To Agree or Not to Agree?
Impact of interviewer speech on survey participation decisions

Frederick Conrad,[1,3] Jessica Broome,[1] Jose Benki,[1,2] Robert Groves,[4] Frauke Kreuter[3] and David Vannette[1]

1 University of Michigan
2 Michigan State University
3 University of Maryland
4 US Census Bureau

Abstract
When potential respondents consider whether or not to participate in a telephone interview, they have very little information about the interviewer, aside from how he/she sounds, speaks and interacts. Yet interviewers vary widely in how often their invitations lead to participation, suggesting that potential respondents may give considerable weight to interviewers' verbal attributes. We examine the impact of three interactional behaviors: disfluencies (*um, uh*), backchannels (*mm-hmm, I see*), and simultaneous speech, on the outcome of specific telephone survey invitations, *agree-to-participate*, *scheduled-callback*, and *refusal,* in a corpus of 1215 audio-recorded survey introductions (contacts). Agreement was highest when interviewers were moderately disfluent. Further, in contacts where "answerers" ultimately agreed to participate, they displayed more backchannels than when they refused. Finally, there was more simultaneous speech in contacts where answerers ultimately refused to participate, but interviewers interrupted answerers more when they scheduled a callback, perhaps reflecting their attempt to salvage the contact. We conclude by discussing next analytic steps, as well as practical implications for interviewer hiring and training.

Key Words: Interviewer, telephone, interaction

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.) The research reported here attempts to explore which verbal attributes of interviewers – primarily speech behaviors – affect participation decisions, as well as how verbal attributes of interviewers interact with those of potential respondents (answerers) in affecting outcome.

1.1 What is known about interviewers' verbal attributes and their success obtaining interviews?

There is a literature on interviewer voice and speech characteristics but its implications are mixed with respect to whether and how particular characteristics are associated with success obtaining interviews. Note that, success is operationalized in numerous ways, for example interviewers' historical response rates or a judge's determination of whether they or "someone" would be willing to participate in an interview with a particular interviewer. This definitional variation makes it hard to compare studies. Moreover, the literature is contradictory, with some studies finding that the presence of an attribute increases success while others finding the opposite relationship.

Some studies have found that success was associated with lower pitch (Oksenberg, Coleman & Cannell, 1986; Oksenberg & Cannell, 1988), while others found an association between success and higher pitch (Sharf & Lehman, 1984; Groves, O'Hare, Gould-Smith, Benki, & Maher 2008). Steinkopf, Bauer, and Best (2010) found a non-linear relationship between pitch and success. Similarly, less *variation* in pitch (Oksenberg, Coleman, & Cannell, 1986; Oksenberg & Cannell, 1988) and more variation in pitch (Sharf and Lehman, 1984) were associated with interviewer success. More recently, Steinkopf and colleagues found a non-linear relationship between variability in pitch and interviewer success, with improvement in success rate seen as variation in pitch increased up to a point, after which increased variation had no effect on success rate. While Oksenberg and Cannell (1988) found an association between falling intonation and low refusal rates (higher response rates), later work by Groves and colleagues (2008) found that rising intonation was associated with higher response rates. Finally, both greater fluency (Van der Vaart, Ongena, Hoogendoorn, & Dijkstra, 2005) and less scripted, i.e., less fluent delivery (Groves, O'Hare, Gould-Smith, Benki, & Maher, 2008) have both been associated with higher interviewer cooperation rates.

Other research has demonstrated an association between greater interviewer success and vocal loudness (Van der Vaart, Ongena, Hoogendoorn, & Dijkstra, 2005), a friendlier "hello" (Groves & Benki, 2006), and a voice which is breathier and less masculine (Groves, O'Hare, Gould-Smith, Benki, & Maher 2008).

In addition to the contradictory findings and variable outcome measures, previous studies vary greatly in the predictors used. Some studies use judges' ratings of, for example, interviewers' pitch or speech rate, while others use physical measurements and still others use both.

In much of the literature, the number of interviewers and the number of contacts are both small. In some studies students, rather than professional interviewers, are used. Many studies present recorded, i.e. staged, interviewer speech to listeners/judges, which makes it impossible to examine the interaction between interviewer and answerer, including interviewers' reactions to answerers.

Finally, published studies have given little or no attention to *how* interviewers speak (for example, measures of fluency or coherence) or *what* interviewers say, i.e., the actual content of their speech.

This study is unique in that it utilizes a large dataset where both answerer and interviewer speech are coded and the actual outcome of the contact is used as a measure of success.

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 1215 audio-recorded telephone survey introductions, we analyzed the relationship between three speech behaviors and answers' participation decisions; independent variables include paralinguistic elements (e.g., simultaneous speech or "overspeech," pauses, fillers, i.e., *uh, um*) and content (e.g., "persuasive" utterances, backchannels).

We explore the impact on contact outcome of three interactive speech phenomena: interviewer disfluencies, primarily fillers such as *um* and *uh*; answerer backchannels, such as *uh huh* or *okay,* which are typically used to signal engagement in a conversation; and overspeech (overlapping or simultaneous speech), with a particular focus on instances where interviewers interrupt answerers.

There is evidence that fillers such as "um" are characteristic of thoughtful, relaxed speakers. Christenfeld (1995) looked at listeners' interpretations of what he calls "filled pauses," such as "um," "uh," and "er." These utterances often go unnoticed, particularly when a listener is focused on the speaker's content, and their presence does not harm ratings of a speaker's eloquence. In addition, use of filled pauses tends to decrease listeners' perceptions of a speaker's anxiety. Christenfeld argues that "ums" indicate that a speaker is contemplating what to say next, or selecting the appropriate word. Clark and Fox Tree (2002) refer to "um" and "uh" as "fillers" and argue that these utterances are indicators that a speaker is anticipating a delay in his or her speech but does not want to give up the floor.

The published research on backchannels indicates that utterances such as "mm-hmm" or "right" indicate that the listener is engaged in the conversation, acknowledging the speaker (Clark & Schaeffer 1984), or "following the speaker's message (Duncan 1974, p. 179)." A dearth of backchannels is therefore interpreted as a lack of understanding or engagement.

Finally, the meaning of overspeech has also been explored within the study of turn-taking, as it represents an imperfect exchange of the floor. Conversational analytic studies have repeatedly demonstrated that speakers normally *take turns*, that is, one person speaks at a time, and the exchange of turns is clearly rule-governed (Saks, Schegloff & Jefferson 1974). Dunne and Ng (1994) observe that this model indicates good coordination between conversational partners; the occurrence of simultaneous speech actually indicates the breakdown of this turn-taking mechanism, that is, an exchange which is problematic (Leighton, Stollak, & Ferguson 1971; Duncan 1972).

Based on our reading of the literature on interactive speech phenomena, we derived three hypotheses that we will explore in the context of survey participation.

1. *Interviewer Disfluencies:* Intermediate levels of interviewer disfluency will produce the highest levels of agreement. Perfectly fluent interviewer speech will sound scripted and not tailored for a particular answerer, thus reducing agree rates while a highly disfluent interviewer will sound inept, also lowering agrees relative to intermediate levels.

2. *Respondent Backchannels:* Answerers will indicate that they are engaged and want the interviewer to continue by producing more backchannels; thus, answerers who ultimately agree will backchannel more than those who refuse.

3. *Overspeech:* When the answerer ultimately refuses there will be more overspeech, indicating interactions gone awry and generally difficult communication.

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]

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.

---

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

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 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.
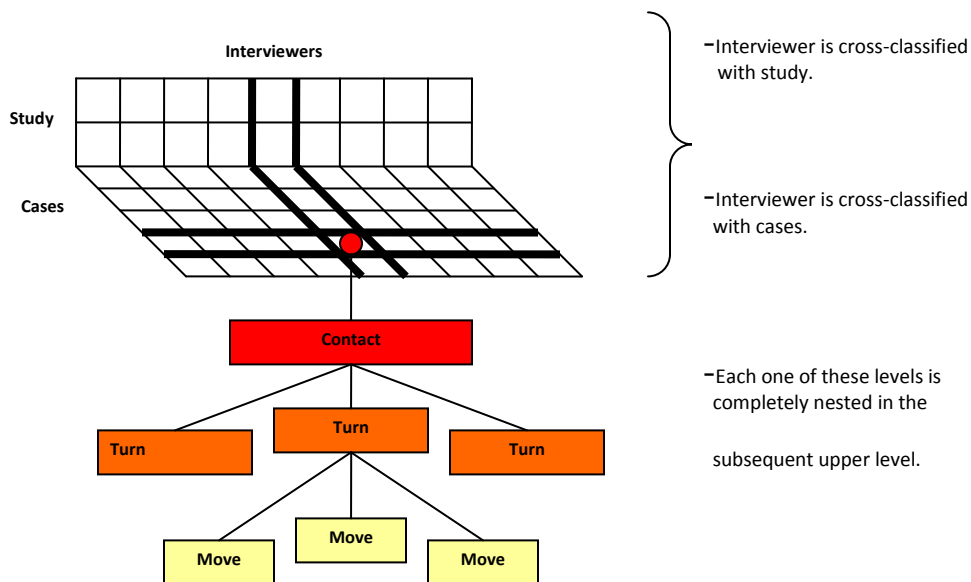
Figure 1: Structure of data set. Analyses reported here are all at the contact level; each contact consists of turns which themselves consists of moves. Although not explicitly taken into account in these analyses, cases can consist of multiple contacts (often by multiple interviewers), and interviewers can be associated with multiple cases as well as multiple studies.

2.1. Sampling Structure
For each study, a sample of households (cases) was selected. The sampling frame consisted of all contacts from the 5 selected studies (558,695 contacts in total). Cases were stratified first by interviewer and then by the outcome of the case's first contact. The convention used to associate one interviewer with a case was to assign the case to the interviewer who made the first contact. For each interviewer, up to eighty cases were sampled: up to forty cases where the first contact was positive (agree) and up to forty cases where the first contact was negative (some outcome other than agree, including refusal). If an interviewer had fewer than forty cases in either the positive or negative strata, all cases in that stratum for that interviewer were included. For each sampled case, all contacts in the case were selected; however, about 30% of contacts could not be

---

[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.

included due to recording failures, although this appeared to be random with respect to our sampling design.

A total of forty-one sample replicates were created. Each of the first forty replicates included all contacts from one positive and one negative case for each interviewer associated with that case, where, again, "positive" was defined as a positive first contact and "negative" as a negative first contact). This was done to assure that for each interviewer there was at least one case whose initial outcome was positive. The same interviewers were included in each of the forty replicates. An additional replicate was constructed for interviewers who had fewer than forty positive or negative cases. This replicate included all contacts from cases assigned to those interviewers. The forty-first (or "take-all") replicate was constructed to include some interviewers who were potentially less experienced than other interviewers. To assign cases to replicates, a systematic selection of cases with a single random start was used within each interviewer set. The sort order of the cases was survey, date of first contact, and time of first contact.

2.2. Transcription and Coding

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*). In the following example transcript, "A" indicates "Answerer" and "I" indicates "Interviewer":

A: Hello?
I: Hello, this is Jane Doe from the University of Michigan?
A: Yeah?
I: Um, you may have gotten a letter from us [P=.52] recently um [P=.48] about a chance for you or someone in your family to earn twenty dollars for participating in a research study about the economy.

During the transcription process, acoustic variables such as fundamental frequency, pause duration, turn duration, and formant frequencies of selected vowels were measured using Praat software (http://fonsg3.hum.uva.nl/praat/).

Following transcription, coding and rating[3] was conducted by seven undergraduate and masters level students in the humanities and social sciences at the University of Michigan. Prior to coding the interaction, coders decomposed each conversational turn into one or more moves. A list of 61 move descriptions was created, including "incentive related question" (answerer), "not enough time" (answerer)," "offer to break interview into pieces" (interviewer), "description of survey" (interviewer), and "backchannel" (interviewer or answerer). The following hypothetical contact illustrates the kinds of codes (*in parens)* that were assigned to each move (segment of the transcripts beginning with "A:" or "I:":

---

[3] Ratings were subjective, global judgments about the contact and were not used in the analyses reported here.

A: Hello? (*formal greeting)*
I: Hello, (*formal greeting*)
I: my name is Jane Doe and I'm calling from the University of Michigan. (*self-identification*)
A: Yeah? (*backchannel*)
I: Um, you may have gotten a letter from us recently (*follow-up statement*)
I: about a chance for you or someone in your family to earn twenty dollars for participating in a research study (*incentive related comment)*
I: about the economy. (*description of survey)*

In addition to classifying each move on the basis of its content (assigning it one of the 61 codes mentioned above), the coders judged the number of fillers, presence or absence of overspeech, and several other speech attributes (twenty-eight altogether). The result was a multivariate description of each move. Different sets of codes were used for interviewers and answerers. The coders entered their judgments in Sequence Viewer (SV) software (http://www.sequenceviewer.nl/), and derived additional contact level variables using SV.

2.3. Reliability of Coding

Two subsets of twenty contacts (audio recordings and transcripts) were randomly selected for a reliability test among the five primary coders. Because decomposing turns into moves is itself a judgment about which there can be disagreement, we tested two subsets of contacts to in order to measure agreement with and without the "turn decomposition" task. One subset of contacts was presented to the coders with transcripts in their original form, that is, not broken into moves. A second subset of twenty contacts for which the transcripts had already been broken into moves was coded by all five coders and tested for reliability. Each of the five coders decomposed the turns in the first subset of twenty contacts into moves and assigned codes to each move. Weighted kappa, as opposed to simple kappa, was used to measure reliability because with the multivariate coding system: the chances of perfect agreement on 28 variables is very remote; weighted kappa takes into account the complexity of the coding structure and essentially gives "partial credit." Weighted kappa for each pair of coders ranged from .53 to .93, mean = .74. For the second set of twenty contacts (already structured by moves) weighted kappas ranged from .77 to .86, mean = .82. We interpret these scores as indicating strong agreement among coders.
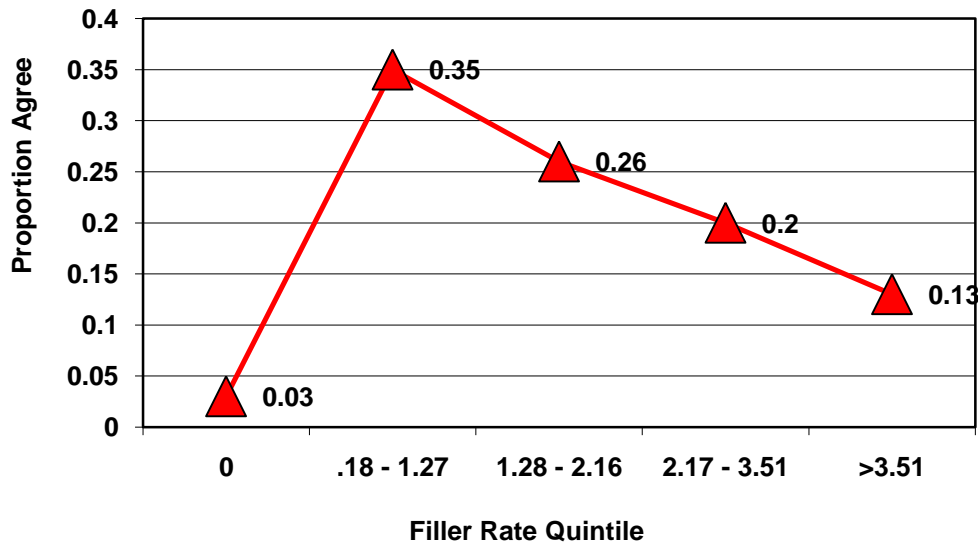
3. Results

We analyzed speech in the corpus in order to test our hypotheses concerning fillers, backchannels and overspeech. In this first presentation of results, we report test statistics 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.

3.1 Fillers

We expected to see a nonlinear relationship between the rate at which interviewers produce fillers (*uh*, *um*, *ah*, and *er*) and the proportion of contacts resulting in agreement in which highly fluent and highly disfluent interviewer speech should lead to lower agree rates than moderately disfluent interviewer speech. Again, the rationale is that if the interviewer is moderately disfluent, this should indicate to the answerer that the interviewer is planning her utterances as she speaks, and is not reading from a script

(which feels robotic) and should give the answerer the sense that the interviewer is competent (in contrast to a highly disfluent interviewer).

To test this interviewer disfluency hypothesis, we computed filler rate for each contact as the number of fillers per 100 words. We then assigned each contact to a filler 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 2. The range of filler rates for each quintile appears on the x-axis. Note that the filler rate for the first quintile is zero. Among contacts where interviewers produced no fillers (the first quintile), the proportion of contacts that result in agreement is extremely low (0.03). The proportion of agrees increases to its highest level (0.36) for the second quintile, where interviewers produce fillers at a rate of .18 to 1.27 per 100 words. The proportion of agrees declines from this point as the filler rate increases until interviewers are at their most disfluent, producing more than 3.51 fillers per 100 words (fifth quintile) and the proportion of agrees is only 0.13. We compared the proportion of contacts resulting in agreement between adjacent filler rate quintiles. All differences were reliable (all $p <$ 0.05) except for quintiles 3 and 4). While relatively few contacts result in agree decisions when interviewers are most disfluent (quintile 5), answerers are more likely to agree to be interviewed by interviewers with this level of disfluency than when interviewers are perfectly fluent. The difference in proportion agreement between the first and fifth quintiles is significant (t [511] = -4.11, $p <$ 0.01).This set of results conforms closely to our first hypothesis.

Figure 2: Contact outcome as a function of interviewer filler rate. Contacts were divided into quintiles according to interviewer filler rate (fillers/100 words). The proportion of contacts resulting in agreement (vs. refusal, scheduled callback, or other outcome) is plotted for each filler rate quintile.

It seems likely that in some of the contacts in which interviewers speak without any fillers they do not sound robotic but fluent and competent. This is particularly likely if they place vocal stress on appropriate syllables, indicating they are thinking about what they are saying much as a good actor animates his or her scripted speech and so communicates effectively without being disfluent. Distinguishing between perfectly fluent contacts that are robotic and others that are animated and effective is an

appropriate next step in understanding the relationship between interviewer disfluency and answerer willingness to be interviewed.

3.2 Backchannels

In our second hypothesis we suggested that the presence of answerer backchannels (e.g., "uh huh," "okay," or "I see") reflects answerer engagement and consequently increased likelihood of agreeing to be interviewed. Consistent with this hypothesis, answerers did produce significantly more backchannels when they ultimately agreed than when they refused. Among those contacts with an outcome of "agree," the mean answerer backchannel rate (calculated as number of answerer backchannels over number of answerer moves) was .19, compared to .10 for scheduled callbacks (t [693]=7.63, $p < 0.01$) and .04 for refusals (t [523]= 11.88, $p < 0.01$).

A reasonable question about this pattern of results is whether answerers who ultimately agree exhibit more backchannels from the beginning of the contact than answerers who refuse or whether there are "turning points" after which answerers seem to have made up their minds and their backchannels increase or decrease accordingly.

To explore this we identified two answerer behaviors (move categories) that on intuitive grounds seemed likely to play the role of turning point in answerer decision making, either toward (positive) or away from (negative) participating. In particular we identified expressions of willingness to participate (which occur in 540 contacts) as a positive turning point and explicit statements of non-interest in participating (which occur in 599 contacts) as a negative turning point. Certainly other answerer behaviors might indicate positive or negative shifts in their thinking but these particular moves seemed promising. We computed the proportion of contacts in which answerers produced backchannels before and the proportion of contacts in which answerers produced backchannels after each of the potential turning points. Note that neither of these statements by itself determines the outcome of the contact because it is possible for an answerer to express willingness and ultimately refuse or non-interest and ultimately agree to participate. The results are displayed in Figure 3.
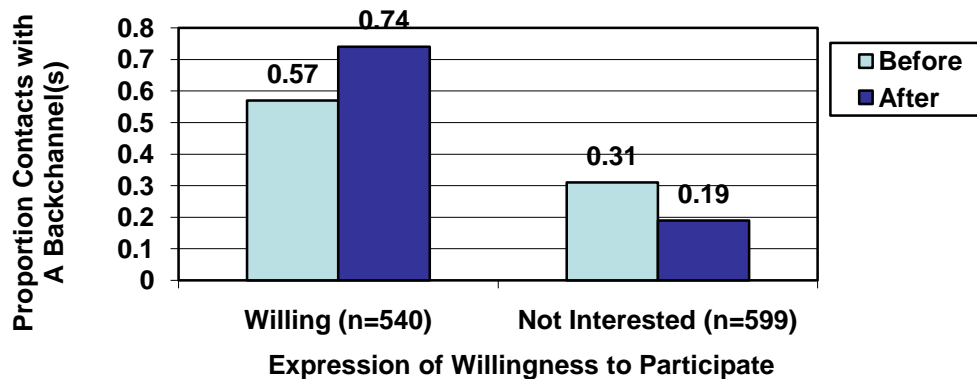
Figure 3: Occurrence of answerer backchannels by expression of willingness to participate. The left-hand pair of bars displays the proportion of contacts containing an answerer backchannel out of all contacts in which answerers express interest in participation. The right-hand pair displays the proportion of contacts containing an answerer backchannel out of all contacts in which answerers express disinterest in participation. In both pairs of bars, the light bar shows contacts in which the backchannel

occurs before the expression of (dis)interest, while the dark bar shows contacts in which the backchannel occurs afterwards

The first feature of the figure to note is that answerers display more backchannels overall (irrespective of position) in contacts where they express willingness (positive turning point) than in contacts where they indicate they are not interested (negative turning point). Second, in contacts where they express willingness to participate, answerers produce more backchannels after the turning point than before, while in contacts in which they indicate non-interest, answerers produce fewer backchannels after the turning point than before. To test these effects we segmented the contacts into (1) all moves before the turning point and (2) all moves after the turning point. We then conducted a logistic regression analysis ($g^2 = 213.60$, $p < .001$) to predict the presence of backchannels in these contact segments based on the type of turning point (positive or negative), the type of contact segment (before or after turning point), and their interaction. Consistent with the pattern in Figure 3, there is a main effect of type of turning point ($Z=11.66$, $p < .001$) indicating that there are more backchannels when there is a positive than negative turning point. There is also an interaction of type of turning point and type of contact segment ($Z= 4.3$, $p < .001$) indicating that there are more backchannels before than after a positive turning point but the opposite for negative turning points. There is also a main effect of type of contact segment ($Z= -4.87$, $p < .001$) indicating that there are generally more backchannels before than after the turning point, probably because there are fewer moves after turning points than before.

3.3 Overspeech
We suggested that contacts in which the communication is labored are at greater risk of refusal than contacts in which the communication is easier, and one indication of the ease or difficulty of the communication is the amount of overspeech: when there is more overspeech we expect lower agreement rates. Consistent with our hypothesis, there is more overspeech (simultaneous speech by the interviewer and answerer) when the answerer ultimately refuses than agrees. The mean proportion of moves containing overspeech is .24 in contacts where the ultimate outcome is a refusal, compared to .19 in contacts resulting in scheduled callbacks (t [762] = 5.29, $p < 0.05$) and .15 (t [523] = -7.42, $p < 0.05$) in contacts where the answerer ultimately agrees to participate.

Despite the generally negative impact of overspeech on participation decisions, interviewers seemed to strategically speak while answerers were speaking. The strategy seemed, in effect, to cut off answerers' attempts to refuse, leading to scheduled callbacks when refusal seemed all but certain. The proportion of interviewer moves that contain such interruptions (the answerer is speaking when the interviewer starts to speak) is .095 in scheduled callbacks, compared to .083 (t [761] = -1.57, $p < 0.10$) in refusals and .064 in agrees (t [693] = -4.30, $p < 0.05$). In the example that follows, overspeech is indicated by bold font enclosed in asterisks.

    I 6 : *Alright*.

     A 7 : *I do.* You know I don't think this is just going to be a good thing because you guys have tried a few times maybe you just might ought to try somebody else **and just take me off the list**.

> I 8 : *Well you know* we~ we would do that ma'am but um we can't do that because of the way the study- it's a scientific study set so once the numbers are- you know once we have a telephone number that household is the one we want to speak with. [breath] Um what would be a better time I certainly do~ do want to call you when it's more convenient for *you*.

> A 9 : *Maybe* you could try on Monday?/

While the answerer suggests the interviewer ought to "try someone else" and as she says "and just take me off the list" the interviewer commiserates and argues for participating on the grounds that the household has been specifically selected, finally asking when she might call back. The answerer, who was asking to be taken off the list when the interviewer started to speak, suggests the interviewer call back on another day by the end of the interviewer's turn.

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 are neither robotic nor highly disfluent. Even the most disfluent interviewer speech seems to lead to more successful contacts than perfectly fluent speech. Because backchannels are generally assumed to signal engagement in the conversation and invite the speaker to continue speaking, the absence of answerer backchannels may be a flag to interviewers that an answerer is headed for refusal. Backchannels seem to be more prevalent after answerers indicate willingness to participate and less frequent after they indicate they are not interested, suggesting that backchannels provide real-time evidence of the participation decision. Finally, high rates of overspeech indicate troubled interactions and are more prevalent in refusals than other outcomes. However, interviewers' strategic interruptions may salvage some contacts headed for refusal, converting them to scheduled callbacks and keeping alive the possibility of obtaining an interview.

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 acoustic attributes, such as fundamental frequency and speech rate; 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; detecting the absence of answerer backchannels 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 such as backchannels 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

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.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology, 23*, 283-292.

Duncan, S. (1974). On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, *3*, 161-180.

Dunne, M. & Ng, S.H. (1994). Simultaneous speech in small-group conversation: all-together-now and one-at-a-time? *Journal of Language and Social Psychology, 13*, 45-71.

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.

Leighton, L.A., Stollak, G.E., & Ferguson, L.R. (1971). Patterns of communication in normal and clinic families. *Journal of Consulting and Clinical Psychology, 36*, 252-256.

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

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.