# Effects of Speech Rate, Pitch, and Pausing on Survey Participation Decisions

José Benkí[1], Jessica Broome[1], Frederick Conrad[1,2], Robert Groves[3], Frauke Kreuter[2]

[1]Survey Research Center, University of Michigan, 426 Thompson St., Ann Arbor, MI 48106-1248
[2]Joint Program in Survey Methodology, 1218 LeFrak Hall, University of Maryland, College Park, MD 20742
[3]US Census Bureau, 4600 Silver Hill Rd., Washington, DC 20233

## Abstract

When potential respondents consider whether or not to participate in a telephone interview, they have very little information about the interviewer, aside from what they hear over the phone. Yet interviewers vary widely in how often their invitations lead to participation, suggesting that potential respondents may give considerable weight not only to the content of such invitations, but the style, rhythm, phrasing, and other prosodic attributes of interviewers. We examine the impact of three prosodic attributes of interviewers: speech rate, pitch, and pausing, on the outcome of specific telephone survey invitations, *agree-to-participate*, *scheduled-callback*, and *refusal,* in a corpus of 1380 audio-recorded survey introductions (contacts). Agreement was highest when interviewers spoke at a moderate rate (3.5 words/sec) and paused at a moderate rate as well, at least once during the invitation but not more than about once every other conversational turn. The median interviewer pitch in successful contacts with both male and female interviewers was significantly lower than in refusals. However, variation in pitch functioned differently for each sex, with increased pitch variability more helpful for female interviewers but hurtful for male interviewers. We interpret the advantage of moderate speaking and pausing rates in this corpus as indicative of respondent preference for extemporaneous and competent deliveries, and dispreference for overly scripted deliveries.

**Key Words:** Interviewer, telephone, speech, prosody

## 1. Introduction

Telephone interviewers' success obtaining interviews is due, at least in part, to what they communicate about themselves, which takes place entirely over the phone. This necessarily includes their voices, the manner and content of their speech, and how they interact with potential respondents. Over the course of their careers, some interviewers are more and others less successful; this implies that differences in what they say and how they say it play an important role in the outcomes of their invitations to participate. Even in particular contacts, an interviewer's voice, speech and interaction surely affect an answerer's decision. (Note that we refer to "answerers" rather than "respondents" as the phone answerer may not necessarily be the selected respondent or may refuse to participate.)

This paper continues an investigation of speech behaviors of both interviewers and answerers in a corpus of 1380 telephone survey invitations (Conrad, Broome, Benkí, Groves, Kreuter, Vanette, & McClain, 2010; Conrad, Broome, Benkí, Groves, Kreuter, & Vanette, under review). In this report, we investigate how the prosody of interviewer speech--specifically speech rate, pitch, and pausing--affects participation decisions.

## 1.1 Speech Prosody and Previous Investigations of its Role in Survey Participation Decisions

Prosody in speech refers to those properties of an utterance that are coincident with but distinct from those speech properties that determine the utterance's phonemic or segmental content (Wagner & Watson, 2010). For example, the greeting "Hello" consists of a low intensity frication noise (the phoneme /h/), followed by three vocalic segments (/ɛlo/) characterized by regular laryngeal vibration filtered through the open oral and pharyngeal cavities. Simultaneously, any given instance of "Hello" has specific prosodic properties, which include rate, intensity, pitch pattern, and phrasing. These properties convey information regarding the physical and emotional state of the speaker as well as her pragmatic intent. Because variation in these simultaneous prosodic or suprasegmental properties does not change the phonemic content of the utterance (at least not in English), but is nevertheless salient to listeners, these properties have potential in explaining differences in interviewer response rates.

The fundamental frequency of laryngeal vibration (f0) is perceived by listeners as pitch. As a key prosodic property of speech, pitch has been investigated in a handful of reports in the survey participation literature, using both subjective ratings as well as objective computer-based acoustic measures. In a study of listener ratings of staged introductions by experienced female telephone interviewers, Oksenberg, Coleman, & Cannell (1986) report that the interviewers with lower historical refusal rates (i.e., more successful interviewers) were rated as having higher pitch, more variable pitch, and a faster speaking rate. An acoustic analysis of the same recordings, reported by Sharf & Lehman (1984), was consistent with the listener ratings of pitch and variability of pitch. However, in a subsequent study which included both new data and a reanalysis of the earlier data, Oksenberg & Cannell (1988) found no consistent relationship between acoustic measures of pitch and interviewer historical success rate. They did find that the high response-rate interviewer introductions were perceived as "speaking relatively rapidly, loudly, and with a standard American accent, and as sounding more confident and more competent (p.265)."

In a more recent study of staged introductions by female interviewers, Steinkopf, Bauer, and Best (2010) found non-linear relationships between historical success rate and acoustic measures of average pitch, variability of pitch and speaking rate. Success rate was highest for all three measures in the middle of their distributions, decreased toward both ends of the distributions of average pitch and speaking rate, and decreased toward the lower end (but not the upper end) of the distribution of variability of pitch.

At least two studies have investigated the relationship between success rate and prosodic properties in recordings of actual (not staged) survey introductions. Van der Vaart, Ongena, Hoogendoorn, & Dijkstra (2005) found listener ratings of fluency to be positively associated with interviewer historical success rates. However, no association was found between success rate and ratings of pitch or rate, or success rate and acoustic measures of pitch.

Finally, in our own study of survey introductions in a smaller corpus than the present report, interviewer historical success rate was found to be positively associated with both listener ratings of pitch as well as the median pitch from acoustic measurement (Groves, O'Hare, Gould-Smith, Benki, & Maher, 2008). No relationship was found between success rate and speaking rate as judged by listener ratings.

## 1.2 Current Research

The current study examines the impact of interviewers' voices, speech and interactions with phone answerers on answerers' decisions to participate (*Agree*), to refuse to participate (*Refuse*), or to defer the decision (*Scheduled Callback*). Using a rich dataset of 1380 audio-recorded telephone survey introductions, we analyzed the relationship between three prosodic properties of interviewer speech—rate, pitch, and pauses—and answerers' participation decisions. We briefly review some of the relevant speech literature on these properties and formulate hypotheses specific to survey participation.

### *1.2.1 Speech rate*

Faster speech rates have been generally found to enhance speakers' credibility and persuasiveness (Lee & Boster, 1992). However, for very high speech rates, both intelligibility and subsequent recall are negatively affected (Foulke & Sticht, 1969). We hypothesize first, that faster speech rates will produce higher levels of agreement to participate. Secondly, we hypothesize that the advantage for faster speech will not extend to the very fastest speech rates in the corpus.

### *1.2.2 Pitch*

If potential respondents consider the vocal attractiveness of the interviewer in their participation decision, then the long-term median or average pitch could predict some variance in participation rates, but in different ways depending on the sex of the interviewer. Pitch is the one consistent objective measure that is correlated with vocal attractiveness in males (Hughes et al., 2008), with lower pitched male voices rated as more attractive than higher pitched male voices by both male and female raters. Vocal attractiveness in females appears to be more complex, however, with mixed reports in the literature. Hughes et al. (2008) do not find a reliable relation between pitch and attractiveness in female voice, while Collins & Missing (2003) report that males judged higher pitched female voices to be more attractive than lower pitched female voices. For median interviewer pitch, therefore, we hypothesize that participation rates will be negatively correlated with pitch in males but positively correlated with pitch in females.

Speakers do actively control their pitch within their pitch range, conveying several types of information, including the pragmatic intent of the utterance, how potentially ambiguous sentences should be disambiguated, and the discourse status of elements along dimensions such as newness or importance (Wagner & Watson, 2010). Thus, speakers who display high variation in their pitch have a greater opportunity than more to be more effective in conveying this information, and are perceived as more enthusiastic and lively (Hincks, 2005), and potentially successful in gaining survey participation. We therefore hypothesize that higher variation in pitch will produce higher rates of agreement for interviewers of both sexes.

*1.2.3 Pauses*

Christenfeld (1995) looked at listeners' interpretations of both pauses and fillers, such as "um," "uh," and "er." These events often go unnoticed, particularly when a listener is focused on the speaker's content. While the presence of fillers do not appear to harm ratings of a speaker's eloquence, the presence of pauses does negatively affect ratings of a speaker's relaxation when raters are attending to content.

In our previous report of the present corpus (Conrad et al., 2010), we found that interviewer speech that contained a moderate amount of fillers was more successful in recruitment than both perfectly fluent speech (i.e., without fillers) and speech with many fillers. We speculate that potential respondents may disprefer perfectly fluent speech without fillers because it sounds overly scripted, and disprefer overly disfluent speech because of the negative impacts on speaker eloquence and relaxation.

For the present investigation of silent pauses, we hypothesize that interviewer speech with a moderate amount of pauses rates will be most successful. Interviewer speech with no pauses or excessive pauses will be less successful.

## 2. Data and Methods

The dataset used for this study consists of 1215 audio recorded survey introductions/invitations from five surveys conducted at the University of Michigan Survey Research Center: "Gujarati" (n=240), "National Study on Medical Decisions" (n=53), "Interests of the General Public" (n=336), "Mississippi Community Study" (n=20), and the "Survey of Consumer Attitudes" (n=566). Three of the studies sampled and recruited respondents from frames generated with Random Digit Dialing techniques which usually involved a within-household respondent selection process; two recruited respondents directly from a list sample.[1] A complete description of the dataset can be found in Conrad et al., (2010).

The data set had a multilevel structure. We conceive of *interviewers* as comprising the highest level (see Figure 1). One hundred different interviewers are represented in the corpus; while most interviewers worked primarily on a single study (survey), 27 worked on more than one study, so interviewers and studies are actually cross-classified. *Cases* – households or individuals sampled for a particular study – are nested within study but may be associated with multiple interviewers: if a case was contacted more than once, different interviewers might make the different *contacts*. Thus cases are nested within study and cross-classified with interviewers. A case consisted of one or more *contacts* – a contact is a call that reached a household member – so contacts are nested within cases. Each contact is comprised of conversational *turns* taken by the interviewer and answerer[2], e.g., the answer's "hello" is one turn followed by an interviewer's turn such as "I am Sally James from the University of Michigan calling about an important research study." Each turn is composed of one or more *move*s, i.e., smallest units of conversation with distinct purposes. In the first move of the example interviewer turn the interviewer

1 Institutional Review Boards at both Michigan State University and the University of Michigan approved analyses of these recorded invitations.

[2] Sometimes there is more than one answerer in a contact. One scenario might be that the initial answerer turns the phone over to the household member selected by the within-household respondent selection procedure.

identifies herself; in the second move she gives her affiliation; and in the third she describes the study. Thus moves are the most fine-grained level in the data set. In the current study we focus on the contact level and the levels it entails, i.e., turns and moves.

Eleven speech-language pathology students at Michigan State University transcribed the sampled, audio-recorded contacts from replicates 1 – 4 and 41 (available resources did not allow analyzing more than this). They transcribed the interactions at the turn level (except for household listing turns because these were not directly related to householders' participation decision) using a set of conventions to capture rising and falling intonation, elongated vowels, and overspeech; they entered the durations of pauses and used standard spellings for fillers (*um* and *uh*) and backchannels (*uh huh*).

## 3. Results

We analyzed speech in the corpus in order to test our hypotheses concerning rate, pitch and pauses. In this first presentation of results, we report univariate analyses without accounting for the complex structure of the dataset, including clustering by interviewers; thus, confidence intervals are likely to be underestimated in these preliminary results. The analyses of pitch and pausing only cover the first 13 interviewer turns (the average length of refusals) in order to avoid any biases due to the increased duration of successful contacts.

### 3.1 Speech Rate

We expected to see a positive relationship between interviewer speech rate and agreement, and that the size of the effect may decrease at high speech rates. To test this interviewer speech rate hypothesis, we computed the mean speech rate for the interviewer speech turns in words/second for each contact. We then assigned each contact to a speech rate quintile and examined the proportion of contacts resulting in agreement for each quintile. The relationship between filler rate and proportion agrees is depicted in Figure 1.

Agreement does indeed to be positively correlated with speech rate consistent with the first rate hypothesis, with agreement increasing from the lowest rate quintile to the 3.5 words/second rate quintile, which has the highest agreement rates. Agreement does not appear to increase beyond this range and may fall off.

### 3.2 Pitch

In our hypothesis regarding median interviewer pitch, we suggested that the median pitch would be negatively correlated agreement for male interviewers and positively correlated with agreement with female interviewers. To test this hypothesis, median pitch values were computed for each turn in Praat and averaged for each contact. We computed the mean pitch for contacts with agreement and refusals, and separated by interviewer sex. There were some interviewers who did not report sex and these contacts are excluded from the male and female groupings. The mean pitch values by interviewer sex are plotted in Figure 2. Successful contacts with male interviewers have a 14 Hz lower mean pitch (f0) than unsuccessful contacts, consistent with our hypothesis regarding pitch in male interviewers. For female interviewers, however, mean pitch in successful contacts are 7 Hz lower than unsuccessful contacts, opposite the prediction of greater attractiveness for higher pitch in females.

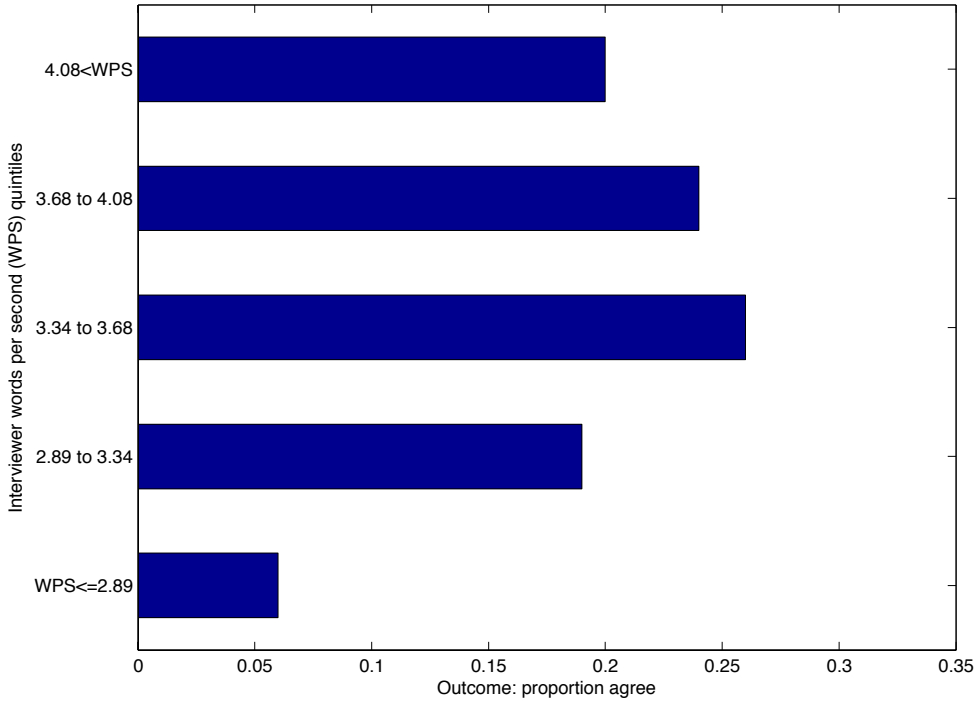


**Figure 1.** Proportion agree for each interviewer rate quintile (words/second).

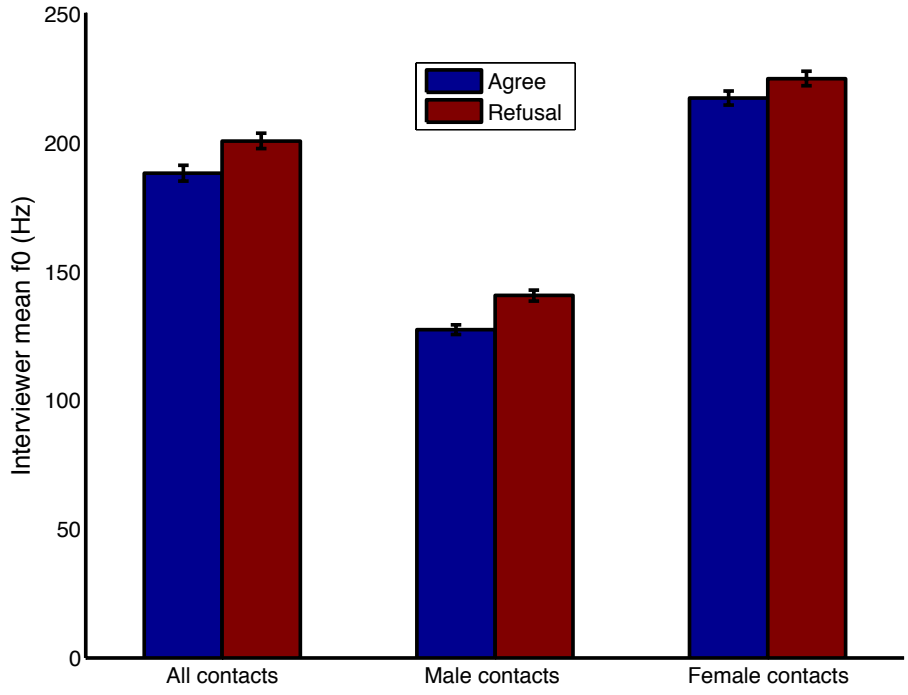


**Figure 2**. Mean pitch for all contacts, contacts with male interviewers, and contacts with female interviewers.

We also examined variation in interviewer pitch by computing F0SDQ, defined as the middle third of the pitch distribution in the pitch measurements for the interviewer speech in each contact. This measure is similar in scale to a standard deviation but since it is a quantile measure it is less sensitive to pitch measurement errors in the computer algorithm measuring pitch. We then pooled the contacts by outcome and sex and computed mean F0SDQ values as shown in Figure 3.

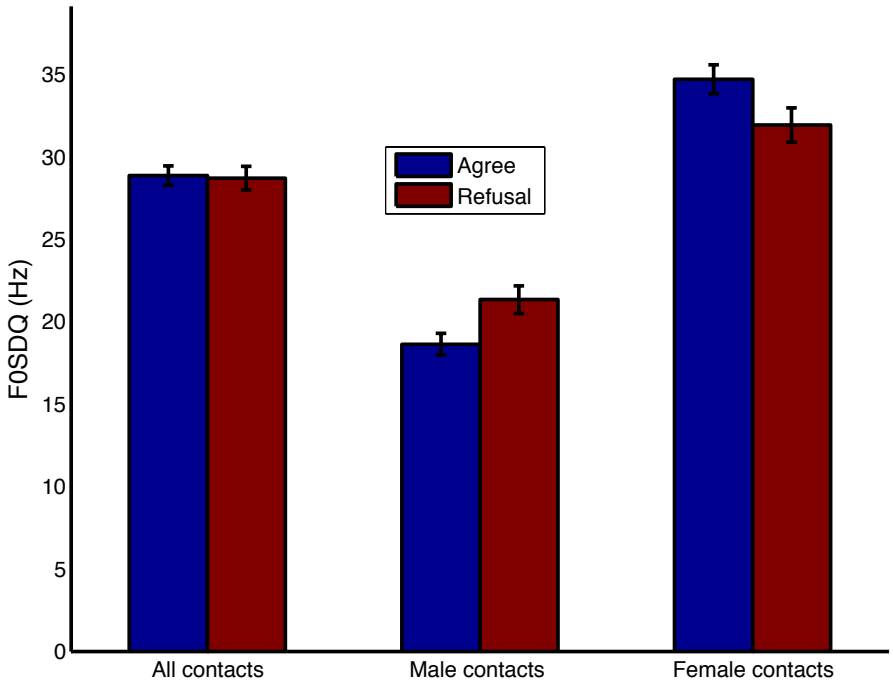


**Figure 3.** Variation in pitch by outcome and interviewer sex as measured by the middle third of the distribution of pitch values in each contact.

As shown in Figure 3, successful contacts with male interviewers had values of pitch variation that were 2.7 Hz lower than in unsuccessful contacts, contrary to the prediction that increased variation in pitch would lead to higher agreement rates. For female interviewers, successful contacts had 2.8 Hz higher values of pitch variation, consistent with our hypothesis of an advantage for increased variation in pitch, although the effect size was small. We speculate that the male interviewers who engaged in variation in pitch raised their overall pitch values significantly and in so doing activated negative stereotypes regarding male voices.

## 3.3 Pausing

Pauses were totalled for each contact. Approximately 40% of the contacts had no pauses at all in the first 13 turns. For the rest of the contacts had a median pause rate of 0.443 pauses/turn. Thus we divided contacts roughly into tertiles, with the first tertile containing no pauses, the second tertile with some pauses but less than or equal to 0.443

pauses/turn, and the upper tertile with pauses at a greater rate than 0.443 pauses/turn. Agreement rates for each tertile are plotted in Figure 4.

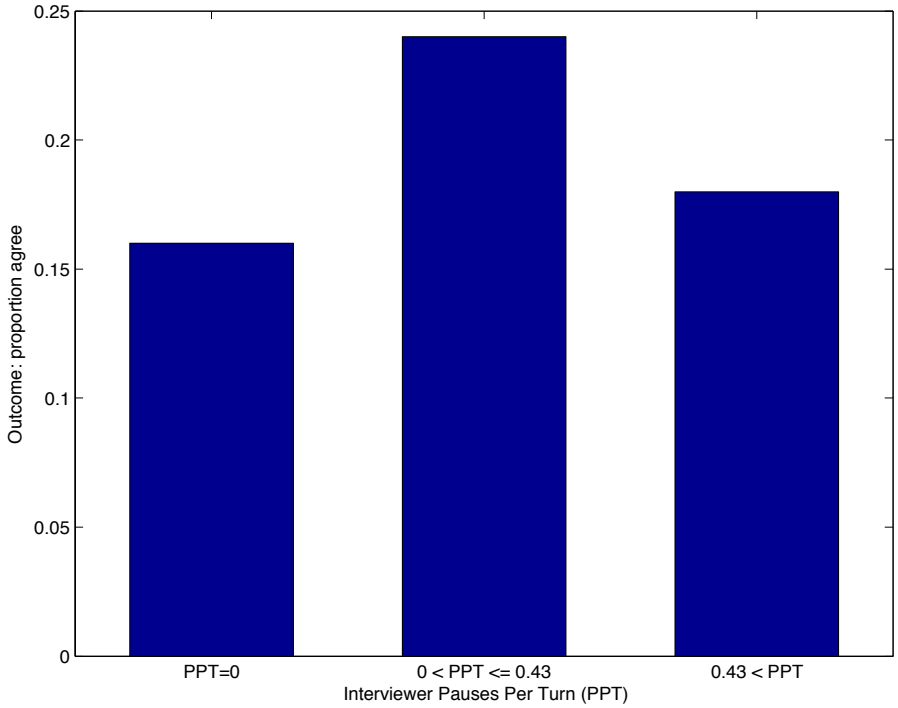


Figure 4. Agreement rate by pause rate tertile.

As shown in Figure 4, the highest agreement rate was obtained for contacts with a moderate amount of pausing, at least 1 pause but no more than approximately 1 pause every other turn. It appears that some pausing is helpful in conveying a less scripted delivery, and is not overly harmful, consistent with our hypothesis on a moderate amount of pausing. Greater pause rates than 0.443 pauses/turn do appear to be hurtful, however.

## 4. Conclusions

The current investigation makes it clear that the way telephone interviewers speak and interact when they invite household members to be interviewed is related to the success of a contact, at least in this corpus. More specifically, it is evident that interviewers are most successful when they speak at a moderate speech rate and are neither robotic nor highly disfluent in their pausing. Even the most disfluent interviewer speech seems to lead to more successful contacts than perfectly fluent speech. Potential respondents appear to be particularly sensitive to the use of pitch by male interviewers.

The current results encourage us that the approach we have used will continue to help us indentify other relationships between what interviewers say and how they say it on the one hand and answerers' participation decisions on the other. However, the multilevel nature of the data needs to be taken into account before final conclusions can be drawn. In addition, models that control for the clustering at the interviewer level will include interviewer covariates that are available in our data set.

The analyses we have reported identify several interviewer behaviors that seem related to more positive outcomes of contacts, but we do not compare particular interviewers. Examining behaviors of more successful interviewers would advance our enterprise by revealing variation across contacts: successful interviewers may apply techniques on different occasions, depending on their assessment of the answerer. Additionally, considering within-interviewer variance may have both statistical and substantive implications: clustering by interviewers may reflect variation of those interviewer behaviors responsible for differences in success. This cannot be observed in the contact level analyses we have done to date.

Future research will analyze subjective ratings, such as animation and coherence; and more detailed analysis of content, including interviewer use of apologies or references to monetary incentives. In addition, analyses at the level of interviewers may enable us to test the hypothesis that interviewers who converge to the vocal characteristics of answerers meet with greater success.

Finally, examining the lifecycle of individual cases across multiple contacts can reveal the interdependence of later on earlier contacts in determining the case's final outcome. Our focus on individual contacts is not sensitive to "historical" effects of this sort.

Although the current work begins to make clear some of the basic processes that operate in survey invitations, there are also practical lessons for survey operations. First, it may be that interviewers can be trained to engage in some of the behaviors that seem to be associated with more successful contacts: avoiding scripted delivery; speaking at a moderate rate and providing opportunities (e.g., pauses) for answers to signal their engagement; and interrupting judiciously. But there may be individual differences in interviewers' abilities to attend to both what they say and how they say it. Monitoring one's fluency may distract some interviewers from the content of their speech, and certainly monitoring paralinguistic aspects of answerers' speech may be hard for some interviewers to do while listening to what answerers say. Nonetheless, we believe our research program will help establish a tighter connection between research on interviews and survey practice.

## Acknowledgements

We are grateful to the following organizations for supporting the research reported here: National Science Foundation (Grant # SES-0819734 and SES-0819725); Survey Research Center, University of Michigan; Dept. of Communicative Sciences & Disorders, Michigan State University; Rensis Likert Fund for Research on Survey Methodology. We also thank the following people for advice and assistance, Pete Batra, Haley Gu, Patty Maher, Joe Matuzak, and Michael Schober. We are indebted to the transcribers/ acoustic analysts at Michigan State University and the coders/ raters at the University of Michigan: Rachel Benner, Kelly Franckowiak, Ben Jarvi, Emily Kordupel, Peter Kotvis, Abby Lincoln, Lacie Linstrom Melissa Littlefield, Daniela Lopez, Colleen McClain, Colleen McCarty, Gabe Moss, Kirsten Mull, Danny Nielsen, Dana Perkins, Fernando Pacheco, Danielle Popielarz, Christine Sheffler, Amanda Tatro, and Dylan Vollans.

## References


Christenfeld, N. (1995). Does it hurt to say um? *Journal of Nonverbal Behavior, 19*, 171-186.

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

Clark, H. & Schaefer, E.F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259-294.

Collins, SA and Missing, C (2004) Vocal and visual attractiveness are related in women. Animal Behavior 65: 997-1004.

Conrad, F.C., Broome, J., Benki, J., Groves, R., Kreuter, F., Vannette, D., and McClain, C. (Under revision). Interviewer speech and the success of survey invitations.

Conrad, F.C., Broome, J., Benki, J., Groves, R., Kreuter, F. and Vannette, D. (2010). "To agree or not to agree? Impact of interviewer speech on survey participation decisions." In *JSM Proceedings*, AAPOR-Section on Survey Research Methods, Alexandria, VA: American Statistical Association.

Foulke, E., Sticht, T., 1969. Review of research on the intelligibility and comprehension of accelerated speech. Psychological Bulletin 72, 50–62.

Groves, R.M., & Benkí, J.R. (2006). 300 hellos: acoustic properties of initial respondent greetings and response propensities in telephone surveys. Paper presented at the 17th International Workshop on Household Survey Nonresponse, Omaha, Nebraska.

Groves, RM, O'Hare, BC, Gould-Smith, D, Benki, J & Maher, P. (2007). Telephone interviewer voice characteristics and the survey participation decision. In Lepkowski, J.M, Tucker, C., Brick, J.M., de Leeuw, E.D., Japec, L., Lavrakas, P.J., Link, M.W., Sangster, R.L. (Eds.), *Advances in telephone survey methodology* (pp. 385-400). New York, NY: John Wiley and Sons.

Houtkoop-Steenstra, H. & van den Bergh, H. (2000). Effects of introductions in large-scale telephone survey interviews. *Sociological Methods and Research*, *28*, 281-300.

Hughes, SM, Pastizzo, MJ, and Gallup Jr., GG (2008): The sound of symmetry revisited: subjective and objective analyses of voice. Journal of Nonverbal Behavior 32: 93-108.

Oksenberg, L. & Cannell, C. (1988). Effects of interviewer vocal characteristics on nonresponse. In Groves, R.M., Biemer, P.B., Lyberg, L.E., Massey, J.T., Nichols II, W.L.,, and Waksberg, J. (Eds.), *Telephone survey methodology* (pp.257-269). New York, NY: John Wiley and Sons.

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

Sharf, D.J. & Lehman, M.E. (1984). Relationship between the speech characteristics and effectiveness of telephone interviewers. *Journal of Phonetics*, *12*, 219-228.

Steinkopf, L., Bauer, G., & Best, H. (2010). Nonresponse in CATI surveys. *Methods, Data, and Analysis, 4*, 3-26.

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*, 488-499.

Michael Wagner & Duane G. Watson (2010): Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25:7-9, 905-945