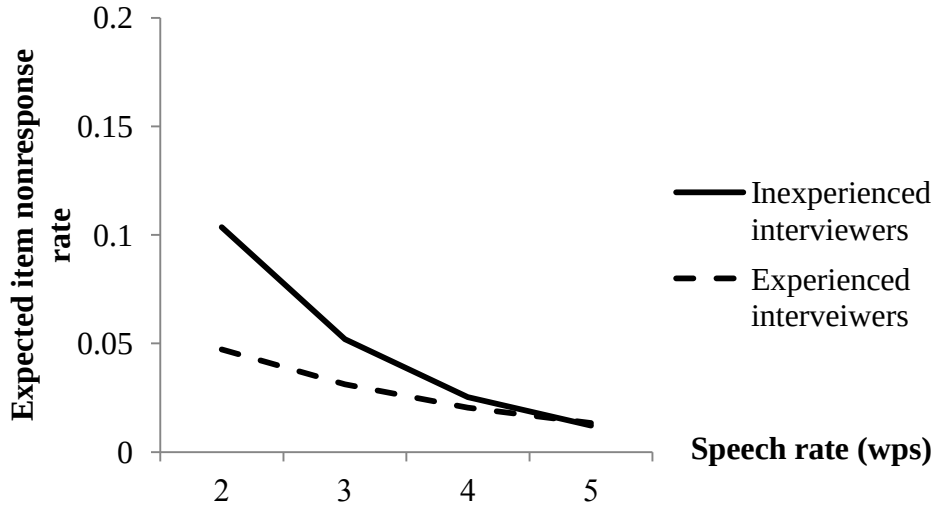


Figure 3.7 Expected item nonresponse rate by speech rate and interviewer experience

As expected, compared to neutral questions, item nonresponse rates were lower in socially desirable questions (coefficient=-1.29, $p<0.01$) and higher in complex questions (coefficient=0.399, $p=0.04$). Unexpectedly, there was not a significant interaction effect between question characteristics and interviewer voice characteristics on item nonresponse. Additionally, there is no statistically significant association between respondent age or education and item nonresponse.

3.3.4 Rounding

Neutral questions. Rounding (e.g., reporting units in multiples of 5 and 10) was used to evaluate data quality in one neutral question, question 21A asking about the number of times that respondents used the internet in the past seven days. On average, 48 percent of respondents rounded their answers on Q21A, and 5.5 percent of the variance in rounding answers resulted from variation across interviewers (χ^2=2.81, p=0.046) (Table C.4 in Appendix C). Results in Table 3.5 show that interviewer voice characteristics and interviewer demographics did not affect the probability that respondents rounded their answers. Respondent age was significantly related to the likelihood of rounding answers on Q21A. Unexpectedly, respondents whose age is 60 or less are more likely to round their answers compared to those who are older than 60 (coefficient= 1.151, p<0.01).

Complex questions. Rounding was also used as a data quality indicator for questions 19 (number of minutes spent on a computer on a typical day) and 20 (number of email messages written or received in the past seven days). Rounding was defined for Q19 as multiples of 60 minutes and for Q20 as multiples of five. On Q19, 56 percent of respondents rounded their answers, and 1.02% of the variance in rounded answers resulted from variation across interviewers (χ^2=0.15, p=0.35). On Q20, 77 percent of respondents rounded their answers, and 6.4% of the variance in rounded answers resulted from variation across interviewers (χ^2=3.01, p=0.04) (Table C.4 in Appendix C).

Results in Table 3.5 shows that speech rate affects the probability that respondents rounded their answers on question 19 but not question 20. However, the effect of speech rate on the probability that respondents rounded their answers on question 19 varied by interviewer sex. As shown in Figure 3.8, male interviewers who

read Q19 with a faster pace obtained fewer rounded answers than those who read the question with a slower pace (coefficient=-0.531, p=0.03). In contrast to male interviewers, female interviewers who read Q19 with a faster pace obtained higher rates of rounded answers than those who read the question with a slower pace (coefficient=0.685, p=0.03). The effect of speech rate for female interviewers (coefficient=0.577, p=0.03) holds when analyzing Q19, Q20 and Q21A together in a hierarchical logistic model (see Table D.1 in Appendix D).

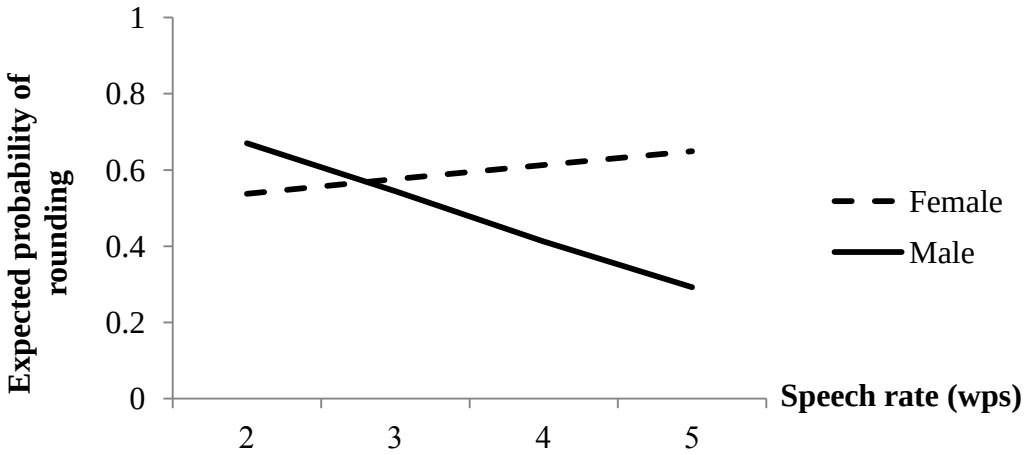


Figure 3.8 Expected probability of rounding answers in question 19 by speech rate and interviewer sex

Respondent characteristics affect the probability of rounding their answers; however, they were in the opposite directions from what I hypothesized. Respondents whose education is high school or less are less likely to round their answers on question 19 and question 20 compared to those whose education is higher than high school (coefficient= -0.682, p=0.03 for question 19; coefficient= -1.283, p<0.01 for question 20). In addition, respondent age was significantly related to likelihood of rounding answers on question 20. Unexpectedly, respondents whose age is 60 or less are more likely to round their answers compared to those who are older than 60 (coefficient= 0.828, p=0.02). I explore this more in the discussion section.

Table 3.5 Hierarchical logistic model predicting proportion of rounded answers by objective voice characteristics

	Q21A (Number of times using the internet)		Q19 (Number of minutes spending on a computer)		Q20 (Number of email messages)	
	coefficient (SE)		coefficient (SE)		coefficient (SE)	
Main effects						
Intercept	-0.520(0.36)		0.301(0.28)		1.526(0.38)	**
Pitch	0.002(0.007)		-0.006(0.005)		0.007(0.007)	
Intonation	0.009(0.01)		-0.001(0.007)		-0.010(0.01)	
Speech rate	0.132(0.29)		-0.531(0.24)	*	0.259(0.31)	
Fillers	-0.056(0.22)		0.018(0.32)		-0.014(0.51)	
Interviewer's experience < 1 year	0.295(0.37)		-0.165(0.29)		-0.083(0.44)	
Female interviewer	-0.435(0.50)		0.472(0.39)		-0.37(0.54)	
R whose education is high school or less	-0.446(0.31)		-0.682(0.28)	*	-1.283(0.33)	**
R whose age is 60 or less	1.151(0.26)	**	0.068(0.25)		0.828(0.31)	*
Interaction btw voice and Iwer char						
Speech rate*female interviewers			0.685(0.34)	*		
Variance components						
2-level variance (interviewers)	0.098(0.15)	**	0		0.186(9.21)	*
Residual variance	3.29		3.29		3.29	
Model fit						
AIC	379.17		408.65		300.57	
n	279		293		287	

Note **p<0.01, *p<0.05; the variance components from the base model is presented in Table C.5. in Appendix C.

3.3.5 The Hypotheses of More/Less is Better

Socially undesirable questions. The directional hypothesis of "more is better," measured through the proportion of "yes" or "at least once" responses, is used as a data quality indicator in socially undesirable questions. Question 5 asks about being fired from jobs, question 21C asks about drinking alcohol, and question 21D asks about sexual behaviors. Respondents who answered that they have engaged in those sensitive behaviors at least one time were considered as providing answers that are less prone to socially desirable bias. Approximately 19 percent of respondents answered that they have ever been fired from a job, 35 percent answered that they had at least one alcoholic drink in the past seven days, and 28 percent reported that they had sex at least one time in the past seven days. There is not a significant interviewer variance effect ($p>0.05$) for these three socially undesirable questions (Table C.6 in Appendix C).

Table 3.6 presents the results of analyses that examined the extent to which interviewer voice characteristics predict better data quality on these sensitive questions. There is no association between interviewer vocal characteristics and responses to Q5 (being fired from a job) and Q21C (alcoholic drink). As shown in Table 3.6, speech rate affects data quality for Q21D (having sex), however, the effect varies by interviewer sex. Figure 3.9 presents the expected probability that a respondent reports that they had sex at least one time in the past seven days by rate of speech and interviewer sex. Respondents are *more likely* to say that they had sex at least one time in the past seven days, i.e. less prone to socially desirable bias, to male interviewers who speak quickly but *less likely* to state this for female interviewers who speak quickly.

Table 3.6 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by objective voice characteristics for socially undesirable questions

	Q5 (Fired from a job)		Q21C (1+ alcohol drinks)		Q21D (Have sex 1+ times)	
	coefficient (SE)		coefficient (SE)		coefficient (SE)	
Main effects						
Intercept	-2.068(0.40)	**	-0.798(0.25)	**	-1.932(0.36)	**
Pitch	-0.001(0.01)		-0.003(0.005)		0.009(0.01)	
Intonation	-0.011(0.01)		0.007(0.006)		-0.005(0.01)	
Speech rate	0.125(0.17)		0.021(0.13)		0.638(0.28)	*
Fillers	0.111(0.813)		-0.035(0.28)		-0.663(0.48)	
Interviewer's experience < 1 year	0.115(0.30)		0.045(0.25)		0.594(0.28)	*
Female interviewer	0.263(0.41)		0.229(0.34)		0.207(0.46)	
R whose education is high school or less	-0.058(0.29)		-0.298(0.23)		-0.636(0.28)	*
R whose age is 60 or less	0.527(0.26)		0.380(0.21)		1.415(0.26)	**
Interaction btw voice and Iwer char						
Speech rate*female interviewers					-0.785(0.35)	*
Variance components						
2-level variance (interviewers)	.		.		0	
Residual variance	3.29		3.29		3.29	
Model fit						
AIC	405.52		551.58		408.90	
n	414		416		377	

Note **p<0.01, *p<0.05; the variance components from the base model is presented in Table C.6. in Appendix C.

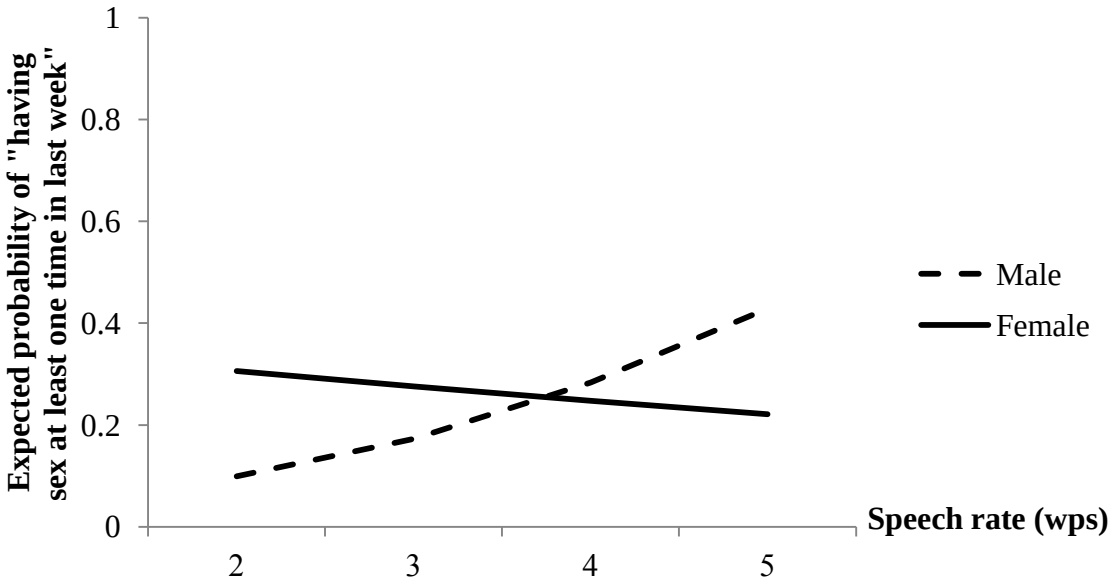


Figure 3.9 Expected probability that respondents reported that they had sex at least one time in the past seven days by speech rate

When analyzing all three undesirable questions (Q5, Q21C, and Q21D) together, interviewer voice characteristics do not predict data quality (See Table D.2 in Appendix D for the result). These three questions may differ in their sensitivity. For example, it is plausible that Q21D (having sex) may be more sensitive than Q5 (fired a job) or Q21C (alcohol drinks). As such, I analyze each question separately.

Interviewer experience affects reporting having sex in past seven days (Q21D). As expected, respondents were more likely to say that they had sex at least one time (i.e., provide an answer less prone to social desirability bias) in the past seven days to inexperienced interviewers than to experienced interviewers (coefficient= 0.594, p=0.05).

With respect to respondent characteristics, respondents whose age is 60 or less tend to report that they had sex more than one time in the past seven days compared to those who are older than 60 (coefficient= 1.415, $p<0.01$). Moreover, respondents whose education is high school or less provided fewer reports that they had sex more than one time in the past seven days than those whose education is higher than high school

(coefficient= -0.636, p=0.04). These results are consistent with the hypothesis that younger and more educated respondents provide better data quality than older and less educated respondents.

Socially desirable questions. The proportion of "no" responses and the directional hypothesis of "less is better" were used as the data quality indicators in socially desirable questions. Question 8 asks about volunteer activities, and question 13A and question 21F ask about reading enjoyment and number of times respondents read a book, magazine, or newspaper in last week respectively. Respondents who answered that they have not done any volunteer activities in the last 12 months in Q8, those who answered that they did not completely enjoy reading a book in Q13A, or those who answered that they read a book, magazine, or newspaper fewer than 10 times[4] in the past seven days are considered to have provided answers that are less influenced by social desirability bias.

Approximately 53% of respondents have not done any volunteer activities in the last 12 months, 45% did not answer that they completely enjoy reading a book, and 71% responded that they read a book, magazine, or newspaper fewer than 10 times in the past seven days. There is no significant variation in reports across interviewers in Q8 and Q13A, but there is for Q21F- 5.6% of the variance in reports of reading a book, magazine, or newspaper less than 10 times in the past seven days was due to interviewers (χ^2=3.65, p=0.03) (Table C.7 in Appendix C).

Table 3.7 presents the results of analyses that examined the extent to which interviewer voice characteristics predict the proportion of answers that are less prone to socially desirable bias, i.e. better data quality. There is no association between pitch and

[4] Mean of number of times reading a book, magazine, or newspaper in the past week = 9.54

intonation and reports on these socially desirable questions. However, there is a U-shaped relationship between speech rate and socially desirable responses on Q13a (coefficient=0.679, p=0.04; Figure 3.10), in which questions read slowly and quickly had better quality data. There is an inverse-U-shaped relationship for number of fillers on Q21F (Figure 3.11, coefficient=2.211, p=0.01 for fillers; coefficient=-1.143, p=0.03 for filler^2). Interviewers who read a question with one filler had better data quality than those reading question with either no fillers or more than one filler.

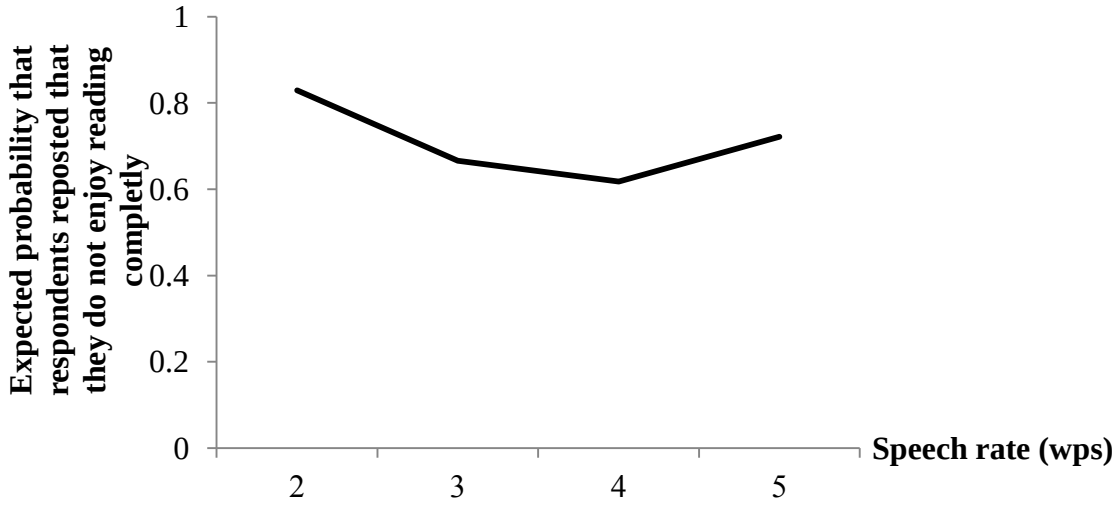


Figure 3.10 Expected probability that respondents reported that they do not enjoy reading completely by speech rate

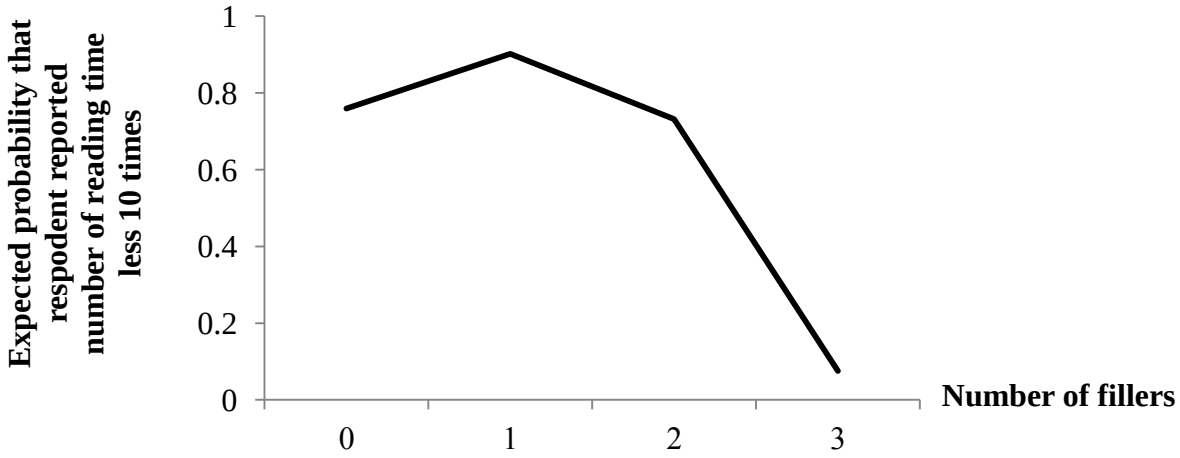


Figure 3.11 Expected probability that respondent reported reading less than 10 times in the past seven days by fillers

Similar to socially undesirable questions, when analyzing the three socially desirable questions (Q8, Q13A, and Q21F) together, the effects of voice characteristics on data quality do not hold (See Table D.3 in Appendix D). As mentioned earlier, I examine each question separately because these three questions may not be identical in their social desirability.

Data quality in responding to socially desirable questions was affected by respondent education and age. Respondents whose education is high school or less were less likely to report engaging in volunteer and reading activities and less likely to report completely enjoy reading compared to respondents whose education is higher than high school (coefficient=1.049, $p<0.01$ for question 8; coefficient= 0.705, $p<0.01$ for question 13A; coefficient= 1.007, $p<0.01$ for question 21F). In addition, respondents whose age is 60 years old or less reported engaging in fewer reading activities and lower enjoyment in reading compared to those whose age is higher than 60 years old (coefficient= 0.848, $p<0.01$ for question 13A; coefficient= 0.563, $p=0.03$ for question 21F).

Table 3.7 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by objective voice characteristics for socially desirable questions

	Q8 (Did not Volunteer Activity)		Q13A (Does not completely Enjoy Reading)		Q21F (Number of Reading times<10)	
	coefficient (SE)		**coefficient (SE)**		**coefficient (SE)**	
Main effects						
Intercept	-0.164(0.29)		-1.097(0.31)	**	0.344(0.29)	
Pitch	0.003(0.005)		-0.005(0.006)		0.011(0.01)	
Intonation	-0.002(0.009)		0.002(0.01)		-0.008(0.01)	
Speech rate	-0.533(0.31)		-0.209(0.22)		-0.119(0.17)	
Speech rate^2			0.679(0.33)	*		
Fillers	0.178(0.16)		0.263(0.20)		2.211(0.85)	**
Fillers^2					-1.143(0.52)	*
Interviewer's experience < 1 year	0.045(0.25)		0.101(0.25)		0.361(0.34)	
Female interviewer	-0.389(0.37)		0.013(0.37)		-0.319(0.42)	
R whose education is high school or less	1.049(0.23)	**	0.705(0.23)	**	1.007(0.28)	**
R whose age is 60 or less	-0.173(0.21)		0.848(0.21)	**	0.563(0.24)	*
Variance components						
2-level variance (interviewers)	0		0		0.113(0.13)	*
Residual variance	3.29		3.29		3.29	
Model fit						
AIC	552.66		564.56		438.61	
N	408		417		413	

Note **p<0.01, *p<0.05; the variance components from the base model is presented in Table C.7. in Appendix C.

3.4 Conclusion and Discussion

In this chapter, I examined how objective measures of interviewer voice characteristics including pitch, intonation, rate of speaking, and disfluencies affect data quality. For all of the outcomes I examined, interviewer voice characteristics affected data quality. However, the effects are inconsistent across questions and seem to vary by interviewer's characteristics. Table 3.8 shows a summary of the effects of interviewer voice characteristics on data quality found in this study.

Table 3.8 Summary results of the effects of interviewer voice characteristics on data quality indicators

	Pitch	Intonation	Speech rate	Disfluencies
Interrupt questions with answers			+ (neutral, complex questions) - (socially undesirable questions)	
Express uncertainty about a question		U-shaped relationship	- (socially undesirable, complex questions)	
Request clarification about a question		U-shaped relationship	- (socially undesirable questions, complex questions) + (inexperienced interviewers)	+ (inexperienced interviewers)
Give qualified answers			+ (neutral questions) - (socially undesirable questions)	- (female interviewers)
Give a response that does not meeting question's objective	+		- (socially undesirable questions)	- (male interviewers); + (female interviewers)
Item nonresponse		U-shaped relationship	-	
Rounding			- (male interviewers); + (female interviewers)	
More is better			+ (male interviewers); - (female interviewers)	
Less is better			U-shaped relationship	Inverse-U-shaped relationship

Note + indicates positive relationship; - indicates negative relationship; variables in parenthesis are interviewer and question characteristics for which the relationship occurs

Pitch. I only found a significant effect of pitch on one data quality indicator. As expected, respondents tend to give a response that does not meet the question's objective to interviewers who read a question with higher pitched voices compared to those who read a question with lower pitched voices. Respondents may perceive interviewers who read a question with higher pitched voices as asking a question that is more sensitive, and thus respondents give an answer that does not meet the question's objective.

Intonation. Respondents are less likely to express uncertainty about a question and to request clarification about a question, and are more likely to respond to a question when interviewers ask a question with moderate intonation (i.e., variation of pitch) (around 40-80 Hz) compared to those asking a question with low or high intonation. It seems to be problematic to speak in a flat, monotone voice as well as to speak with too much variation in pitch. As such, telephone interviewers should be instructed to ask questions with moderate variation in pitch. Intonation did not affect any other data quality outcomes.

Rate of speaking. The voice characteristic with the largest effects on data quality is speech rate. The effects of speech rate on data quality vary by question type and interviewer's characteristics. The results for neutral and socially desirable questions are consistent with the survey methodological literature. Respondents are less likely to engage in problematic interview behaviors and provide better data quality when interviewers read the neutral and socially desirable questions at a speech rate of 2 wps, compared to faster speech rates. When interviewers ask questions with the recommended speech rate of 2 wps - which is slower than the speech rate in the ordinary conversation

(Tauroza & Allison 1990) - respondents may perceive that interviewers would like them to take time to think about their responses attentively, leading to better data quality.

In contrast to neutral and socially desirable questions, the results for socially undesirable questions are consistent with paralinguistic studies. Respondents are less likely to engage in problematic behaviors when interviewers read socially undesirable questions *faster* than 2 wps. In addition, interviewers who read a question with faster speech rates had lower item nonresponse rates than those who read a question with slower speech rates. Respondents may perceive interviewers who read a question with a faster pace as more credible and more trustworthy than interviewers who read a question with a slower pace. As a result, respondents are less likely to engage in problematic behaviors and are more likely to provide better quality data to interviewers whom they perceived as more credible and trustworthy, which are interviewers who read a question with a faster pace. However, the effect of speech rate on survey responses in undesirable questions varies by interviewer sex. Respondents are more likely to provide better data quality (higher undesirable answers) to male interviewers who ask questions quickly compared to those who ask questions slowly, but the opposite was found for female interviewers. Consistent with previous paralinguistic research, male interviewers in this study have slightly faster speech rate compared to female interviewers. Voice characteristics deviate from the gender stereotype may result in negative perceptions of an interviewer, for example, when male interviewers read questions slowly and female interviewers read questions quickly. Thus, at a slow speech rate, respondents provided better data quality to female interviewers. In contrast, at a faster speech rate, respondents provided better data quality to male interviewers.

For complex questions, the results are mixed. Respondents are less likely to interrupt questions with answers but are more likely to express uncertainty about a question and request clarification about a question when interviewers read complex questions at the speech rate of 2 wps compared to faster speech rates. It may be hard to keep information about complex questions with long introductions in a respondent's working memory when interviewers read the questions at the speech rate of 2 wps. At the judgment stage, respondents may forget information in the introduction. As a result, respondents are more likely to express uncertainty and request clarification when interviewers read questions at slower pace. In addition to respondent behaviors, I found that the interviewer's rate of speaking is significantly associated with the probability of rounding an answer in one complex question, but the effect varies by interviewer gender. Male interviewers who ask questions with a faster pace receive fewer rounded answers, but the opposite is true for female interviewers. This is consistent with gender differences previously found in paralinguistic research.

Results from this study imply that the speech rate can affect respondents at different cognitive stages depending on the type of question. As mentioned earlier, results in socially undesirable questions are consistent with linguistic theory- rapid speech is perceived as more credible. However, results in neutral and desirable questions are consistent with survey methodology theory- respondents may not have adequate time to think about their answers when interviewers read questions quickly. These findings suggest that speech rate mainly affects respondents at the perception stage for socially undesirable questions, but at other stages for neutral and socially desirable questions. Results from this study suggest that interviewers should read neutral and socially

desirable questions with the recommended speech rate of 2 wps, but interviewers (especially males) should read socially undesirable questions more quickly. Yet this recommendation may not be practical, or may change the nature of the survey interview in other, unanticipated ways. However, as observed in this study, interviewers read socially undesirable questions faster than other types of question. Interviewers may find it awkward to ask sensitive questions, thus, they read the questions quickly.

Disfluencies. The effects of fillers on respondent behaviors are mixed and vary by interviewer's characteristics. Interviewers tend to receive better data quality in socially desirable questions when they read the question with one filler rather than too many or no disfluencies. This is consistent with previous research that found interviewers who speak with neither robotic speech nor are highly disfluent have the highest participation rates (Conrad et al. 2013).

Overall, objective voice characteristics do have an effect on data quality. As long suspected, but little analyzed, the rate at which interviewers ask survey questions has a profound effect on the quality of survey answers. Interestingly, this effect of pace varies by type of question, by gender of the interviewer, and by data quality outcome. Other characteristics, such as pitch, intonation, and disfluencies, also have an effect on data quality outcomes, but not as consistently as speech rate.

Interestingly, most of the effects on these measures of data quality are on respondent behaviors, not on the more conventional measures of data quality such as item nonresponse, rounding, or the directional hypotheses. Respondent may reveal problems in comprehension and in their ability to provide answers through their behaviors. The

behaviors thus may be a more direct reflection of problems that respondents are having with the questions, although may or may not manifest in lower data quality.

Additionally, although there are consistent effects of speech rate on a wide variety of data quality indicators, the effect of interviewer voice characteristics on data quality may not be revealed through these indirect measures of data quality created from the survey responses. Answers that are not rounded or answers that are less influenced by socially desirable bias may be inaccurate. It is well known that it is hard to distinguish true values from measurement error without gold standard data available (Groves 1989). As such, I did not know whether fewer rounded answers, more undesirable responses, and fewer desirable responses as that respondents reported are from their true values or the measurement error.

In this study, I found that interviewer experience affects data quality only in the question asking about the number of times respondents had sex. As expected, interviewers with less experience obtained higher rates of undesirable answers than interviewers with more experience. More experienced interviewers tend to be careless in conducting a survey compared to less experienced interviewers (Groves, et al. 2009). As a result, more experienced interviewers obtain lower data quality than less experienced interviewers (e.g., Olson and Bilgen 2011).

Data quality is related to respondent age and education. However, the direction of the relationships varies by question type. The relationship between respondent age and data quality is in the hypothesized direction in socially undesirable questions. Younger and more educated respondents are more likely to give responses that are less likely prone to socially desirable bias compared to older and less educated respondents. These

results are consistent with previous research (e.g., Holbrook, et al. 2006; Knauper 1999; Belli, et al. 1999; Narayan & Krosnick 1996).

Regarding complex and neutral questions, older respondents and respondents whose education is high school or less are less likely to round their answers compared to younger respondents and more educated respondents. This result is opposite what I hypothesized. Question topic may confound these results. Rounding was examined in questions asking about computer and internet use. Previous research found that older respondents and less educated respondents are less likely to use computers and the internet than younger respondents and respondents with more education (Tourangeau, et al. 2013; Teo, et al. 1999; Dyck & Smither 1995). Individuals are more likely to recall events that they do less frequently than events that they do more frequently (Means & Loftus 1990). In addition, respondents who can recall the event distinctly tend to enumerate the events in frequency questions, and thus are less likely to round their answers (Conrad, et al. 1998). In contrast, respondents who cannot recall the events distinctly are more likely to use an estimation strategy, which is more likely to yield rounded responses. As such, as older and less educated respondents are less likely to engage in computer usage, they may be more likely to remember their behaviors and thus less likely to round their answers than younger and more educated respondents.

Similar to complex and neutral questions, question content may also confound the effect of respondent education on data quality in socially desirable questions. Less educated respondents are less likely to be involved in volunteer activities and read a book than more educated respondents (Smith 1994; Sharon 1973). As a result, they may report fewer of these activities because of their true responses. In fact, as discussed earlier,

because gold standard data is not available, true values and measurement error are hard to distinguish. As such, I do not know whether fewer desirable responses from less educated respondents are their true values or measurement error.

3.5 Limitations and future research

This study has limitations. The largest limitation is that the data are from a landline telephone survey. Respondents in this study are more likely to be female, white, and older than the general population. Models in this study only controlled for respondent's age and education. As such, respondent's gender and race may affect the results in this study. Respondents from the minority groups (e.g., Hispanic and non-white) may have difficulty comprehending survey questions, and thus may affect data quality (Holbrook, et al. 2006). Additionally, significant interaction effects of interviewer experience and voice characteristics occurred on one data quality indicator, perhaps as a result of Type I error.

Additionally, only a sample of voice files is analyzed for each interviewer on each question. More questions analyzed will increase statistical power in this study. There also is a quite strong correlation (r=0.56) between pitch and intonation, leading to potential multicollinearity problems. Furthermore, I did not know whether fewer rounded answers, higher rates of undesirable responses, and lower rates of desirable responses in fact indicate "better" data quality. The responses could be either true values or measurement error.

Previous research found subjective ratings of interviewers' voice characteristics are more useful for predicting response rates compared to objective ratings of interviewer voices (e.g. Van der Vaart, et al. 2005; Oksenberg & Cannell 1988). This study only

focused on the effect of objective measures of interviewer voice characteristics on data quality. As such, future research should examine whether subjective ratings of interviewer voice characteristics also affect data quality. This is what I examine in the next chapters of this dissertation.

CHAPTER 4: SUBJECTIVE VOICE CHARACTERISTICS, INTERVIEWER PERSONALITY TRAITS, AND DATA QUALITY

Introduction

An interviewer's voice is an important part of a telephone survey because respondents only receive audio cues from interviewers (Groves 1990). An interviewer's voice can convey much more information than simply the meaning of words or sentences themselves (Groves, et al. 2008). Voices can be reliably judged as indicating certain personality traits of a speaker (Ketrow 1990; Apple, et al. 1979). In chapter 2, I found that listeners could perceive interviewers' personality traits (e.g., expertise, trustworthiness, and reliability) from interviewer's voices. However, the relationship between interviewer voice characteristics and perception of an interviewer's personality traits varied for male and female interviewers and across question types.

It is important to understand factors related to data quality in telephone surveys because it is a primary mode of data collection in many large national surveys. Previous research found that interviewer behaviors (e.g., probing) and interviewer's demographic characteristics (e.g., gender) can affect data quality (Fowler 2011; Dykema, et al. 1997; Kane & Macaulay 1993). However, little research has examined whether interviewer voice characteristics affect data quality. In chapter 3, I found that objective measures of interviewer voice characteristics affect data quality. Because listeners could perceive interviewers' personality traits from their voices (See Chapter 2), I hypothesized that perceptions of interviewer's personality traits may mediate the relationships between objective voice characteristics and data quality. This is the question addressed in this chapter of this dissertation.

Figure 4.1 presents the conceptual model of the relationship between interviewer voice characteristics and data quality examined in this dissertation. This chapter has three objectives. Objective 1 of this chapter is to examine the effects of subjective measures of interviewer voice characteristics on data quality. Objective 2 is to investigate how perceived interviewers' personality traits affect data quality. Objective 3 is to examine whether subjectively perceived interviewers' personality traits mediate the relationship between objective voice characteristics and data quality.

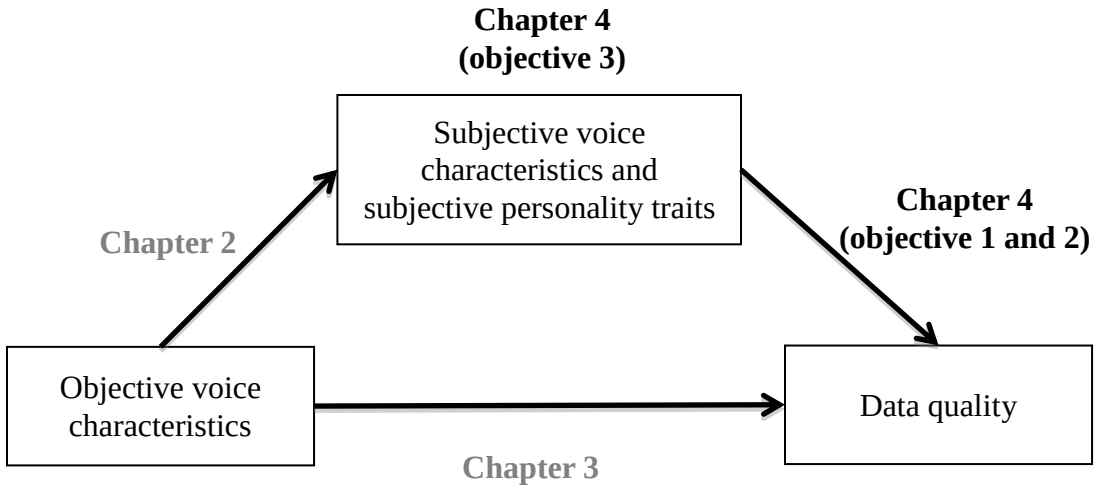


Figure 4.1 Conceptual model for the relationship between interviewer voice characteristics and data quality

4.1 Literature review

4.1.1 Objective 1: Effects of subjective interviewer's voice characteristics on data quality

Previous research has found that subjective ratings of characteristics of interviewers' voices (e.g. intonation and fluency) are better predictors of unit nonresponse than the objective measurement of the same voice characteristics (Van der Vaart, et al. 2005). In general, interviewers who are rated as having higher pitched voices, faster rates of speaking, greater loudness, falling intonation, and clearer and more

distinct pronunciation have higher response rates than those who are rated as having lower pitched voices, slower rates of speaking, lower loudness, rising intonation, and less distinct pronunciation (Groves et al. 2008; Oksenberg & Cannell 1988; Oksenberg et al. 1986). However, there is no empirical research that examines relationships between subjective ratings of interviewer voice characteristics and the quality of answers provided to survey questions.

In this chapter, I investigate whether subjective ratings of interviewer's voices (rated pitch, rated intonation, rated speech rate, and rated disfluency) have an effect on data quality measured by item nonresponse, problematic respondent behaviors, rounded answers, and directional reporting for socially (un)desirable questions. Hypotheses for testing the effects of *subjective* interviewer's voice characteristics on data quality are the same as the hypotheses for testing the effects of *the same objective* interviewer's voice characteristics on data quality as presented in Chapter 3. However, I expect that the effect of subjective interviewer voice characteristics on data quality will be stronger than the effect of objective interviewer voice characteristics. A comparison between the effect of objective and subjective voice characteristics on data quality will be shown in Chapter 5.

4.1.2 Objective 2 and 3: Effects of interviewers' personality traits on data quality and mediation effects of interviewers' personality traits on the effects of objective voice characteristics on data quality

Perceptions of interviewer's personality traits can affect unit nonresponse. Interviewers judged as being more pleasant, cheerful, friendly, enthusiastic, interested, intelligent, educated, professional, and confident have higher response rates than those who were judged as being less pleasant, cheerful, friendly, enthusiastic, interested,

intelligent, educated, professional, and confident (Oksenberg & Cannell 1988; Oksenberg, et al. 1986). However, there is no empirical research that examines whether perceptions of interviewer personality traits affect data quality, and whether these perceptions mediate the relationship between objective voice characteristics and data quality.

In this dissertation, I examine five interviewer personality traits including expertise, confidence, reliability, trustworthiness, and easiness to understand. Expertise is the extent to which an interviewer is good at his/her job in asking a survey question. Confidence is the extent to which the interviewer is self-assured and conducts the interview with poise. Reliability is the extent to which an interviewer says something that can be believed. Trustworthiness is the degree of confidence in an interviewer to ask a valid survey questions and keep respondents' answers confidential. Lastly, easiness to understand is the extent to which an interviewer's voice is easy to understand. In Chapter 2, I examined how the objectively measured voice characteristics predict these perceived personality traits.

Previous research has found that reliability, trustworthiness, expertise, and confidence all reflect an underlying credibility construct (Figure 4.2; Sah, et al. 2013; Ohanian 1990; Hovland, et al. 1953). Speakers who are rated as being more reliable, more trustworthy, and having more expertise are perceived as being more credible than those who are rated as being less reliable, less trustworthy, and having less expertise. In addition, more confident speakers are perceived as being more credible than less confident speakers (Sah, et al. 2013, Anderson, et al. 2012, Price & Stone, 2004). In this

chapter, I will examine whether these perceptions of expertise, reliability, trustworthiness, and confidence reflect a single underlying construct of credibility.

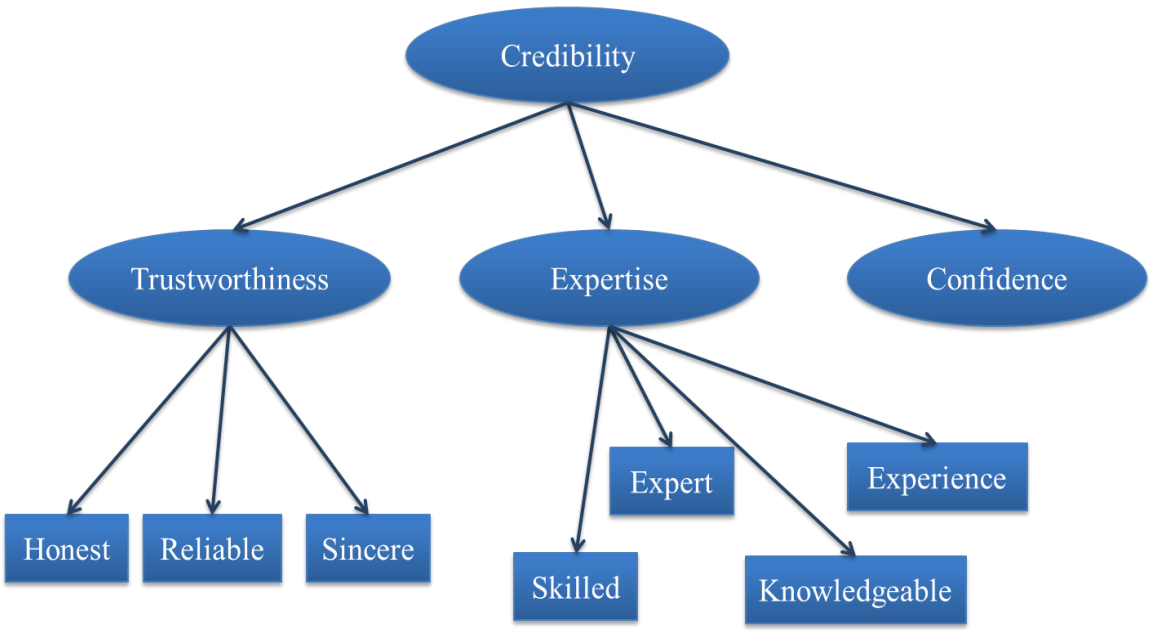


Figure 4.2 Diagram of the construct of credibility, trustworthiness, and expertise adapted from Sah, et al. (2013), Ohanian (1990), and Hovland, et al. (1953)

Respondents are more likely to provide better data quality to interviewers whom they perceived as being more credible (Groves, 1990). As such, I expect that interviewers who are perceived as being more confident, more reliable, more trustworthy, and having more expertise (i.e. more credible) will receive better quality answers (less rounded answers, fewer answers prone to socially desirable bias, fewer item nonresponse rates, and fewer problematic respondent behaviors) than those who are perceived as being less confident, less reliable, less trustworthy, and having less expertise (i.e. less credible). In addition, I expect that perceived expertise, confidence, trustworthiness, and reliability will mediate the effects of objective measures of pitch, intonation, speech rate, and disfluencies on data quality (Lai 2010; Bortfeld et al. 2001; Smith & Shaffer 1995;

Ketrow 1990; Oksenberg & Cannell 1988; Apple et al. 1979; Miller et al. 1976; Pearce & Conking 1971).

In addition to credibility, respondents are more likely to give better quality answers when they are more easily comprehend questions read by interviewers (Japec 2008). Easiness to understand may play an especially important role for complex questions because understandability can affect a listener's comprehension of these questions (Miller et al. 1976). As such, I expect higher data quality among interviewers whose voices are perceived as being easier to understand, and that understandability may mediate the effects of objective measures of pitch, intonation, speech rate, and disfluencies on data quality.

Similar to Chapter 3, interviewer demographic characteristics (interviewer sex and experience) and question characteristics (complex, socially desirable, and socially undesirable questions) may moderate effects of interviewer voice characteristics on data quality. As such, interaction effects of interviewer voice characteristics with interviewer demographic characteristics and question characteristics are examined. I also examined nonlinear relationships between interviewer voice characteristics and data quality indicators. In addition, analyses in this study controlled for respondent age and education because data quality tends to be lower for older and less educated respondents compared to younger and more educated respondents (Knauper 1999; Narayan & Krosnick 1996; Groves 1989).

4.2 Data and methods

4.2.1 Data

Data in this study come from the Work and Leisure today survey. It is a landline RDD CATI survey conducted by 22 interviewers with 450 completed interviews (AAPOR RR1=4.7%). To increase the stability of the analyses, interviewers who conducted a survey less than 10 interviews were removed from this study. As such, I analyze 432 interviews conducted by 19 interviewers (9 female and 10 male interviewers).

As mentioned above, four types of questions including socially desirable, undesirable, complex, and neutral questions are examined. In this study, twelve questions (three of each of the four types of questions - Q8, Q13A, and Q21F for socially desirable questions; Q5, Q21C, and Q21D for socially undesirable questions; Q13E, Q19, and Q20 for complex questions; Q2, Q13D, and Q21A for neutral questions- were selected based on the criteria that the questions contain both item nonresponse and sufficient variability in responses (See question wording in Appendix A).

4.2.2 Measures of voice characteristics

In this dissertation, interviewer's voice characteristics are measured subjectively by raters and objectively by computer program. Details on these measurements are described in Chapters 2 and 3 of this dissertation.

4.2.3 Data quality analysis

There are three main sets of analyses corresponding to the three objectives in this chapter. First, I examine the effects of subjective interviewer voice characteristics on data quality. Second, I investigate the associations between perceptions of interviewer

personality traits and data quality. Third, I examine mediation effects of perceptions of interviewer personality traits on the associations between objective voice characteristics and data quality.

For the three main analyses, I examine four sets of data quality indicators: 1) respondent behaviors associated with data quality, 2) item nonresponse, 3) rounding responses, and 4) the directional hypotheses of "more/less is better." Five respondent behaviors examined in this study are the respondent 1) interrupts questions with an answer, 2) expresses uncertainty about a question, 3) requests clarification, 4) gives qualified answers, and 5) gives a response that does not meet the question's objective (See Kirchner & Olson (2014) for more detail about behavior coding process). The data quality indicators are all dichotomous variables (See Table C.1 in Appendix C for a summary of the data quality indicators in this study and Table C.2 for descriptive statistics of the data quality indicators).

In addition, for the three main analyses, dichotomous variables of interviewer sex and experience and respondent age and education were included in the models—interviewer experience: one year of experience or less (25%) and more than one year of experience (75%); respondent age: respondents whose age is 60 years old or less (47%) versus greater than 60 years old (53%); and respondent education: respondents whose education is high school or less (30%) versus higher than high school (70%).

Analysis I: Effects of subjective interviewer's voice characteristics on data quality

Data in this study are nested - each respondent was interviewed by one interviewer, and interviewers obtained responses from multiple respondents. Consequently, multi-level modeling is used for analysis. Respondent behaviors and item

nonresponse were analyzed for all 12 questions in one model each. However, rounding was used as a data quality indicator for the subset of neutral and complex questions and the directional hypotheses was used as a data quality indicator for socially (un)desirable questions. As such, rounding and the directional hypotheses were analyzed for each question separately.

Respondent behaviors and item nonresponse analyses

As in Chapter 3, three-level multi-level models were estimated to account for variability due to questions, respondents, and interviewers (O'Muircheartaigh & Campanelli 1998; 1999). Item nonresponse and five problematic respondent behaviors are coded as dichotomous variables where 1 indicates that respondent did not answer a question versus 0 for answering the question and where 1 indicates that respondents engaged in problematic behaviors versus 0 for not having this behavior, respectively. A logit link with a binary distribution was used to estimate the model as shown below.

$$logit(y_{tij}) = \gamma_{000} + \gamma_{001} IChar_j + \gamma_{010} RChar_{ij} + \sum_{a=1}^{A} \gamma_{a00} ratedVoiceChar_{aij}$$
$$+ \sum_{a=1}^{A} \gamma_{a01} IChar_j \times ratedVoiceChar_{cij} + \sum_{b=A+1}^{B} \gamma_{b00} QChar_{bij}$$
$$+ \sum_{c=B+1}^{C} \gamma_{d00} ratedVoiceChar_{ij} \times QChar_{ij} + V_{00j} + U_{0ij} + e_{tij}$$

where y_{tij} = respondent behaviors and item nonresponse,

ratedVoiceChar$_{ij}$ = rated pitch (centered at 3), rated intonation (centered at 3), rated speech rate (centered at 3), and rated disfluencies (centered at 3),

QChar$_{ij}$= neutral (reference group), complex, socially undesirable, and socially desirable questions,

IChar$_j$ = interviewer's sex and experience,

RChar$_j$ = respondent's age and education,

V_{ooj} = random interviewer effect,

U_{0ij} = random respondent effect, and

e_{tij} = residual.

Rounding and Directional Hypotheses

Rounding is measured using a dichotomous variable indicating whether the respondent gave a prototypical answer (e.g., 5, 10, 15 or multiple of 60 minutes for question 19 asking about number of minutes that respondents spend on computer on a typical day). In addition, the directional "more/less is better" hypotheses are tested using the responses to the survey questions (e.g., engage in socially undesirable behaviors at least one time or engage in socially desirable behaviors few times); the "yes" response is predicted for the socially undesirable questions (e.g., respondents answered "yes" they have ever been fired from a job) and the "no" response is predicted for the socially desirable questions (e.g., respondents answered "no" they have not done any volunteer activities in the last 12 months).

Two-level hierarchical models predicting data quality were estimated to account for variation due to respondents and interviewers (O'Muircheartaigh & Campanelli 1999). Because the data quality indicators for the directional hypotheses and rounding are dichotomous variables, a logit link with a binary distribution was used to estimate the models. In general, using the terms defined above, the model is:

$$logit(y_{ij}) = \gamma_{00} + \gamma_{01} Ichar_j + \sum_{a=1}^{A} \gamma_{a0} ratedVoiceChar_{ij}$$

$$+ \sum_{a=1}^{A} \gamma_{a1} ratedVoiceChar_{ij} \times Ichar_j + \sum_{b=A+1}^{B} \gamma_{b0} RChar_{ij} + V_{0j} + e_{ij}$$

where y_{ij} = rounded answers and whether respondents gave answers that are less influenced by socially desirable bias

Because interviewer voice characteristics may have a nonlinear relationship with the data quality indicators (e.g. Conrad, et al. (2013) found fillers have inverse-U shaped associations with the likelihood of survey participation), I also examined the squared terms of pitch, intonation, speech rate, and number of fillers. Results of the effects of subjective voice characteristics on data quality are reported in this chapter, but a comparison between the effects of objective and subjective voice characteristics on data quality will be reported in Chapter 5.

Analysis II and III: Interviewer's personality traits as mediators for the relationship between objective voice characteristics and data quality

Mediation effects of perception of interviewer personality traits on the relationships between objective interviewer voice characteristics and data quality occur if the following four conditions exist (Baron & Kenny 1986). First, listeners can perceive interviewer personality traits from interviewer voice characteristics (Chapter 2 of this dissertation, See Figure 4.1). Second, there is a relationship between objective voice characteristics and data quality (Chapter 3 of this dissertation, See Figure 4.1). Third, there is a statistically significant effect of interviewer personality traits on data quality. Fourth, the effects of objective voice characteristics on data quality indicators are reduced or eliminated when controlling for interviewer personality traits. Perfect mediation holds if objective voice characteristics have no effect on data quality indicators when controlling for interviewer personality traits, but the personality traits maintain statistical significance.

In this chapter, the third and fourth condition to test for mediation effects (Baron & Kenny, 1986) are examined. Data quality indicators that fulfill all four conditions are

evaluated to observe the direct effects of objective voice characteristics on data quality indicators and the indirect effects through interviewer's personality traits.

Analysis II: The third condition to test the mediation effect: Effect of perceptions of an interviewer's personality traits on data quality

Based on previous paralinguistic studies, interviewer personality traits observed in this study may be highly correlated. As such, I examined the association among the five rated interviewer's personality traits - confidence, easiness to understand, reliability, trustworthiness, and expertise. As expected, there are moderate-to-high correlations among the five perceived interviewer personality traits ($p<0.01$, Table 4.1). The highest correlation was found for the association between reliability and expertise ($r=0.84$), followed by the association between confidence and expertise ($r=0.82$), and the associations between trustworthiness and reliability, and confidence and reliability ($r=0.72$). As shown in Table 4.1, the associations between easiness to understand and other personality traits ($0.42\leq r \leq 0.65$) are lower than the associations among the other traits.

Table 4.1 Pearson's correlation matrix of rated confidence, easiness to understand, reliability, trustworthiness, and expertise

	Confidence	Easiness to understand	Reliability	Trustworthiness	Expertise
Confidence	1	0.42	0.72	0.66	0.82
Easiness to understand	0.42	1	0.58	0.48	0.65
Reliability	0.72	0.58	1	0.72	0.84
Trustworthiness	0.66	0.48	0.72	1	0.75
Expertise	0.82	0.65	0.84	0.75	1

Note all correlations are significant at $p<0.01$

Because interviewer personality traits observed in this study are highly correlated, a multicollinearity problem may arise if all of the personality traits are used

simultaneously to predict data quality. As such, I conducted a principal components factor analysis to examine whether these traits can be combined into a single factor. Table 4.2 presents the eigenvalues from the principal components factor analysis. By using the eigenvalue-one criterion, i.e. retaining component with an eigenvalue greater than 1 (Kaiser 1960), there is only one underlying component in this study. This component accounts for 73.6% of the total variance.

Table 4.2 Eigenvalues of the Correlation matrix of interviewers' personality traits

	Eigenvalue	Difference	Proportion	Cumulative
1	3.682	3.065	0.736	0.736
2	0.617	0.267	0.123	0.859
3	0.351	0.113	0.070	0.930
4	0.238	0.126	0.048	0.978
5	0.112		0.022	1

According to the factor loadings, the factor is highly correlated with confidence (loading=0.85), reliable (loading=0.90), trustworthiness (loading=0.85), and expertise (loading=0.95). The correlation between the factor and easiness to understand is lower (loading=0.71). In addition, communalities, which explain how much variance in an observed variable is accounted for by the factor, indicate that only 51% of the variation in easiness to understand is explained by the factor (Table 4.3). However, the results suggest that the factor explain about 73% of the variation in confidence, 82% of the variation in reliability, 71% of the variation in trustworthiness, and 91% of the variation in expertise. As such, in this study, confidence, reliable, trustworthiness, and expertise are considered to be one factor (alpha=0.92). In fact, previous work has found that these characteristics create a more general factor of *"credibility."* Easiness to understand is also examined in this study as one observed variable because it may affect data quality as discussed in literature review.

Table 4.3 Communalities from the Principal Components Factor analysis

	Communality
Rated confidence	0.73
Rated easiness to understand	0.51
Rated reliability	0.82
Rated trustworthiness	0.71
Rated expertise	0.91

In summary, two interviewer personality traits - credibility and easiness to understand - are used to examine the effects of interviewer personality traits on data quality. Credibility is calculated as the sum of ratings of confidence, reliable, trustworthiness, and expertise.

To examine whether perceived interviewer personality traits affect data quality, three-level multi-level models were estimated for item nonresponse and problematic respondent behaviors to account for variability due to questions (level-1), respondents (level-2), and interviewers (level-3) (O'Muircheartaigh & Campanelli 1998; 1999). In general, using the terms defined above, the model is:

$$logit(y_{tij}) = \gamma_{000} + \gamma_{001} ICha r_j + \gamma_{010} RChar_{ij} + \sum_{a=1}^{A} \gamma_{a00} Interveiwer\ traits_{aij}$$

$$+ \sum_{a=1}^{A} \gamma_{c01} ICha r_j \times Interveiwer\ traits_{cij} + \sum_{b=A+1}^{B} \gamma_{b00} QChar_{bij}$$

$$+ \sum_{c=B+1}^{C} \gamma_{d00} Interveiwer\ traits_{ij} \times QChar_{ij} + V_{00j} + U_{0ij} + e_{tij}$$

where y_{tij} = respondent behaviors and item nonresponse,

Interviewer traits = credibility and easiness to understand (centered at their mean values).[5]

In addition, two-level multi-level models were estimated for rounding and the hypothesis of "more/less is better" to account for variability due to respondents (level-1)

[5] Mean $_{credibility}$=23, s.d. $_{credibility}$=1.78; Mean $_{easiness\ to\ understand}$=6, s.d. $_{easiness\ to\ understand}$=0.49

and interviewers (level-2) (O'Muircheartaigh & Campanelli 1998; 1999). The estimated model is:

$$logit(y_{ij}) = \gamma_{00} + \gamma_{01} Ichar_j + \sum_{a=1}^{A} \gamma_{a0} interveiwer\ traits_{ij} + \sum_{b=A+1}^{B} \gamma_{b0} RChar_{ij} + \sum_{a=1}^{A} \gamma_{a1} interveiwer\ traits_{ij} \times Ichar_j + V_{0j} + e_{ij}$$

where y_{ij} = rounded answers and whether respondents gave answers that less prone to social desirability bias.

Analysis III: The fourth condition to test the mediation effects of an interviewer's personality traits on the effects of objective voice characteristics on data quality

The fourth condition to test for mediation is whether the effects of objective voice characteristics on data quality are reduced after controlling for interviewer personality traits (Baron & Kenny, 1986). Three-level multi-level models with a logit link were estimated for the respondent behaviors and item nonresponse analyses. The estimated model is:

$$logit(y_{tij}) = \gamma_{000} + \gamma_{001} IChar_j + \gamma_{010} RChar_{ij} + \sum_{a=1}^{A} \gamma_{a00} VoiceChar_{aij} + \sum_{a=1}^{A} \gamma_{a01} IChar_j \times VoiceChar_{aij} + \sum_{b=A+1}^{B} \gamma_{b00} QChar_{bij} + \sum_{c=B+1}^{C} \gamma_{c00} VoiceChar_{ij} \times QChar_{ij} + \sum_{d=C+1}^{D} \gamma_{d00} Interveiwer\ traits_{dij} + \sum_{d=C+1}^{D} \gamma_{d01} IChar_j \times Interveiwer\ traits_{eij} + \sum_{e=D+1}^{E} \gamma_{e00} Interveiwer\ traits_{ij} \times QChar_{ij} + V_{00j} + U_{0ij} + e_{tij}$$

where y_{tij} = respondent behaviors and item nonresponse,

$VoiceChar_{ij}$ = pitch (centered at 165 Hz), intonation (centered at 40 Hz), rate of speaking (centered at 3.5 wps), and number of fillers.

In addition, a two-level multi-level model with a logit link was estimated for rounding answers and answers that are less prone to socially desirable bias. The estimated model is:

$$logit(y_{ij}) = \gamma_{00} + \gamma_{01} Ichar_j + \sum_{a=1}^{A} \gamma_{a0} VoiceChar_{ij} + \sum_{a=1}^{A} \gamma_{a1} VoiceChar_{ij} \times Ichar_j$$
$$+ \sum_{b=A+1}^{B} \gamma_{b0} interveiwer\ traits_{ij} + \sum_{b=A+1}^{B} \gamma_{b1} interveiwer\ traits_{ij} \times Ichar_j$$
$$+ \sum_{c=B+1}^{C} \gamma_{b0} RChar_{ij} + V_{0j} + e_{ij}$$

where y_{ij} = rounded answers and whether respondents gave answers that less prone to social desirability bias.

Whether or not the effect of objective voice characteristics on data quality is reduced after controlling for interviewer's personality traits is examined by comparing results in this chapter with the results of the effects of objective voice characteristics on data quality as reported in Table 3.3 – 3.7 in Chapter 3.

Finally, I attempted to examine mediation effects by using a moderated mediation multi-level model (Rabe-Hesketh & Skrondal 2012; Bauer, Preacher, & Gill 2006). This study has interviewer experience and sex as moderators and interviewer's personality traits as mediators. Generalized Structural Equation Modeling (GSEM) was used to estimate the moderated mediation multi-level models (Rabe-Hesketh & Skrondal 2012). Unfortunately, the models failed to converge (more details are presented in the result section).

4.3 Results

4.3.1 Objective 1: Subjective voice characteristics and data quality

Descriptive statistics of each data quality indicator and the variance components from the base model for each data quality indicator are presented in Appendix C. I now

turn to examining whether subjective evaluations of voice characteristics predict data quality.

Respondent behaviors

Table 4.4 shows results from the hierarchical logistic regression models of subjective interviewer's voice characteristics predicting five problematic respondent behaviors. Overall, subjective voice characteristics affect respondent behaviors. However, the effects of rated voice characteristics on respondent behaviors varied by question types. Interviewer demographic characteristics do not moderate the effect of subjective measures of interviewer voices on data quality.

Rated pitch. Rated pitch is positively associated with respondents **expressing uncertainty about a question** (coefficient=0.648, $p<0.01$). As expected, respondents are more likely to express uncertainty about a question when interviewers read a question with voices rated as having higher pitch compared to voices with lower rated pitch. The positive effect of rated pitch on the expected probability that respondents express uncertainty about a question was modestly stronger in neutral questions than in complex questions; there was no statistical difference in this association for desirable or undesirable questions compared to neutral questions (Figure 4.3). The positive association between pitch and problematic respondent behaviors cannot be explained by perception of question sensitivity (higher pitched voice are perceived as asking sensitive questions, leading to higher rates of problematic respondent behaviors) as I hypothesized because I did not find an association between pitch and respondent behaviors in sensitive questions.

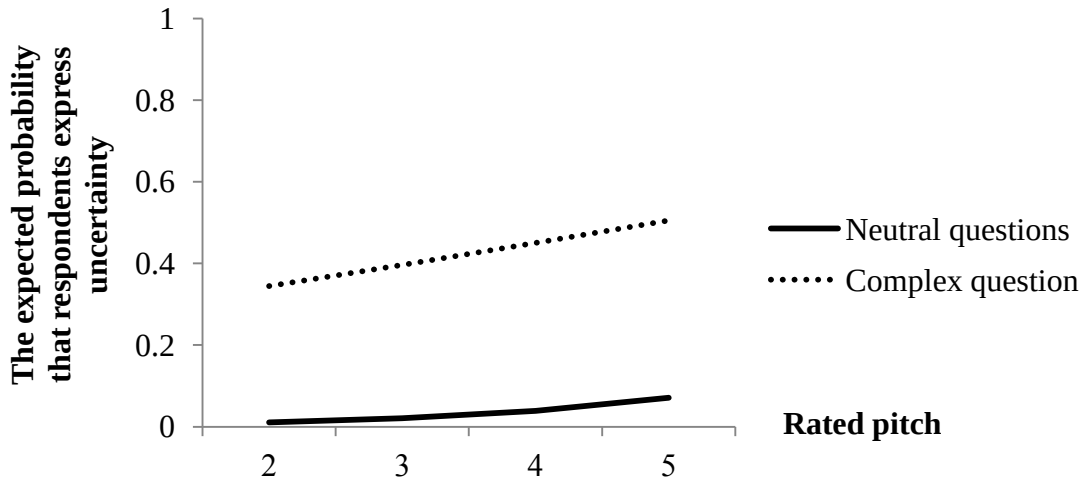


Figure 4.3 Expected probability that respondents express uncertainty about a question in neutral and complex questions by rated pitch

Rated intonation. Rated intonation is negatively associated with respondents expressing **uncertainty about a question** (coefficient=-0.364, $p<0.01$) and respondents giving a **response that does not meet the question's objectives** (coefficient=-0.168, $p=0.04$); however, the effects were modest. Respondents are less likely to express uncertainty about a question and give a response that does not meet question's objectives for interviewer voices perceived to have more intonation (more pitch variation) compared to those with less intonation (less pitch variation). In addition, there is a negative association between intonation and respondents **requesting clarification** for socially desirable and complex questions (Figure 4.4). Respondents are less likely to request clarification about a question in socially desirable and complex questions to an interviewer with a voice rated to have more intonation (more pitch variation) than less intonation (less pitch variation). This is consistent with the hypothesis that voice with higher intonation is perceived as more credible, leading to fewer problematic respondent behaviors.

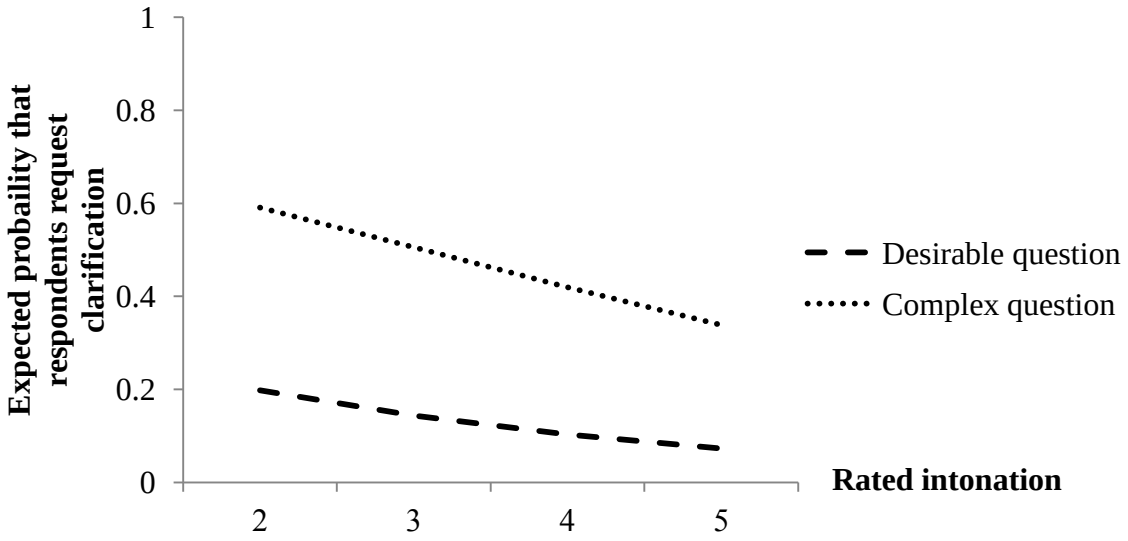


Figure 4.4 Expected probability that respondents request clarification about a question in socially desirable and complex questions by rated intonation

Rated speech rate. Rated speech rate is associated with all respondent behaviors, although the direction of the association varies by question type. There is a positive association between rated speech rate and problematic respondent behaviors in neutral questions. Respondents are more likely to **request clarification about a question** (coefficient=0.282, p=0.04), **give qualified answers** (coefficient=0.309, p<0.01), and **give a response that does not meet question's objectives** (coefficient for speech rate=0.038, p=0.79; coefficient for speech rate squared=0.118, p=0.02) for interviewers perceived to read questions more quickly compared to questions read more slowly (See an example of the positive association between rated speech rate and respondents requesting clarification in neutral questions in Figure 4.5). This is consistent with the hypothesis from survey practice that respondents may not understand a question asked more quickly, leading to higher rates of problematic respondent behaviors.

In contrast, I found a negative association between rated speech rate and problematic respondent behaviors in socially desirable and complex questions. In socially

desirable questions, respondents are less likely to **interrupt questions with answers** (coefficient=-0.505, p<0.01), **request clarification about a question** (coefficient=-0.378, p=0.02), and **give a response that does not meet the question's objective** (coefficient=-0.286, p=0.04) for interviewer voices perceived to be read more quickly. In complex questions, respondents are less likely to **express uncertainty about a question** (coefficient=-0.889, p<0.01) and **request clarification about a question** (coefficient=-0.551, p<0.01) when interviewer voices are perceived to be fast (See an example of the negative associations between speech rate and respondents requesting clarification in socially desirable and complex questions in Figure 4.5). This is consistent with the hypothesis from paralinguistic study that interviewer voices with fast speech rates are perceived as more credible, leading to fewer problematic respondent behaviors.

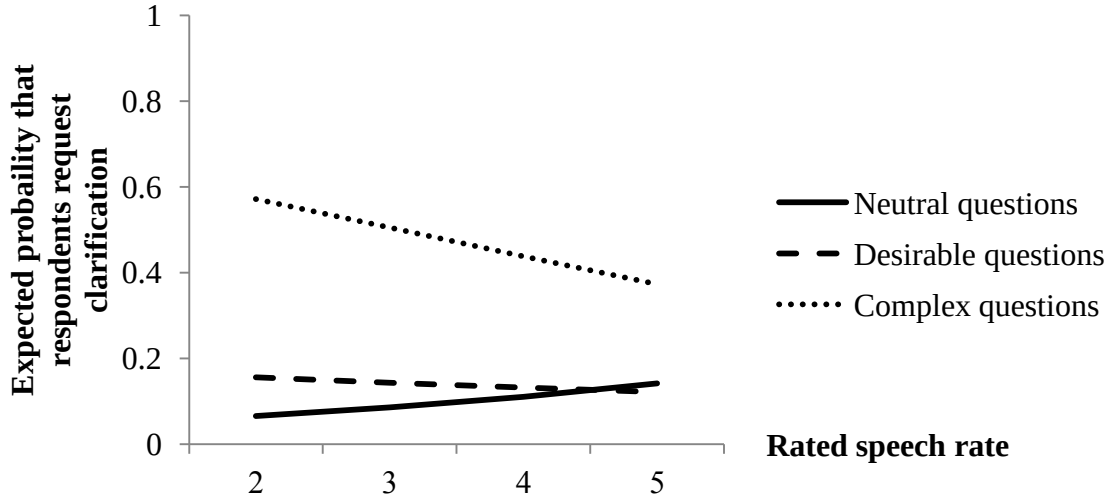


Figure 4.5 Expected probability that respondents request clarification about a question in neutral, socially desirable, and complex questions by rated speech rate

Rated disfluencies. Rated disfluencies are negatively associated with respondents giving a **response that does not meet the question's objective** (coefficient for disfluency=-0.189, p=0.05; coefficient for disfluency squared=-0.142, p=0.02). Respondents are less likely to give a response that does not meet question's objective to

interviewers whose voices are perceived to have more disfluencies (Figure 4.6). This is consistent with hypotheses from survey research that disfluencies have a disfluency advantage that allows respondents to have more time thinking about their answer, leading to fewer responses that do not meet the question's objective.

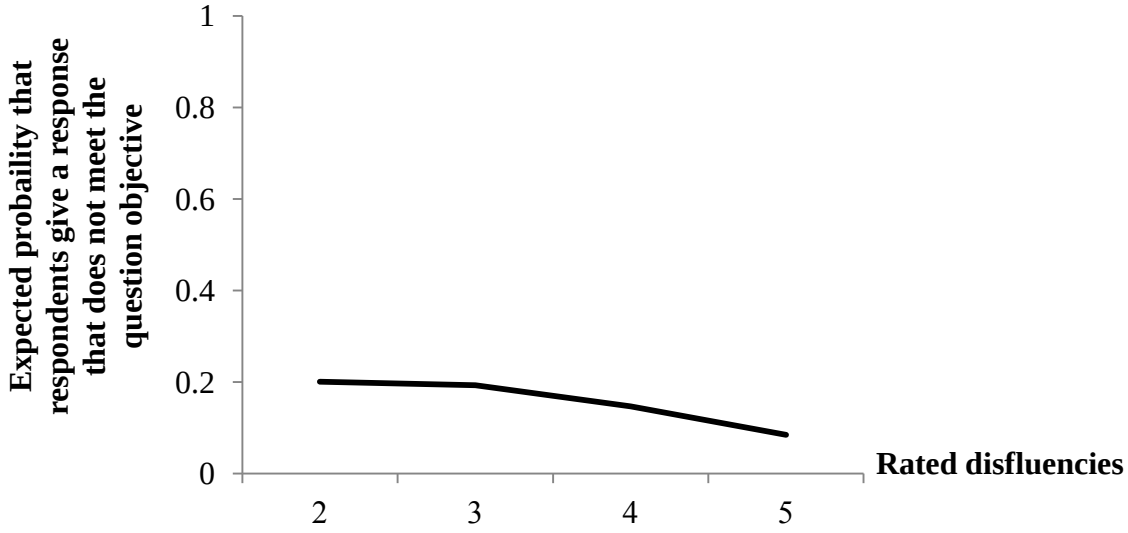


Figure 4.6 Expected probability that respondents give a response that does not meet question objective by rated disfluencies

In addition, for socially desirable questions, more perceived disfluencies are associated with fewer **qualified answers** (coefficient=-0.537, p=0.01). However, as can be seen from Figure 4.7, the effect is modest. Moreover, there is a modest U-shaped association between disfluencies and respondents **expressing uncertainty** about a question (coefficient for disfluency=-0.023, p=0.85; coefficient for disfluency squared=0.171, p=0.03). Respondents are slightly less likely to **express uncertainty** to interviewers perceived to have moderate levels of disfluencies (rated disfluencies=3) compared to either with fewer disfluencies (rated disfluencies=2) or with more disfluencies (rated disfluencies higher than 3) (Figure 4.8).

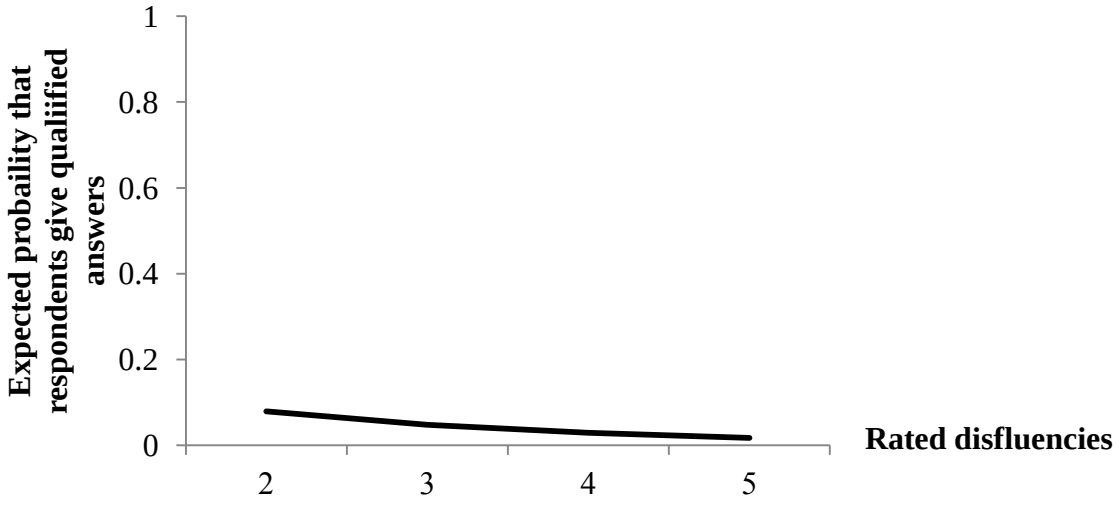


Figure 4.7 Expected probability that respondents give a qualified answer for socially desirable questions by rated disfluencies

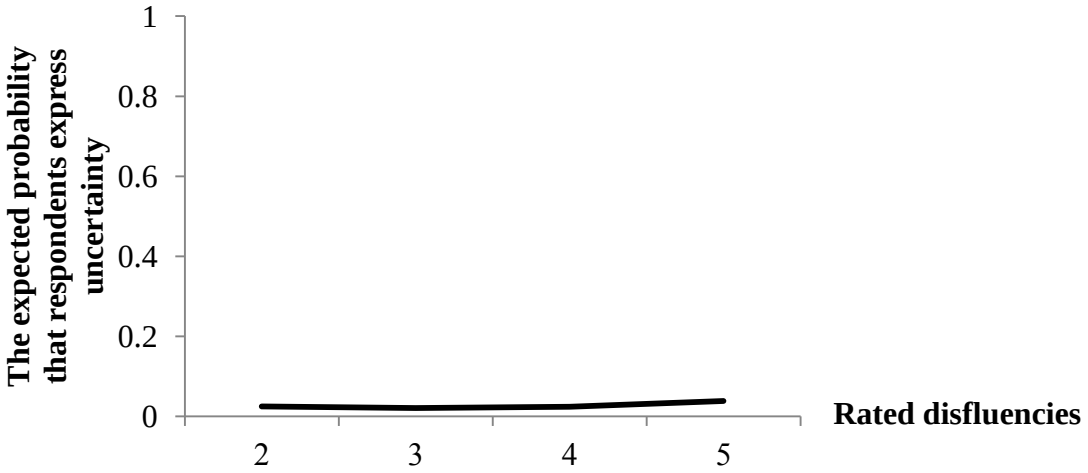


Figure 4.8 Expected probability that respondents express uncertainty about a question by rated disfluencies

Table 4.4 Hierarchical logistic model predicting respondent behaviors by subjective voice characteristics

	Interrupt questions with answers		Express uncertainty about a question		Request clarification about a question		Give a qualified answer		Give a response that does not meet the question's objective	
	coefficient (SE)		coefficient (SE)		coefficient (SE)		coefficient (SE)		coefficient (SE)	
Main effects										
Intercept	-2.392(0.26)	**	-3.424(0.31)	**	-2.170(0.19)	**	-2.057(0.20)	**	-1.235(0.16)	**
Pitch	-0.144(0.17)		0.648(0.23)	**	0.215(0.14)		-0.140(0.15)		0.088(0.12)	
Intonation	0.007(0.12)		-0.364(0.13)	**	0.079(0.15)		0.033(0.10)		-0.168(0.08)	*
Speech rate	0.252(0.17)		0.243(0.21)		0.282(0.14)	*	0.309(0.07)	**	0.038(0.15)	
Speech rate^2									0.118(0.05)	*
Disfluencies	-0.010(0.08)		-0.023(0.13)		-0.120(0.07)		0.008(0.12)		-0.189(0.10)	*
Disfluencies^2			0.171(0.08)	*					-0.142(0.06)	*
Interviewer's experience < 1 year	0.090(0.26)		-0.276(0.24)		-0.153(0.15)		0.118(0.14)		0.044(0.13)	
Female interviewer	0.188(0.34)		-0.446(0.33)		-0.150(0.24)		0.407(0.25)		-0.070(0.21)	
Desirable question	0.946(0.21)	**	0.349(0.36)		0.582(0.19)	*	-1.236(0.27)	**	0.099(0.16)	
Complex question	0.948(0.22)	**	3.443(0.28)	**	2.386(0.18)	**	0.417(0.23)		0.720(0.16)	**
Undesirable question	-0.530(0.29)		-0.965(0.51)		-1.299(0.29)	**	-1.291(0.34)	**	-0.934(0.20)	**
R education is high school or less	-0.079(0.13)		0.136(0.12)		0.005(0.10)		0.029(0.13)		0.344(0.10)	**
R whose age is 60 or less	-0.658(0.13)	**	-0.296(0.11)	*	-0.091(0.09)		0.064(0.12)		-0.648(0.10)	**
Interaction btw voice and Q char										
Pitch*desirable question			-0.091(0.23)							
Pitch*complex question			-0.427(0.18)	*						
Pitch*undesirable question			-0.005(0.29)							
Intonation*desirable question					-0.462(0.19)	**				
Intonation*complex question					-0.425(0.18)	*				
Intonation*undesirable question					-0.320(0.27)					

Note **p<0.01, *p<0.05; n = 4,689

Table 4.4 continued.

	Interrupt questions with answers	Express uncertainty about a question	Request clarification about a question	Give a qualified answer	Give a response that does not meet the question's objective
	coefficient (SE)	**coefficient (SE)**	**coefficient (SE)**	**coefficient (SE)**	**coefficient (SE)**
Speech rate*desirable question	-0.505(0.19) **	-0.363(0.26)	-0.378(0.16) *		-0.286(0.14) *
Speech rate*complex question	-0.217(0.21)	-0.889(0.23) **	-0.551(0.16) **		0.063(0.15)
Speech rate*undesirable question	-0.327(0.24)	-0.183(0.32)	-0.016(0.20)		-0.181(0.16)
Fillers*desirable question				-0.537(0.21) *	
Fillers*complex question				0.106(0.17)	
Fillers*undesirable question				0.105(0.24)	
Variance components					
2-level variance (respondents)	0.423(0.10) **	0	0.092(0.05)	0.582(0.09)	0.389(0.07) **
3-level variance (interviewers)	0.162(0.09) **	0.14(0.08)	0.041(0.03)	0	0.011(0.02) **
Residual variance	3.29	3.29	3.29	3.29	3.29
Model fit					
Generalized Chi-square	3747.58	4726.41	4465.61	3788.03	4065.04

Note **p<0.01, *p<0.05; n = 4,689

Item nonresponse

The overall item nonresponse rate across these 12 questions is 4.6 percent. Table 4.5 presents the results of analyses that examined the extent to which subjective voice characteristics predict item nonresponse. Unexpectedly, rated interviewer intonation and fillers are not associated with item nonresponse. However, rated interviewer pitch and speech rate *are* associated with item nonresponse, but only for particular interviewer and question characteristics (Table 4.2).

Table 4.5 Hierarchical logistic model predicting item nonresponse by subjective voice characteristics

	Coefficient(SE)	
Main effects		
Intercept	-3.248(0.35)	**
Pitch	0.377(0.29)	
Intonation	-0.097(0.18)	
Speech rate	0.114(0.15)	
Fillers	0.002(0.12)	
Interviewer's experience less than 1 year	1.152(0.44)	*
Female interviewers	-0.657(0.51)	
Socially desirable questions	-1.167(0.31)	**
Complex questions	0.772(0.21)	**
Socially undesirable questions	0.101(0.23)	
R whose education is high school or less	0.074(0.17)	
R whose age is 60 or less	-0.129(0.16)	
Interaction between voice and Iwer char		
Speech rate*interviewer experience < 1 year	-0.772(0.26)	**
Interaction between voice and question char		
Pitch*Socially desirable questions	-0.689(0.37)	
Pitch*Complex questions	-0.137(0.21)	
Pitch*Socially undesirable questions	-0.661(0.25)	**
Variance components		
2-level variance (respondents)	0.177(0.15)	**
3-level variance (interviewers)	0.431(0.21)	
Residual variance	3.29	
Model fit		
Generalized Chi-square	3707.59	

Note **p<0.01, *p<0.05; n = 4,689

The effect of speech rate on item nonresponse was only found for inexperienced interviewers. Consistent with the hypothesis from paralinguistic research (faster speech rate is perceived as more credible, leading to lower item nonresponse rates), inexperienced interviewers who read a question more quickly obtained lower item nonresponse rates than those who read a question with a slower pace (coefficient=-0.772, $p<0.01$; Figure 4.9).

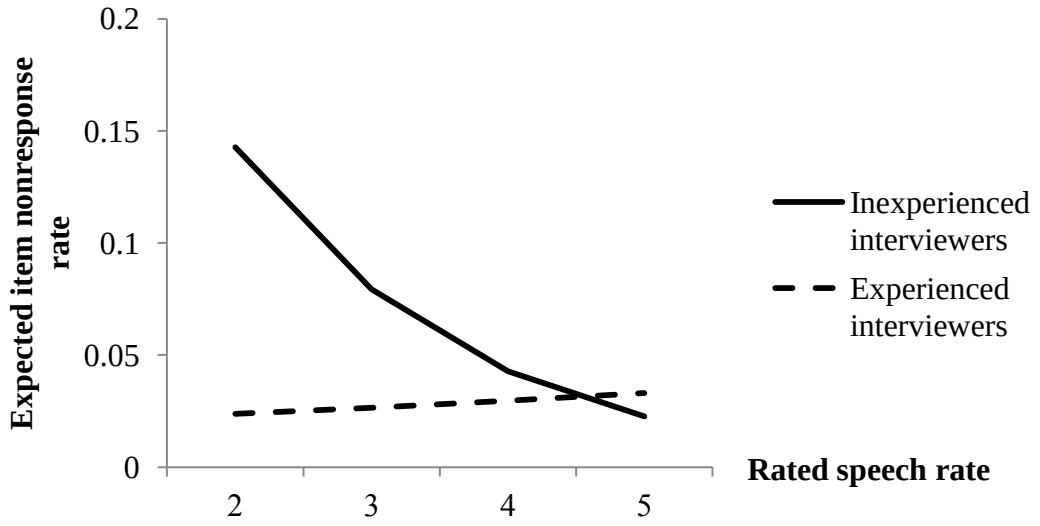


Figure 4.9 Expected item nonresponse rate by rated speech rate and interviewer experience

In addition, rated pitch affects item nonresponse rates only for socially undesirable questions. Consistent with the hypothesis (higher pitch is perceived as more attractive and thus more likely to receive responses from respondents), interviewers whose voices were rated to have higher pitch obtain lower item nonresponse rates to socially undesirable questions than those who read the questions with lower pitched voices (coefficient=-0.661, $p<0.01$; Figure 4.10).

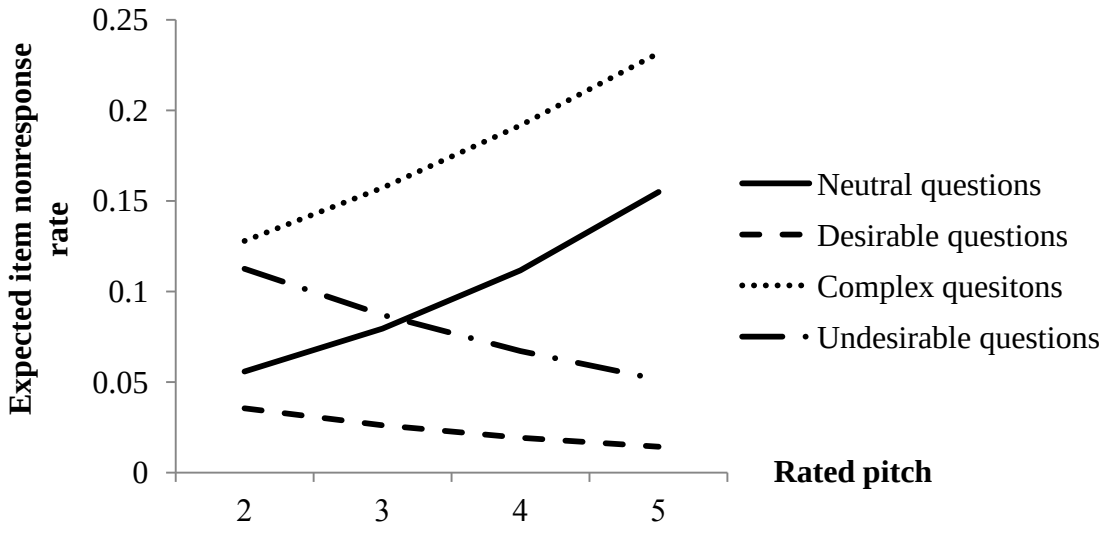


Figure 4.10 Expected item nonresponse rate by rated pitch and question type

Rounding

Table 4.6 presents the extent to which subjective voice characteristics predict the probability that respondents rounded their answers. As shown in Table 4.6, subjective voice characteristics do not affect the probability that respondents rounded their answers on question 21A, 19, and 20. When analyzing these three questions together in one model, I also found that subjective voice characteristics do not affect the probability that respondents rounded their answers (See Table E.1 in appendix E).

Table 4.6 Hierarchical logistic model predicting proportion of rounded answer by subjective voice characteristics

	Q21A (Number of times using the internet)		Q19 (Number of minutes spending on a computer)		Q20 (Number of email messages)	
	coefficient (SE)		**coefficient (SE)**		**coefficient (SE)**	
Main effects						
Intercept	-0.788(0.39)		0.771(0.37)		0.751(0.48)	
Pitch	0.696(0.43)		0.535(0.37)		-0.031(0.48)	
Intonation	-0.259(0.27)		-0.164(0.27)		0.179(0.33)	
Speech rate	0.204(0.21)		-0.230(0.18)		0.579(0.31)	
Fillers	-0.255(0.19)		0.156(0.21)		-0.101(0.28)	
Interviewer's experience < 1 year	0.070(0.39)		-0.123(0.30)		0.027(0.48)	
Female interviewer	-1.231(0.70)		-0.498(0.59)		-0.073(0.82)	
R whose education is high school or less	-0.475(0.31)		-0.643(0.28)	*	-1.257(0.33)	**
R whose age is 60 or less	1.11(0.26)	**	0.031(0.24)		0.773(0.32)	*
Variance components						
2-level variance (interviewers)	0.112(0.16)	*	0		0.251(0.24)	
Residual variance	3.29		3.29		3.29	
Model fit						
AIC	373.64		408.58		296.8	
N	279		293		287	

Note **p<0.01, *p<0.05

The hypotheses of more/less is better

Socially undesirable questions. Table 4.7 presents the results of analyses that examined the extent to which subjective voice characteristics predict answers that are less prone to socially desirable bias, i.e. better data quality, in socially undesirable questions. Rated speech rate affects data quality on question 21D (having sex). Consistent with the paralinguistic hypothesis (faster speech rates are perceived as more credible, leading to better data quality), respondents are more likely to report that they had sex at least one time in the past seven days, i.e. less prone to socially desirable bias, to interviewers perceived as reading question 21D more quickly than those perceived as reading the question with a slower pace (Figure 4.11). However, when analyzing three undesirable questions (Q5, Q21C, and Q21D) together in one model, the effect of speech rate only hold for inexperienced interviewers (coefficient=0.614, $p<0.01$; Table E.2 in Appendix E).

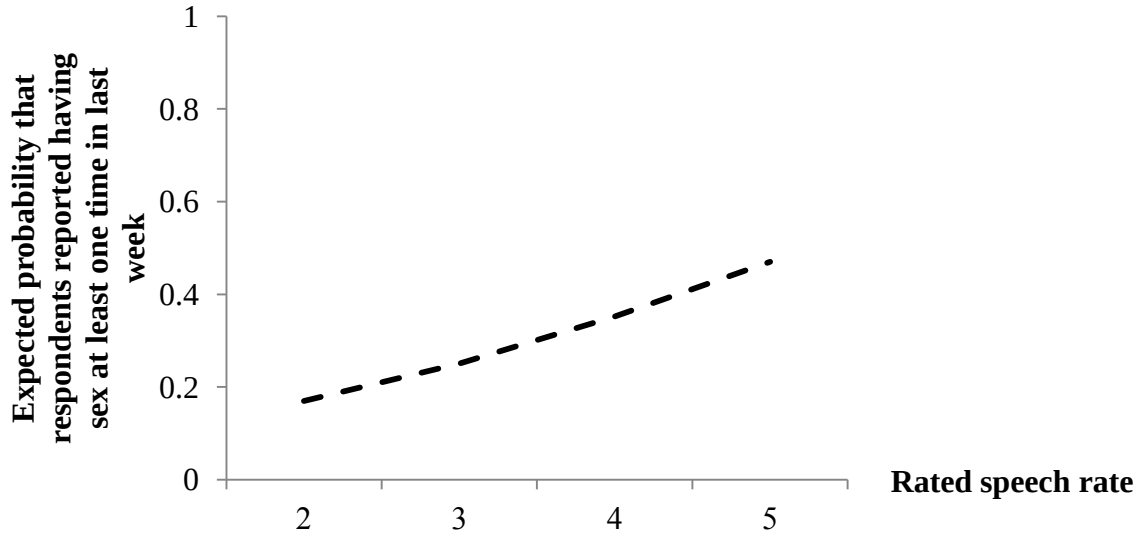


Figure 4.11 Expected probability that respondents reported that they had sex in the past seven days by rated speech rate

Table 4.7. Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by subjective voice characteristics for socially undesirable questions

	Q5 (Fired from a job)		Q21C (1+ alcohol drinks)		Q21D (Have sex 1+ times)	
	coefficient (SE)		coefficient (SE)		coefficient (SE)	
Main effects						
Intercept	-2.089(0.57)	**	-0.696(0.32)	**	-2.644(0.47)	**
Pitch	-0.623(0.38)		0.126(0.32)		-0.602(0.41)	
Intonation	-0.086(0.27)		0.158(0.21)		0.033(0.26)	
Speech rate	0.455(0.27)		0.078(0.13)		0.488(0.16)	**
Fillers	0.245(0.33)		0.086(0.17)		0.168(0.21)	
Interviewer's experience < 1 year	0.105(0.31)		0.140(0.26)		0.667(0.29)	*
Female interviewer	1.025(0.61)		-0.092(0.53)		1.526(0.68)	*
R whose education is high school or less	-0.070(0.29)		-0.351(0.23)		-0.533(0.28)	
R whose age is 60 or less	0.595(0.26)	*	0.366(0.21)		1.434(0.26)	**
Variance components						
2-level variance interviewers	0		0		0	
Residual variance	3.29		3.29		3.29	
Model fit						
AIC	403.86		550.94		404.62	
n	413		416		377	

Note **p<0.01, *p<0.05

Socially desirable questions. Table 4.8 shows the extent to which voice characteristics predict better data quality in socially desirable questions. Rated pitch and fillers affect data quality in question 13A. Respondents provide better data quality answers - i.e. report that they do not completely enjoy reading - to interviewers who read question 13A with more perceived intonation and more fillers compared to those perceived as reading the question with lower intonation and fewer fillers (coefficient for intonation=0.415, p=0.03; coefficient for fillers=0.432, p=0.03). Figure 4.12 presents the positive association between rated intonation and expected probability that respondents reported that they do not enjoy reading completely. However, when analyzing three socially desirable questions together (Q8, Q13A, and Q21F), the effects of rated pitch and fillers on data quality do not hold (See Table E.3 in Appendix E).

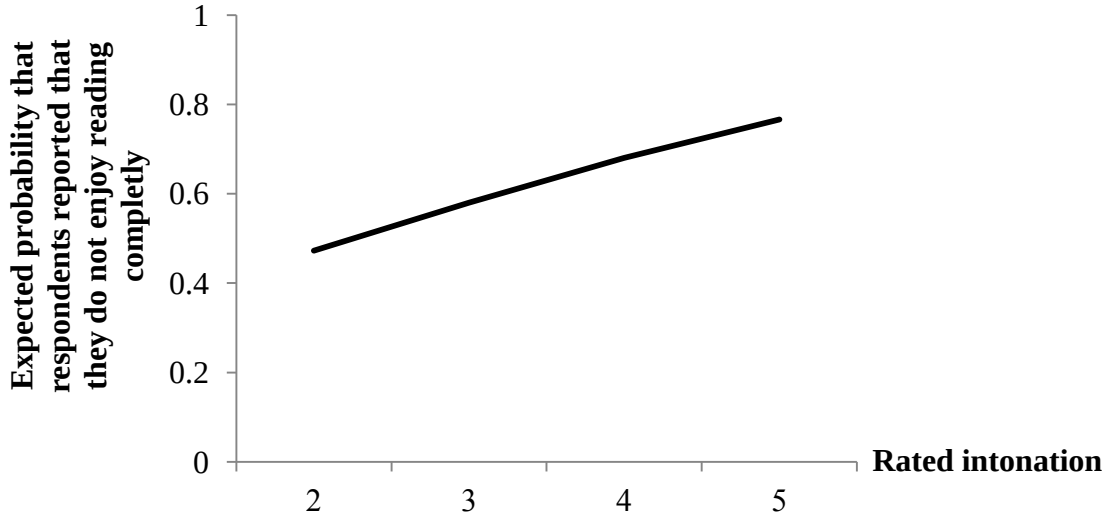


Figure 4.12 Expected probability that respondents reported that they do not enjoy reading completely by rated intonation

In addition, a negative association between rated speech rate and data quality on Q21F was only found for inexperienced interviewers (coefficient=-0.936, p=0.02). Respondents are more likely to report that they read books, magazines, or newspapers fewer than 10 times in the past 7 days to inexperienced interviewers who were perceived

as reading the question with more slowly (Figure 4.13). This association holds when analyzing three socially desirable questions together (coefficient=-0.513, p=0.02; Table E.3 in Appendix E).

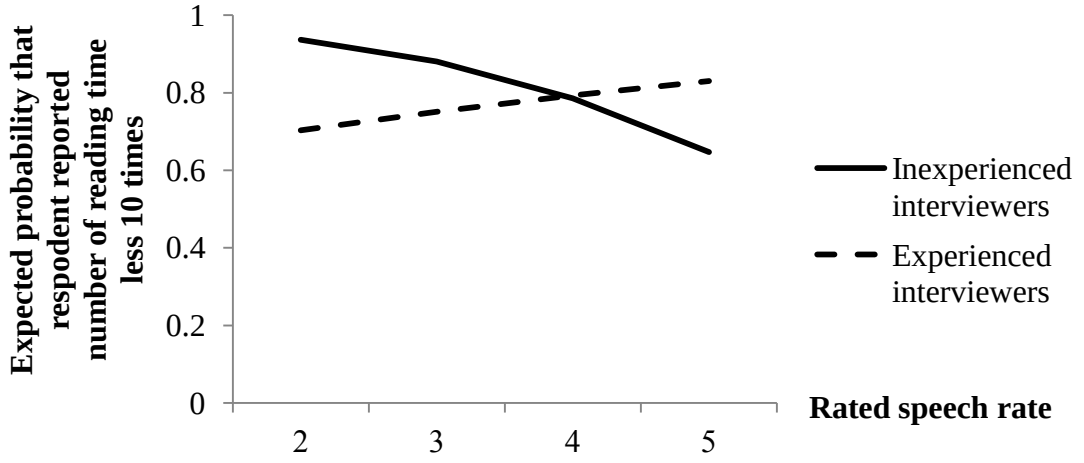


Figure 4.13 Expected probability that respondents reported reading less than 10 times in the past seven days by rated speech rate and interviewer experience

Table 4.9 presents a summary of the effects of subjective interviewer voice characteristics on data quality. Overall, subjectively perceived interviewer voice characteristics affected data quality, except for rounding. However, the effects varied by question and interviewer demographic characteristics.

Whether subjective voice characteristics mediate the relationships between objective voice characteristics and data quality is also examined in this study (See Appendix F). Results show that the mediation effects of subjective voice characteristics are found in respondents expressing uncertainty about a question and having a response in Q21F that is less prone to social desirability bias (See a summary Table F.6 in Appendix F). That is, rated speech rate and rated disfluencies mediate the effect of objective voice characteristics on respondents expressing uncertainty about a question. In addition, rated speech rate mediates the effect of objective voice characteristics on respondents reporting that they read books less than 10 times in the past 7 days.

Table 4.8 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by subjective voice characteristics for socially desirable questions

	Q8 (Did not Volunteer Activity)		Q13A (Does not completely Enjoy Reading)		Q21F (Number of Reading times<10)	
	coefficient (SE)		**coefficient (SE)**		**coefficient (SE)**	
Main effects						
Intercept	-0.017(0.26)		-0.616(0.32)		0.246(0.39)	
Pitch	-0.137(0.31)		-0.120(0.32)		0.207(0.39)	
Intonation	0.019(0.20)		0.415(0.20)	*	-0.419(0.23)	
Speech rate	-0.016(0.18)		0.154(0.14)		0.242(0.18)	
Fillers	0.099(0.14)		0.432(0.19)	*	0.19(0.22)	
Interviewer's experience < 1 year	0.026(0.26)		0.431(0.27)		0.246(0.39)	
Female interviewer	0.093(0.50)		-0.091(0.51)		0.091(0.63)	
R whose education is high school or less	1.068(0.23)	**	0.664(0.22)	**	1.073(0.28)	**
R whose age is 60 or less	-0.174(0.21)		0.875(0.21)	**	0.554(0.24)	*
Interaction btw voice and Iwer char						
Speech rate*inexperienced interviewers					-0.936(0.41)	*
Variance components						
2-level variance (interviewers)	0		0		0.104(0.14)	
Residual variance	3.29		3.29		3.29	
Model fit						
AIC	556.08		558.02		485.35	
n	408		417		413	

Note **p<0.01, *p<0.05

Table 4.9 Summary results of the effects of subjective interviewer voice characteristics on data quality indicators

	Perceived Pitch	Perceived Intonation	Perceived Speech rate	Perceived Disfluencies
Interrupt questions with answers			- (socially desirable Qs)	
Express uncertainty about a question	+ (neutral, complex Qs)	-	- (complex Qs)	U-shaped association
Request clarification about a question		- (socially desirable Qs, complex Qs)	- (socially desirable, complex Qs) + (neutral Qs)	
Give qualified answers			+	- (desirable Qs)
Give a response that does not meet question objective		-	+(neutral Qs) - (socially desirable Qs)	-
Item nonresponse	- (socially undesirable Qs)		- (inexperienced interviewers)	
Rounding				
More is better			+	
Less is better		+	- (inexperienced interviewer)	+

Note + indicates positive relationship; - indicates negative relationship; variables in parenthesis are interviewer and question characteristics for which the relationship occurs

I now turn to examining whether perceptions of interviewers' personality trait mediate the relationships between objective voice characteristics and data quality.

4.3.2 Objective 2 and 3: whether interviewers' personality traits mediate the relationship between objective voice characteristics and data quality

This study first follows Baron and Kenny (1986) to examine whether perceived interviewer personality traits mediate the association between objective voice characteristics and data quality. As mentioned earlier, mediation occurs if four conditions are fulfilled. The first condition (i.e., listeners could perceive interviewer's personality traits from interviewer voices) and second condition (i.e. objective voice characteristics affect data quality) were examined in Chapter 2 and 3 of this dissertation respectively. As

such, in this chapter, I examine whether there is an association between perceived interviewer personality traits and data quality (the third condition) and whether the effects of objective voice characteristics on data quality are reduced when controlling for the interviewer personality traits (the fourth condition). The analysis for the fourth condition was performed only for those data quality indicators that fulfill the third condition. The data quality indicators that fulfill all four conditions are evaluated to observe the direct effects of objective voice characteristics on data quality and the indirect effects through interviewer personality traits.

4.3.2.1 Objective 2: Interviewers' personality traits and data quality

Tables 4.10-4.14 show the results from analyses that examined the extent to which credibility and easiness to understand predict data quality indicators including problematic respondent behaviors (Table 4.10), item nonresponse (Table 4.11), rounded answers (Table 4.12), and responses that are less prone to socially desirable questions in socially undesirable questions (Table 4.13) and in socially desirable questions (Table 4.14). Overall, there are associations between the interviewers' personality traits and respondent behaviors, item nonresponse, and report on sexual behavior (Q21D). Unexpectedly, there is no association between interviewers' personality traits and rounded answers and responses in socially desirable questions.

Respondent behaviors

Credibility. Credibility is *positively* associated with respondents **expressing uncertainty about a question** (coefficient=0.157, $p<0.01$) and **giving a qualified answer** (coefficient=0.097, $p=0.01$). Respondents are more likely to express uncertainty about a question and give a qualified answer to interviewers who are perceived as more credible than those perceived as less credible (Figure 4.14). In contrast, credibility is

negatively associated with respondents interrupting questions with answers in socially desirable questions (coefficient=-0.164, p=0.04). Respondents are less likely to interrupt more credible interviewers on socially desirable questions (Figure 4.15). However, from Figure 4.14 and 4.15, the effects of credibility on respondent behaviors are modest. For every one score increase in credibility, the odds that respondents express uncertainty increased by $e^{0.157}$=1.170, or by 17%, and the odds that respondents give a qualified answers increased by $e^{0.097}$=1.102, or by 10.2%. However, the odds that respondents interrupt questions with answers to more credible interviewers are $e^{-0.164}$=0.849, or about 15%, lower compared to less credible interviewers.

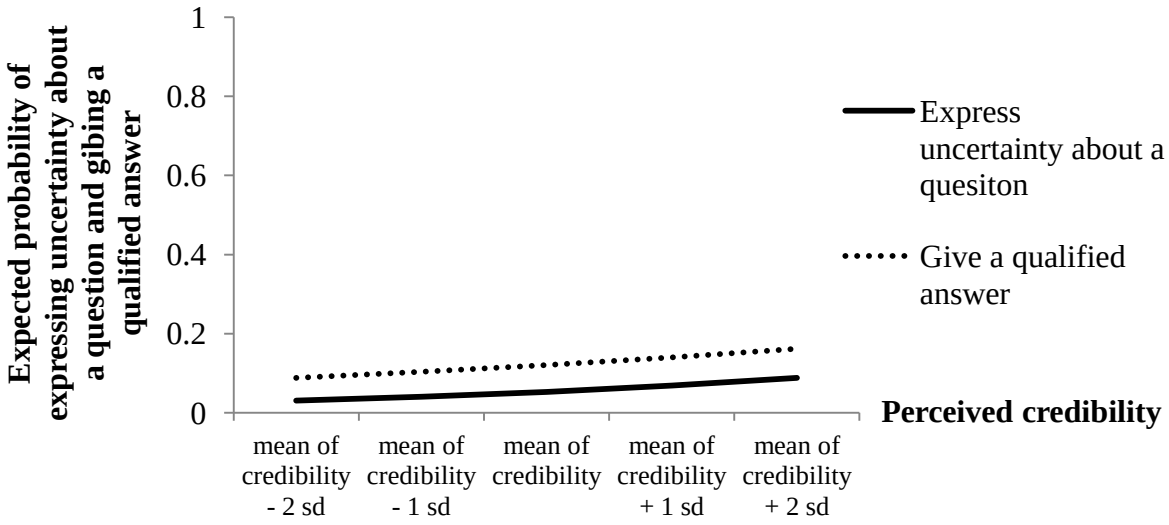


Figure 4.14. Expected probability that respondents express uncertainty about a question and give qualified answers by perceived credibility

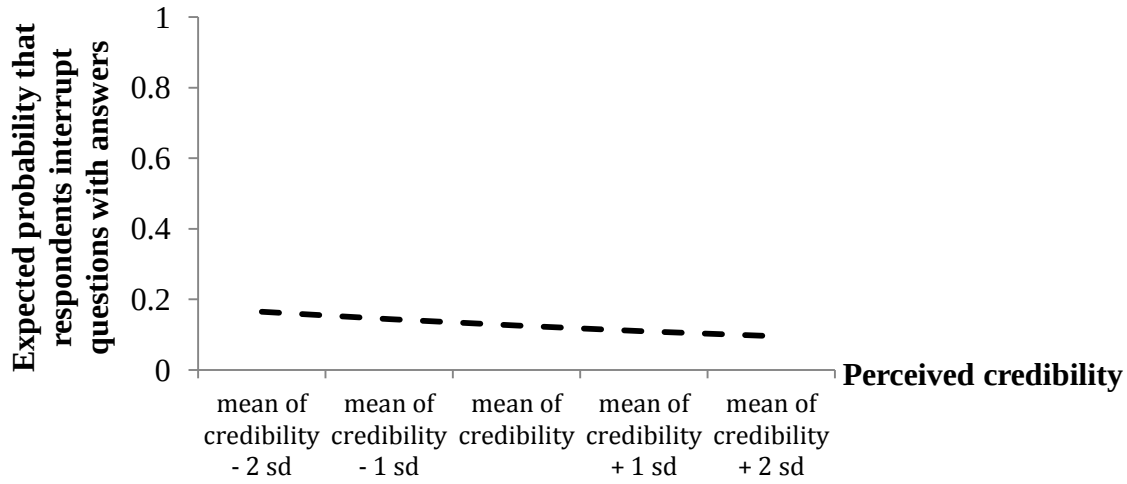


Figure 4.15 Expected probability that respondents interrupt questions with answers in socially desirable questions by perceived credibility

Easiness to understand. Perceived easiness to understand only affects respondent behaviors in complex questions; however, the direction of the effects varied by respondent behaviors (Figure 4.16). While easiness to understand is *negatively* associated with respondents **expressing uncertainty** about a question (coefficient=-1.928, p<0.01) and **requesting clarification** about a question (coefficient=-1.418, p<0.01), it is *positively* associated with respondents giving a **qualified answer** (coefficient=0.659, p=0.01) and giving a response that **does not meet question's objective** (coefficient=0.521, p=0.02).

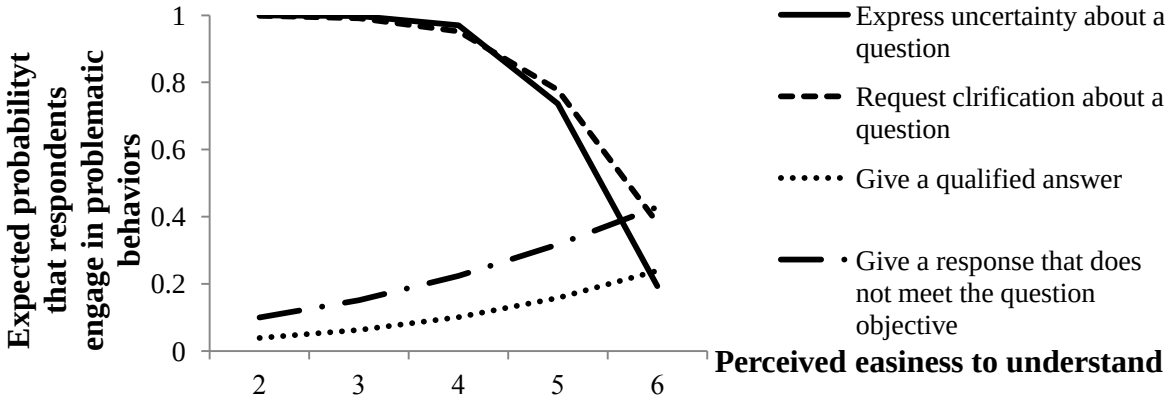


Figure 4.16 Expected probability of problematic respondent behaviors in complex questions by perceived easiness to understand

Table 4.10 Hierarchical logistic model predicting respondent behaviors by subjective ratings of interviewer personality traits

	Interrupt questions with answers		Express uncertainty about a question		Request clarification about a question		Give a qualified answer		Give a response that does not meet the question's objective	
Main effects	**coefficient (SE)**		**coefficient (SE)**		**coefficient (SE)**		**coefficient (SE)**		**coefficient (SE)**	
Intercept	-2.188(0.20)	**	-2.945(0.19)	**	-1.903(0.14)	**	-1.765(0.13)	**	-1.096(0.12)	**
Credibility	0.076(0.07)		0.157(0.05)	**	0.056(0.04)		0.097(0.04)	*	0.001(0.03)	
Easiness to understand	0.012(0.14)		-0.532(0.37)		-0.315(0.25)		-0.145(0.24)		-0.044(0.22)	
Interviewer's experience < 1 year	0.112(0.26)		0.028(0.21)		0.024(0.17)		0.121(0.14)		0.082(0.14)	
Female interviewer	-0.009(0.23)		0.114(0.19)		0.128(0.15)		0.089(0.12)		-0.050(0.12)	
Desirable question	0.569(0.14)	**	-0.197(0.19)		0.167(0.12)		-0.616(0.12)	**	-0.108(0.10)	
Complex question	0.810(0.15)	**	1.514(0.17)	**	1.384(0.12)	**	0.457(0.12)	**	0.938(0.11)	**
Undesirable question	-0.807(0.18)	**	-1.091(0.25)	**	-1.227(0.16)	**	-1.331(0.15)	**	-1.056(0.13)	**
R education is high school or less	-0.077(0.13)		0.096(0.14)		-0.043(0.10)		0.013(0.13)		0.339(0.10)	**
R whose age is 60 or less	-0.663(0.13)	**	-0.235(0.13)		-0.031(0.10)		0.075(0.12)		-0.629(0.10)	**
Credibility*desirable questions	-0.164(0.08)	*								
Credibility*complex questions	-0.097(0.08)									
Credibility*undesirable questions	0.105(0.11)									
Easiness to understand*desirable questions			0.248(0.46)		0.029(0.29)		-0.189(0.31)		0.025(0.26)	
Easiness to understand*complex questions			-1.928(0.37)	**	-1.418(0.26)	**	0.659(0.26)	*	0.521(0.23)	*
Easiness to understand*undesirable questions			-0.197(0.53)		-0.254(0.36)		-0.280(0.35)		-0.313(0.29)	
Variance components										
2-level variance (respondents)	0.420(0.10)		0.263(0.11)		0.167(0.06)		0.569(0.09)		0.370(0.07)	
3-level variance (interviewers)	0.174(0.09)		0.079(0.05)		0.054(0.03)		0		0.021(0.02)	
Residual variance	3.29		3.29		3.29		3.29		3.29	
Model fit										
Generalized Chi-square	3694.31		3846.45		4272.64		3792.76		4068.67	

Note **p<0.01, *p<0.05; n = 4,689

Item nonresponse

Table 4.11 presents the results of analyses that examined the extent to which interviewer personality traits predict item nonresponse rates. As expected, there is a negative association between easiness to understand and item nonresponse rates (coefficient=-0.897, $p<0.01$). Item nonresponse rates are lower for interviewers whose voices are easier to understand than those whose voices are less easy to understand. Credibility does not affect item nonresponse rates.

Table 4.11 Hierarchical logistic model predicting item nonresponse by subjective ratings of interviewer personality traits

	coefficient (SE)	
Main effects		
Intercept	-3.087(0.30)	**
Credibility	-0.065(0.06)	
Easiness to understand	-0.897(0.17)	**
Interviewer's experience < 1 year	0.673(0.41)	
Female interviewer	-0.495(0.37)	
Desirable question	-1.445(0.29)	**
Complex question	0.264(0.20)	
Undesirable question	-0.299(0.20)	
R whose education is high school or less	0.049(0.17)	
R whose age is 60 or less	-0.088(0.16)	
Variance components		
2-level variance (respondents)	0.234(0.15)	**
3-level variance (interviewers)	0.469(0.22)	
Residual variance	3.29	
Model fit		
Generalized Chi-square	3785.24	

Note **p<0.01, *p<0.05; n = 4,689

Rounding

Results from Table 4.12 show credibility and easiness to understand do not predict the probability that respondents rounded their answers in question 21A, 19 and 20.

Table 4.12 Hierarchical logistic model predicting proportion of rounded answer by subjective ratings of interviewer personality traits

	Q21A (Number of times using the internet)		Q19 (Number of minutes spending on a computer)		Q20 (Number of email messages)	
	coefficient (SE)		coefficient (SE)		coefficient (SE)	
Main effects						
Intercept	-0.592(0.29)		0.289(0.25)		1.328(0.37)	**
Credibility	0.149(0.11)		-0.148(0.11)		0.018(0.14)	
Easiness to understand	0.264(0.45)		0.279(0.44)		0.415(0.58)	
Interviewer's experience < 1 year	0.318(0.38)		-0.221(0.29)		-0.120(0.50)	
Female interviewer	-0.363(0.33)		0.412(0.26)		-0.103(0.44)	
R whose education is high school or less	-0.409(0.30)		-0.616(0.28)	*	-1.274(0.33)	**
R whose age is 60 or less	1.159(0.26)	**	0.015(0.24)		0.844(0.31)	*
Variance components						
2-level variance (interviewers)	0.156(0.18)		0.016(0.11)		0.416(0.29)	
Residual variance	3.29		3.29		3.29	
Model fit						
Generalized Chi-square	270.59		292.33		256.58	
n	279		293		287	

Note **p<0.01, *p<0.05

The Hypothesis of More/Less is Better

Socially undesirable questions. Table 4.13 presents the extent to which interviewer personality traits predict better data quality on socially undesirable questions. There is no association between an interviewer's personality traits and reports in Q5 (being fired from a job) and Q21C (alcoholic drink). As shown in Table 4.13, the perception of an interviewer's credibility affects data quality for Q21D (having sex), but the effect varies by interviewer sex. Figure 4.17 presents the expected probability that a respondent reports that they had sex at least one time in the past seven days by perceived credibility. Respondents are *more likely* to report that they had sex at least one time in the past seven days, i.e. less prone to socially desirable bias, to male interviewers whose voices are perceived as more credible but *slightly less likely* to state this for female interviewers whose voices are perceived as more credible.

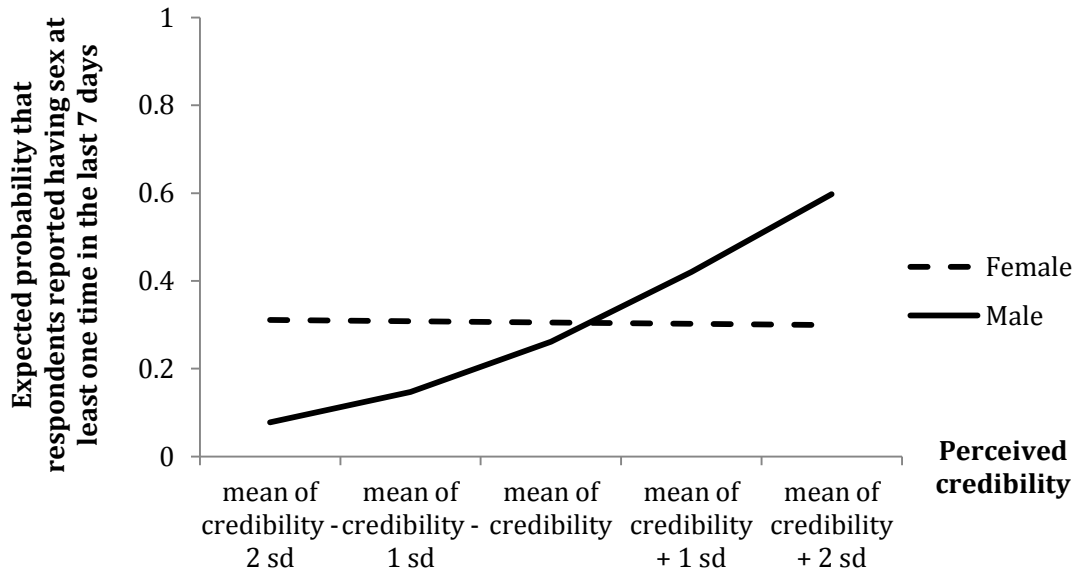


Figure 4.17 Expected probability that respondents reported having sex at least one time in the last 7 days by perceived credibility

Table 4.13 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by subjective ratings of interviewer personality traits for socially undesirable questions

	Q5 (Fired from a job)		Q21C (1+ alcohol drinks)		Q21D (Have sex 1+ times)	
	coefficient (SE)		coefficient (SE)		coefficient (SE)	
Main effects						
Intercept	-1.801(0.25)	**	-0.760(0.20)	**	-1.926(0.30)	**
Credibility	-0.043(0.10)		0.108(0.07)		0.405(0.13)	**
Easiness to understand	-0.206(0.45)		-0.134(0.30)		-0.570(0.38)	
Interviewer's experience < 1 year	0.184(0.30)		0.101(0.25)		0.918(0.33)	
Female interviewer	0.011(0.27)		0.095(0.22)		0.215(0.29)	
R whose education is high school or less	-0.067(0.28)		-0.358(0.23)		-0.628(0.29)	*
R whose age is 60 or less	0.587(0.26)	*	0.347(0.21)		1.485(0.26)	**
Interaction btw interviewer traits and Iwer char						
Credibility*Female					-0.413(0.16)	*
Variance components						
2-level variance interviewers)	0.003(0.11)		0		0.066(0.14)	
Residual variance	3.29		3.29		3.29	
Model fit						
Generalized Chi-square	410.09		414.63		362.43	
n	413		416		377	

Note **p<0.01, *p<0.05

Table 4.14 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by subjective ratings of interviewer personality traits for socially desirable questions

	Q8 (Did not Volunteer Activity)		Q13A (Does not completely Enjoy Reading)		Q21F (Number of Reading times<10)	
	coefficient (SE)		coefficient (SE)		coefficient (SE)	
Main effects						
Intercept	-0.043(0.19)		-0.739(0.20)	**	0.255(0.29)	
Credibility	-0.093(0.08)		0.038(0.08)		-0.040(0.10)	
Easiness to understand	0.347(0.38)		-0.107(0.33)		-0.079(0.38)	
Interviewer's experience < 1 year	-0.058(0.26)		0.212(0.27)		0.130(0.42)	
Female interviewer	-0.058(0.22)		-0.198(0.22)		0.207(0.36)	
R whose education is high school or less	1.069(0.23)	**	0.678(0.22)	**	1.041(0.28)	**
R whose age is 60 or less	-0.177(0.21)		0.783(0.21)	**	0.557(0.24)	*
Variance components						
2-level variance (interveiwers)	0		0.015(0.07)		0.33(0.22)	
Residual variance	3.29		3.29		3.29	
Model fit						
Generalized Chi-square	407.88		415.39		383.15	
n	408		417		413	

Note **p<0.01, *p<0.05

Socially desirable questions. Results from Table 4.14 shows that there is no association between interviewer personality traits and reports that are less prone to socially desirable bias in Q8, Q13A, and Q21C.

Table 4.15 presents a summary of the effects of interviewer personality traits on data quality found in this study. As mentioned earlier, to examine whether perceived interviewers' personality traits mediate the association between objective measures of interviewer voice characteristics and data quality, an association between interviewer personality traits and data quality indicators must be held (the third condition to examine the mediation effects proposed by Baron & Kenny 1986). As such, data quality indicators that are not associated with perceived interviewer personality traits will not be used for mediation analyses. As shown in Table 4.15, there may be potential mediation effects of interviewer personality traits on the relationships between interviewer voice characteristics and 7 data quality indicators including 5 problematic respondent behaviors, item nonresponse, and response in question 21D.

Table 4.15 Summary of the effects of interviewer personality traits on data quality

		Credibility	**Easiness to understand**
Respondent behaviors	Interrupt questions with answers	**-** (Desirable Qs)	
	Express uncertainty about a question	+	- (Complex Qs)
	Request clarification about a question		- (Complex Qs)
	Give a qualified answer	+	+ (Complex Qs)
	Give a response that does not meet question objective		+ (Complex Qs)
Item nonresponse			**-**
Rounding	Q19		
	Q20		
	Q21A		
More is better	Q5		
	Q21C		
	Q21D	+ (male)/ - (female)	
Less is better	Q8		
	Q13A		
	Q21F		

4.3.2.2 Objective 3: Effects of interviewer voice and interviewers' personality traits on data quality

The fourth condition to examine whether mediation occurs is that the effects of interviewer voices on data quality must be reduced when controlling for interviewer personality traits (Baron & Kenny 1986). In addition, effects of interviewer personality traits on data quality must be significant. As discussed above, there may be potential mediation effects of interviewer personality traits on the effect of interviewer voice characteristics on seven data quality indicators including five problematic respondent behaviors, item nonresponse, and responses to question 21D. As such, these seven data quality indicators are evaluated for potential mediation effects.[6]

Respondent behaviors

Table 4.16 includes both perceived interviewer personality traits and objectively measured voice characteristics predicting problematic respondent behaviors. By comparing Table 4.16 and 3.3, I can examine whether the effect of objective voice characteristics on response behaviors is reduced after controlling for interviewer personality traits in the model.

***Interrupt question with answers*.** Comparing Table 3.3 and Table 4.16, the effect of objectively measured speech rate is reduced when perceived interviewer personality traits are included. In addition, the negative association between credibility and respondents interrupting questions with answers is still significant (coefficient=-0.163, p=0.04). As such, the fourth condition to test whether the mediation effect exists (Baron

[6] I also examined the potential mediation effects for rounded answers and responses less prone to socially desirable bias. I found effects of interviewer personality traits on these data quality indicators are not significant, implying that there is no mediation effects of interviewer personality traits on these data quality indicators (Results are in Appendix G).

& Kenny 1986) is fulfilled. Thus, credibility mediates the association between speech rate and respondents interrupting questions with answers.

Express uncertainty about a question. Comparing Table 3.3 and Table 4.16, when interviewer personality traits were included in the model, the effects of speech rate and intonation on respondents expressing uncertainty about a question are reduced overall, but the interaction effect between speech rate and socially undesirable questions increases (the coefficient changes from -0.982 ($p<0.01$; Table 3.3) to -1.052 ($p<0.01$; Table 4.16)). As such, the fourth condition to test whether the mediation effect exists (Baron & Kenny 1986) is *not* fulfilled.

Request clarification about a question. Comparing Table 3.3 and Table 4.16, the effects of speech rate and intonation on respondents requesting clarification become *stronger*. For example, the effect of speech rate on respondents requesting clarification for inexperienced interviewers changes from 0.312 ($p<0.01$; Table 3.3) to 0.352 ($p<0.01$; Table 4.16). As such, the fourth condition to test whether the mediation effect exists (Baron & Kenny 1986) is *not* fulfilled.

Give a qualified answer. Comparing Table 3.3 and Table 4.16, the effects of speech rate, intonation, and fillers on respondents giving a qualified answer are reduced when perceived interviewer personality traits are included in the model. However, the effects of easiness to understand on data quality in complex questions becomes insignificant (coefficient=0.395, $p=0.16$). As such, the fourth condition to test whether the mediation effect exists (Baron & Kenny 1986) is *not* fulfilled.

Table 4.16. Hierarchical logistic model predicting respondent behaviors by objective voice characteristics and subjective ratings of interviewer personality traits

	Interrupt questions with answers coefficient (SE)		Express uncertainty about a question coefficient (SE)		Request clarification about a question coefficient (SE)		Give a qualified answer coefficient (SE)		Give a response that does not meet the question's objective coefficient (SE)	
Main effects										
Intercept	-2.067(0.23)	**	-3.206(0.26)	**	-2.062(0.17)	**	-1.719(0.15)	**	-1.343(0.17)	**
Pitch	0.002(0.003)		-0.003(0.003)		-0.001(0.003)		0.0001(0.002)		0.007(0.002)	**
Intonation	-0.001(0.004)		0.001(0.007)		0.002(0.005)		-0.004(0.003)		-0.005(0.003)	
Intonation^2			0.0004(0.0001)	**	0.0002(0.00001)	**				
Speech rate	0.493(0.17)	**	0.445(0.22)	*	0.227(0.14)		0.275(0.13)	*	0.182(0.12)	
Speech rate^2							-0.110(0.05)	*		
Fillers	-0.078(0.11)		-0.020(0.14)		-0.094(0.105)		0.110(0.13)		0.174(0.14)	
Credibility	0.050(0.07)		0.132(0.06)	*	0.053(0.04)		-0.582(0.23)	*	-0.009(0.04)	
Easiness to understand	-0.035(0.16)		-0.164(0.40)		-0.200(0.26)		-0.094(0.25)		-0.023(0.23)	
Interviewer's experience < 1 year	0.139(0.28)				-0.033(0.19)		0.100(0.14)		0.134(0.17)	
Female interviewer	-0.093(0.30)				0.294(0.21)		0.225(0.18)		0.504(0.19)	*
Desirable question	0.534(0.14)	**			0.193(0.12)		-0.665(0.13)	**	-0.067(0.11)	
Complex question	0.799(0.15)	**			1.368(0.13)	**	0.487(0.12)	**	0.968(0.11)	**
Undesirable question	-0.717(0.19)	**			-1.141(0.18)	**	-1.211(0.17)	**	-0.928(0.14)	**
R whose education is high school or less	-0.072(0.14)				-0.036(0.11)		0.014(0.13)		0.340(0.11)	**
R whose age is 60 or less	-0.680(0.13)	**			-0.022(0.10)		0.080(0.12)		-0.653(0.10)	**

Note **p<0.01, *p<0.05; n = 4,689

Table 4.16 Continued.

Interaction btw voice and Q char												
Intonation*desirable question				-0.004(0.01)			-0.015(0.06)	*				
Intonation*complex question				-0.014(0.007)	*		-0.015(0.01)	**				
Intonation*undesirable question				-0.020(0.01)			-0.021(0.01)	**				
Speech rate*desirable question	-0.141(0.22)			-0.370(0.31)			-0.300(0.19)		0.096(0.21)		0.106(0.16)	
Speech rate*complex question	-0.334(0.20)			-1.222(0.23)	**		-0.791(0.16)	**	-0.214(0.16)		0.146(0.14)	
Speech rate*undesirable question	-0.944(0.23)			-1.052(0.30)	**		-0.718(0.19)	**	-0.527(0.19)		-0.540(0.15)	**
Interaction btw voice and Iwer char												
Speech rate*interwer exp < 1 yr							0.352(0.12)	**				
Fillers*interwer exp < 1 yr							0.515(0.05)	*				
Fillers*female interviewer									-0.582(0.23)		-0.475(0.18)	**
Interaction btw trait and Q char												
Credibility*desirable question	-0.163(0.08)		*									
Credibility*complex question	-0.097(0.08)											
Credibility*undesirable question	0.171(0.11)											
Easiness to understand*desirable question				0.147(0.49)			0.019(0.30)		-0.185(0.32)		0.082(0.26)	
Easiness to understand*complex question				-1.375(0.40)	**		-1.020(0.28)	**	0.395(0.28)		0.222(0.24)	
Easiness to understand*undesirable question				-0.292(0.55)			-0.272(0.37)		-0.360(0.35)		-0.335(0.29)	
Variance components												
2-level variance (respondents)	0.426(0.10)		**				0.179(0.07)		0.595(0.09)		0.393(0.07)	**
3-level variance (interviewers)	0.204(0.11)		**				0.07(0.05)		0		0.048(0.04)	**
Residual variance	3.29			3.29			3.29		3.29		3.29	
Model fit												
Generalized Chi-square	3704.06						4267.14		3734.83		4027.23	

Note **p<0.01, *p<0.05; n = 4,689

Give a response that does not meet the question's objective. Comparing Table 3.3 and Table 4.16, the effects of pitch and speech rate on respondent giving a response that does not meet the question's objective stay the same, but the effect of fillers on respondent giving a response that does not meet the question's objective is reduced. However, the effect of easiness to understand on respondents giving a response that does not meet the question's objective in complex question becomes insignificant (coefficient=0.222, p=0.36). As such, the fourth condition to test whether the mediation effect exists (Baron & Kenny 1986) is *not* fulfilled.

Item nonresponse. Table 4.17 examines the extent to which interviewer personality traits and objective voice characteristics predict item nonresponse rates. Comparing between Table 3.4 and Table 4.17, the effect of intonation on item nonresponse rates becomes insignificant while the effect of speech rate on item nonresponse rates is reduced, i.e. the coefficient changes from -0.433 (Table 3.4) to -0.303 (Table 4.17). In addition, the effect of easiness to understand on item nonresponse rates is still significant (coefficient=-0.556, $p<0.01$). As such, the fourth condition to test whether the mediation effect exists (Baron & Kenny 1986) is fulfilled. Easiness to understand mediates the effect of speech rate and intonation on item nonresponse.

Table 4.17 Hierarchical logistic model predicting item nonresponse by objective voice characteristics and subjective ratings of interviewer personality traits

	coefficient (SE)	
Main effects		
Intercept	-3.190(0.34)	**
Pitch	-0.0002(0.004)	
Intonation	-0.009(0.006)	
Intonation^2	0.0002(0.0001)	
Speech rate	-0.303(0.10)	**
Fillers	-0.040(0.18)	
Credibility	-0.063(0.06)	

Easiness to understand	-0.556(0.19)	**
Interviewer's experience < 1 year	0.452(0.44)	
Female interviewer	-0.441(0.46)	
Desirable question	-1.359(0.29)	**
Complex question	0.149(0.21)	*
Undesirable question	-0.027(0.22)	
R whose education is high school or less	0.066(0.17)	
R whose age is 60 or less	-0.068(0.16)	
Interaction btw voice and Iwer char		
Speech rate*interviewer experience < 1 yr	-0.268(0.15)	
Variance components		
2-level variance (respondents)	0.237(0.15)	
3-level variance (interviewers)	0.526(0.15)	**
Residual variance	3.29	
Model fit		
Generalized Chi-square	3699.3	

Note **p<0.01, *p<0.05; n = 4,689

Response in Q21D (sexual behaviors). Table 4.18 examines the extent to which interviewer voice characteristics and perceived interviewer personality traits predict a report of having sex at least one time in the past seven days (Q21D). Comparing between Table 3.6 and Table 4.18, the effect of speech rate becomes slightly stronger when interviewer personality traits are included in the model, i.e. the coefficient changes from 0.638 (p=0.02) (Table 3.6) to 0.647 (p=0.04) (Table 4.18). As such, the fourth condition to test whether the mediation effect exists (Baron & Kenny 1986) is *not* fulfilled.

Table 4.18 Hierarchical logistic model predicting proportion of reporting having sex at least one time in the past seven days by objective voice characteristics and subjective ratings of interviewer personality traits

	coefficient (SE)	
Main effects		
Intercept	-1.889(0.41)	**
Pitch	0.008(0.01)	
Intonation	-0.009(0.01)	
Speech rate	0.647(0.32)	*
Fillers	-0.383(0.51)	
Credibility	0.204(0.11)	

Easiness to understand	0.932(0.63)	
Interviewer's experience < 1 year	0.751(0.36)	
Female interviewer	-0.148(0.55)	
R whose education is high school or less	-0.568(0.29)	
R whose age is 60 or less	1.498(0.27)	
Interaction btw voice and Iwer char		
Speech rate*female interviewers	-0.841(0.39)	*
Interaction btw Iwer trait and Iwer char		
Easiness to understand*female interviewers	-2.421(0.70)	**
Variance components		
2-level variance interviewers	0.148(0.19)	
Residual variance	3.27	
Model fit		
Generalized Chi-square	362.19	
n	377	

Note **p<0.01, *p<0.05

In summary, only two data quality indicators - respondents interrupting questions with an answer and item nonresponse rates - fulfill all four conditions for a mediation effect (Baron & Kenny 1986). Results from Chapter 3 indicate that only speech rate affects respondents interrupting questions and answers while intonation and speech rate affect item nonresponse rates. As such, credibility mediates the effect of speech rate on respondent interrupting questions with answers. In addition, easiness to understand mediates the effects of intonation and speech rate on item nonresponse rates.

Next, I examined a wide variety of formal tests for mediation effects for these two outcomes. There are multiple approaches to testing for mediation. First, I attempted to use the moderated mediation model proposed by Bauer, Preacher, and Gill (2006), estimated in SAS. However, the Bauer, Preacher and Gill moderated mediation multilevel model proposed only accounts for a 2-level multi-level model and only applies to one mediator and one moderator. Because the models for respondents interrupting questions

with answer and item nonresponse rates use 3-level models with two moderators, i.e. interviewer sex and experience, this analysis procedure cannot be used in this study.

Next, I also examined using a Sobel test for mediation via Stata; however, it is only developed for a single-level model, and thus cannot be used for this dissertation. Finally, Generalized Structural Equation Modeling (GSEM) was used to estimate 3-level multilevel models with one mediator and two moderators in Stata. Unfortunately, the models failed to converge. As such, the direct effect of objective voice characteristics on data quality indicators as well as the indirect effect through interviewer traits cannot be computed in this study. Future research should examine the direct and indirect effects further.

4.4 Conclusion and discussion

There were three objectives in this chapter. The first objective of this chapter was to examine how subjective measures of interviewer voice characteristics including rated pitch, intonation, speech rate, and disfluencies affect data quality. The second objective of this chapter was to examine potential association between perceived interviewer personality traits and data quality. Finally, the third objective was to examine whether these perceived interviewer personality traits mediate the relationship between objective voice characteristics and data quality.

Subjective voice characteristics and data quality

Overall, interviewer voice characteristics affected data quality, except for rounding. However, the effects varied by question and interviewer characteristics (Table 4.6).

Pitch. There is a positive association between perceived pitch and respondents expressing uncertainty about a question in socially desirable and complex questions. In addition, consistent with the hypothesis, I found lower item nonresponse rates among interviewers perceived to have higher pitched voices. As found in Chapter 2, high-pitched voices are perceived as more credible than lower-pitched voices. As such, respondents may trust these more credible interviewers and perhaps find them more attractive (Ketrow 1990, Oksenberg, et al. 1986), and thus be more comfortable expressing uncertainty about a complex or socially desirable question and more likely to provide an answer overall.

Intonation. Interviewers obtain higher quality responses when their question reading is perceived to have more intonation. Respondents are less likely to express uncertainty about a question, request clarification about a question (only for socially desirable and complex questions), give a response that does not meet question's objective, and more likely to report that they did not completely enjoy reading books, magazine, and newspapers to interviewers perceived to have more intonation versus less intonation. Respondents may perceive voices with higher intonation (more pitch variability) as easier to understand (as I found within an intonation ranging from 20 to 80 Hz which is account for 95% of voice files in Chapter 2). As such, they are less likely to express uncertainty and request clarification about questions as I found in this chapter. In addition, within an intonation ranging from 20 to 80 Hz, respondents perceived interviewer's voices with higher intonation as more credible (See Chapter 2). Thus, they are more likely to provide better data quality in their responses (give response that meet question's objective and responses that less prone to socially desirable bias) to

interviewers whom they perceive as more credible, i.e. have higher intonation (Groves 1990; Addington 1968). However, in this chapter, I did not find a direct effect of perceived credibility on respondents giving a response that does not meet question's objective and data quality in socially desirable questions.

Speech rate. The effect of perceptions of interviewer voices on data quality is the most prominent for speech rate. The effects of speech rate on data quality vary by question type and interviewer experience. The results for neutral questions are consistent with the survey methodological literature. Respondents are less likely to engage in problematic interview behaviors when interviewers are perceived to read questions more slowly. Respondents may perceive interviewers who read a question with a slower pace as they would like them to take time to think about their responses attentively, leading to lower proportion of requesting clarification about a question and giving a response that does not meet question's objective.

In contrast to neutral questions, the results for socially undesirable and complex questions are consistent with paralinguistic studies. Respondents are less likely to express uncertainty about a complex questions and are more likely to provide better data quality, i.e. provide higher report of having sex at least one time in the past seven days, to interviewers who are perceived to read questions *more quickly*. In addition, inexperienced interviewers perceived as reading a question with faster speech rates had lower item nonresponse rates than those who read a question with slower speech rates. Results from Chapter 2 of this dissertation indicate that listeners perceive interviewers who read a question with a faster pace as more credible than interviewers who read a question with a slower pace. Respondents are more likely to provide better data quality to interviewers

perceived as more credible, which are interviewers who read a question with a faster pace. For example, in this chapter, I found more credible male interviewers received higher reports of having sex at least one time in the past seven days than those perceived as less credible. In addition, because respondents may not be able to keep all of the information about complex questions in working memory when interviewers read complex questions with a slower pace, respondents are more likely to express uncertainty about a question to interviewers who read a question with a slower pace. Moreover, in complex questions, a faster speech rate is evaluated as easier to understand than slower speech rate (result from chapter 2), leading to less uncertainty about a complex questions as I found in this chapter.

Results for socially desirable questions are mixed. Respondents are less likely to interrupt questions with answers, request clarification about a question, and give a response that does not meet the question's objective when they perceived that interviewers read questions *faster* compared to when they perceived that interviewers read questions more slowly. However, respondents provide better data quality, i.e. reporting the number of reading times fewer than 10 times, when inexperienced interviewers are perceived to read the question at slower speech rates.

Results from this study suggest that speech rate may affect respondents at different cognitive stages depending on types of questions. Speech rate mainly affects respondents at the perception stage for socially undesirable and complex questions, but at other stages for neutral questions. As such, the "ideal" reading rate for questions may vary by the type of question being asked.

Disfluencies. Consistent with the hypothesis from survey practice, respondents provide better data quality when interviewers are perceived to read questions with higher disfluency rates, especially for socially desirable questions. Respondents are less likely to give a response that does not meet the question's objective, give qualified answers (only for socially desirable questions), and are more likely to report that they do not completely enjoy reading when interviewers are perceived to read questions with higher disfluency rates than to interviewers perceived to read questions with lower disfluency rates. Disfluencies have a "disfluency advantage" that allows respondents to have more time to think about their answers, thus increasing data quality.

In addition, fewer respondents express uncertainty about a question when the interviewer reads questions with moderate perceived disfluencies rather than having too many or too few disfluencies. This is consistent with previous research that found interviewers who speak with neither robotic speech nor are highly disfluent have the highest participation rates (Conrad et al. 2013).

Interviewer demographic characteristics. Unexpectedly, I found that inexperienced interviewers obtained higher item nonresponse than experienced interviewers. However, the effect was opposite when inexperienced interviewers are perceived as reading question faster - inexperienced interviewers who are perceived as reading question *faster* received *lower* item nonresponse rates than those who are perceived as reading questions slower. In addition, as expected, I found that inexperienced and female interviewers received higher reports that respondents had sex at least one time in the past seven days compared to experienced and male interviewers. More experienced interviewers tend to be careless in conducting a survey compared to

less experienced interviewers, thus, more experienced interviewers received lower data quality (Olson and Bilgen 2011; Groves, et al. 2009). Moreover, previous research found respondents are more likely to report more sexual information to female interviewers compared to male interviewers (Catania, et al. 1996).

Mediation effect of interviewer's personality traits on the effects of objective voice characteristics on data quality

As discussed earlier, to establish mediation, four conditions proposed by Baron & Kenny (1986) must hold. The first and second conditions are examined in Chapter 2 and 3 respectively. In this chapter, I examined the third condition that interviewer personality traits must affect data quality indicator and the fourth condition that the effect of interviewer voice characteristics on data quality must be reduced after controlling for interviewer personality traits.

Interviewer personality traits and data quality indicators. I examined whether two interviewer personality traits - credibility and easiness to understand - affect data quality. I found that interviewer personality traits affect respondents behaviors, item nonresponse, and reports on Q21D (having sex) (Table 4.15).

Credibility. Respondents are more likely to express uncertainty about a question and give a qualified answer, but are less likely to interrupt socially desirable questions with answers to interviewers whom they perceived as more credible compared to those whom they perceived as less credible. Respondents trust credible interviewers; thus, they are less likely to interrupt questions with answers but they are more likely to express uncertainty about a question in order to get a clarification about a question. In addition, they may try their best to give an answer to credible interviewers even though they are

not confident in their answer and give a qualified answer. Moreover, respondents are more likely to report that they had sex at least one time in the past seven days to male interviewers who are perceived as more credible compared to less credible male interviewers. This is consistent to Groves (1989) who stated that respondents are more likely to give better data quality to interviewer whom they perceived as more credible. The opposite direction was found for female interviewers, however, the effect is modest.

Easiness to understand. As expected, item nonresponse rates are higher among interviewers whose voices are perceived as less easy to understand. When respondents do not understand questions, they may not respond to survey questions (Beatty & Herrmann, 2002). Perceived easiness to understand only affects respondent behaviors for complex questions. Respondents are less likely to express uncertainty and request clarification about a question, but are more likely to give a qualified answer and give a response that does not meet the question's objective to interviewers whose voices are perceived as easier to understand compared to those whose voices are perceived as less easy to understand. Because respondents are more likely to comprehend questions read by interviewers whose voices are easier to understand, they are less likely to express uncertainty and request clarification about a question. However, because respondents may experience difficulty answering complex questions, they may try their best to answer the questions by giving qualified answers or giving answers that do not meet question's objective.

Mediation effects. I examined whether there are mediation effects for perceptions of an interviewer's personality traits, i.e. credibility and easiness to understand, on data quality indicators including five problematic respondent behaviors, item nonresponse

rates, and reports on question 21D (having sex). I found mediation effects of interviewer personality traits on two data quality indicators. That is, interviewer credibility mediates the effects of objective measure of speech rate on respondents interrupting questions with answers and interviewer whose voices are easy to understand mediate the effects of objective measure of speech rate and intonation on item nonresponse rates. Unfortunately, direct and indirect effects cannot be estimated through the moderated mediation multi-level model.

4.5 Limitations and future research

This study has limitations. First, the main assumption in this study is that undergraduate raters' subjective perceptions are a good proxy for respondent's subjective perceptions. However, this may not be the case. In addition, a sample of questions was analyzed in this study. Including more questions analyzed in this study will increase statistical power in the analyses in this study. Finally, this study examined only a limited number of highly correlated interviewer personality traits as perceived by raters. Due to multicollinearity issues, four of these traits had to be combined into a single "credibility" factor. Although this is consistent with theory about how these traits should be related, it limits the ability to make conclusions about any single trait.

Because the moderated mediation multi-level model to examine the direct effects of interviewer voice characteristics on data quality indicator and the indirect effects through interviewer personality traits failed to converge in this study, future studies should examine these effects further.

CHAPTER 5: CONCLUSION

In this dissertation, I examined the effects of interviewer voice characteristics including pitch, intonation, speech rate, and disfluencies, on survey data quality. Voices can be either objectively measured by a computer program or subjectively evaluated by coders. In this study, I used the Praat computer program to extract objective voice information. In addition, I recruited six undergraduate students (4 males and 2 females) to subjectively evaluate the same interviewer voice characteristics as measured objectively and interviewer personality traits (expertise, confidence, reliability, trustworthiness, and easiness to understand) based on a seven-point scale.

In telephone surveys, interviewers are typically instructed to read a question with proper phrasing and inflection and to read a question at a speech rate of 2 words per second (wps) which is slower than the speech rate in the ordinary conversation (Tauroza & Allison 1990; Guenzel, et al. 1983; Cannell, et al. 1981). However, no prior studies had examined whether these voice characteristics affect data quality.

Because gold standard data is not available in this dissertation, data quality was measured through item nonresponse, rounded answers, the hypothesis of "more/less is better" in reports on sensitive or socially desirable items, and five respondent behaviors associated with data quality: the respondent 1) interrupts questions with an answer, 2) expresses uncertainty about a question, 3) requests clarification, 4) gives qualified answers, and 5) gives a response that does not meet the question's objective (Tourangeau, et al. 2000).

I examined socially desirable, undesirable, and complex questions because a speaker's vocal pattern changes for these types of questions and because there are known

problems with data quality on these types of questions. In addition, I also examined neutral questions (not complex and not socially desirable/undesirable) as a comparison for both socially (un)desirable and complex question.

Specifically, this dissertation had three objectives. The *first* was to examine whether listeners could perceive interviewer voice characteristics and interviewer personality traits from interviewer voices (Objective 1; Figure 5.1). I hypothesized that there are high associations between objective and subjective measures of interviewer voice characteristics. In addition, I hypothesized that listeners can perceive five interviewer personality traits (expertise, confidence, reliability, trustworthiness, and easiness to understand) from interviewer voices.

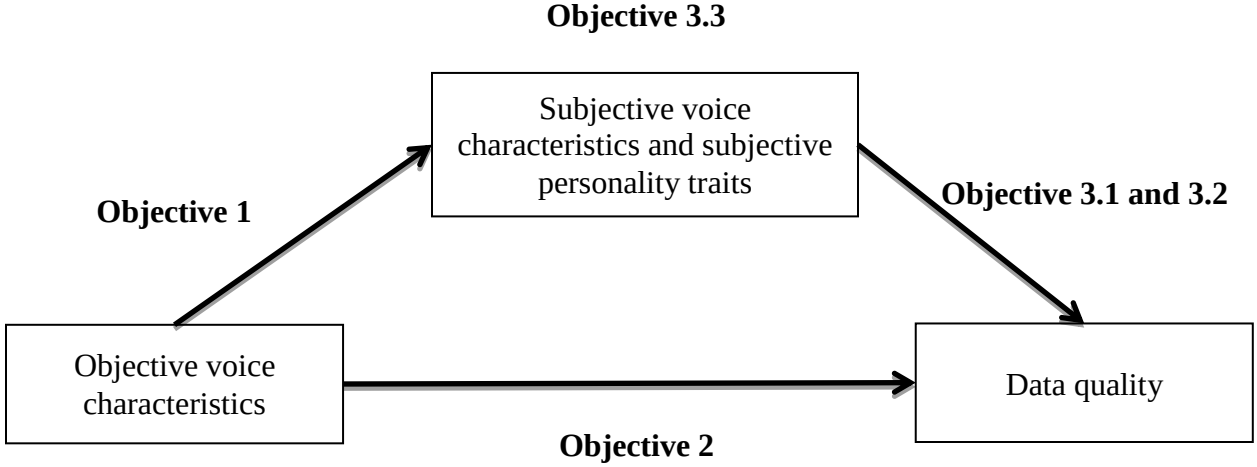


Figure 5.1 Three objectives of this dissertation

The *second* objective in this dissertation was to examine the associations between objective voice characteristics and data quality (Objective 2; Figure 5.1). Based on previous research from paralinguistic study and survey practice, I hypothesized that interviewer voices affect data quality. An interviewer's voice may affect respondents at

the perception stage, and affect respondents' ability to comprehend a question, retrieve relevant information, make judgments, and report answers, and thus affect data quality.

My *third* objective was to investigate the associations between subjective voice characteristics and data quality (Objective 3.1; Figure 5.1). In addition, I examine the associations between perceptions of an interviewer's personality traits and data quality (Objective 3.2; Figure 5.1), and whether subjective perceptions of an interviewer's personality traits mediate the relationships between objective voice characteristics and data quality (Objective 3.3; Figure 5.1). I created a scale measuring credibility from rated confidence, reliable, trustworthiness, and expertise (alpha = 0.91). As such, two interviewer personality traits examined in the third objective of this dissertation are **credibility and easiness to understand**. I hypothesized that subjective voice characteristics and perceptions of interviewers' personality traits are associated with data quality. In addition, I hypothesized that subjective perceptions of interviewers' personality traits mediate the relationship between objective voice characteristics and data quality.

In this chapter, I synthesize results from the three main objectives of this dissertation. Results from this dissertation provide suggestions for interviewer training and how to select interviewers based on voice characteristics to maximize data quality.

5.1 Summary of Findings and Implications

5.1.1 Pitch

I hypothesized that there would be a positive correlation between perceptions of pitch and measured pitch. I found listeners can perceive pitch as measured by a computer program, i.e. listeners perceived high (low) pitched voices as having a high (low) pitch.

In addition, listeners infer an interviewer's credibility from his/her voice pitch. I hypothesized that coders would perceive interviewers who read a question with higher pitched voices as less reliable, less trustworthy, less easy to understand, and more confident than those who read a question with lower pitched voices, and that pitch would be associated with expertise, but the direction could not be anticipated. I found that interviewers who read questions with higher pitched voices were rated as being more reliable, more confident, more trustworthy and having more expertise (i.e., more credible) than those who read questions with lower pitched voices, and easiness to understand was not related to an interviewer's pitch. Thus, associations between pitch and perceived reliability and perceived trustworthiness are in contrast with the hypothesized direction. This may be because low pitched voices are also associated with undesirable personality traits such as dishonesty and dominance while high pitched voices are related to positive personality traits such as friendliness. Listeners may consider these positive and negative traits when evaluating trustworthiness and reliability (Boehme 2014). As such, they rated higher pitched voices as more reliable and trustworthy.

As shown in Table 5.1, pitch and rated pitch affect different data quality indicators. The objective measure of pitch only affects one respondent behavior - giving a response that does not meet question's objective. Consistent with my hypothesis, interviewers who read a question with higher pitched voices had respondents who were more likely to give an answer that does not meet the question's objective, perhaps because they perceived the question as being more sensitive or complex.

For rated pitch, respondents are more likely to express uncertainty about neutral and complex questions and to answer socially undesirable questions to interviewers with

perceived higher pitched voices compared to interviewers with perceived lower pitched voices. I found that higher pitched voices are perceived as more credible than lower pitched voices. As trustworthiness is part of credibility, I infer that respondents trust credible interviewers; thus, they are more likely to express uncertainty about a question in order to get a clarification about a question. In addition, as expected, respondents provide responses to those interviewers with higher pitched voices, perhaps because they find them more attractive or credible (Ketrow 1990, Oksenberg, et al. 1986). Results in this study suggest that interviewers should read questions with higher pitched voices to be perceived more positively. However, effects of pitch on data quality indicators are inconclusive. Respondents are more likely to express uncertainty about neutral and complex questions and give a response that does not meet question objective to interviewer with higher pitched voice, but they are more likely to respond to them.

5.1.2 Intonation

Intonation is defined as a variation in the pitch of the interviewer's voice over the question reading. As expected, intonation and rated intonation are moderately positively correlated (r=0.49) - listeners perceived high (low) variation in pitch voices as having a varied (flat) intonation. In addition, listeners can perceive interviewer personality traits from interviewer voice. Overall, interviewers who read questions with moderate intonation (around 40 Hz for male and 60-80 Hz for female) were rated as being more reliable, more confident, more trustworthy and having more expertise (i.e. more credible) than those who read questions with an intonation that was either too high or too low. In addition, interviewer voices are perceived as easier to understand when interviewers read questions with moderate (objective) intonation. The effects of intonation on the rating of

Table 5.1 A summary of the effects of objective and subjective voice characteristics on data quality

	Pitch		Intonation		Speech rate		Disfluencies	
	Pitch	**Perceived Pitch**	**Intonation**	**Perceived Intonation**	**Speech rate**	**Perceived Speech rate**	**Disfluencies**	**Perceived Disfluencies**
Interrupt questions with answers					+ (neutral, complex Qs) - (socially undesirable Qs)	- (socially desirable Qs)		
Express uncertainty about a question		+ (neutral, complex Qs)	U-shaped relationship	-	- (socially undesirable, complex Qs)	- (complex Qs)		U-shaped association
Request clarification about a question			U-shaped relationship	- (socially desirable Qs, complex Qs)	- (socially undesirable, complex Qs) + (inexperienced interviewers)	- (socially desirable, complex Qs) + (neutral Qs)	+ (inexperienced interviewers)	
Give qualified answers					+ (neutral Qs) - (socially undesirable Qs)	+	- (female interviewers)	- (desirable Qs)
Give a response that does not meet question objective	+			-	- (socially undesirable Qs)	+(neutral Qs) - (socially desirable Qs)	- (male interviewers); + (female interviewers)	-

Note + indicates positive relationship; - indicates negative relationship; variables in parenthesis are interviewer and question characteristics for which the relationship occurs

Table 5.1 Continued.

	Pitch		Intonation		Speech rate		Disfluencies	
	Pitch	**Perceived Pitch**	**Intonation**	**Perceived Intonation**	**Speech rate**	**Perceived Speech rate**	**Disfluencies**	**Perceived Disfluencies**
Item nonresponse		- (socially undesirable Qs)	U-shaped relationship		-	- (inexperienced interviewers)		
Rounding					- (male interviewers); + (female interviewers)			
More is better					+ (male interviewers); - (female interviewers)	+		
Less is better				+	U-shaped relationship	- (inexperienced interviewer)	Inverse-U-shaped relationship	+

Note + indicates positive relationship; - indicates negative relationship; variables in parenthesis are interviewer and question characteristics for which the relationship occurs

easiness to understand vary across question types. To be perceived as easier to understand, interviewers should read complex question with higher intonation levels (around 80 Hz) than neutral questions (around 60 Hz).

Within a range of 20 to 80 Hz for the objective measure of intonation, accounting for 95% of voice files, I found a weak positive association between perception of interviewer personality traits and the objective measure of intonation. The positive association of intonation and perceived reliability and perceived trustworthiness is in contrast with my hypothesis. Similar to pitch, low intonation is associated with negative personality traits such as dishonesty while high intonation is associated with positive traits such as friendliness, which may be indicative of reliability and trustworthiness (Boehme 2014).

The effects of intonation on interviewer personality traits are consistent with gendered stereotypes of voice characteristics (Yuan, et al. 2006; Kent & Read 2002). Negative stereotypes can occur if males/females deviate from their expected voice pattern, for example, when male interviewers read questions with high intonation and female interviewers read questions with low intonation, thus leading to negative impressions of the interviewers (Benki, et al. 2011; Rubin 1992). In this study, I found that perceptions of interviewer personality traits for male interviewers were more positive at *lower* intonation levels. In contrast, perceptions of interviewer personality traits for female interviewer were more positive at *higher* intonation levels.

Table 5.1 shows a summary of the effects of objective and subjective measures of intonation on data quality. The objective measure of intonation has a U-shaped relationship with respondents expressing uncertainty and requesting clarification about a

question and item nonresponse. For rated intonation, interviewer voices perceived as having higher intonation obtain better data quality (fewer respondents expressing uncertainty, requesting clarification about a question, or giving responses that do not meet the question's objective, and higher report that respondents did not completely enjoy reading books, magazine, and newspapers).

For objective measures of intonation, interviewers obtain better data quality (lower item nonresponse rates and fewer respondents expressing uncertainty and requesting clarification) when they read a question with moderate intonation (around 40-80 Hz) compared to those asking a question with either low or high intonation. I found that interviewer voices with moderate intonation are perceived as easier to understand than those voices with high or low intonation, perhaps leading to lower item nonresponse rates, less uncertainty about questions, and fewer requests for clarification about a question. I found that these perceived interviewer personality traits mediated the effects of intonation on item nonresponse, but did not have sufficient evidence for personality traits mediating intonation for the other outcomes. Future research should examine this further.

As mentioned earlier, within an intonation ranging from 20 to 80 Hz (the range for 95% of the voice files), I found a weak positive association between objective intonation and the perception of interviewer credibility and easiness to understand. This may explain why interviewers obtain higher quality responses when their question reading is perceived to have more intonation. As expected, respondents perceive voices with higher intonation (more pitch variability) as easier to understand, leading to fewer expressions of uncertainty and requests for clarification about questions. In addition, as I

found, within an intonation ranging from 20 to 80 Hz, respondents perceived interviewer voices with higher intonation as more credible, leading to more responses that met the question's objective and responses that were less prone to socially desirable bias. However, unexpectedly, in this study, I did not find a direct effect of perceived credibility on these data quality outcomes.

Results in this study suggest that interviewers should read questions with moderate intonation (40-80 Hz) to be perceived more positively and to maximize data quality.

5.1.3 Speech rate

Consistent with the hypothesis, listeners perceived voices with a faster (slower) speech rate as speaking quickly (slowly). In addition, as expected, interviewers who read questions with a faster pace were rated as being more reliable, more confident, more trustworthy and having more expertise (i.e. more credible) than those who read questions with a slower pace. However, unexpectedly, interviewer voices were perceived as easier to understand when interviewers read questions with a speech rate between 3 and 4 wps. Effects of speech rate on rated easiness to understand differed across question types. To gain the highest rating of easiness to understand, interviewers should read neutral questions and socially desirable questions with a speech rate of 3 wps which is in the range of the typical conversational speech rate (Tauroza & Allizon 1990), but read socially undesirable questions and complex questions with a speech rate faster than 3 wps. It may be hard for listeners to hold information about complex questions in working memory when interviewers read questions slowly. In addition, it may be awkward to listen to someone asking sensitive questions with a slow pace. Thus, listeners may find it

more pleasant (or less awkward) to listen to complex questions and sensitive questions quickly, and thus perceiving a faster speech as easier to understand.

As shown in Table 5.1, the effects of objective and subjective measures of speech rate on data quality vary by question types.

For ***neutral questions,*** both measures of speech rate are *positively* associated with problematic respondent behaviors. Respondents are less likely to engage in problematic behaviors when interviewers read questions at a slower pace than faster speech rates. When interviewers ask questions more slowly, respondents infer that interviewers would like them to take time to think about their responses attentively, leading to better data quality. Results in this study imply that speech rate *does not* affect data quality through respondents' perception on interviewer personality traits in *neutral questions*.

For ***socially undesirable questions***, only the objective measure of speech rate is *negatively* associated with problematic respondent behaviors. Respondents are less likely to engage in problematic behaviors when interviewers read the questions faster than 2 wps. In addition, for both measures of speech rate, respondents provide higher rates of socially undesirable answers to interviewers who ask questions quickly (the effect only holds for male interviewer in objective measure of speech rate). In addition, as expected, I found that respondents perceive interviewers who read a question with a faster pace as more credible than interviewers who read a question with a slower pace. Moreover, consistent with my hypothesis, I found that more credible male interviewers received higher reports of having sex at least one time in the past seven days than those perceived as less credible. Thus, speech rate *does affect* respondents' perception of *interviewer credibility in socially desirable questions*, which in turn affects data quality.

For ***complex questions***, both measures of speech rate are *negatively* associated with respondents expressing uncertainty and requesting clarification about a question. In addition, objective speech rate is *negatively* associated with the probability of rounding in one complex question (Q19) for male interviewers. It may be hard for respondents to keep all of the information about complex questions in their working memory when interviewers read the questions at a speech rate of 2 wps, slower than the typical conversational speech rate (Tauroza & Allizon 1990). Thus, respondents are more likely to express uncertainty, request clarification, and to round their answers when interviewers read questions at slower pace. In addition, unexpectedly, I found that faster speech rates are perceived as easier to understand in complex question than slower speech rates, possibly leading to fewer expressions of uncertainty and fewer requests for clarification about a question. Consistent with the hypothesis, I also found that interviewers who read questions with faster objective speech rates obtain *higher* rates of respondents interrupting questions with answers than those who read question with slower speech rate. Results in this study imply that speech rate plays an important role on respondents' perception of *easiness to understand in complex questions*, which consequentially affects data quality.

Easiness to understand also plays an important role for the effect of speech rate on item nonresponse. As expected, I found interviewers who ask questions quickly are perceived as easier to understand, and those interviewer voices perceived as easier to understand have fewer item nonresponses. The effect only holds for inexperienced interviewers for rated speech rate.

For ***socially desirable questions***, the results are *mixed* (Table 5.1). For both measures of speech rate, respondents are more likely to provide better data quality when interviewers read questions more slowly (the effect only holds for inexperienced interviewers for rated speech rate). When interviewers ask questions with a slower objective speech rate, respondents may perceive that interviewers would like them to take time to think about their responses attentively, leading to better data quality.

In contrast, for rated speech rate, respondents are less likely to engage in problematic interview behaviors when interviewers are perceived as reading questions more quickly. Speech rate may affect the perception of interviewer credibility for socially desirable questions, leading to better data quality. For example, as I found in this study, in socially desirable questions, interviewer voices with faster speech rate are perceived as more credible, leading to fewer interruptions of questions with answers.

The effects of objective measures of speech rate on data quality are consistent with gendered stereotypes of voice characteristics (Yuan, et al. 2006; Kent & Read 2002). Negative stereotypes can occur when male interviewers read questions slowly and female interviewers read questions quickly, thus affect data quality (Benki, et al. 2011; Rubin 1992). That is, at a slow speech rate, respondents provided better data quality to female interviewers. In contrast, at a faster speech rate, respondents provided better data quality to male interviewers. Respondents are more likely to provide better data quality (higher rates of undesirable answers and fewer rounding) to male interviewers who ask questions quickly as measured objectively compared to those who ask questions slowly, but the opposite was found for female interviewers.

Results in this study suggest that interviewers should read questions with a speech rate faster than 2 wps to be perceived more positively. Speech rate affects perceptions of interviewer's credibility and easiness to understand, which in turn affects data quality. However, the relationship between perceptions of interviewer personality traits and data quality varies by question types and data quality indicators. That is, listeners perceive interviewer voices with faster speech rates as more credible, and this may affect reports and problematic respondent behaviors in socially desirable questions. In addition, listeners perceived interviewer voices with faster speech rate as easier to understand, and this may affect data quality in complex questions. Results for mediation effects of interviewer personality traits are reported below.

Results from this study also suggest that interviewers should read neutral questions with a slower pace, i.e. at a recommended speech rate of 2 wps, but read socially undesirable and complex questions more quickly to obtain better data quality.

5.1.4 Disfluencies

As expected, listeners perceived interviewers with high (low) number of fillers as having high (low) disfluencies. In addition, interviewers who read questions with fewer disfluencies were rated as being more reliable, more confident, more trustworthy and having more expertise than those who read questions with more disfluencies. In addition, interviewer voices are more likely to be perceived as easier to understand when interviewers read questions with fewer disfluencies. The associations between objective voice characteristics and perceived interviewers' personality traits are consistent with the hypotheses.

As shown in Table 5.1, the effects of subjective and objective measures of disfluencies on data quality are mixed and vary by interviewer's demographic characteristics.

Consistent with the hypothesis from survey practice that disfluencies have a "disfluency advantage" that allows respondents to have more time to think about their answers, thus increasing data quality, for both measures of disfluencies, I found respondents provide better data quality answers (fewer qualified answers, fewer responses that do not meet the question's objective, lower rates of endorsement for socially desirable questions) when interviewers are perceived to read questions with higher disfluency rates, especially for socially desirable questions. However, the association between objective disfluencies and respondents providing responses that do not meet the question's objective vary by interviewer sex. The result is consistent with gendered speech patterns in which females use fewer fillers relative to males (Bortfeld et al. 2001). Negative stereotypes can occur when male interviewers read questions with fewer disfluencies and female interviewers read questions with more disfluencies, thus affecting data quality (Benki, et al. 2011; Rubin 1992). I found higher rates of responses that do not meet the question's objective when male interviewers use fewer disfluencies and female interviewers use more disfluencies.

In addition, for rated disfluencies, fewer respondents express uncertainty about a question when the interviewer reads questions with moderate perceived disfluencies rather than having too many or too few disfluencies. This is consistent with previous research that found interviewers who speak with neither robotic speech nor are highly disfluent have the highest participation rates (Conrad et al. 2013). Results from this

dissertation suggest that interviewers should read questions with moderate disfluencies to be perceived more positively and to maximize data quality.

As previous research has found that subjective ratings of characteristics of interviewers' voices (e.g. intonation and fluency) are better predictors of unit nonresponse than the objective measurement of the same voice characteristics (Van der Vaart, et al. 2005), I expect that the effect of subjective interviewer voice characteristics on data quality will be stronger than the effect of objective interviewer voice characteristics. From the results discussed above, this hypothesis is inconclusive because subjective and objective voice characteristics tend to affect different data quality indicators. However, the effects from both objective and subjective voice characteristics are in a consistent direction. Both measures of voice characteristics affect data quality; however, the effects vary by question type and interviewer demographic characteristics.

5.1.5 Potential mediation effects of interviewer personality traits

As I found, interviewers who read questions with higher pitched voices, moderate intonation, faster speech rates, and fewer fillers are perceived as being more reliable, more confident, more trustworthy, and having more expertise (i.e. more credible) than those who read questions with lower pitched voice, either too high or too low intonation, slower speech rates, and higher fillers (objective 1; Figure 5.2). In addition, I found that pitch, intonation, speech rate, and disfluencies affect data quality (objective 2; Figure 5.2). Because I also found associations between credibility and data quality (objective 3; Figure 5.2), I hypothesized that credibility may mediate the effects of interviewer voice characteristics on data quality.

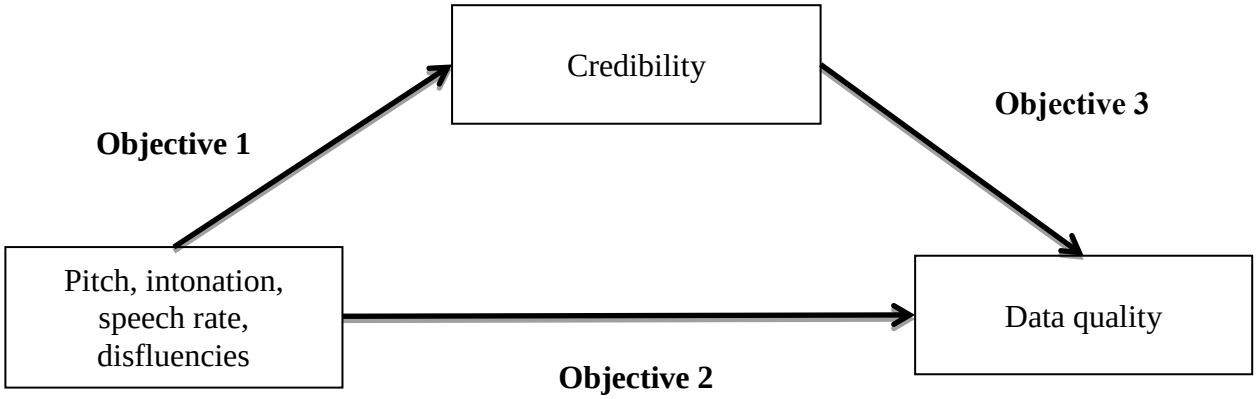


Figure 5.2 Possible mediation effect of credibility on the effect of objective voice characteristics on data quality

This study also found that interviewers who read questions with moderate intonation, a faster speech rate, and fewer fillers are perceived as easier to understand than those who read questions with either lower or higher intonation, a slower speech rate, and more fillers. There is no association between pitch and perception of easiness to understand. In addition, I found that objective measures of intonation, speech rate, and disfluencies affect data quality (Objective 2; Figure 5.3). Because I also found the associations between easiness to understand and data quality (objective 3; Figure 5.3), I hypothesized that there may be a mediation effect of easiness to understand on the effects of interviewer voice characteristics on data quality.

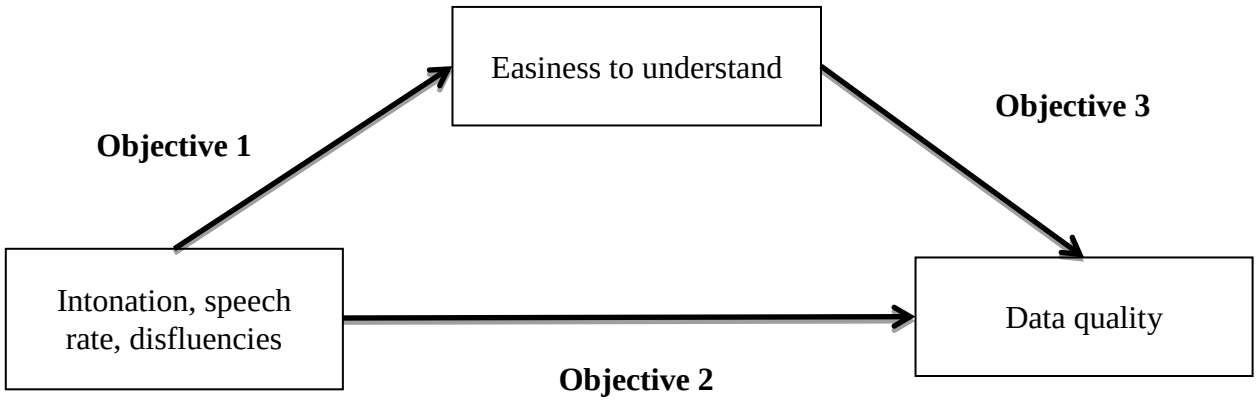


Figure 5.3 Possible mediation effect of easiness to understand on the effect of objective voice characteristics on data quality

Mediation effect of interviewer personality traits on the effects of objective voice characteristics on data quality

By following Baron & Kenny's (1986) procedures for examining whether interviewer personality traits (credibility and easiness to understand) mediate the effects of objective interviewer characteristics on data quality, I found that interviewer credibility mediates the effects of objective measure of speech rate on respondents interrupting questions with answers and interviewers whose voices are rated as easy to understand mediates the effect of objective measures of intonation and speech rate on item nonresponse rates. This implies that, for other data quality indicators, interviewer voice characteristics independently affect data quality. Interviewer personality traits do not mediate the effect of pitch, intonation, speech rate, and disfluencies on data quality indicators, except for respondents interrupting questions with answers and item nonresponse.

Unfortunately, the moderated mediation multi-level models for examining the direct effect of interviewer voice on data quality and the indirect effect through interviewer personality traits failed to converge. As such, direct and indirect effects cannot be estimated in this study.

Results from this study partially explain contrast results that Barath & Cannell (1976) and Blair (1977) found. Barath and Cannell (1976) found that interviewers whose voices were rated as rising at the end of questions received higher rates of acquiescent reports of sensitive health conditions than those whose voices were rated as falling. Blair (1977), looking at nonsensitive questions, found the opposite. Because I defined intonation as variation of pitch, not rising or falling at the end of the sentence as Barath &

Cannell (1976) defined, results from this study cannot exactly explain whether perception of interviewers' personality traits mediated the relationship between rising/falling intonation on data quality. However, results from this study provide evidence that intonation can affect listener's perception of interviewer personality traits, and these perceptions can possibly affect data quality. The mediation effect of interviewer personality traits may vary by question type, leading to the contrasting results that Barath & Cannell (1976) and Blair (1977) found.

5.2 Limitation and future research

This study has many limitations. Only a sample of questions is analyzed in this study. Including more questions will increase statistical power in this study. Moreover, respondents in this study are more likely to be female, white, and older than the general population. Models in this study only controlled for respondent age and education. As such, respondent gender and race may affect the results in this study. Future research should analyze more questions and control for respondent gender and race.

In addition, data in this study is only from landline telephone surveys, limiting the ability to generalize findings to mobile telephone surveys. Currently, 43.1% of US households live in cell-only households (Blumberg & Luke, 2014). Findings in this study may not be the same if cell phone respondents are included in this study. When answering questions, respondents in cell phone surveys are more likely to be multitasking and concentrate less on questions compared to respondents in landline telephone surveys (Lavrakas et al., 2010). As such, better data quality resulting from interviewers reading neutral questions at slower speech rates may be lower when respondents in cell phone surveys are included in the survey sample. Future research should analyze respondents

from both landline and cell phone surveys to generalize findings of the effects of interviewer voice characteristics on data quality to telephone surveys.

I do not know whether fewer rounded answers, higher rates of undesirable responses, and lower rates of desirable responses in fact indicate "better" data quality. The responses could be either true values or measurement error. In addition, respondent behaviors examined in this study may indicate better data quality rather than worse data quality. For example, Belli et al. (1999) found a positive association between expressing uncertainty and data accuracy for younger respondents. As such, future research should compare gold standard data with survey estimates to measure the data quality (Olson 2006; Biemer & Lyberg 2003; Groves 1989), instead of using data quality indicators.

If gold standard data is still not available, future research should examine other data quality indicators such as interviewer behaviors. For example, interviewers may be anxious when reading a question, leading to an increase in their pitch, but also may change behaviors during the interview. Interviewers may reflect their anxiety through less probing which may decrease data quality (Fowler & Mangione 1990).

The age of the coders may have affected results in this study because listeners at different age cohorts judge voices based on different criteria (Ketrow 1990). Results in this study indicate that interviewers only account for 35% of variation in the perception of interviewer personality traits. Each listener may perceive interviewer personality traits differently. Unfortunately, because the number of coders (6 coders) in this study is quite small, variability of coders is not large enough to be included as a separate level in the multilevel analysis. In addition, to examine the associations between subjective voice characteristics and data quality, I assumed that the undergraduate raters' subjective

perceptions are a good proxy for respondents' subjective perceptions. For future research, researchers should collect subjective ratings of interviewer voice and personality traits from respondents themselves.

Another analytic limitation is that the moderated mediation multi-level model failed to converge. As such, future studies should examine this further.

One application of this research could be recording interviewer voices in audio computer-assisted self-interviewing (ACASI) to maximize data quality. There has been an increase using text-to-speech (TTS) voice, a computer-generated voice, in ACASI surveys instead of a human-recorded voice (Geisen, et al. 2015). Geisen et al (2015) found no significant differences between TTS voice and human recorded voice on participants' comprehension of questions. With respect to data quality, DiLoreto et al. (2015) found higher reports of sensitive behaviors in an empathetic human voice than in a TTS voice. However, they only examined female voices. Future research should account for pitch, intonation, and rate of speech differences when examining whether TTS voice affects data quality in socially (un)desirable questions and complex questions. In addition, whether TTS voices affect listeners' perceptions of interviewer personality traits should be examined because it may mediate the effect of objective measure of TTS voices on data quality. Also, perception of interviewer personality traits from TTS voices and human voice may differ.

This study focused on individual questions, not how voice characteristics change over an entire questionnaire. Future research should examine whether interviewer voice characteristics change during a survey interview, and whether these changes affect data quality. For example, an interviewer may start reading questions at a slower pace at the

beginning of a survey and then increase his/her speech rate at the end of the interview, and this may affect data quality. As found in this study, interviewers who read socially undesirable questions with faster speech rates received better data quality compared to those who read the questions more slowly. As such, data quality may be improved when socially undesirable questions are put at the end of questionnaire (as is normally done in surveys) (Dillman, et al. 2014).

5.3 Conclusion

Understanding the factors that affect data quality in telephone surveys is important because the telephone mode has been widely used in many national surveys. In this dissertation, I examined how interviewer voice characteristics affect data quality in socially desirable, undesirable, and complex questions. Findings in this dissertation indicate that interviewer voice characteristics affect data quality; however, the effects are inconsistent across data quality indicators.

Pitch. The effect of pitch on data quality is inconclusive. I found listeners perceive high pitched voices as more credible than low pitched voice. In addition, respondents are more likely to respond to interviewer whose voice was perceived as having higher pitched voice than those whose voice was perceived as having lower pitched voice. However, pitch is positively associated with respondents giving answers that do not meet questions' objective and respondents expressing uncertainty about a question.

Intonation. Interviewers who ask questions with moderate intonation are perceived as being more credible and having voice that are easier to understand and obtain better data quality than those who ask questions with low or high intonation.

Speech rate. Speech rate has the largest effect on perception on interviewer personality traits and data quality. Reading questions at a recommended speech rate of 2 wps leads to negative perception of interviewer personality traits. The effects of speech rate on data quality vary by question types. For neutral questions, interviewers obtained better data quality when they read at a speech rate of 2 wps compared to a faster pace. However, for socially undesirable questions, better data quality was found among interviewers who read socially undesirable questions faster than 2 wps.

Disfluencies. Interviewers who ask questions with fewer disfluencies are perceived as being more credible and having voices that are easier to understand than those who ask questions with more disfluencies. In addition, those who ask questions with moderate disfluencies obtain better data quality than those who ask questions with low or high disfluencies.

In line with previous paralinguistic studies, gendered stereotypes of voices are found for speech rate, intonation, and disfluencies. Male and female interviewers whose voices deviate from their expected voice pattern received negative impressions and lower data quality.

Results in this dissertation provide suggestions for interviewer voice training to maximize data quality. Interviewers should be trained to read questions with moderate intonation and moderate disfluencies. In addition, to maximize data quality, interviewers should read neutral questions with the recommended speech rate of 2 wps, but interviewers should read socially undesirable questions more quickly.

REFERENCES

Chapter 1

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*(5), 715-727.

Axinn, W. G. (1991). The influence of interviewer sex on responses to sensitive questions in Nepal. *Social Science Research, 20*(3), 303-318.

Bachorowski, J. A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science, 8*(2), 53-57.

Barath, A., & Cannell, C. F. (1976). Effect of interviewer's voice intonation. *Public Opinion Quarterly, 40*, 370–373.

Belli, R. F., Weiss, P. S. & Lepkowski J. M. (1999). Dynamics of survey interviewing and the quality of report: Age comparisons. In N. Schwarz, D. Park, B. Knäuper, & S. Sudman (Eds.), *Cognition, aging, and self-reports* (pp. 303-326). Philadelphia: Psychology Press.

Benki, J., Broome, J., Conrad, F., Groves, R., & Kreuter, F. (2011, May). *Effects of speech rate, pitch, and pausing on survey participation decisions.* Paper presented at the 66th annual conference of the American Association for Public Opinion Research, Phoenix, AZ.

Biemer & Lyberg (2003). *Introduction to survey quality.* Hoboken, NJ: Wiley.

Bilgen, I. & Belli, R. F. (2010). Comparison of verbal behaviors between calendar and standardized conventional questionnaires. *Journal of Official Statistics, 26*(3), 481-505.

Blair, E. (1977). More on the effects of interviewer's voice intonation. *Public Opinion Quarterly, 41*(4), 544-548.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(2), 123-147.

Bradburn, N.M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236*, 157-161.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*(2), 274-296.

Brooke, M. E. & Ng, S. H. (1986). Language and social influence in small conversational groups. *Journal of Language and Social Psychology, 5*(3), 201-210.

Broome, J. S. (2012). *Vocal characteristics, speech, and behavior of telephone interviewers* (Doctoral dissertation). Retrieved from http://deepblue.lib.umich.edu

Castro, L. & de Moraes, J. A. (2008). The temporal structure of professional speaking styles in Brazilian Portugueses. *Proceedings of Experimental Linguistics, ExLing 2008,* 57-60.

Clark, H. H. & Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73-111.

Clark, H. H. (2002). Speaking in time. *Speech Communication, 36*, 5-13.

Conrad, F. G., Schober, M., & Dijkstra, W. (2008). Cues of communication difficulty in telephone interviews. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 212-230). New York: John Wiley & Sons.

Dey, A., Feinberg, D., & Kim J. (2006). Effect of voice pitch on content comprehension. Retreived 26 March 2015 from http://cll.mcmaster.ca/sotl/pdf/3024_Group1/3024_2Dey.pdf

De la Puente, M., & McKay, R. (1995). Developing and testing race and ethnic origin questions for the current population survey supplement on race and ethnic origin. In *Proceedings of the 1994 Annual Meeting of the American Statistical Association: Section on Survey Research Methods* (Vol. 1).

Dillman, D. A., Smyth, J. D., Christian, L. M. (2014). *Internet, mail and mixed-mode surveys: The tailored design method* (4th ed.). New York: Wiley.

Dillman, D. A., Sanster, R. L., Tarnai J., Rockwood, T. H. (1996). Understanding differences in people's answers to telephone and mail surveys. *New Directions for Evaluation, 70*, 45-61.

Dijkstra, W. (1987). Interviewing style and respondent behavior: An experimental study of the survey-interview, *Sociological Methods & Research, 16*(2), 309-334.

Dykema, J., Lepkowski, J. & Blixt, S. (1997). The effect of interviewer and respondent behavior on data quality: Analysis of interaction coding in a validation study. In L. E. Lyberg, P. Biemer, M. Collins, E. D. Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 287-310). New York: Wiley.

Ehlen, P., Schober, M. F., & Conrad, F. G. (2007). Modeling speech disfluency to predict conceptual misalignment in speech survey interfaces. *Discourse Processes, 44*(3), 245-265.

Fowler, F. J. (1966). *Education, interaction, and interview performance.* (Doctoral dissertation). University of Michigan.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing; minimizing interviewer- related error.* Newbury Park, Calif.: Sage Publications.

Fowler, F. J. (1992). How unclear terms affect survey data. *Public Opinion Quarterly, 56*(2), 218-231.

Fowler, F. J., & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N. Schwarz & S. Sudman (Eds.), *Answering questions. Methodology for determining cognitive and communicative processes in survey research* (pp. 15-36). San Francisco: Jossey-Bass.

Fowler, F. J. (2011). Coding the behavior of interviewers and respondents to evaluate survey questions. In J. Madans, K. Millers, A. Maitland, & G. Willis (Eds.), *Question evaluation methods: Contributing to the science of data quality* (pp.7-21). Hoboken, NJ: John Wiley & Sons.

Groves, R. M., & Fultz, N. H. (1985). Gender effects among telephone interviewers in a survey of economic attitudes. *Sociological Methods and Research, 14*(1), 31-52.

Groves, R. M., & Magilavy, L. J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly, 50*(2), 251-266.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Groves R. M. (1990). Theories and methods of telephone surveys. *Annual Review of Sociology, 16*, 221-240.

Groves, R. M., O'Hare, B. C., Gould-Smith, D., Benkí, J., & Maher, P. (2008). Telephone interviewer voice characteristics and the survey participation decision. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology*, (pp. 385–400). New York: Wiley.

Groves, R. M., Fowler, F. J, Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology.* Hoboken, NJ: Wiley.

Guenzel, P. J., Berckmans, T. R., & Cannell, C. F. (1983). *General interviewing techniques: a self-instructional workbook for telephone and personal interviewer training.* Ann Arbor, Michigan: Institute for Social Research, Survey Research Center.

Harms (1961). Listener judgments of status cues in speech. *Quarterly Jornal of Speech, 47*(2), 164-168.

Holbrook, A., Cho, Y. I., & Johnson, T. (2006). The impact of question and respondent characteristics on comprehension and mapping difficulties. *Public Opinion Quarterly, 70*(4), 565-595.

Hovland, C. I., Irving, K. J., & Harold H. K. (1953). *Communication and persuasion*, New Haven, CT: Yale University Press.

Japec, L. (2008). Interviewer error and interviewer burden. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in Telephone Survey Methodology* (pp. 187-221). New York: Wiley.

Jans, M. E. (2010). *Verbal paradata and survey error: Respondent speech voice, and question-answering behavior can predict income item nonresponse* (Doctoral dissertation). Retrieved from http://deepblue.lib.umich.edu

Kane, E. W., & Macaulay, L. J. (1993). Interviewer gender and gender attitudes. *Public Opinion Quarterly, 57,* 1–28.

Kent, R. D. & Read, C. (2002). *Acoustic analysis of speech.* (2nd ed). Canada: Thomson Learning.

Ketrow, S. M. (1990). Attributes of a telemarketer's voice and persuasiveness: A review and synthesis of the literature. *Journal of Direct Marketing, 4,* 7–21.

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science,* 1-10.

Kirchner, A. & Olson, K. (2014). *Commitment, concealment, and confusion? Exploring interviewer and respondent behaviors in survey interviews across different types of questions.* Paper presented at the 2014 American Association for Public Opinion Research (AAPOR) conference. Anaheim, CA.

Knauper, B., Belli, R. F., Hill, D. H., & Herzog, A. R. (1997). Question difficulty and respondents' cognitive ability: the effect on data quality. *Journal of Official Statistics, 13*(2), 181-199.

Knauper, B. (1999). The impact of age and education on response order effects in attitude measurement. *The Public Opinion Quarterly*, *63*(3), 347-370.

Kreuter, F., Presser, S. & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and Web surveys. *Public Opinion Quarterly*, 72(5), pp. 847–865.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213-236.

Krysan, M., & Couper, M. P. (2003). Race in the live and the virtual interview: Racial deference, social desirability, and activation effects in attitude surveys. *Social Psychology Quarterly*, *66*(4), 364-383.

Lai, C. (2010). What do you mean, you're uncertain?: The interpretation of cue words and rising intonation in dialogue. *Proceedings of the Annual Conference of*

International Speech Communication Association, Makuhari, Chiba, Japan.

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology, 34*(4), 615-624.

Mille, G. R. & Hewgill, M. A. (1964). The effect of variations in nonfluency on audience ratings of source credibility. *The Quarterly Journal of Speech, 50*(1), 36-44.

Narayan, S &. Krosnick, J. A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly, 60,* 58-89.

Natsumi, O. (2013). *How is comprehension affected by intonation?* (Graduation Thesis). Notre Dame Seishin University

Ohanian, R. (1990). Construction and Validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attraction, *Journal of Advertising, 19*(3), 39-52.

Oksenberg, L., Coleman, L., & Cannell, C. (1986). Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly, 50*(1), 97–111.

Oksenberg, L., & Cannell, C. (1988). Effects of interviewer vocal characteristics on nonresponse. In R. M. Groves, P. B. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, and J. Waksberg (Eds.), *Telephone Survey Methodology* (pp. 252-272). New York: John Wiley and Sons.

Olson, K. & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly, 75*, 99-114.

Olson, K. & Peytchev, A. (2007). Effect of interviewer experience on interview pace and interviewer attitudes. *Public Opinion Quarterly, 71*, 273-286.

O'Muircheartaigh, C., & Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society Series A, 161*, 63-77.

O'Muircheartaigh, C., & Campanelli, P. (1999). A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society Series A, 162*(Part 3), 437–446.

Ongena, Y. P. & Dijkstra, W. (2007). A model of cognitive process and conversational principles in survey interview interaction, Applied Cognitive Psychology, *21*, 145-163.

Pearce, W. B. & Conking, F. (1971). Nonverbal vocalic communication and perceptions of a speaker. *Speech Monographs, 38*, 235-241.

Pytko, J. L. & Reese, L. O. (2013). The effect of using "Um" and "Uh" on the perceived intelligence of a speaker. *Journal of the Behavioral Sciences, Spring,*1-21.

Sah, S, Moore, D. A., & MacCoun, R. J. (2013). Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiceness. *Organizational Behavior and Human Decision Processes, 121,* 246-255.

Sakshaug, J., W., Yan, T., & Tourangeau, R. (2010). Nonresponse error, measurement, error, and mode of data collection. Tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly, 74*, 907-933.

Schaeffer, N. C., & Dykema, J. (2011). Response 1 to Fowler's chapter: Coding the behavior of interviewers and respondents to evaluate survey questions. In J. Madans, K. Miller, A. Maitland & G. Willis (Eds.), *Question evaluation methods: Contributing to the science of data quality* (pp. 23-39). Hoboken, NJ: John Wiley & Sons.

Schnell, R. & Kreuter, F. (2005). Separating interviewer and sampling-point effects. *Journal of Official Statistics, 21,* 389-410.

Schuman, H. & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Sage.

Scherer, K. R., Londom, H., & Wolf J. J. (1973). The voice of confidence: paralinguistic cues and audience evaluation. *Journal of Research in Personality, 7,* 31-44.

Smith, P. M. (1979). Sex markers in speech. *Social markers in speech, 109,* 146.

Smith, S. M., & Shaffer, D. R. (1995). Speed of speech and persuasion: Evidence for multiple effects. *Personality and Social Psychology Bulletin, 21,* 1051–1060.

Steeh, C. (2008). Telephone surveys. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 221-238). New York: Psychology Press, Taylor & Francis.

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11*(1), 90-105.

Tigue, C. C., Borak, D. J., O'Connor, J. J., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior, 33,* 210-216.

Traunmuller, H., & Eriksson, A. (1993). The frequency range of the voice fundamental in the speech of male and female adults. Stockholm: Institutionen för Lingvistik, Stockholms University Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.64.5133&rep=rep1&type=pdf

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859-883.

Tourangeau, R., Groves, R. M., and Redline, C. D. (2010). Sensitive topics and reluctant respondents demonstrating a link between nonresponse bias and measurement error. *Public Opinion Quarterly, 74*(3): 413-432.

Van der Vaart, W., Ongena, Y., Hoogendoorn, A., & Dijkstra, W. (2005). Do interviewers' voice characteristics influence cooperation rates in telephone surveys? *International Journal of Opinion Research*, *18*(4), 488–499.

Webb, J. T. (1969). Subject speech rates as a function of interviewer behavior. *Language & Speech*, *12*(1), 54-67.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. *Proceedings of the Ninth International Conference on Spoken Language Processing.*

Chapter 2

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*(5), 715-727.

Bachorowski, J. A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *The Journal of the Acoustical Society of America, 106*(2), 1054-1063.

Barath, A., & Cannell, C. F. (1976). Effect of interviewer's voice intonation. *Public Opinion Quarterly, 40,* 370–373.

Benki, J., Broome, J., Conrad, F., Groves, R., & Kreuter, F. (2011, May). *Effects of speech rate, pitch, and pausing on survey participation decisions.* Paper presented at the 66th annual conference of the American Association for Public Opinion Research, Phoenix, AZ.

Blair, E. (1977). More on the effects of interviewer's voice intonation. *Public Opinion Quarterly, 41*(4), 544-548.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(2), 123-147.

Boehme, B. (2014). How trustworthy is your voice? The effects of voice manipulation on the perceived trustworthiness of novel speakers. Retreived 13 June 2015 from http://www.did.stu.mmu.ac.uk/MMU_Psychology_Dissertations_UK/2014dissertations/GtoK/Glasgow/Boehme%20%28Bibi%29%202014%20%28Glasgow%29.pdf

Brennan, S. E. & Williams, M. (1995). The felling of another's knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34,* 383-398.

Broome, J. S. (2012). *Vocal characteristics, speech, and behavior of telephone interviewers* (Doctoral dissertation). Retrieved from http://deepblue.lib.umich.edu

Brooke, M. E. & Ng, S. H. (1986). Language and social influence in small conversational groups. *Journal of Language and Social Psychology, 5*(3), 201-210.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4, retrieved 4 October 2014 from http://www.praat.org

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques, In S. Leinhardt (Ed.), *Sociological methodology* (pp. 289-437). SanFrancisco: Jossey-Bass.

Castro, L. & de Moraes, J. A. (2008). The temporal structure of professional speaking styles in Brazilian Portugueses. *Proceedings of Experimental Linguistics, ExLing 2008,* 57-60.

Conrad, F. G., Schober, M., & Dijkstra, W. (2008). Cues of communication difficulty in telephone interviews. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 212-230). New York: John Wiley & Sons.

Dey, A., Feinberg, D., & Kim J. (2006). Effect of voice pitch on content comprehension. Retreived 26 March 2015 from http://cll.mcmaster.ca/sotl/pdf/3024_Group1/3024_2Dey.pdf

Ehlen, P., Schober, M. F., & Conrad, F. G. (2007). Modeling speech disfluency to predict conceptual misalignment in speech survey interfaces. *Discourse Processes, 44*(3), 245-265.

Groves, R. M., O'Hare, B. C., Gould-Smith, D., Benkí, J., & Maher, P. (2008). Telephone interviewer voice characteristics and the survey participation decision. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology,* (pp. 385–400). New York: Wiley.

Guenzel, P. J., Berckmans, T. R., & Cannell, C. F. (1983). *General interviewing techniques: a self-instructional workbook for telephone and personal interviewer training.* Ann Arbor, Michigan: Institute for Social Research, Survey Research Center.

Hayati, A. (2010). The effect of speech rate on listening comprehension of EFL learners. *Creative Education, 2,* 107-114.

Johnson (2003). *Acoustic and auditory phonetics.* MA: Blackwell Publishing.

Kent, R. D. & Read, C. (2002). *Acoustic analysis of speech.* (2nd ed). Canada: Thomson Learning.

Ketrow, S. M. (1990). Attributes of a telemarketer's voice and persuasiveness: A review and synthesis of the literature. *Journal of Direct Marketing, 4,* 7–21.

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science,* 1-10.

Mackay, I. R. (1987). *Phonetics: the science of speech production.* (2nd Ed). A College-Hill Publication, New York.

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology, 34*(4), 615-624.

Mille, G. R. & Hewgill, M. A. (1964). The effect of variations in nonfluency on audience ratings of source credibility. *The Quarterly Journal of Speech, 50*(1), 36-44.

Munro, B. H. (2005). *Statistical Methods for Health Care Research.* Lippincott Williams & Wilkins, London, UK.

Natsumi, O. (2013). *How is comprehension affected by intonation?* (Graduation Thesis). Notre Dame Seishin University

Oksenberg, L., Coleman, L., & Cannell, C. (1986). Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly, 50*(1), 97–111.

Oksenberg, L., & Cannell, C. (1988). Effects of interviewer vocal characteristics on nonresponse. In R. M. Groves, P. B. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, and J. Waksberg (Eds.), *Telephone Survey Methodology* (pp. 252-272). New York: John Wiley and Sons.

Olson, K. & Peytchev, A. (2007). Effect of interviewer experience on interview pace and interviewer attitudes. *Public Opinion Quarterly, 71*, 273-286.

O'Muircheartaigh, C., & Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society Series A, 161*, 63-77.

O'Muircheartaigh, C., & Campanelli, P. (1999). A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society Series A, 162*(Part 3), 437–446.

Pearce, W. B. & Conking, F. (1971). Nonverbal vocalic communication and perceptions of a speaker. *Speech Monographs, 38,* 235-241.

Pytko, J. L. & Reese, L. O. (2013). The effect of using "Um" and "Uh" on the perceived intelligence of a speaker. *Journal of the Behavioral Sciences, Spring,*1-21.

Rabe-Hesketh, S. and Skrondal, A. (2012). *Multilevel and longitudinal modeling using stata* (3rd ed.). College Station, TX: Stata Press.

Scherer, K. R., Londom, H., & Wolf J. J. (1973). The voice of confidence: paralinguistic cues and audience evaluation. *Journal of Research in Personality, 7,* 31-44.

Schober, M. F, & Bloom, J. E. (2004). Discourse cues that respondents have misunderstood survey questions. *Discourse Process, 38*(3), 287-308.

Smith, S. M., & Shaffer, D. R. (1995). Speed of speech and persuasion: Evidence for multiple effects. *Personality and Social Psychology Bulletin, 21,* 1051–1060.

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11*(1), 90-105.

Tigue, C. C., Borak, D. J., O'Connor, J. J., Schandl, C., & Feinberg, D. R. (2012). Voice

pitch influences voting behavior. *Evolution and Human Behavior, 33,* 210-216.

Traunmuller, H., & Eriksson, A. (1993). The frequency range of the voice fundamental in the speech of male and female adults. Stockholm: Institutionen för Lingvistik, Stockholms University Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.64.5133&rep=rep1&type=pdf

Van der Vaart, W., Ongena, Y., Hoogendoorn, A., & Dijkstra, W. (2005). Do interviewers' voice characteristics influence cooperation rates in telephone surveys? *International Journal of Opinion Research*, *18*(4), 488–499.

Webb, J. T. (1969). Subject speech rates as a function of interviewer behavior. *Language & Speech, 12*(1), 54-67.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. *Proceedings of the Ninth International Conference on Spoken Language Processing.*

Chapter 3

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*(5), 715-727.

Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs, 35*, 492-503.

Axinn, W. G. (1991). The influence of interviewer sex on responses to sensitive questions in Nepal. *Social Science Research, 20*(3), 303-318.

Bachorowski, J. A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science, 8*(2), 53-57.

Belli, R. F., Weiss, P. S. & Lepkowski J. M. (1999). Dynamics of survey interviewing and the quality of report: Age comparisons. In N. Schwarz, D. Park, B. Knäuper, & S. Sudman (Eds.), *Cognition, aging, and self-reports* (pp. 303-326). Philadelphia: Psychology Press.

Benki, J., Broome, J., Conrad, F., Groves, R., & Kreuter, F. (2011, May). *Effects of speech rate, pitch, and pausing on survey participation decisions.* Paper presented at the 66th annual conference of the American Association for Public Opinion Research, Phoenix, AZ.

Biemer & Lyberg (2003). *Introduction to survey quality.* Hoboken, NJ: Wiley.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4, retrieved 4 October 2014 from http://www.praat.org

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(2), 123-147.

Bradburn, N.M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236,* 157-161.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*(2), 274-296.

Broome, J. S. (2012). *Vocal characteristics, speech, and behavior of telephone interviewers* (Doctoral dissertation). Retrieved from http://deepblue.lib.umich.edu

Campanelli, P. C., Martin, E. A., & Rothgeb, J. (1991). The use of respondent and interviewer debriefing studies as a way to study response error in survey data. *The Statistician, 40,* 253-264.

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques, In S. Leinhardt (Ed.), *Sociological methodology* (pp. 289-437). SanFrancisco: Jossey-Bass.

Clark, H. H. & Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73-111.

Clark, H. H. (2002). Speaking in time. *Speech Communication, 36*, 5-13.

Conrad, F. G., Broome, J. S., Benki, J. R., Kreuter, F., Groves, R. M., Vannette, D., & McClain C. (2013). Interviewer speech and the success of survey invitations. *Journal of the Royal Statistical Society, 176*, 191-210.

Conrad, F. G., Brown, N. R., & Cashman, E. R. (1998). Strategies for estimating behavioural frequency in survey interviews. *Memory, 6*(4), 339-366.

Conrad, F. G., Schober, M., & Dijkstra, W. (2008). Cues of communication difficulty in telephone interviews. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 212-230). New York: John Wiley & Sons.

Dillman, D. A., Smyth, J. D., Christian, L. M. (2014). *Internet, mail and mixed-mode surveys: The tailored design method* (4th ed.). New York: Wiley.

De la Puente, M., & McKay, R. (1995). Developing and testing race and ethnic origin questions for the current population survey supplement on race and ethnic origin. In *Proceedings of the 1994 Annual Meeting of the American Statistical Association: Section on Survey Research Methods* (Vol. 1).

Dijkstra, V. (2009). Sequence Viewer version 5. Retrieved from http://sequenceviewer.nl/downloads/Reference%20manual%20version%205.0.pdf

Dyck & Smither (1995). Age differences in computer anxiety: The role of computer experience, gender and education. *Journal of Educational Computing Research, 10*(3), 239-248.

Dykema, J., Diloreto, K., Price, J. L., White, E., & Schaeffer, N. C. (2012). ACASI gender-of-interviewer voice effects on reports to questions about sensitive behaviors among young adults. *Public Opinion Quarterly, 76*(2), 311–325.

Dykema, J., Lepkowski, J. & Blixt, S. (1997). The effect of interviewer and respondent behavior on data quality: Analysis of interaction coding in a validation study. In L. E. Lyberg, P. Biemer, M. Collins, E. D. Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 287-310). New York: Wiley.

Fowler, F. J. (1966). *Education, interaction, and interview performance.* (Doctoral dissertation). University of Michigan.

Fowler, F. J. (2011). Coding the behavior of interviewers and respondents to evaluate survey questions. In J. Madans, K. Millers, A. Maitland, & G. Willis (Eds.), *Question evaluation methods: contributing to the science of data quality* (pp.7-21). Hoboken, NJ: John Wiley & Sons.

Fowler, F. J. (1992). How unclear terms affect survey data. *Public Opinion Quarterly, 56*, 218-231.

Fowler, F. J., & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N. Schwarz, & S. Sudman (Eds.), *Answering questions: methodology for determining cognitive and communicative processes in survey research* (pp.15-36). San Francisco, CA: Jossey-Bass.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing; minimizing interviewer- related error.* Newbury Park, Calif.: Sage Publications.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Groves R. M. (1990). Theories and methods of telephone surveys. *Annual Review of Sociology, 16*, 221-240.

Groves, R. M., Fowler, F. J, Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology.* Hoboken, NJ: Wiley.

Groves, R. M., & Magilavy, L. J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly, 50*(2), 251-266.

Groves, R. M., O'Hare, B. C., Gould-Smith, D., Benkí, J., & Maher, P. (2008). Telephone interviewer voice characteristics and the survey participation decision. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology,* (pp. 385–400). New York: Wiley.

Guenzel, P. J., Berckmans, T. R., & Cannell, C. F. (1983). *General interviewing techniques: a self-instructional workbook for telephone and personal interviewer training.* Ann Arbor, Michigan: Institute for Social Research, Survey Research Center.

Holbrook, A., Cho, Y. I., & Johnson, T. (2006). The impact of question and respondent characteristics on comprehension and mapping difficulties. *Public Opinion Quarterly, 70*(4), 565-595.

Hovland, C. I., Irving, K. J., & Harold H. K. (1953). *Communication and persuasion*, New Haven, CT: Yale University Press.

Jans, M. E. (2010). *Verbal paradata and survey error: Respondent speech voice, and question-answering behavior can predict income item nonresponse* (Doctoral dissertation). Retrieved from http://deepblue.lib.umich.edu

Japec, L. (2008). Interviewer error and interviewer burden. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in Telephone Survey Methodology* (pp. 187-221). New York: Wiley.

Jenkins and Dillman (1997). Towards a theory of self-administered questionnaire design. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin.(Eds.), *Survey measurement and process quality* (pp. 165-196). New York: John Wiley and Sons.

Johnson (2003). *Acoustic and auditory phonetics.* MA: Blackwell Publishing.

Kane, E. W., & Macaulay, L. J. (1993). Interviewer gender and gender attitudes. *Public Opinion Quarterly, 57,* 1–28.

Kent, R. D. & Read, C. (2002). *Acoustic analysis of speech.* (2nd ed). Canada: Thomson Learning.

Ketrow, S. M. (1990). Attributes of a telemarketer's voice and persuasiveness: A review and synthesis of the literature. *Journal of Direct Marketing, 4,* 7–21.

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science,* 1-10.

Kish, L. (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association, 57,* 92-115.

Kirchner, A. & Olson, K. (2014). *Commitment, concealment, and confusion? Exploring interviewer and respondent behaviors in survey interviews across different types of questions.* Paper presented at the 2014 American Association for Public Opinion Research (AAPOR) conference. Anaheim, CA.

Knauper, B. (1999). The impact of age and education on response order effects in attitude measurement. *Public Opinion Quarterly, 63,* 347-370.

Knauper, B., Belli, R. F., Hill, D. H., & Herzog, A. R. (1997). Question difficulty and respondents' cognitive ability: the effect on data quality. *Journal of Official Statistics, 13*(2), 181-199.

Kreuter, F., Presser, S. & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and Web surveys. *Public Opinion Quarterly,* 72(5), pp. 847–865.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5,* 213-236.

Means, B. & Loftus, E. F. (1991). When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology,* 5, 297-318.

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology, 34*(4), 615-624.

Narayan, S &. Krosnick, J. A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly, 60,* 58-89.

Ohanian, R. (1990). Construction and Validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attraction, *Journal of Advertising, 19*(3), 39-52.

O'Muircheartaigh, C., & Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society Series A, 161*, 63-77.

O'Muircheartaigh, C., & Campanelli, P. (1999). A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society Series A, 162*(Part 3), 437–446.

Oksenberg, L., & Cannell, C. (1988). Effects of interviewer vocal characteristics on nonresponse. In R. M. Groves, P. B. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, and J. Waksberg (Eds.), *Telephone Survey Methodology* (pp. 252-272). New York: John Wiley and Sons.

Oksenberg, L., Coleman, L., & Cannell, C. (1986). Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly, 50*(1), 97–111.

Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly, 70*, 737-758.

Olson, K. & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly, 75*, 99-114.

Olson, K. & Peytchev, A. (2007). Effect of interviewer experience on interview pace and interviewer attitudes. *Public Opinion Quarterly, 71*, 273-286.

Ongena, Y. P. & Dijkstra, W. (2007). A model of cognitive process and conversational principles in survey interview interaction, Applied Cognitive Psychology, *21*, 145-163.

Pearce, W. B. & Conking, F. (1971). Nonverbal vocalic communication and perceptions of a speaker. *Speech Monographs, 38*, 235-241.

Rabe-Hesketh, S. and Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata* (3rd ed.). College Station, TX: Stata Press.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511-531.

Sakshaug, J., W., Yan, T., & Tourangeau, R. (2010). Nonresponse error, measurement, error, and mode of data collection. Tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly, 74*, 907-933.

Schaeffer, N. C., & Dykema, J. (2011). Response 1 to Fowler's chapter: Coding the behavior of interviewers and respondents to evaluate survey questions. In J. Madans, K. Miller, A. Maitland & G. Willis (Eds.), *Question evaluation methods: Contributing to the science of data quality* (pp. 23-39). Hoboken, NJ: John Wiley & Sons.

Schnell, R. & Kreuter, F. (2005). Separating interviewer and sampling-point effects. *Journal of Official Statistics, 21*, 389-410.

Sharon, A. T. (1973). What do adults read? *Reading Research Quarterly,9*(2), 148-169.

Smith, D. H. (1994). Determinants of voluntary association participation and volunteering: A literature review. *Nonprofit and Voluntary Sector Quarterly, 23*, 243-263.

Smith, S. M., & Shaffer, D. R. (1995). Speed of speech and persuasion: Evidence for multiple effects. *Personality and Social Psychology Bulletin, 21*, 1051–1060.

Street, R. L., Brady, R. M., & Putman W. B. (1983). The influence of speech rate sperotype and rate similarity or listeners' evaluations of speakers. *Journal of Language and Social Psychology, 2*(1), 37-56.

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11*(1), 90-105.

Teo, T. S., Lim, V. K., & Lai, R. Y. (1999). Intrinsic and extrinsic motivation in internet usage. *Omega International Journal of Management Science, 27*, 25-37.

Tourangeau, R., Conrad, F. G., & Couper, M. (2013). *The Science of Web Surveys*. New York, NY: Oxford University Press.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859-883.

Traunmuller, H., & Eriksson, A. (1993). The frequency range of the voice fundamental in the speech of male and female adults. Stockholm: Institutionen för Lingvistik, Stockholms University Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.64.5133&rep=rep1&type=pdf

Van der Vaart, W., Ongena, Y., Hoogendoorn, A., & Dijkstra, W. (2005). Do

interviewers' voice characteristics influence cooperation rates in telephone surveys? *International Journal of Opinion Research, 18*(4), 488–499.

Webb, J. T. (1969). Subject speech rates as a function of interviewer behavior. *Language & Speech, 12*(1), 54-67.

West, B. T. & Olson, K. (2010). How much of interviewer variance is really nonresponse error variance? *Public Opinion Quarterly, 74*, 1004-1026.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. *Proceedings of the Ninth International Conference on Spoken Language Processing.*

Chapter 4

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*(5), 715-727.

Anderson, C., Brion, S., Moore, D., A., & Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology*, *103*(4), 718-735.

Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs, 35*, 492-503.

Bachorowski, J. A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science, 8*(2), 53-57.

Baron, R. M., & Kenny, D. A. (1986).The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173-1182.

Bauer, D. J., Preacher, K. J., & Gill, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in Multilevel models: New procedures and recommendations. *Psychological methods, 11*(2), 142-163.

Benki, J., Broome, J., Conrad, F., Groves, R., & Kreuter, F. (2011, May). *Effects of speech rate, pitch, and pausing on survey participation decisions.* Paper presented at the 66th annual conference of the American Association for Public Opinion Research, Phoenix, AZ.

Beatty, P., & Herrmann, D. (2002). To answer or not to answer: Decision processes related to survey item nonresponse. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 71–85). New York: Wiley Series in Survey Methodology.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4, retrieved 4 October 2014 from http://www.praat.org

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(2), 123-147.

Broome, J. S. (2012). *Vocal characteristics, speech, and behavior of telephone interviewers* (Doctoral dissertation). Retrieved from http://deepblue.lib.umich.edu

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., & Hauck, W.(1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly, 60*(3), 345-375.

Conrad, F. G., Broome, J. S., Benki, J. R., Kreuter, F., Groves, R. M., Vannette, D., & McClain C. (2013). Interviewer speech and the success of survey invitations. *Journal of the Royal Statistical Society, 176,* 191-210.

Conrad, F. G., Schober, M., & Dijkstra, W. (2008). Cues of communication difficulty in telephone interviews. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 212-230). New York: John Wiley & Sons.

Dey, A., Feinberg, D., & Kim J. (2006). Effect of voice pitch on content comprehension. Retreived 26 March 2015 from http://cll.mcmaster.ca/sotl/pdf/3024_Group1/3024_2Dey.pdf

De Leeuw, E. D., How, J. J. & Dillman, D. A. (Eds.) (2008). *International handvook of survey methodology.* New York: Psychology Press.

Dijkstra, V. (2009). Sequence Viewer version 5. Retrieved from http://sequenceviewer.nl/downloads/Reference%20manual%20version%205.0.pdf

Durrant, G. B., Groves, R. M., Staetsky, L., & Steele, F. (2010). Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys. *Public Opinion Quarterly, 74*(1), 1–36.

Dykema, J., Lepkowski, J. & Blixt, S. (1997). The effect of interviewer and respondent behavior on data quality: Analysis of interaction coding in a validation study. In L. E. Lyberg, P. Biemer, M. Collins, E. D. Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 287-310). New York: Wiley.

Fowler, F. J. (2011). Coding the behavior of interviewers and respondents to evaluate survey questions. In J. Madans, K. Millers, A. Maitland, & G. Willis (Eds.), *Question evaluation methods: contributing to the science of data quality* (pp.7-21). Hoboken, NJ: John Wiley & Sons.

Fowler, F. J. (1992). How unclear terms affect survey data. *Public Opinion Quarterly, 56,* 218-231.

Fowler, F. J., & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N. Schwarz, & S. Sudman (Eds.), *Answering questions: methodology for determining cognitive and communicative processes in survey research* (pp.15-36). San Francisco, CA: Jossey-Bass.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing; minimizing interviewer- related error.* Newbury Park, Calif.: Sage Publications.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Groves R. M. (1990). Theories and methods of telephone surveys. *Annual Review of Sociology, 16*, 221-240.

Groves, R. M., Fowler, F. J, Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology.* Hoboken, NJ: Wiley.

Groves, R. M., O'Hare, B. C., Gould-Smith, D., Benkí, J., & Maher, P. (2008). Telephone interviewer voice characteristics and the survey participation decision. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology,* (pp. 385–400). New York: Wiley.

Hovland, C. I., Janis, I. J., & Kelley H. H. (1953). *Communication and persuasion*, New Haven, CT: Yale University Press.

Japec, L. (2008). Interviewer error and interviewer burden. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in Telephone Survey Methodology* (pp. 187-221). New York: Wiley.

Johnson (2003). *Acoustic and auditory phonetics.* MA: Blackwell Publishing.

Kane, E. W., & Macaulay, L. J. (1993). Interviewer gender and gender attitudes. *Public Opinion Quarterly, 57*, 1–28.

Kent, R. D. & Read, C. (2002). *Acoustic analysis of speech.* (2nd ed). Canada: Thomson Learning.

Ketrow, S. M. (1990). Attributes of a telemarketer's voice and persuasiveness: A review and synthesis of the literature. *Journal of Direct Marketing, 4*, 7–21.

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science,* 1-10.

Kirchner, A. & Olson, K. (2014). *Commitment, concealment, and confusion? Exploring interviewer and respondent behaviors in survey interviews across different types of questions.* Paper presented at the 2014 American Association for Public Opinion Research (AAPOR) conference. Anaheim, CA.

Knauper, B. (1999). The impact of age and education on response order effects in attitude measurement. *Public Opinion Quarterly, 63*, 347-370.

Kreuter, F., Presser, S. & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and Web surveys. *Public Opinion Quarterly*, 72(5), pp. 847–865.

Krosnick, J. A. (2002). "The causes of no-opinion responses to attitude measures in surveys: They are rarely what they appear to be." In R. M. Groves, D. A. Dillman,

J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 87-100). New York: Wiley.

Lai, C. (2010). What do you mean, you're uncertain?: The interpretation of cue words and rising intonation in dialogue. *Proceedings of the Annual Conference of International Speech Communication Association,* Makuhari, Chiba, Japan.

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology, 34*(4), 615-624.

Narayan, S &. Krosnick, J. A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly, 60,* 58-89.

Ohanian, R. (1990). Construction and Validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attraction, *Journal of Advertising, 19*(3), 39-52.

O'Muircheartaigh, C., & Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society Series A, 161,* 63-77.

O'Muircheartaigh, C., & Campanelli, P. (1999). A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society Series A, 162*(Part 3), 437–446.

Oksenberg, L., & Cannell, C. (1988). Effects of interviewer vocal characteristics on nonresponse. In R. M. Groves, P. B. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, and J. Waksberg (Eds.), *Telephone Survey Methodology* (pp. 252-272). New York: John Wiley and Sons.

Oksenberg, L., Coleman, L., & Cannell, C. (1986). Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly, 50*(1), 97–111.

Olson, K. & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly, 75,* 99-114.

Olson, K. & Peytchev, A. (2007). Effect of interviewer experience on interview pace and interviewer attitudes. *Public Opinion Quarterly, 71,* 273-286.

Pearce, W. B. & Conking, F. (1971). Nonverbal vocalic communication and perceptions of a speaker. *Speech Monographs, 38,* 235-241.

Price, P. C. & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making, 17,* 39-57.

Rabe-Hesketh, S. and Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata* (3rd ed.). College Station, TX: Stata Press.

Sakshaug, J. W., Yan, T., & Tourangeau, R. (2010). Nonresponse error, measurement, error, and mode of data collection. Tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly, 74*, 907-933.

Sah, S, Moore, D. A., & MacCoun, R. J. (2013). Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiceness. *Organizational Behavior and Human Decision Processes, 121*, 246-255.

Schaeffer, N. C., & Dykema, J. (2011). Response 1 to Fowler's chapter: Coding the behavior of interviewers and respondents to evaluate survey questions. In J. Madans, K. Miller, A. Maitland & G. Willis (Eds.), *Question evaluation methods: Contributing to the science of data quality* (pp. 23-39). Hoboken, NJ: John Wiley & Sons.

Smith, S. M., & Shaffer, D. R. (1995). Speed of speech and persuasion: Evidence for multiple effects. *Personality and Social Psychology Bulletin, 21*, 1051–1060.

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11*(1), 90-105.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859-883.

Van der Vaart, W., Ongena, Y., Hoogendoorn, A., & Dijkstra, W. (2005). Do interviewers' voice characteristics influence cooperation rates in telephone surveys? *International Journal of Opinion Research, 18*(4), 488–499.

Webb, J. T. (1969). Subject speech rates as a function of interviewer behavior. *Language & Speech, 12*(1), 54-67.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. *Proceedings of the Ninth International Conference on Spoken Language Processing.*

Chapter 5

Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs, 35*, 492-503.

Baron, R. M., & Kenny, D. A. (1986).The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Belli, R. F., Weiss, P. S. & Lepkowski J. M. (1999). Dynamics of survey interviewing and the quality of report: Age comparisons. In N. Schwarz, D. Park, B. Knäuper, & S. Sudman (Eds.), *Cognition, aging, and self-reports* (pp. 303-326). Philadelphia: Psychology Press.Benki, J., Broome, J., Conrad, F., Groves, R., & Kreuter, F. (2011, May). *Effects of speech rate, pitch, and pausing on survey participation decisions.* Paper presented at the 66th annual conference of the American Association for Public Opinion Research, Phoenix, AZ.

Biemer & Lyberg (2003). *Introduction to survey quality.* Hoboken, NJ: Wiley.

Blumberg, S. J. & Luke, J. V. (2014). Wireless Substitution: Early Release of estimates from the National Health Interview Survey National Center for Health Statistics. Available from http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201412.pdf

Boehme, B. (2014). How trustworthy is your voice? The effects of voice manipulation on the perceived trustworthiness of novel speakers. Retreived 13 June 2015 from http://www.did.stu.mmu.ac.uk/MMU_Psychology_Dissertations_UK/2014dissertations/GtoK/Glasgow/Boehme%20%28Bibi%29%202014%20%28Glasgow%29.pdf

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques, In S. Leinhardt (Ed.), *Sociological methodology* (pp. 289-437). SanFrancisco: Jossey-Bass.

Conrad, F. G., Broome, J. S., Benki, J. R., Kreuter, F., Groves, R. M., Vannette, D., & McClain C. (2013). Interviewer speech and the success of survey invitations. *Journal of the Royal Statistical Society, 176*, 191-210.

Dey, A., Feinberg, D., & Kim J. (2006). Effect of voice pitch on content comprehension. Retreived 26 March 2015 from http://cll.mcmaster.ca/sotl/pdf/3024_Group1/3024_2Dey.pdf

Dillman, D. A., Smyth, J. D., Christian, L. M. (2014). *Internet, mail and mixed-mode surveys: The tailored design method* (4th ed.). New York: Wiley.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing; minimizing interviewer- related error.* Newbury Park, Calif.: Sage Publications.

Geisen, E. M. et al. (2015). *Comparison of Text-To-Speech with human voice recordings on comprehension of survey questions in Audio Computer-Assisted Self-Interviewing.* Paper presented at the 2015 American Association for Public Opinion Research (AAPOR) conference. Hollywood, FL.

Groves R. M. (1990). Theories and methods of telephone surveys. *Annual Review of Sociology, 16*, 221-240.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Guenzel, P. J., Berckmans, T. R., & Cannell, C. F. (1983). *General interviewing techniques: a self-instructional workbook for telephone and personal interviewer training.* Ann Arbor, Michigan: Institute for Social Research, Survey Research Center.

Kent, R. D. & Read, C. (2002). *Acoustic analysis of speech.* (2nd ed). Canada: Thomson Learning.

DiLoreto, K. et al. (2015). *Effects of ACASI voice choice and voice persona on reports to questions about sensitive behaviors among young adults.* Paper presented at the 2015 American Association for Public Opinion Research (AAPOR) conference. Hollywood, FL.

Ketrow, S. M. (1990). Attributes of a telemarketer‘s voice and persuasiveness: A review and synthesis of the literature. *Journal of Direct Marketing, 4*, 7–21.

Lavrakas, P. J. et al. (2010). AAPOR Cell Phone Task Force 2010 New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. with Respondents Reached via Cell Phone Numbers, available at http://www.aapor.org

Oksenberg, L., Coleman, L., & Cannell, C. (1986). Interviewers’ voices and refusal rates in telephone surveys. *Public Opinion Quarterly, 50*(1), 97–111.

Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly, 70*, 737-758.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates’ judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511-531.

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11*(1), 90-105.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. *Proceedings of the Ninth International Conference on Spoken Language Processing.*

APPENDICES

APPENDIX A: QUESTION WORDING OF THE TWELVE QUESTIONS

Neutral questions

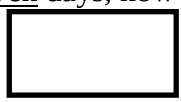
Q2 Compared to 10 years ago (2013), do you think people have more leisure time, less leisure time or about the same amount?

1 More
2 Same amount
3 Less
8 DK
9 REF

Q13D (On a scale from 1 to 5 where 5 means you enjoy the activity completely and 1 means you do not enjoy the activity at all, please tell me how much you enjoy)- Fishing or hunting ?

5 Enjoy completely
4 Enjoy a lot
3 Enjoy somewhat
2 Enjoy a little
1 Do not enjoy at all
8 DK
9 REF

Q21A People do a number of different types of activities for leisure. Thinking about the past seven days, how many times did you use the internet?



88888 DK
99999 REF

Complex questions

Q13E (On a scale from 1 to 5 where 5 means you enjoy the activity completely and 1 means you do not enjoy the activity at all, please tell me how much you enjoy)- Kaninhop?

5 Enjoy completely
4 Enjoy a lot
3 Enjoy somewhat
2 Enjoy a little
1 Do not enjoy at all
8 DK
9 REF

Q19 On a typical day, how many minutes do you spend on a computer?

DEFINITION (ON THE SAME SCREEN AS THE QUESTION): There are 1440 minutes in a 24 hour day.

8888 DK
9999 REF

Q20 In the past week, how many email messages, if any, have you written or received?

88888 DK
99999 REF

Socially undesirable questions

Q5 Have you ever been fired from a job?

1 Yes
2 No
8 DK
9 REF

Q21C (Thinking about the past seven days, how many times did you)- drink alcohol? [INTERVIEWER: We are interested in the total number of drinks.]

88888 DK
99999 REF

Q21D (Thinking about the past seven days, how many times did you)- have sex?

88888 DK
99999 REF

Socially desirable questions

Q8 We are interested in volunteer activities for which people are not paid, except perhaps expenses. We only want you to include volunteer activities that you did through or for an organization, even if you only did them once in a while. In the last 12 months, that is since July of last year (2012), have you done any volunteer activities through or for an organization?

[PROBE: IF HAVE NOT VOLUNTEERED, ASK:] Sometimes people don't think of activities they do infrequently or activities they do for children's schools or youth organizations as volunteer activities. Since July of last year, have you done any of these types of volunteer activities?

1 Yes
2 No
8 DK
9 REF

Q13A On a scale from 1 to 5 where 5 means you enjoy the activity completely and 1 means you do not enjoy the activity at all, please tell me how much you enjoy the following leisure activities. First, how about reading?

5 Enjoy completely
4 Enjoy a lot
3 Enjoy somewhat
2 Enjoy a little
1 Do not enjoy at all
8 DK
9 REF

Q21F (Thinking about the past seven days, how many times did you)-Read a book, magazine or newspaper

88888 DK
99999 REF

APPENDIX B: CODER INSTRUCTION

In this task, we would like you to rate interviewer voice characteristics and interviewer traits, and identify interviewer gender from voice files.

Interviewer voice characteristics. You will rate 4 interviewer voice characteristics including rate of speech, pitch, intonation, and disfluency on a 7-point scale.

- *Rate of speech* is how quickly you think that the interviewer is reading a survey question. You will rate how fast an interviewer speaks from slow (1) to fast (7).
- *Pitch* is your perception of an interviewer's voice from low-pitched to high-pitched. You will rate interviewer's pitch on a range from low (1) to high (7).
- *Intonation* refers to the rise and fall of voice pitch. You will rate interviewer's intonation from flat (1), i.e. no variation in voice pitch, to varied (7), i.e. large variation in voice pitch.
- *Disfluency* is the parts of speech that are not words, such as stutters, saying "um" and "uh," and pauses. Interviewer's disfluency will be rated from low (1) to high (7).

Interviewer traits. From each voice file, you will rate 5 interviewer traits including confidence, ease of understanding, reliability, trustworthiness, and expertise on a 7-point scale.

- *Confidence* -You will rate whether the interviewer is self-assured and conducts the interview with poise from low (1) to high (7).
- *Ease of understanding* -You will rate whether the interviewer's voice is easy to understand from low (1) to high (7).
- *Reliability* occurs when interviewers say something that can be believed. You will rate interviewer's reliability from low (1) to high (7).
- *Trustworthiness* is the degree of confidence in an interviewer to ask a valid survey question and to keep respondents' answers confidential. You will rate interviewer's trustworthiness from low (1) to high (7).
- *Expertise* is the extent to which an interviewer is good at her/his job in asking a survey question. You will rate interviewer's expertise from low (1) to high (7).

Interviewer gender. You determine whether you think an interviewer is male or female from interviewer's voice. You code "m" for male interviewers and "f" for female interviewers or leave blank if you cannot identify interviewer gender from his/her voice.

APPENDIX C: DATA QUALITY INDICATORS

Table C.1 Summary of data quality indicators

Types of questions	Questions	Data quality indicators	
		Data quality indicators	**Definition**
Neutral questions	Q2 (Leisure time)	Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior
	Q13D (Fishing and hunting enjoyment)	Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior
	Q21A (Number of times using the internet)	Rounding	1=a respondent reported units in multiples of 5 0=a respondent did not report units in multiples of 5
		Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior

Table C.1 (cont.)

Types of questions	Questions	Data quality indicators	
		Data quality indicators	**Definition**
Complex questions	Q13E (Kaninhop enjoyment)	Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior
	Q19 (Number of minutes spending on a computer)	Rounding	1=a respondent reported units in multiples of 60 minutes 0=a respondent did not report units in multiples of 60 minutes
		Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior
	Q20 (Number of email messages)	Rounding	1=a respondent reported units in multiples of 5 0=a respondent did not report units in multiples of 5
		Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior

Table C.1 (cont.)

Types of questions	Questions	Data quality indicators	
		Data quality indicators	**Definition**
Undesirable questions	Q5 (Fired from a job)	Proportion "yes"	1=a respondent reported 'yes' they have ever been fired from a job 0=a respondent reported 'no' they have never been fired from a job
		Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior
	Q21C (alcohol drinks)	"More is better" hypothesis	1=a respondent reported that they had at least one alcoholic drink in the past seven days 0=a respondent reported that they did not have alcoholic drink in the past seven days
		Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior
	Q21D (Number of times having sex)	"More is better" hypothesis	1=a respondent reported that they had sex at least one time in the past seven days 0=a respondent reported that they did not have sex at least one time in the past seven days
		Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior

Table C.1 (cont.)

Types of questions	Questions	Data quality indicators	
		Data quality indicators	**Definition**
Desirable questions	Q8 (Volunteer Activity)	Proportion "no"	1= a respondent reported that they have not done any volunteer activities in the last 12 months 0= a respondent reported that they have done any volunteer activities in the last 12 months
		Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior
	Q13A (Reading Enjoyment)	"Less is better" hypothesis	1=a respondent reported that they did not completely enjoy reading a book 0=a respondent reported that they completely enjoyed reading a book
		Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior
	Q21F (Number of times reading books)	"Less is better" hypothesis	1= a respondent reported that they read a book, magazine, or newspaper fewer than 10 times 0=a respondent reported that they read a book, magazine, or newspaper 10 times or more
		Item nonresponse	1=a respondent did not answer a question 0=a respondent answered a question
		Problematic respondent behaviors	1=a respondent engaged in the behavior 0=a respondent did not engage in the behavior

Table C.2 Descriptive statistics of data quality indicators

	Percentage of data quality indicator
1. Item nonresponse	4.56%
2. Problematic respondent behaviors	
2.1 Interrupting questions with answers	10.45%
2.2 Expressing uncertainty about a question	11.32%
2.3 Requesting clarification about a question	20.11%
2.4 Giving a qualified answer	13.67%
2.5 Giving a response that does not meet the question's objective	22.01%
3. Rounded answer	
3.1 Rounded answer in question 21A	47.67%
3.2 Rounded answer in question 19	55.63%
3.3 Rounded answer in question 20	77.00%
4. Hypothesis of "more is better"	
4.1 Percentage of respondents reported that they have ever been fired from a job in question 5	18.60%
4.2 Percentage of respondents reported that they had at least one alcohol in the past seven day in question 21C	35.58%
4.3 Percentage of respondent reported that they had sex at least one time in the past seven days in question 21D	28.38%
5. Hypothesis of "less is better"	
5.1 Percentage of respondents reported that they have not done any volunteer activities in the last 12 month in question 8	52.94%
5.2 Percentage of respondents reported that they did not completely enjoy reading a book in question 13A	45.08%
5.3 Percentage of respondent reported that they read a book, magazine, or newspaper fewer than 10 times in the past seven days in question 21F	71.19%

Table C.3 Variance components in a base model for problematic respondent behaviors

	Interrupt questions with answers		Express uncertainty about a question	Request clarification about a question	Give a qualified answer	Give a response that does not meet the question's objective	
	coefficient (SE)		coefficient (SE)	coefficient (SE)	coefficient (SE)	coefficient (SE)	
Variance components							
2-level variance (respondents)	0.409(0.09)		0(.)	0(.)	0(.)	0.341(0.06)	
3-level variance (interviewers)	0.161(0.08)		0.033(0.02)	0026(0.02)	0.504(0.08)	0.016(0.02)	
Residual variance	3.29		3.29	3.29	3.29	3.29	
ICC							
Respondents							
Value	0.106		0	0	0.133	0.094	
Chi-square	35.86	**	.	.	.	60.98	**
Interviewers							
Value	0.042		0.010	0.008	0	0.004	
Chi-square	5.74	**	.	.	.	24.94	**
Model fit							
Generalized Chi-square	3716.75		4635.33	4659.41	3860.46	4222.10	

Note **p<0.01, *p<0.05; n = 4,689

Table C.4 Variance components in a base model for item nonresponse

	Item Nonresponse	
	coefficient (SE)	
Variance components		
2-level variance (respondents)	0.063(0.13)	**
3-level variance (interviewers)	0.424(0.18)	**
Residual variance	3.29	
ICC		
Respondents		
Value	0.017	
Chi-square	0.24	
Interviewers		
Value	0.112	
Chi-square	5.59	**
Model fit		
Generalized Chi-square	4131.54	

Note **p<0.01, *p<0.05; n = 4,689

Table C.5 Variance components in a base model for rounding as the data quality indicator

	Q21A (Number of times using the internet)		Q19 (Number of minutes spending on a computer)		Q20 (Number of email messages)	
	coefficient (SE)		**coefficient (SE)**		**coefficient (SE)**	
Variance components						
2-level variance (interviewers)	0.1899		0.034		0.225	
Residual variance	3.29		3.29		3.29	
ICC						
Value	0.055		0.010		0.064	
Chi-square	2.81	*	0.15		3.01	*
Model fit						
AIC	387.36		406.31		310.51	
n	279		293		287	

Note **p<0.01, *p<0.05

Table C.6 Variance components in a base model for proportion of answers that are less prone to socially desirable bias for socially undesirable questions

	Q5 (Fired from a job)	Q21C (1+ alcohol drinks)	Q21D (Have sex 1+ times)
	coefficient (SE)	**coefficient (SE)**	**coefficient (SE)**
Variance components			
2-level variance (interviewers)	0(.)	0(.)	0.109(0.14)
Residual variance	3.29	3.29	3.29
ICC			
Value	0	0	0.032
Chi-square	1	1	1.04
Model fit			
AIC	399.74	543.59	452.74
n	414	416	377

Note **p<0.01, *p<0.05

Table C.7 Variance components in a base model for proportion of answers that are less prone to socially desirable bias for socially desirable questions

	Q8 (Did not Volunteer Activity)	Q13A (Does not completely Enjoy Reading)	Q21F (Number of Reading times<10)	
	coefficient (SE)	**coefficient (SE)**	**coefficient (SE)**	
Variance components				
2-level variance (interviewers)	0(.)	0(.)	0.195(0.16)	
Residual variance	3.29	3.29	3.29	
ICC				
Value	0	0	0.056	
Chi-square	0	0	3.65	*
Model fit				
AIC	566.20	576.05	496.34	
N	408	417	413	

Note **p<0.01, *p<0.05

APPENDIX D: RESULTS OF HIRARCHICAL LOGISTIC MODELS TO EXAMINE ASSOCIATIONS BETWEEN OBJECTIVE VOICE CHARACTERISTICS AND DATA QUALITY INDICATORS FOR THREE QUESTIONS ANALYZING TOGETHER

Table D.1 Hierarchical logistic model predicting proportion of rounded answers by objective voice characteristic for three questions (Q21A, Q19, and Q20)

	coefficient (SE)	
Main effects		
Intercept	0.976(0.24)	**
Pitch	-0.002(0.004)	
Intonation	-0.001(0.005)	
Speech rate	-0.236(0.19)	
Fillers	0.020(0.16)	
Interviewer's experience < 1 year	-0.049(0.24)	
Female interviewer	0.135(0.31)	
Question 21A	-1.395(0.19)	**
Question 19	-0.938(0.19)	**
R whose education is high school or less	-0.678(0.19)	**
R whose age is 60 or less	0.572(0.16)	**
Interaction btw voice and Iwer char		
Speech rate*female interviewer	0.577(0.27)	*
Variance components		
2-level variance (respondents)	0.231(0.15)	
3-level variance (interviewers)	0.075(0.07)	
Residual variance	3.29	
Model fit		
Generalized Chi-square	809.73	

Note **p<0.01, *p<0.05; Question 20 is a reference group for Question variable

Table D.2 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by objective voice characteristics for three socially undesirable questions (Q5, Q21C, and Q21D)

	coefficient (SE)	
Main effects		
Intercept	-1.301(0.19)	**
Pitch	0.001(0.003)	
Intonation	-0.002(0.004)	
Speech rate	0.105(0.09)	
Fillers	-0.191(0.23)	
Interviewer's experience < 1 year	0.225(0.16)	
Female interviewer	0.138(0.23)	
Question 5	-0.675(0.19)	**
Question 21C	0.389(0.16)	*
R whose education is high school or less	-0.328(0.16)	
R whose age is 60 or less	0.740(0.14)	**
Variance components		
2-level variance (respondents)	0	
3-level variance (interviewers)	0(0.14)	
Residual variance	3.29	
Model fit		
Generalized Chi-square	1110.62	

Note **p<0.01, *p<0.05; Question 21D is a reference group for question variable

Table D.3 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by objective voice characteristics for three socially desirable questions (Q8, Q13A, and Q21F)

	coefficient (SE)	
Main effects		
Intercept	1.426(3.55)	
Pitch	0.001(0.003)	
Intonation	-0.002(0.004)	
Speech rate	0.025(0.11)	
Fillers	0.048(0.11)	
Interviewer's experience < 1 year	0.097(0.14)	
Female interviewer	-0.036(0.20)	
Question 8	-1.013(0.16)	**
Question 13A	-1.097(0.15)	**
R whose education is high school or less	0.132(0.13)	
R whose age is 60 or less	0.479(0.12)	**
Variance components		
2-level variance (respondents)	0	
3-level variance (interviewers)	0	
Residual variance	3.29	
Model fit		
Generalized Chi-square	1236.01	

Note **p<0.01, *p<0.05; Question 21F is a reference group for question variable

APPENDIX E: RESULTS OF HIRARCHICAL LOGISTIC MODELS TO EXAMINE ASSOCIATIONS BETWEEN SUBJECTIVE VOICE CHARACTERISTICS AND DATA QUALITY INDICATORS FOR THREE QUESTIONS ANALYZING TOGETHER

Table E.1 Hierarchical logistic model predicting proportion of rounded answers by subjective voice characteristics for three questions (Q21A, Q19, and Q20)

	coefficient (SE)	
Main effects		
Intercept	1.079(0.27)	**
Rated pitch	0.464(0.24)	
Rated intonation	-0.097(0.16)	
Rated speech rate	0.094(0.13)	
Rated fillers	-0.082(0.12)	
Interviewer's experience < 1 year	0.041(0.23)	
Female interviewer	-0.647(0.40)	
Question 21A	-1.404(0.20)	**
Question 19	-1.105(0.19)	**
R whose education is high school or less	-0.742(0.19)	**
R whose age is 60 or less	0.582(0.16)	**
Variance components		
2-level variance (respondents)	0.257(0.16)	
3-level variance (interviewers)	0.041(0.06)	
Residual variance	3.29	
Model fit		
Generalized Chi-square	778.34	

Note **p<0.01, *p<0.05; Question 20 is a reference group for Question variable

Table E.2 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by subjective voice characteristics for three socially undesirable questions (Q5, Q21C, and Q21D)

	coefficient (SE)	
Main effects		
Intercept	-1.516(0.27)	**
Rated pitch	-0.415(0.21)	
Rated intonation	0.150(0.14)	
Rated speech rate	0.122(0.10)	
Rated fillers	0.109(0.12)	
Interviewer's experience < 1 year	-0.380(0.30)	
Female interviewer	0.831(0.35)	
Question 5	-0.494(0.18)	*
Question 21C	0.363(0.16)	*
R whose education is high school or less	-0.322(0.16)	
R whose age is 60 or less	0.780(0.14)	**
Interaction between voice and interviewer characteristics		
Rated speech rate* Interviewer's experience < 1 year	0.614(0.23)	**
Variance components		
2-level variance (respondents)	0.195(0.14)	
3-level variance (interviewers)	0	
Residual variance	3.29	
Model fit		
Generalized Chi-square	1114.84	

Note **p<0.01, *p<0.05; Question 21D is a reference group for question variable

Table E.3 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by subjective voice characteristics for three socially desirable questions (Q8, Q13A, and Q21F)

	coefficient (SE)	
Main effects		
Intercept	0.486(0.23)	
Rated pitch	0.001(0.24)	
Rated intonation	0.038(0.12)	
Rated speech rate	0.171(0.11)	
Rated fillers	0.165(0.11)	
Interviewer's experience < 1 year	0.522(0.24)	
Female interviewer	-0.011(0.33)	
Question 8	-0.902(0.16)	**
Question 13A	-1.203(0.16)	**
R whose education is high school or less	0.904(0.16)	**
R whose age is 60 or less	0.391(0.14)	*
Interaction between voice and interviewer characteristics		
Rated speech rate* Interviewer's experience < 1 year	-0.513(0.23)	*
Variance components		
2-level variance (respondents)	0.529(0.15)	
3-level variance (interviewers)	0.003(0.04)	
Residual variance	3.29	
Model fit		
Generalized Chi-square	1080.07	

Note **p<0.01, *p<0.05; Question 21F is a reference group for question variable

APPENDIX F: RESULTS OF HIRAROCHICAL LOGISTIC MODELS TO EXAMINE WHETHER SUBJECTIVE VOICE CHARACTERISTICS MEDIATE RELATIONSHIPS OF OBJECTIVE VOICE CHARACTERISTICS ON DATA QUALITY

Table F.1 Hierarchical logistic model predicting respondent behaviors by subjective and objective voice characteristics

	Interrupt questions with answers	Express uncertainty about a question	Request clarification about a question	Give a qualified answer	Give a response that does not meet the question's objective
	coefficient (SE)	coefficient (SE)	coefficient (SE)	coefficient (SE)	coefficient (SE)
Intercept	-1.967(0.30)**	-3.702(0.32)**	-2.334*(0.25)**	-1.884(0.24)**	-0.973(0.21)**
Pitch	0.003(0.003)	-0.002(0.0004)	-0.002(0.003)	0.001(0.002)	0.007(0.002)**
Intonation	-0.002(0.004)	0.002(0.007)	0.002(0.005)	-0.001(0.003)	-0.004(0.003)
Intonation^2		0.0004(0.0001)**	0.0003(0.00001)**		
Speech rate	0.508(0.21)*	0.212(0.21)	0.109(0.17)	0.195(0.14)	0.058(0.14)
Speech rate^2				-0.143(0.05)**	
Fillers	-0.096(0.12)	0.112(0.15)	0.014(0.11)	0.141(0.14)	-0.311(0.12)*
Rated pitch	-0.140(0.18)	0.228(0.21)	0.163(0.15)	-0.123(0.15)	0.027(0.13)
Rated intonation	-0.038(0.21)	-0.270(0.15)	-0.139(0.10)	0.018(0.10)	-0.203(0.09)*
Rate speech rate	0.004(0.12)	0.266(0.12)*	-0.102(0.21)	0.240(0.09)*	0.244(0.15)
Rated speech rate^2			0.127(0.06)*		
Rate fillers	0.047(0.09)	-0.266(0.11)*	-0.165(0.08)*	0.043(0.13)	-0.122(0.10)
Rate fillers^2					-0.125(0.06)*
Interviewer's experience < 1 year	0.100(0.27)	-0.264(0.31)	-0.132(0.19)	0.098(0.15)	0.008(0.16)
Female interviewer	0.046(0.37)	0.049(0.43)	0.184(0.29)	0.435(0.27)	-0.513(0.26)
Desirable question	0.982(0.25)**	-0.078(0.20)	0.632(0.23)**	-1.288(0.28)**	0.336(0.20)
Complex question	0.826(0.27)**	1.833(0.17)**	1.219(0.23)**	0.526(0.24)*	1.161(0.20)**
Undesirable question	-0.892(0.31)**	-0.916(0.26)**	-1.483(0.31)**	-1.460(0.37)**	-1.117(0.22)**
R whose education is highschool or less	-0.088(0.14)	0.168(0.14)	0.00004(0.11)	0.033(0.13)	0.340(0.11)**
R whose age is 60 or less	-0.649(0.13)**	-0.294(0.14)*	-0.062(0.10)	0.066(0.12)	-0.660(0.10)**

Note **p<0.01, *p<0.05; n = 4,689

Table F.1 Continued

	Interrupt questions with answers	Express uncertainty about a question	Request clarification about a question	Give a qualified answer	Give a response that does not meet the question's objective
	coefficient (SE)	**coefficient (SE)**	**coefficient (SE)**	**coefficient (SE)**	**coefficient (SE)**
Intonation*desirable question		-0.001(0.01)	-0.013(0.01) *		
Intonation*complex question		-0.021(0.01) **	-0.019(0.01) **		
Intonation*undesirable question		-0.022(0.01) *	-0.020(0.01) **		
Speech rate*desirable question	0.193(0.27)	-0.459(0.31)	-0.001(0.24)	-0.071(0.21)	0.416(0.21) *
Speech rate*complex question	-0.365(0.23)	-1.473(0.22) **	-1.176(0.19) **	-0.097(0.16)	0.339(0.16) **
Speech rate*undesirable question	-0.963(0.27) **	-0.954(0.29) **	-0.862(0.24) **	-0.519(0.21) **	-0.599(0.19) **
Rated speech rate*desirable question	-0.534(0.24) *		-0.428(0.21) *		-0.434(0.18) *
Rated speech rate*complex question	-0.046(0.25)		0.372(0.21)		-0.255(0.19)
Rated speech rate*undesirable question	0.217(0.28)		0.316(0.25)		0.215(0.19)
Rated fillers*desirable question				-0.563(0.22) *	
Rated fillers*complex question				0.076(0.18)	
Rated fillers*undesirable question				-0.210(0.27)	
Speech rate*iwer exp < 1 yr			0.345(0.12) **		
Fillers*iwer exp < 1 yr			0.497(0.26)		
Fillers*female interviewer				-0.622(0.24) **	0.492(0.18) **
Variance components					
2-level variance (respondents)	0.438(0.10) **	0.193(0.11)	0.162(0.06)	0.605(0.1)	0.405(0.07) **
3-level variance (interviewers)	0.184(0.10) **	0.255(0.13)	0.074(0.05)	0	0.043(0.04) **
Residual variance	3.29	3.29	3.29	3.29	3.29
Model fit					
Generalized Chi-square	3708.42	4210.83	4302.54	3693.14	4026.97

Note **p<0.01, *p<0.05; n = 4,689

Table F.2 Hierarchical logistic model predicting item nonresponse by subjective and objective voice characteristics

	coefficient (SE)
Intercept	-3.646(0.40) **
Pitch	-0.002(0.004)
Intonation	-0.011(0.01)
Intonation^2	0.0002(0.0001) *
Speech rate	-0.520(0.11) **
Fillers	0.064(0.20)
Rated pitch	-0.032(0.26)
Rated intonation	-0.039(0.18)
Rate speech rate	0.297(0.15) *
Rate fillers	-0.068(0.13)
Interviewer's experience < 1 year	0.384(0.45)
Female interviewer	-0.251(0.58)
Desirable question	-1.322(0.30) **
Complex question	0.329(0.20)
Undesirable question	0.089(0.22)
R whose education is highschool or less	0.107(0.17)
R whose age is 60 or less	-0.114(0.16)
Interaction btw voice and Iwer char	
Speech rate*interwer exp < 1 yr	-0.319(0.15) *
Variance components	
2-level variance (respondents)	0.216(0.15)
3-level variance (interviewers)	0.559(0.26) **
Residual variance	3.29
Model fit	
Generalized Chi-square	3784.1

Note **p<0.01, *p<0.05; n = 4,689

Table F.3 Hierarchical logistic model predicting rounding by subjective and objective voice characteristics

	Q21A (Number of times using the internet)		Q19 (Number of minutes spending on a computer)		Q20 (Number of email messages)	
	coefficient (SE)		**coefficient (SE)**		**coefficient (SE)**	
Main effects						
Intercept	-1.080(0.51)	*	0.645(0.46)		0.620(0.56)	
Pitch	0.0002(0.006)		-0.008(0.005)		0.005(0.008)	
Intonation	0.014(0.01)		-0.005(0.008)		-0.016(0.01)	
Speech rate	-0.524(0.43)		-0.335(0.31)		-0.380(0.41)	
Fillers	0.079(0.24)		-0.018(0.33)		0.226(0.58)	
Rated pitch	0.609(0.44)		0.713(0.40)		0.051(0.48)	
Rated intonation	-0.475(0.29)		-0.031(0.30)		0.126(0.35)	
Rate speech rate	0.482(0.28)		-0.204(0.23)		0.780(0.38)	*
Rate fillers	-0.278(0.22)		0.082(0.25)		-0.202(0.31)	
Interviewer's experience < 1 year	0.194(0.36)		-0.155(0.31)		0.053(0.45)	
Female interviewer	-1.348(0.73)		0.393(0.62)		-0.363(0.85)	
R whose education is high school or less	-0.472(0.31)		-0.676(0.29)	*	-1.235(0.33)	
R whose age is 60 or less	1.107(0.26)	**	0.092(0.25)		0.712(0.32)	
Interaction btw voice and Iwer char						
Speech rate*female interviewers			0.635(0.35)			
Variance components						
2-level variance (interviewers)	0.016(0.13)	**	0		0.131(0.20)	*
Residual variance	3.29		3.29		3.29	
Model fit						
AIC	378.7		412.73		302.17	
n	279		293		287	

Note **p<0.01, *p<0.05

Table F.4 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by subjective and objective voice characteristics for socially undesirable questions

	Q5 (Fired from a job)		Q21C (1+ alcohol drinks)		Q21D (Have sex 1+ times)	
	coefficient (SE)		coefficient (SE)		coefficient (SE)	
Intercept	-2.552(0.70)	**	-0.841(0.40)		-2.151(0.58)	
Pitch	0.001(0.006)		-0.006(0.006)		0.015(0.007)	*
Intonation	-0.007(0.009)		0.006(0.006)		-0.002(0.008)	
Speech rate	0.114(0.22)		-0.056(0.17)		0.366(0.32)	
Fillers	-0.114(0.87)		-0.160(0.31)		-0.732(0.50)	
Rated pitch	-0.497(0.40)		0.192(0.33)		-0.634(0.44)	
Rated intonation	-0.010(0.30)		0.178(0.22)		-0.193(0.30)	
Rate speech rate	0.841(0.41)	*	0.151(0.16)		0.260(0.24)	
Rate fillers	0.308(0.38)		0.097(0.19)		0.160(0.24)	
Interviewer's experience < 1 year	0.114(0.31)		0.130(0.26)		-0.503(0.57)	
Female interviewer	1.818(0.79)	*	0.035(0.55)		1.042(0.73)	
R education is high school or less	-0.063(0.29)		-0.327(0.23)		-0.601(0.29)	
R whose age is 60 or less	0.567(0.27)	*	0.362(0.21)		1.472(0.27)	
Interaction btw voice and Iwer						
Speech rate*female interviewers					-0.784(0.37)	*
Rated speech rate*female iwers	-0.990(0.50)	*				
Rated speech rate*inwer exp< 1 year					0.961(0.41)	*
Variance components						
2-level variance interviewers	0		0		0	
Residual variance	3.29		3.29		3.29	
Model fit						
AIC	408.36		557.17		403.06	
n	414		416		377	

Note **p<0.01, *p<0.05

Table F.5 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by subjective and objective voice characteristics for socially desirable questions

	Q8 (Did not Volunteer Activity)	Q13A (Does not completely Enjoy Reading)	Q21F (Number of Reading times<10)
	coefficient (SE)	coefficient (SE)	coefficient (SE)
Intercept	-0.577(0.45)	-1.269(0.44) *	0.205(0.46)
Pitch	0.005(0.006)	-0.011(0.006)	0.014(0.01) *
Intonation	0.0007(0.01)	-0.003(0.01)	-0.002(0.01)
Speech rate	-0.758(0.48) *	-0.178(0.32)	-0.145(0.21)
Speech rate^2		0.683(0.34) *	
Fillers	0.187(0.17)	0.062(0.21)	2.104(0.86) *
Fillers^2			-1.024(0.52) *
Rated pitch	-0.225(0.33)	0.023(0.34)	-0.074(0.39)
Rated intonation	-0.083(0.24)	0.572(0.23) *	-0.583(0.24)
Rate speech rate	0.323(0.26)	0.226(0.19)	0.197(0.20)
Rate fillers	-0.042(0.16)	0.385(0.21)	-0.017(0.24)
Interviewer's experience < 1 year	0.064(0.26)	0.404(0.27)	0.909(0.47)
Female interviewer	-0.238(0.56)	0.245(0.57)	-0.171(0.60)
R whose education is high school or less	1.067(0.23) **	0.666(0.23)	1.049(0.28) **
R whose age is 60 or less	-0.204(0.21)	0.950(0.22)	0.549(0.24) *
Rated speech rate*interviewer exp < 1 yr			-0.3806(0.38) *
Variance components			
2-level variance (interviewers)	0	0	0
Residual variance	3.29	3.29	3.29
Model fit			
AIC	558.14	561.18	480.74
n	408	417	413

Note **p<0.01, *p<0.05

		Mediation effect of sub voice characteristics	Note
Respondent behaviors	Interrupt questions with answers	No mediation	Effect of objective speech rate*undesirable Q on data quality becomes stronger
	Express uncertainty about a question Request clarification about a question	**Have mediation effect**	
	Give a qualified answer	No mediation	Effect of objective speech rate*inexp iwers on data quality becomes stronger
	Give a response that does not meet question objective	No mediation	Effect of objective speech rate*undesirable Q on data quality becomes stronger
		No mediation	Effect of objective pitch on data quality becomes stronger
Item nonresponse		No mediation	Effect of objective speech rate on data quality becomes stronger
Rounding	Q19	No mediation	Subjective voice characteristics do not affect data quality
	Q20	No mediation	Subjective voice characteristics do not affect data quality
	Q21A	No mediation	Subjective voice characteristics do not affect data quality
More is better	Q5	No mediation	Subjective voice characteristics do not affect data quality
	Q21C	No mediation	Subjective voice characteristics do not affect data quality
	Q21D	No mediation	Effect of objective pitch on data quality becomes stronger
Less is better	Q8	No mediation	Subjective voice characteristics do not affect data quality
	Q13A	No mediation	Effect of objective speech rate^2 on data quality becomes stronger
	Q21F	**Have mediation effect**	

Table F.6 A summary table examining the mediate effect of subjective voice characteristics

APPENDIX G: RESULTS OF HIRAROCHICAL LOGISTIC MODELS TO EXAMINE WHETHER INTERVIEWER PERSONALITY TRAITS MEDIATE RELATIONSHIPS OF OBJECTIVE VOICE CHARACTERISTICS ON DATA QUALITY FOR ROUNDED ANSWERS AND ANSWERS THAT ARE LESS PRONE TO SOCIALLY DESIRBLE BIAS

Table G.1 Hierarchical logistic model predicting proportion of rounded answer by objective voice characteristics and subjective ratings of interviewer's personality traits

	Q21A (Number of times using the internet)		Q19 (Number of minutes spending on a computer)		Q20 (Number of email messages)	
	coefficient (SE)		coefficient (SE)		coefficient (SE)	
Main effects						
Intercept	-0.588(0.35)		0.349(0.29)		1.537(0.39)	*
Pitch	0.0005(0.01)		-0.004(0.01)		0.007(0.01)	
Intonation	0.005(0.01)		0.0001(0.01)		-0.012(0.01)	
Speech rate	0.134(0.31)		-0.498(0.24)	*	0.257(0.32)	
Filler	-0.018(0.22)		0.015(0.32)		0.073(0.52)	
Credibility	0.140(0.12)		-0.106(0.12)		-0.022(0.15)	
Easiness to understand	0.311(0.49)		0.198(0.46)		0.542(0.58)	
Interviewer's experience < 1 year	0.370(0.35)		-0.199(0.30)		-0.142(0.45)	
Female interviewer	-0.479(0.47)		0.438(0.39)		-0.363(0.55)	
R whose education is high school or less	-0.424(0.31)		-0.672(0.29)	*	-1.286(0.33)	**
R whose age is 60 or less	1.170(0.26)	**	0.061(0.25)		0.821(0.32)	*
Speech rate*Female interviewers			0.671(0.35)			
Variance components						
2-level variance (interviewers)	0.04(0.14)		0		0.18(0.21)	
Residual variance	3.29		3.29		3.29	
Model fit						
Generalized Chi-square	274.96		293.66		267.23	
n	279		293		287	

Note **p<0.01, *p<0.05

Table G.2 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by objective voice characteristics and subjective ratings of interviewer's personality traits for socially undesirable questions

	Q5 (Fired from a job)		Q21C (1+ alcohol drinks)	
	coefficient (SE)		**coefficient (SE)**	
Main effects				
Intercept	-2.114(0.41)	**	-0.882(0.26)	**
Pitch	-0.001(0.01)		-0.006(0.01)	
Intonation	-0.011(0.01)		0.005(0.01)	
Speech rate	0.130(0.18)		-0.067(0.14)	
Filler	0.088(0.82)		-0.026(0.19)	
Credibility	0.008(0.12)		0.144(0.09)	
Easiness to understand	-0.325(0.47)		-0.263(0.33)	
Interviewer's experience < 1 year	0.153(0.31)		0.098(0.26)	
Female interviewer	0.298(0.42)		0.310(0.35)	
R whose education is high school or less	-0.062(0.29)		-0.333(0.23)	
R whose age is 60 or less	0.529(0.26)		0.344(0.21)	
Variance components				
2-level variance interviewers)	0		0	
Residual variance	3.29		3.29	
Model fit				
Generalized Chi-square	408.64		415.12	
n	413		416	

Note **p<0.01, *p<0.05

Table G.3 Hierarchical logistic model predicting proportion of answers that are less prone to socially desirable bias by objective voice characteristics and subjective ratings of interviewer's personality traits for socially desirable questions

	Q8 (Did not Volunteer Activity)		Q13A (Does not completely Enjoy Reading)		Q21F (Number of Reading times<10)	
	coefficient (SE)		coefficient (SE)		coefficient (SE)	
Main effects						
Intercept	-0.104(0.31)		-1.196(0.32)		0.356(0.30)	
Pitch	0.004(0.01)		-0.007(0.01)		0.012(0.01)	
Intonation	-0.002(0.01)		0.001(0.01)		-0.007(0.01)	
Speech rate	-0.437(0.36)		-0.288(0.24)		-0.103(0.19)	
Speech rate^2			0.760(0.34)	*		
Filler	0.178(0.16)		0.256(0.20)		2.105(0.86)	*
Filler^2					-1.080(0.52)	*
Credibility	-0.052(0.10)		0.094(0.09)		-0.041(0.10)	
Easiness to understand	0.258(0.43)		-0.151(0.35)		-0.097(0.38)	
Interviewer's experience < 1 year	-0.002(0.27)		0.160(0.26)		0.353(0.33)	
Female interviewer	-0.378(0.37)		0.076(0.37)		-0.313(0.41)	
R education is high school or less	1.055(0.23)	**	0.681(0.23)	**	1.011(0.28)	
R whose age is 60 or less	-0.175(0.21)		0.844(0.21)	**	0.561(0.24)	
Variance components						
2-level variance (interviewers)	0		0		0.09(0.13)	
Residual variance	3.29		3.29		3.29	
Model fit						
Generalized Chi-square	409.58		416.66		385.42	
n	408		417		413	

Note **p<0.01, *p<0.05