



Interviewer speech and the success of survey invitations

Frederick G. Conrad,

University of Michigan, Ann Arbor, and University of Maryland, College Park, USA

Jessica S. Broome and José R. Benkí,

University of Michigan, Ann Arbor, USA

Frauke Kreuter,

University of Maryland, College Park, USA

Robert M. Groves,

US Census Bureau, Washington DC, USA

and David Vannette and Colleen McClain

University of Michigan, Ann Arbor, USA

[Received April 2011. Final revision May 2012]

Summary. When potential survey respondents decide whether or not to participate in a telephone interview, they may consider what it would be like to converse with the interviewer who is currently inviting them to respond, e.g. how he or she sounds, speaks and interacts. In the study that is reported here, we examine the effect of three interactional speech behaviours on the outcome of survey invitations: interviewer fillers (e.g. ‘*um*’ and ‘*uh*’), householders’ backchannels (e.g. ‘*uh huh*’ and ‘*I see*’) and simultaneous speech or ‘overspeech’ between interviewer and householder. We examine how these behaviours are related to householders’ decisions to participate (*agree*), to decline the invitation (*refusal*) or to defer the decision (*scheduled call-back*) in a corpus of 1380 audiorecorded survey invitations (contacts). Agreement was highest when interviewers were moderately disfluent—neither robotic nor so disfluent as to appear incompetent. Further, household members produced more backchannels, a behaviour which is often assumed to reflect a listener’s engagement, when they ultimately agreed to participate than when they refused. Finally, there was more simultaneous speech in contacts where householders ultimately refused to participate; however, interviewers interrupted household members more when they ultimately scheduled a call-back, seeming to pre-empt householders’ attempts to refuse. We discuss implications for hiring and training interviewers, as well as the development of automated speech interviewing systems.

Keywords: Interaction between interviewers and household members; Interviewers; Interviewer success; Survey recruitment; Telephone interviews

1. Introduction

Telephone interviewers’ success in obtaining interviews is due, at least in part, to what they communicate about themselves, which takes place entirely over the phone. This necessarily includes

Address for correspondence: Frederick G. Conrad, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI 48104, USA.
E-mail: fconrad@umich.edu

the sound of their voices, the manner and content of their speech and how they interact with potential respondents. Over the course of their careers, some telephone interviewers are more and others less successful; this implies that differences in what they say, how they say it and how it sounds all play important roles in the outcomes of their invitations to participate.

Even in particular contacts, an interviewer's voice, speech and interaction surely affect a householder's decision to participate or not. (We refer to 'householders' rather than 'respondents' because the phone answerer has not agreed—and may ultimately refuse—to participate or even if willing may not be selected from the household roster to be the respondent.) The research that is reported here explores how verbal attributes of interviewers—primarily speech behaviours—affect householders' participation decisions, as well as how interviewers' speech behaviours interact with those of householders in affecting the outcome of the invitation.

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

The literature on how interviewers' verbal attributes affect participation decisions is inconclusive but has at least identified several attributes that can affect the decision. The particular attributes that have been investigated are almost exclusively vocal, e.g. pitch and intonation contours, not speech characteristics (e.g. fluency) and interaction with household members (for example do they communicate attentiveness and engagement through what they say?), which are our focus here. Nonetheless, the research on interviewers' voices provides useful context for our study.

The role of vocal pitch, i.e. the perceptual correlate of laryngeal vibration of the human voice and more colloquially how high or low the voice sounds, has been investigated in survey recruitment. Oksenberg and Cannell (1988) found that success in recruitment is associated with lower pitch of the interviewer's voice, whereas other studies have found higher pitch more related to success (Sharf and Lehman, 1984; Oksenberg *et al.*, 1986; Groves *et al.*, 2008). Steinkopf *et al.* (2010) found a non-linear relationship between pitch and success. Similarly, the relationship between *variation* in pitch and interviewer success is difficult to interpret: more variation has been associated with increased interviewer success (Oksenberg *et al.*, 1986; Sharf and Lehman, 1984); Steinkopf and colleagues observed that success increased with variation in pitch up to a point, after which changes in pitch variation had no effect on success; Oksenberg and Cannell (1988) found no relationship between pitch variation and success. Intonation patterns have also produced inconsistent results: Oksenberg and Cannell (1988) found an association between falling intonation and higher response rates, whereas later work by Groves and colleagues (Groves *et al.*, 2008) found that a rising intonation was associated with higher response rates. Finally, both greater fluency (van der Vaart *et al.*, 2006) and less fluent (less scripted) delivery (Groves *et al.*, 2008) have both been associated with higher interviewer co-operation rates.

This area of study is prone to contradictory results because it lacks consistency in how key variables are operationalized and how they are measured. For example, *success* in obtaining interviews has been operationalized as interviewers' historical response rates as well as a judge's determination of whether they or 'someone' would be willing to participate in an interview with a particular interviewer. In some studies, the pitch or rate of speech is based on judges' ratings of interviewers' speech, whereas in other studies these are based on physical measurements and in others both are used.

In much of the literature, the number of interviewers is small; for example, Sharf and Lehman (1984) collected pitch ratings of six interviewers and in the study in Oksenberg *et al.* (1986) the number of interviewers was also 6. In addition, the number of telephone contacts is often relatively small; for example Huefken and Schaefer (2003) examined 219 contacts, and Groves

and Benki (2006) examined 300. In some studies students, rather than professionals, serve as interviewers, e.g. Huefken and Schaefer (2003), Steinkopf *et al.* (2010) and van der Vaart *et al.* (2005). And many studies ask listeners or judges to rate or make a hypothetical participation decision about staged and prerecorded invitations (e.g. Steinkopf *et al.* (2010)). This approach facilitates precise sampling and control of particular speech attributes but makes it impossible to examine the *interaction* between interviewer and householder, including interviewers' reactions to householders. In addition, staged invitations make it impossible to assess the influence of speech on actual participation decisions.

Here we report analyses of a much larger data set than has been previously used to study the verbal and vocal origins of success in survey invitations. We analyse telephone contacts in which professional interviewers in several surveys call sample members in their households. The speech of both the interviewer and the householder in each contact is transcribed and coded in detail and the actual outcome of the contact is used as a measure of success. Thus we can examine how the interviewer's speech and interaction with householders affect the likelihood that the householder agrees to be interviewed. Because our data set is derived from contacts in actual surveys, we have not experimentally manipulated independent variables, potentially limiting our ability to conclude that variation in these speech behaviours *causes* particular outcomes. An experimental approach might make causal attribution more straightforward but might also lack realism: asking interviewers to vary their speech parametrically runs the risk of creating unnatural speech and, as a result, unnatural interaction. Our study was observational by design, examining a large number of actual telephone contacts in which those with successful and unsuccessful outcomes were systematically sampled from several different studies.

2. Speech and interviewer success

The current study examines the influence of telephone interviewers' speech on household members' decisions to participate in a survey interview. Householders can decide to participate (*agree*), not to participate (*refuse*) or to defer the decision (*scheduled call-back*). We analyse the relationship between three interactive speech phenomena and participation decisions. In particular we investigate interviewers' disfluencies, primarily fillers (e.g. 'uh' and 'um'), householders' backchannels (e.g. 'uh huh' and 'okay') and 'overspeech' in which both parties speak simultaneously. In some cases we also examine the content of what has been said (e.g. householders' statements that they do not have enough time) and acoustic attributes such as pitch variation to help to interpret the use of particular speech behaviours.

We have explored these particular speech behaviours (fillers, backchannels and overspeech) because there is considerable agreement in the literature about how they operate in everyday conversation. This allowed us to derive relatively straightforward predictions about how they might affect the outcome of survey invitations. Other interviewer speech behaviours are also likely to have regular and predictable effects on the outcome of survey invitations, as might interviewers' voices. We do not examine voice properties in detail here but have done so elsewhere (Benki *et al.*, 2011).

2.1. Fillers

Clark and Fox Tree (2002) argued that speakers produce fillers to indicate that they anticipate a delay in their speech, perhaps because of planning or word finding difficulty; by producing fillers, speakers may also communicate their desire to keep the floor while they are not directly contributing to the discourse. The evidence for this view comes from Clark and Fox Tree's

analyses of spontaneous speech in a corpus of transcribed dialogues. In the corpus, speakers are most likely to produce fillers at the start of ‘intonation units’, which are stretches of speech under a single intonation contour that are believed to require complete planning before they are uttered. Immediately after producing these fillers, speakers paused more often and for longer than later in the utterance when presumably the demands of planning were reduced (Clark and Fox Tree (2002), pages 93–94). Smith and Clark (1993) observed a similar pattern of filler use: participants in their study produced more fillers when answering questions on topics about which they felt relatively uninformed and these fillers were followed by frequent long pauses relative to the shorter pauses produced in filler-free contexts. Smith and Clark interpreted fillers as speakers’ signals that they anticipate pausing while preparing what they will ultimately say.

In the context of survey invitations, it seems plausible that interviewer fillers seem to indicate planning activities that are similar to those observed in these psycholinguistic studies. If so, a household member may interpret interviewer fillers as evidence that the interviewer is not reciting a scripted invitation but is instead planning an utterance in realtime as part of a survey invitation designed specifically for him or her. We propose that interviewer fillers may reflect a version of ‘tailoring’ behaviour (Groves and Couper, 1998), in which skilled interviewers purposefully adapt their survey invitations to individual householders or at least give the impression of such tailoring.

Christenfeld (1995) presented evidence that fillers in actual speech can positively affect listeners’ impression of speakers, despite general negative impressions of hypothetical speakers who produce fillers. When asked directly in a questionnaire about their impressions of speakers who produce fillers, respondents overwhelmingly indicated negative impressions of such speakers. These negative impressions were quite inconsistent with how the respondents rated actual recorded speech that was presented in one of three versions. One version included fillers; a second version had been edited so that the fillers were replaced with silence (pauses) of equal duration; in the third version, the fillers were simply excised (no pauses). Listeners rated the speech that included fillers as more eloquent and produced by a more relaxed speaker than when it included pauses; no pauses elicited the highest eloquence ratings but the same level of relaxation as elicited by fillers. So, compared with pauses, fillers seemed largely to improve listeners’ impressions of the speaker. These patterns held whether listeners were instructed to attend to *content* or the *style* of the speech, suggesting they do not need to notice fillers (as when they attended to content) to be affected by them.

2.2. Backchannels

Almost all dialogue is punctuated by utterances such as ‘uh huh’, ‘mmm’, ‘yes’, ‘I see’ and ‘okay’. These have been labelled ‘backchannels’ (e.g. Duncan and Fiske (1977), Duncan (1973, 1974) and Goodwin (1981), ‘acknowledgements’ (Clark and Schaeffer, 1989), ‘uh huh, yeah and the like’ (Schegloff, 1982) and ‘ways listeners signal their attentiveness to speakers’ (Wardhaugh, 1985) among others. By almost all accounts, backchannels indicate that the listener is paying attention and understanding the speaker’s message, possibly agreeing with the message; because backchannels are mostly free of content, they indicate that the listener is foregoing an opportunity to take the floor and so implicitly endorses the speakers’ continued speech. It would follow that listeners who do not wish the conversation to continue will produce few if any backchannels. Backchannels may be particularly informative in phone conversations because comparable visual cues of engagement such as nods and smiles are not available (e.g. Brunner (1979)).

In survey invitations, householders’ backchannels presumably operate in much the same way: their production indicates that the householder is attending to the interviewer and consents to

the interviewer's continuing to speak but, if backchannels are rare, this may reflect householders' wishes to discourage continued conversation which could indicate that refusal, or at least a scheduled call-back, is imminent. We shall treat the number of backchannels produced by the householder as an indication of his or her engagement, with greater engagement increasing the likelihood of agreement and lower engagement increasing the chances of refusal, with scheduled call-backs in between.

2.3. Overspeech

A defining feature of conversation is that participants take turns. Speakers become listeners and listeners become speakers, and such turn taking usually occurs smoothly and in an orderly way. A simple set of turn taking rules appears in Sacks *et al.* (1974) and the proposal is still widely accepted. The key ideas are that

- (a) the speaker 'projects' the completion of the current turn largely through prosodic and syntactic cues so that the listener can anticipate when he or she might have an opportunity to speak and that
- (b) at these turn relevant junctures the speaker may select the listener to speak (often by asking a question of the listener), the speaker may self-select, i.e. continue to speak, possibly encouraged by a backchannel, or the listener might self-select, i.e. begin speaking when possible. Sacks *et al.* (1974) noted that, in ordinary conversation, turn taking is finely co-ordinated: there are very few gaps between turns and overlaps at turn relevant junctures are rare. Turn taking would seem to be smoother and to involve less overlapping speech (which we call overspeech) in face-to-face than in telephone conversations because visual cues are available in the former; for example the current speaker can look at a partner to select him or her as the next speaker (Duncan and Fiske, 1977), but in fact turn transitions are smoother on the phone (Hopper (1992), page 126).

We propose that a telephone conversation with large amounts of overspeech involves relatively troubled communication: the turn taking system to which co-operative conversationalists voluntarily adhere is not working. Evidence in support of this view includes the observation that overspeech is more frequent in the interaction of families that are referred for clinical treatment than control families (Leighton *et al.*, 1971) and for children who are classified as 'language impaired' rather than 'normally developing' (Fujiki *et al.*, 1990). One possibility is that householders who do not want to be interviewed feel no obligation to follow turn taking conventions that promote effective and potentially prolonged communication.

2.4. Hypotheses

On the basis of the body of literature about interactive speech phenomena, we derived three hypotheses about these behaviours in survey participation.

- (a) *Interviewer fillers*: intermediate levels of interviewer fillers will produce the highest levels of agreement. Perfectly fluent interviewer speech (no fillers) will sound scripted and not tailored for a particular householder, thus reducing agreement rates, whereas a highly disfluent interviewer (many fillers) will sound inept, also lowering agreements relative to intermediate levels of disfluency.
- (b) *Respondent backchannels*: householders will indicate that they are engaged and want the interviewer to continue by producing more backchannels; thus, householders who ultimately agree will produce more backchannels than those who refuse.

- (c) *Overspeech*: in contacts in which householders refuse there will be more overspeech, indicating interactions gone awry and generally difficult communication.

3. Data

The data were constructed by coding and acoustically measuring speech in a corpus of 1380 audiorecorded survey invitations delivered by interviewers in five surveys conducted at the University of Michigan Survey Research Center: the ‘Gujarati community survey’ (GCS), the ‘National study on medical decisions’ (NSMD), the ‘Interests of the general public’ (IGP) survey, the ‘Mississippi community study’ (MCS) and the ‘Survey of consumer attitudes’ (SCA). In three of the studies, households were sampled and participants recruited from frames that were generated with random-digit dialling techniques; recruitment in these studies usually involved a within-household respondent selection process. In two studies, respondents were directly selected and recruited from a list sample, one (the GCS) used a list of households and one (the IGP survey) used a list of named sample members. Institutional review boards at both Michigan State University and the University of Michigan approved analyses of these recorded invitations.

The data set consisted of 1380 records (one for each contact) and had a multilevel structure (Fig. 1): interviewer, study, case, contact, turn and move. 100 different interviewers were represented in the corpus; 27 worked on more than one of the five studies, 27 cases—unique households or individuals—were sampled in total from across the five studies. (We think of study as a level because cases are entirely nested within it but we do not treat it hierarchically in our models; it is simply a covariate). Cases were contacted between one and 13 times (median, 3); between one and 11 interviewers contacted each case (median number of interviewers per case, 2.0); individual interviewers contacted between one and 46 cases (median, between 10 and 11). Each contact is comprised of conversational turns taken by the interviewer and householder; for example, the householder’s ‘hello’ is one turn followed by an interviewer’s turn such as ‘I am

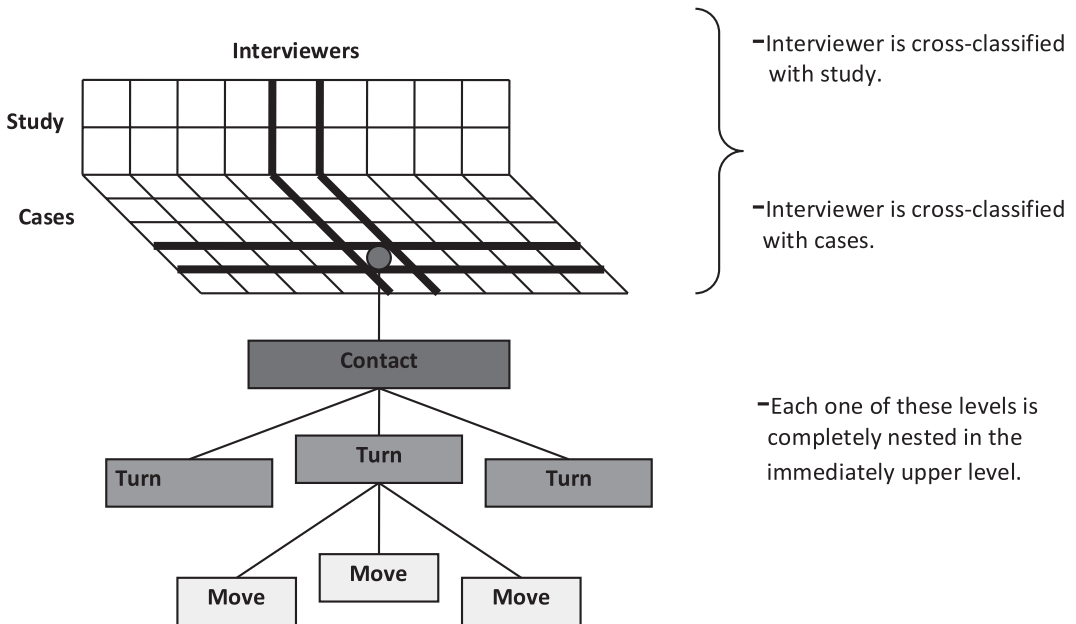


Fig. 1. Multilevel structure of the data set

Sally James from the University of Michigan calling about an important research study'. Each turn is composed of one or more moves, i.e. the smallest units of conversation with distinct purposes. In the first move of this example interviewer turn, the interviewer identifies herself, in the second move she gives her affiliation and in the third she describes the study.

Our primary interest is in the outcome of contacts (agree, refuse or scheduled call-back) based on how interviewers speak and interact with household members, captured at the move level. It is possible that the relationship between the speech and outcome of contacts differs across the five surveys (studies) but because some of the studies contributed only a few contacts to the corpus (20 and 56 contacts from the two smallest contributors) we control for the effects of study in the models that are reported below. We do not explore interviewer level effects in detail, in part because some interviewers made only a small number of contacts (range, 1–76 contacts; median, 4) and some worked on more than one study, leading to a relatively complex cross-classification.

The data set contained relatively little information about the household or householder beyond the number of calls leading up to and including the contact. Ideally, we would have had access to information about the householder's chronic tendencies to participate or not (response propensity) such as prior survey participation, attitudes towards science and attitudes towards government (e.g. Groves *et al.* (2000)). A household member's response propensity may overwhelm the effects of interviewers' speech. For example, someone who simply does not participate in surveys may be unaffected by how fluently an interviewer speaks. To the extent that we observe relationships between the speech variables and participation in the absence of any householder information, this would underscore the importance of speech variables in the decision.

3.1. Sampling structure

A sample of households or individuals (cases) was selected from the five studies (558 695 contacts in total). Cases were clustered within the interviewer who made the first contact in each case and then stratified by the outcome of the first contact. For each of the 100 interviewers, up to 40 cases were sampled for which the first contact resulted in agreement and up to 40 for which first contact resulted in a negative outcome (something other than agreement including refusal). This was done to assure that for each interviewer there was at least one case whose initial outcome was positive; this guaranteed that we had at least one successful contact for each interviewer in the corpus. For each sampled case, all contacts in the case were selected. Note that the outcome of any one contact can be independent of the outcome for the overall case. For example, it is possible that, even though a contact ends in 'refuse', an interview may be obtained from the household in a later contact.

Every attempt was made to select all contacts for sampled cases, but about 30% of contacts could not be transcribed owing to random recording failures. Further, a small number of contacts were sampled but, after review, were not included in the analyses primarily because of eligibility problems (e.g. children had been selected), householder language barriers or non-English contacts.

3.2. Agreement rates in the five surveys

The percentages of contacts in the corpus for which the outcome was 'agree' ranged from 11% to 41% across the five surveys (studies); $F(1, 1379) = 7.37$, $p < 0.01$ (see the fourth column of Table 1). Note that the percentage for the IGP study is considerably higher than for the other studies. It is possible that this is because the IGP is the only study for which householders are

Table 1. Agreement rates within the corpus and response rate in the surveys

<i>Study</i>	<i>Number of contacts</i>	<i>Number of agrees</i>	<i>Agree rate in corpus (%)</i>	<i>Response rate[†] in survey (%)</i>
1, GCS	240	26	11	55
2, IGP	56	23	41	34
3, MCS	20	5	25	43
4, NSMD	358	68	19	53
5, SCA	706	141	20	41

[†]American Association for Public Opinion Research second response rate.

specifically recruited by name: this might reduce the amount of persuasion that is required by interviewers, possibly making recruitment easier than in the ‘cold call’ which is typical of the other studies. We cannot know for sure whether this difference in design is responsible for the different agreement rates because a single interviewer made 29 of the 56 IGP contacts so, even if the conversational interaction differed from what is observed in the other studies, it might simply be due to idiosyncrasies of this one interviewer and not the fact that interviewers ask for the intended respondents by name. Because the response rate in the actual survey (the rightmost column in Table 1), as opposed to the corpus, is actually lower than in the other studies, it seems unlikely that the list frame used in the IGP led to fundamentally different interactions from those in the studies with other frames. Whatever the explanation, the number of IGP contacts is small so they cannot have a large effect on any of our analyses.

3.3. Transcription and coding

11 undergraduate and Masters level speech–language pathology students at Michigan State University transcribed the sampled, audiorecorded contacts. They transcribed the interactions at the turn level (excluding household listing turns, which were not directly related to householders’ *decisions* to participate or not) by 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, ‘H’ indicates the householder and ‘I’ indicates the interviewer.

H: ‘Hello?’

I: ‘Hello, this is Jane Doe from the University of Michigan.’

H: ‘Okay?’

I: ‘Um, you may have gotten a letter from us [$P = 0.52$] recently um [$P = 0.48$] 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 and pause duration (‘ P ’ followed by the duration in seconds, above) were measured by using version 5 of the Praat program (Boersma and Weenik, 2007).

Following transcription, seven undergraduate and Masters level students in the humanities and social sciences at the University of Michigan coded the content and paralinguistic behaviour in each transcript. Before 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’ (householder), ‘not enough time’ (householder), ‘offer to break interview into pieces’ (interviewer), ‘description of survey’ (interviewer) and ‘backchannel’ (interviewer

or householder); the codes are available from the authors. The following hypothetical contact illustrates the kinds of codes (*in parentheses*) that were assigned to each move (the segment of the transcripts beginning with ‘H:’ or ‘I:’).

H: ‘Hello?’ (*formal greeting*).

I: ‘Hello’ (*formal greeting*).

I: ‘My name is Jane Doe and I’m calling from the University of Michigan’ (*self-identification*).

H: ‘Okay?’ (*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 that were mentioned above), the coders judged the number of fillers, the presence or absence of overspeech and several other speech attributes (28 altogether). The result was a multivariate description of each move. Different sets of codes were used for interviewers and householders. The coders entered their judgements in the Sequence Viewer software version 5 (<http://www.sequenceviewer.nl/>) and derived additional contact level variables by using Sequence Viewer.

3.4. Reliability of coding

Two subsets of 20 contacts (audio recordings and transcripts) were randomly selected for a reliability test among the five coders who coded more than 90% of the contacts. Because decomposing turns into moves is itself a judgement about which there can be disagreement, we tested two subsets of contacts 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, i.e. not broken into moves. A second subset of 20 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 20 contacts into moves and assigned codes to each move. Weighted κ -coefficients for each pair of coders ranged from 0.53 to 0.93 (mean, 0.74). For the second set of 20 contacts (already structured by moves) the weighted κ -coefficients ranged from 0.77 to 0.86 (mean, 0.82). We interpret these scores as indicating strong agreement between coders.

3.5. Ratings

In addition to coding small segments of speech (moves) within a contact, we obtained global ratings of interviewer and householder speech on several attributes for each contact. We include ratings of two speech attributes in our analyses: the degree to which the interviewer spoke with an accent (regional, ethnic or foreign), and how similar the interviewer and householder accents were. We viewed these two judgements as more likely to be reliable than rating particular accents or dialects, and they helped to address the general questions of whether interviewers who spoke with an accent were less likely to succeed than interviewers who spoke standard, non-accented English, and whether contacts in which the interviewer and householder spoke with the same accent—including no accent—were more likely to end with the householder agreeing to be interviewed.

The contacts were divided between two judges who rated them independently: one judge rated

70% of the contacts and the other rated 30%. They both rated a subset of 25 contacts and we computed κ for each of the ratings of this subset: κ for interviewer accent, 0.72, $p = 0.000$; κ for accent similarity, 0.36, $p < 0.013$. We include both interviewer accent and accent similarity in the models that are reported below (i.e. the main effect on the likelihood of ‘agree’) and, because it is possible that they might affect the relationship between the speech behaviours (fillers, backchannels and overspeech) and outcome, we also tested interaction terms involving these. Because the raters rarely used some levels of accent similarity we dichotomized the measure for subsequent analyses: the lower two levels of the original variable formed one level of the new variable and the three higher levels formed the other.

4. Analytic methods

Almost all of the analyses that we report examine the relationship between a particular speech behaviour in the telephone contact and the outcome of the contact. We primarily model this relationship by using a cross-classified multilevel logistic regression model for a binary dependent variable, in which the outcome is predicted on the basis of a particular speech behaviour. (The models were computed with `xtmelogit` in Stata release 12 (StataCorp, 2011).)

More specifically, we denote by $y_{i(jk)}$ the outcome of a contact i made by interviewer j to a household k , where the parentheses indicate the cross-classification of interviewers with households. The dependent variable is coded as

$$y_{i(jk)} = \begin{cases} 1 & \text{agree,} \\ 0 & \text{refuse or scheduled call-back.} \end{cases}$$

The cross-classified multilevel logistic model for the outcome of contacts (taking ‘refuse and scheduled call-back’ as the reference category) can be written as

$$\log\left(\frac{\pi_{i(jk)}}{1 - \pi_{i(jk)}}\right) = \beta^T x_{i(jk)} + u_j + v_k \tag{1}$$

where $\pi_{i(jk)}$ denotes the probability of agreement, i.e. $\pi_{i(jk)} = \Pr(y_{i(jk)} = 1)$, $x_{i(jk)}$ is a vector of contact and interviewer covariates, β is a vector of coefficients and u_j and v_k are random effects, representing unobserved interviewer and household effects respectively. The random effects are assumed to follow normal distributions, i.e. $u_j \sim N(0, \sigma_u^2)$ and $v_k \sim N(0, \sigma_v^2)$. The variance parameters σ_u^2 and σ_v^2 are respectively the residual between-interviewer and between-household variances in the log-odds of agreement *versus* refusal or scheduled call-back.

A separate model conforming to equation (1) was fitted for each of the three speech behaviours (filler rate, backchannel rate and overspeech) measured for a contact. In addition, the models included several interviewer level covariates: interviewer experience (more or less than 90 days at the time of contact), interviewer accent and similarity of interviewer and household accent. In the model building step interaction terms were included between interviewer covariates and the speech behaviours—the main question was whether the effect of the speech behaviour on participation was moderated by interviewer attributes—but only significant interactions are included in the models that are presented here. All main effects, whether significant or not, are included in the final models because each was theoretically motivated. A categorical variable for study is included to test differences in agreement rate between the studies (the Gujarati study is the reference category). The fit was tested for all models with Wald χ^2 and was significant at $p < 0.001$ for all the models that we report, indicating that they fit the data well.

For one analysis we used a linear instead of logistic link function to predict a continuous

dependent variable (the proportion of moves containing a householder backchannel) on the basis of the same covariates and random effect as used in logistic regression models.

Another approach would have been to model all three outcomes discretely (instead of pooling scheduled call-backs and refusals) by using multinomial logistic regression. This might have allowed for more nuanced comparisons because surely scheduled call-backs and refusals are qualitatively different from each other. We adopted the current approach to simplify the interpretation of results.

5. Results

5.1. Fillers

Interviewers' use of fillers in our corpus clearly affected householders' likelihood of participating. This relationship is evident in Fig. 2, which displays the proportion of agree decisions as a function of filler rate (fillers per 100 words) divided into quintiles. Slightly more than 20% of the contacts involve zero interviewer fillers so the lowest quintile is somewhat larger than the others. Of those contacts that were free of interviewer fillers, only about 3% result in agree decisions but a small increase in filler rate—up to 1.28 fillers per 100 words—produces a sharp increase in agreement to about 36%. However, as the filler rate increases beyond this level, agreement drops continuously until at the highest filler rates it is about 12%.

To account for this pattern we fit a logistic regression model predicting the log-odds of agreement based on the filler rate (the number of fillers per 100 words in the contact) and a quadratic term, filler rate squared. By including the quadratic term we can test the change in slope from positive to negative as the filler rate increases. The parameter estimates are presented in Table 2. The key finding is that the filler rate is positively and significantly related to the likelihood of agreement and filler rate squared is negatively and significantly related to the likelihood of agreement, reflecting the curvilinear pattern that is apparent in Fig. 2.

Although it is clear that moderate levels of filler use are associated with a higher likelihood of survey participation, a key piece of the story is which extreme—frequent use or no use of fillers—is associated with lower participation. The agreement rate for zero filler contacts was 0.03 and for the top quintile (3.49 fillers or more per 100) the agreement rate was 0.12. To test the

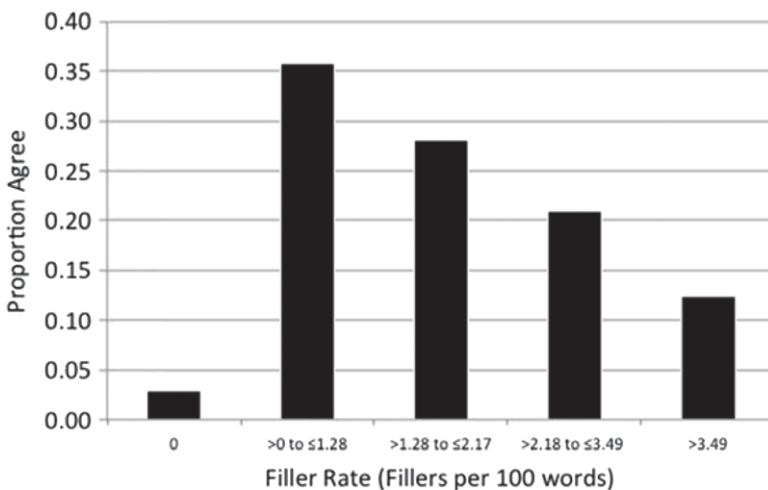


Fig. 2. Percentage agreement for each interviewer filler rate quintile

Table 2. Probability of agreement (*versus* refused and scheduled call-back) based on the filler rate and filler rate squared[†]

	<i>Coefficient</i>	<i>Standard error</i>	<i>Z</i>	<i>P > z </i>
Filler rate	0.89	0.16	5.74	0.001
Filler rate ²	-0.20	0.03	-5.97	0.001
Interviewer experience	0.38	0.23	1.64	0.101
Interviewer's accent	-0.00	0.10	-0.04	0.969
Accent similarity	-0.11	0.16	-0.68	0.496
Study 2	1.87	0.50	3.76	0.000
Study 3	0.64	0.67	0.95	0.340
Study 4	0.64	0.32	2.04	0.042
Study 5	0.51	0.31	1.61	0.107
Constant	-3.18	0.48	-6.62	0.000

[†]Household random effect: estimate of household standard deviation, 0.522; standard error, 0.233; confidence interval, 0.217–1.252. Interviewer random effect: estimate of interviewer standard deviation, 0.738; standard error, 0.146; confidence interval, 0.176–0.807.

difference, we estimated a cross-classified multilevel logistic regression model that predicted the likelihood of agreement on the basis of the same interviewer and study covariates as included in the previous analysis, and filler quintiles as an additional covariate. The difference between the first and fifth quintile is significant ($z = 3.40$, $p < 0.001$) suggesting that perfectly fluent speech may be more harmful to participation than highly disfluent speech.

We have suggested that the low rate of agreement when the survey invitation is free of fillers occurs because the interviewer sounds more like a machine programmed to deliver scripted text than a person talking to a person. If so, then other attributes may co-occur with this highly fluent speech contributing to its machine-like character. One such attribute is a monotonic delivery, characterized by relatively little variation in the pitch or fundamental frequency, often abbreviated F0 (pronounced 'F-zero'). Acoustically, F0 is the lowest frequency of the periodic vibration of a speaker's vocal cords; listeners experience F0 as vocal pitch. If monotonous interviewer speech co-occurs with highly fluent speech, we should see a reduced variation in F0 in the first filler quintile than in the other quintiles. Alternatively, interviewer speech that is perfectly fluent may be accompanied by *more* variation in pitch than observed at the other levels of filler use, e.g. the irregular intonation pattern when spoken computer-generated digits are inserted into automated speech dialogue as in telephone reservation systems. The main point is that zero-filler speech may seem machine like because, in addition to lacking evidence of planning (no fillers), its pitch contours are atypical. (See Hincks (2005) for a related discussion of pitch variation.)

To explore this possibility in our corpus, F0 was computed for all interviewer speech at intervals of 0.01 (10 ms) by using the auto-correlation algorithm (Boersma, 1993) in Praat. Pitch variation was measured for each interviewer turn by computing the range of the middle third of the distribution of measured F0-values. The pitch variation measure for each contact was the median of its turn level pitch variation values. Pitch variation was highest in the first quintile (zero fillers) and decreased almost monotonically across the filler quintiles: 31.9; 30.5; 26.5; 28.7; 26.6. We compared F0 at the first quintile and each of the other quintiles in a logistic regression model (which is not shown) that included the same interviewer and study covariates as in the previous analyses. The F0-measure was reliably lower for all filler quintiles than the first ($p < 0.001$ in all cases). We propose that this pattern reflects a 'singsong' delivery in the first

filler quintile in which interviewers stressed (increased pitch) on more syllables and different syllables than appropriate in ordinary conversation. The combination of filler-free speech and relatively high variation in pitch may have sounded particularly unnatural and linguistically odd, which may in turn have discouraged participation.

5.2. Backchannels

In our second hypothesis we suggested that householders produce more backchannels (e.g. ‘uh huh’, ‘okay’ or ‘I see’) when they are more engaged and thus more likely to agree to be interviewed. The data are consistent with this hypothesis. Householders produced more backchannels when they ultimately agreed to participate than when they refused or scheduled a call-back. When householders ultimately agreed, 0.14 of their moves were backchannels but when they refused only 0.04 of their moves were backchannels ($t[594] = 11.26, p < 0.001$), and when they scheduled a call-back 0.09 of their moves were backchannels (comparison with agreements: $t[798] = 5.37, p < 0.001$). Finally, householders produced more backchannels in scheduled call-backs than in refusals ($t[868] = -7.04, p < 0.001$). We modelled the likelihood of agreement *versus* refusal and scheduled call-backs on the basis of the proportion of moves that were backchannels. The main effect of the proportion of backchannels was significant (Table 3), reflecting the higher proportion of householder backchannels in agreements than contacts with other outcomes. The interaction of interviewer accent and backchannels is also significant, indicating that backchannels are particularly predictive of agreement when interviewers speak with an accent. This could be because accents are generally harmful to participation (the coefficient for the main effect of interviewer accent is negative although the effect is not significant) so when a respondent is engaged—indicated by more backchannelling—despite accented interviewer speech, he or she is particularly likely to participate.

One question that is raised by these results is whether householders who ultimately agree to participate exhibit more backchannels from the beginning of the contact than those who refuse or whether their backchannels increase or decrease after they have made up their minds one way or the other. In other words does frequent backchannelling reflect a general predisposition to respond or does it reflect the result of a decision to respond? To explore this we identified

Table 3. Probability of agreement (*versus* refuse and scheduled call-backs) on the basis of the proportion of householder backchannels†

	Coefficient	Standard error	Z	$P > z $
Proportion backchannels	6.01	1.25	4.82	0.000
Interviewer experience	0.58	0.26	2.23	0.026
Interviewer accent	-0.23	0.18	-1.27	0.203
Accent similarity	-0.02	0.18	-0.13	0.893
Study 2	1.55	0.54	2.87	0.004
Study 3	1.16	0.76	1.54	0.124
Study 4	1.09	0.37	2.97	0.003
Study 5	0.95	0.36	2.66	0.008
Backchannel × interviewer accent	1.89	0.88	2.15	0.031
Constant	-3.81	0.47	-8.14	0.000

†Household random effect: estimate of household standard deviation, 0.624; standard error, 0.263; confidence interval, 0.273–1.427. Interviewer random effect: estimate of interviewer standard deviation, 0.417; standard error, 0.149; confidence interval, 0.207–0.841.

Table 4. Proportion of householder backchannels by turning point statement and position relative to turning point statements

Turning point statement	Position of backchannels		Mean proportion householder backchannels
	Before	After	
Willing	0.183	0.293	0.199
Not interested	0.095	0.045	0.068
Mean proportion of backchannels	0.10	0.13	0.11

two behaviours that seemed to be plausible turning points in the conversation: one towards and the other away from participation. In particular we treated the 310 expressions of willingness to participate (e.g. ‘I fully intend to co-operate’ or ‘All right, let’s do it’) as indicating that the householder was leaning towards agreeing and the 361 statements of non-interest (e.g. ‘I really don’t like to do these things’ or ‘I’m not interested, thank you’) as indicating unlikely participation. Note that expressions of willingness do not guarantee that householders ultimately agree to participate and expressions of non-interest do not necessarily portend refusals or scheduled call-backs. We computed the proportion of moves that were householder backchannels before and after the two turning point statements (Table 4).

As is evident in Table 4, householders display more backchannels after than before a *willingness* statement but fewer after a *not interested* statement than before. To test this pattern, we split the contacts into the collection of moves before and after the turning point statement; we refer to these as ‘splits’. We then predicted the proportion of moves in each split that were householder backchannels, using ordinary least squares regression (implemented in `xtmixed` in Stata 11 (StataCorp, 2009), on the basis of the type of turning point statement, the position of the split relative to the statement and the interaction of the statement and position. The results are presented in Table 5. The main effect of position with respect to the turning point statement was significant, reflecting the large increase in householder backchannels after willingness statements. The interaction between statement and position was also significant, reflecting the increase after willingness and the decrease after not interested turning point statements.

We interpret these analyses as evidence that householders produce backchannels in survey invitations to authorize the interviewer to keep speaking, much as they do in ordinary conversation with ordinary speakers; they do this more in contacts that they wish to extend than to truncate. By extending a contact, householders are at least keeping open the option of agreeing. The ebbs and flows in backchannel production seem to reflect changes in the state of the householder’s decision making, increasing and decreasing as their thinking tilts towards agreement or refusal respectively.

5.3. Overspeech

We suggested that contacts in which the communication is laboured are at greater risk of refusal than contacts in which the communication is easier, and one indication of the ease or difficulty of communication is the amount of simultaneous speech or overspeech: greater amounts of overspeech almost surely reflect interaction in which both parties struggle to express themselves as turn taking is breaking down. Under these conditions we expect lower rates of agreement. To test this we computed the proportion of moves containing overspeech, excluding backchannels,

Table 5. Probability of backchannel based on type of turning point statement, position of move relative to the statement and interaction of statement and position†

	<i>Coefficient</i>	<i>Standard error</i>	<i>Z</i>	<i>P > z </i>
Turning point statement	0.06	0.04	1.51	0.130
Position	0.27	0.04	7.05	0.000
Statement × position	-0.16	0.02	-6.80	0.000
Interviewer accent	-0.01	0.01	-0.89	0.375
Accent similarity	-0.00	0.01	-0.03	0.977
Interviewer experience	-0.02	0.02	-1.11	0.266
Study 2	0.05	0.04	1.27	0.202
Study 3	-0.04	0.06	-0.69	0.489
Study 4	-0.01	0.02	-0.33	0.741
Study 5	-0.04	0.02	-1.56	0.120
Constant	0.09	0.07	1.28	0.200

†Household random effect: estimate of household standard deviation, 0.049; standard error, 0.016; confidence interval, 0.026–0.093. Interviewer random effect: estimate of interviewer standard deviation, 0.029; standard error, 0.011; confidence interval, 0.013–0.063.

Table 6. Probability of refusals (*versus* agree and scheduled call-backs) on the basis of overspeech†

	<i>Coefficient</i>	<i>Standard error</i>	<i>Z</i>	<i>P > z </i>
Proportion overspeech	-3.28	0.659	-4.97	0.000
Interviewer experience	0.460	0.238	-4.97	0.000
Interviewer accent	-0.021	0.101	-0.21	0.833
Accent similarity	-0.046	0.162	-0.28	0.776
Study 2	1.75	0.510	3.43	0.000
Study 3	1.05	0.749	1.41	0.159
Study 4	0.871	0.338	2.58	0.010
Study 5	0.701	0.330	2.12	0.034
Constant	-2.142	0.350	-6.11	0.000

†Household random effect: estimate of household standard deviation, 0.800; standard error, 0.185; confidence interval, 0.209–0.789. Interviewer random effect: estimate of interviewer standard deviation, 0.406; standard error, 0.138; confidence interval, 0.209–0.789.

in each contact. (We removed backchannels from all overspeech analyses on the grounds that authorizing the speaker to continue, as backchannels are typically assumed to do, is fundamentally different from introducing content while there is on-going speech by another. The patterns of results are identical whether backchannels are included or not.) Consistent with our hypothesis, the mean proportion of moves containing overspeech in refusals is 0.24, compared with 0.17 for contacts resulting in scheduled call-backs and 0.12 in contacts resulting in agreement. We modelled the likelihood of agreeing (*versus* refusing and scheduling a call-back) on the basis of the proportion of moves that involved overspeech and the standard covariates (Table 6). The main effect of overspeech is highly significant: as overspeech increases the chances of refusal increase.

Despite the generally negative influence of overspeech on participation decisions, interviewers seemed to interrupt householders strategically. The strategy essentially cuts off householders'

attempts to refuse, leading to scheduled call-backs when refusal seemed all but certain. We defined interviewer interruptions as overspeech in a householder move, i.e. interviewer speech whose onset occurred when the householder was already speaking. We computed the proportion of moves containing interviewer interruptions by dividing the total number of these events by the total number of opportunities to interrupt, i.e. householder turns. Consistent with the idea that interviewers interrupt to head-off contacts headed for refusal, the proportion of interviewer interruptions is 0.142 in scheduled call-backs, compared with 0.126 in refusals and 0.104 in agreements. In a logistic regression model (which is not shown) that predicts scheduled call-backs (*versus* refusals and agreements), the probability of a scheduled call-back is greater as interviewer interruptions increase (coefficient 1.39; $z = 3.19$, $p = 0.001$); the model includes our standard covariates.

In the example that follows, the interviewer interrupts while the householder seems to be in the act of refusing (overspeech is enclosed in asterisks). More specifically, just after the householder says, in turn H7, ‘this just isn’t going to be a good thing’ and as she is requesting that the interviewer take her ‘off the list’, the interviewer points out (turn I8) that sampled phone numbers cannot legitimately be eliminated from the study, and asks when it would be better to call. The householder, who seemed on the verge of refusing in her previous turn, then (turn H9) suggests that the interviewer calls back on another day.

I6: ‘*Alright*.’

H7: ‘*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*.’

I8: ‘*Well you know* we [pause] we would do that ma’am but um we can’t do that because of the way the study—it’s a scientific study 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 [pause] do want to call you when it’s more convenient for *you*.’

H9: ‘*Maybe* you could try on Monday?’

The usual assumption is that speakers who are interrupted experience the interruption as negative, e.g. reminding them that they are less powerful than the interrupter (Smith-Loven and Brody, 1989; Zimmerman and West, 1975). However, this is not always so. Some interruptions are helpful; for example they complete the turn that was started by the original speaker, or are ‘supportive’, indicating agreement or acceptance (Dunne and Ng, 1994; Smith-Loven and Brody, 1989). In scheduled call-backs, the interviewer interruptions may have been as successful as they were at pre-empting refusal because they were relatively supportive. Interviewers most often interrupted householders in scheduled call-backs as the householders were stating that they did not have enough time to participate. In the 275 scheduled call-backs in which householders expressed lack of time as a reason for not participating, interviewers interrupted during the ‘not enough time’ moves 372 times (some contacts involved more than one such statement and more than one interruption). The most frequent interviewer move in such interruptions was an attempt to find a time that would better fit householders’ schedules.

6. Discussion and conclusions

The current investigation makes it clear that the way that telephone interviewers speak and interact when they invite household members to be interviewed is related to the success of those invitations, at least in our corpus. We examined three speech behaviours: interviewer fillers,

householder backchannels and overspeech; each was related to the outcome of the telephone invitations in regular and predictable ways.

For fillers, it is evident that interviewers are most successful when they are neither perfectly fluent nor highly disfluent. One reason that perfectly fluent speech led to the lowest levels of agreement is that this kind of speech is difficult to process. Fox Tree (2001) demonstrated that listeners recognized target words faster when preceded by 'uh' than when the 'uhs' had been removed. Clark and Fox Tree (2002) argued that fillers help listeners by alerting them to the complexity of upcoming speech, enabling them to allocate attention accordingly. Thus the absence of fillers may place listeners at a disadvantage. If we overlay already impoverished (i.e. filler-free) speech with potentially distracting pitch contours (highly variable F0), householders may reason that continued interaction is more difficult than they are willing to endure.

In addition, the absence of fillers is likely to be interpreted by householders as symptomatic of speech that is not being planned 'live' and is instead being read from a text or script. When interviewers produce fillers, even at a moderate rate, they provide the impression that they are individually tailoring their delivery for the householder during the survey invitation itself. This type of tailoring is similar in spirit, if not form, to the tailoring that was identified by Groves and Couper (1998): both types of tailoring serve to convince the householder that the survey invitation is designed for her personally, thereby increasing the appeal of the invitation. Whether householders in the present corpus are responding to the fillers themselves or to the customization that simultaneously occurs cannot be determined on the basis of the current data.

We also observed that householders produced more backchannels when they ultimately agreed to participate than when they did not. And the rate at which they produced backchannels seemed to change as their decision became firm, increasing after statements of willingness and decreasing after statements of non-interest. These patterns suggest that backchannels follow and reflect householders' decision making, but it may be possible for interviewers to reverse the causality, increasing householders' willingness to be interviewed by increasing their backchannels. For example, if interviewers pause longer at turn transitions than they might have spontaneously, some householders may acknowledge that they are listening and that they wish for the interviewers to continue speaking although they might not have done so during a shorter pause. The question is whether the act of producing backchannels, even when deliberately elicited by the interviewer, can increase householders' engagement, ultimately leading to agreement when refusal would otherwise have been more likely.

Finally, high rates of overspeech were more prevalent in refusals than other outcomes, presumably indicating troubled interaction. However, interviewers' strategic interruptions may salvage some contacts that were otherwise headed for refusal, converting them to scheduled call-backs and keeping alive the possibility of obtaining an interview. We suggested that these interruptions successfully pre-empted refusals, in part, because they addressed householders' reasons for non-participation (most frequently, discussing a better time for the interview). Although we do not advocate interrupting the householder to be responsive, Groves and McGonagle (2001) demonstrated the effectiveness of training interviewers to recognize the concerns of potential respondents and to deliver one of a set of associated relevant responses (e.g. responding to 'not enough time' by 'this only takes a few minutes'). Interviewers who were trained in these techniques obtained more interviews than their untrained counterparts.

That we observed such regular relationships between interviewer speech and participation decisions in the absence of any information about householders' chronic tendencies to agree or refuse is testimony to the importance of these regularities in the decision process. Indeed, very

small differences in rates of some speech behaviours can have large effects on the outcome (for example the difference between the first and second quintile of filler rates is only 1.28 fillers per 100 words but coincides with a tenfold increase in rates of agreement). We might not have been able to detect these relationships in a smaller corpus but they certainly exist.

This was an observational study, using actual survey introductions. As a result, we could not randomly assign interviewers to households: for example, as indicated earlier, interviewers who are believed to be more successful are often assigned to households that have persisted in not agreeing to be interviewed. In addition, we could not control whether the same or different interviewers contacted each household. Most contacts were preceded by a variable number of contacts, rendering many contacts not perfectly comparable, and some interviewers worked in multiple studies. An experimental design would have allowed us to eliminate many of these limitations but at the cost of reduced ecological validity.

A critical next step is incorporating response propensity indicators at the household and householder level into future models of interviewer speech and participation decisions. In addition, testing the implications of the current work in a factorial field experiment will support causal attributions of the behaviours that we have examined observationally. For example, an experimental design might include three levels of interviewer filler rate (high, medium and low), interviewers who were trained or not trained to pause sufficiently long for the householder to produce backchannels, interviewers who were trained or not trained to interrupt strategically and householders with high and low response propensity. It remains to be seen whether interviewers can implement this type of training in a way that sounds realistic.

Future work will be of most value if it can lead to practices that increase response rates, and to do so consistently across interviewers. As West and Olson (2010) have shown, differences in whom interviewers recruit can contribute to what appear to be interviewer effects in measurement. For example, they observed differences in the distributions of answers across interviewers to a question about age at divorce, despite a relatively interpenetrated design. They determined that at least some of this variation is due to differential success in which types of householders interviewers can recruit. They suggested, as an example, that if some interviewers do not adjust their speech for older householders—perhaps speaking louder or more slowly—they may recruit predominantly younger respondents, i.e. fail to recruit older respondents. Whether due to recruitment or measurement, it reduces the precision of survey estimates. Applications of the current work and its extensions will be of greatest value to the extent that they promote comparably high levels of success across interviewers.

An open question is which speech behaviours are trainable and whether they are equally trainable. For example, monitoring one's fluency while speaking is likely to be distracting for most interviewers but may be especially distracting for some, e.g. those with shorter working memory capacity such as older interviewers; similarly, monitoring paralinguistic aspects of householders' speech such as backchannels is likely to be challenging for most interviewers to do while listening to what householders say, and especially difficult for some.

One promising application domain for the current work and follow-up studies is the design of automated interviewing systems such as speech interactive voice response (e.g. Bloom (2007)) in which household members receive a recorded invitation and can speak to the system. Under one approach, the speech in the invitation would have characteristics that are optimal, e.g. moderately disfluent speech, in general across respondents. Under another approach its characteristics could adapt over the course of the invitation, adjusting filler rate, speech rate, pitch variation and loudness on the basis of what is learned about the user's speech. In addition, the system could be programmed to stop speaking and to start listening when the householder begins to speak (known as 'barge-in') so that overspeech is minimized.

Clearly there is much to learn about speech and decision making in general, and in survey introductions in particular. The current work is a first step in that direction.

Acknowledgements

We are grateful to the following organizations for supporting the research that is reported here: the National Science Foundation (grant SES-0819734 to the University of Michigan and grant SES-0819725 to Michigan State University); the Survey Research Center, University of Michigan; Department of Communicative Sciences and Disorders, Michigan State University; Charles Cannell Fund in Survey Methodology; Rensis Likert Fund for Research on Survey Methodology.

We also thank the following people for advice and assistance: Pete Batra, Carolina Casas-Cordero, Wil Dijkstra, Haley Gu, Patty Maher, Joe Matuzak and Michael Schober. We are indebted to the transcribers and acoustic analysts at Michigan State University and the coders and 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 McCarty, Gabe Moss, Kirsten Mull, Danny Nielsen, Dana Perkins, Fernando Pacheco, Danielle Popielarz, Christine Sheffler, Amanda Tatro and Dylan Vollans.

References

- Benki, J., Broome, J., Conrad, F., Groves, R. and Kreuter, F. (2011) Effects of speech rate, pitch, and pausing on survey participation decisions. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*
- Bloom, J. (2008) The speech IVR as Survey Interviewing Methodology. In *Envisioning the Survey Interview of the Future* (eds F. G. Conrad and M. F. Schober), pp. 119–136. Hoboken: Wiley.
- Boersma, P. (1993) Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proc. Inst. Phonet. Sci.*, **17**, 97–110.
- Boersma, P. and Weenink, D. (2007) Praat: doing phonetics by computer (Version 5.0). University of Amsterdam, Amsterdam. (Available from <http://www.praat.org>.)
- Brunner, L. J. (1979) Smiles can be backchannels. *J. Personality Soc. Psychol.*, **37**, 728–734.
- Christenfeld, N. (1995) Does it hurt to say um? *J. Nonverb. Behav.*, **19**, 171–186.
- Clark, H. and Fox Tree, J. E. (2002) Using *uh* and *um* in spontaneous speaking. *Cognition*, **84**, 73–111.
- Clark, H. and Schaefer, E. F. (1989) Contributing to discourse. *Cogn. Sci.*, **13**, 259–294.
- Duncan, S. (1973) Toward a grammar for dyadic conversation. *Semiotica*, **9**, 29–47.
- Duncan, S. (1974) On the structure of speaker-auditor interaction during speaking turns. *Lang. Soc.*, **3**, 161–180.
- Duncan, S. and Fiske, D. W. (1977) *Face-to-face Interaction*. Hillsdale: Erlbaum.
- Dunne, M. and Ng, S. H. (1994) Simultaneous speech in small-group conversation: all-together-now and one-at-a-time? *J. Lang. Soc. Psychol.*, **13**, 45–71.
- Fox Tree, J. E. (2001) Listener's uses of *uh* and *um* in speech comprehension. *Memory Cogn.*, **29**, 320–326.
- Fujiki, M., Brinton, B. and Sonnenberg, E. A. (1990) Repair of overlapping speech in the conversations of specifically language-impaired and normal developing children. *Appl. Psycholing.*, **11**, 201–215.
- Goodwin, C. (1981) *Conversational Organization: Interaction between Speakers and Hearers*. New York: Academic Press.
- Groves, R. M. and Benki, J. R. (2006) 300 hellos: acoustic properties of initial respondent greetings and response propensities in telephone surveys. *17th Int. Wrkshp Household Survey Nonresponse, Omaha*.
- Groves, R. M. and Couper, M. P. (1998) *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, R. M. and McGonagle, K. (2001) A theory-guided interviewer training protocol regarding survey participation. *J. Off. Statist.*, **17**, 249–265.
- Groves, R. M., O'Hare, B. C., Gould-Smith, D., Benki, J., Maher, P. and Hansen, S. E. (2008) Telephone interviewer voice characteristics and the survey participation decision. In *Advances in Telephone Survey Methodology* (eds J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japek, P. J. Lavrakas, M. W. Link and R. L. Sangster), pp. 385–400. New York: Wiley.
- Groves, R. M., Singer, E. and Corning, A. (2000) Leverage-saliency theory of survey participation—description and an illustration. *Publ. Opin. Q.*, **64**, 299–308.
- Hincks, R. (2005) Measures and perceptions of liveliness in student oral presentation speech: a proposal for an automatic feedback mechanism. *System*, **33**, 575–591.

- Hopper, R. (1992) *Telephone Conversation*. Bloomington: Indiana University Press.
- Huefken, V. and Schaefer, A. (2003) Zum Einfluss Stimmlicher Merkmale und Ueberzeugungsstrategien der Interviewer auf die Teilnahme in Telefonumfragen. *Koehn. Zeits. Soziol. Sozpsychol.*, **55**, 321–339.
- Leighton, L. A., Stollak, G. E. and Ferguson, L. R. (1971) Patterns of communication in normal and clinic families. *J. Consult. Clin. Psychol.*, **36**, 252–256.
- Oksenberg, L. and Cannell, C. (1988) Effects of interviewer vocal characteristics on nonresponse. In *Telephone Survey Methodology* (eds R. M. Groves, P. B. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II and J. Waksberg), pp. 257–269. New York: Wiley.
- Oksenberg, L., Coleman, L. and Cannell, C. F. (1986) Interviewers' voices and refusal rates in telephone surveys. *Publ. Opin. Q.*, **50**, 97–111.
- Sacks, H., Schegloff, E. A. and Jefferson, G. (1974) A simplest systematics for the organization of turn-taking for conversation. *Language*, **50**, 696–735.
- Schegloff, E. A. (1982) Discourse as an interactional achievement: some uses of “uh huh” and other things that come between sentences. In *Analyzing Discourse: Text and Talk, 32nd Georgetown University Roundtable on Languages and Linguistics 1981* (ed. D. Tannen), pp. 71–93. Washington DC: Georgetown University Press.
- Sharf, D. J. and Lehman, M. E. (1984) Relationship between the speech characteristics and effectiveness of telephone interviewers. *J. Phon.*, **12**, 219–228.
- Smith, V. L. and Clark, H. H. (1993) On the course of answering questions. *J. Mem. Lang.*, **32**, 25–38.
- Smith-Loven, L. and Brody, C. (1989) Interruptions in group discussions: the effect of gender and group composition. *Am. Sociol. Rev.*, **54**, 424–435.
- StataCorp (2009) *Stata Statistical Software: Release 11*. College Station: StataCorp.
- StataCorp (2011) *Stata Statistical Software: Release 12*. College Station: StataCorp.
- Steinkopf, L., Bauer, G. and Best, H. (2010) Nonresponse in CATI surveys. *Meth. Data Anal.*, **4**, 3–26.
- van der Vaart, W., Ongena, Y., Hoogendoorn, A. and Dijkstra, W. (2006) Do interviewers' voice characteristics influence cooperation rates in telephone surveys? *Int. J. Publ. Opin. Res.*, **18**, 488–499.
- Wardhaugh, R. (1985) *How Conversation Works*. Oxford: Basil Blackwell.
- West, B. and Olson, K. (2010) How much of interviewer variance is really nonresponse error variance? *Publ. Opin. Q.*, **74**, 1004–1026.
- Zimmerman, D. and West, C. (1975) Sex roles, interruptions and silences in conversation. In *Language and Sex: Differences and Dominance* (eds B. Thorne and E. Henley), pp. 105–129. Rowley: Newbury House.