Who Is Willing to Use Audio and Voice Inputs in Smartphone Surveys, and Why?

Lenzner, Timo; Höhne, Jan Karem

Who Is Willing to Use Audio and Voice Inputs in Smartphone Surveys, and Why?

Timo Lenzner (ID)
GESIS – Leibniz Institute for the Social Sciences, Germany

Jan Karem Höhne
University of Duisburg-Essen, Germany; RECSM-Universitat Pompeu Fabra, Spain

Abstract

The ever-growing number of respondents completing web surveys via smartphones is paving the way for leveraging technological advances to improve respondents' survey experience and, in turn, the quality of their answers. Smartphone surveys enable researchers to incorporate audio and voice features into web surveys, that is, having questions read aloud to respondents using pre-recorded audio files and collecting voice answers via the smartphone's microphone. Moving from written to audio and voice communication channels might be associated with several benefits, such as humanizing the communication process between researchers and respondents. However, little is known about respondents' willingness to undergo this change in communication channels. Replicating and extending earlier research, we examine the extent to which respondents are willing to use audio and voice channels in web surveys, the reasons for their (non)willingness, and respondent characteristics associated with (non)willingness. The results of a web survey conducted in a nonprobability online panel in Germany (N = 2146) reveal that more than 50% of respondents would be willing to have the questions read aloud (audio channel) and about 40% would also be willing to give answers via voice input (voice channel). While respondents mostly name a general openness to new technologies for their willingness, they mostly name preference for written communication for their nonwillingness. Finally, audio and voice channels in smartphone surveys appeal primarily to frequent and competent smartphone users as well as younger and tech-savvy respondents.

Keywords

audio communication channel, automatic question reading, respondent willingness, smartphone, voice answers, voice communication channel

Corresponding author:
Timo Lenzner, Department of Survey Design and Methodology, GESIS – Leibniz Institute for the Social Sciences, B6, 4-5, 68159 Mannheim, Germany.
Email: timo.lenzner@gesis.org

Introduction

The number of respondents completing web surveys on mobile devices, particularly smartphones, has increased substantially over recent years (see Gummer, Quoß, & Roßmann, 2019; Peterson, Griffin, LaFrance, & Li, 2017; Revilla, Toninelli, Ochoa, & Loewe, 2016). For example, the smartphone rate in the probability-based German Internet Panel (GIP) increased from 4% in September 2012 (first regular GIP wave) to 37% in November 2021 (last GIP wave available). This trend has also been detected in the German Respondi online access panel (Gummer et al., 2019), in commercial and academic surveys in the US and several other countries, such as the Netherlands and Spain (Peterson et al., 2017), and in surveys of the Netquest online access panel in Argentina, Brazil, Chile, Colombia, Mexico, Portugal, and Spain (Revilla et al., 2016). Given that answering a web survey on a smartphone differs from answering it on a PC due to device characteristics, such as screen size and operation, a variety of studies have investigated how the use of different devices affects survey data quality (see, for example, Couper & Peterson, 2017; de Bruijne & Wijnant, 2013; Krebs & Höhne, 2021; Lugtig & Toepoel, 2016; Mavletova, 2013; Revilla et al., 2016; Struminskaya, Lugtig, Keusch, & Höhne, 2020; Wells, Bailey, & Link, 2014). For example, these studies have found that smartphone respondents are more likely to drop out of a survey (Callegaro, 2010; Poggio, Bosnjak, & Weyandt, 2015) and need longer to complete it (Antoun, Couper, & Conrad, 2017; Couper & Peterson, 2017; Gummer & Roßmann, 2015) than PC respondents. Consequently, survey researchers and practitioners have been advised to rethink the established survey design principles for web surveys (which were commonly based on PCs) and to consider optimizing web surveys for smartphones and other mobile devices (Antoun, Katz, Argueta, & Wang, 2018; Callegaro, 2010; de Bruijne & Wijnant, 2014; Revilla et al., 2016).

However, the increasing use of smartphones for completing web surveys does not only pose a challenge for web survey research. It also offers new opportunities to improve respondents' survey experience and thus the quality of their answers. For instance, smartphones enable respondents to have survey questions read aloud to them using pre-recorded audio files and to answer survey questions using the voice recording feature (or microphone). Incorporating audio and voice communication channels would also humanize the communication process between researchers and respondents (Tourangeau, Rips, & Rasinski, 2000, chapter 10). It also simulates daily conversation, allowing respondents to engage in open and in-depth narrations. For example, Gavras and Höhne (2020) estimated the criterion validity of written and spoken answers to open-ended web survey questions using sentiment scores and found (somewhat) higher criterion validity for voice than for text answers.

This methodology is related to self-administered (mobile) web surveys. It is not to be confused with "Audio Computer-Assisted Self-Interviewing (ACASI)" in face-to-face surveys or "Interactive Voice Response (IVR)" in telephone surveys[1]. All respondents are exposed to the same web survey environment that includes an audio play function for reading survey questions aloud to respondents and a recording function for recording respondents' voice answers. In general, it follows the notion of the audio and voice features of instant messaging services, such as WhatsApp and WeChat.

To date, there is only limited evidence on whether and to what extent respondents are willing to participate in web surveys with audio and voice communication channels (see Höhne, 2021; Revilla, Couper, Ochoa, 2018). The purpose of this study is to extend our current knowledge on respondents' willingness to have questions read aloud to them (audio channel) and to record their answers via the built-in microphone of smartphones (voice channel). In addition, we ask respondents to provide reasons for their (non)willingness by employing follow-up probes and examine the association of respondent characteristics with (non)willingness.

In what follows, we review previous research on respondents' willingness to use their smartphones' audio and voice recording features for completing web surveys and outline our research questions. We then report on a study that examines respondents' willingness to engage in web surveys with audio and voice communication channels. Finally, we close with a discussion of the practical implications of our study and suggest perspectives for future research.

Background and Research Questions

The majority of today's web surveys rely on text-based communication channels. Researchers mostly ask survey questions in text form by providing a question stem and answer options (closed answer format) or text boxes (open answer format). These communication channels require respondents to read the survey questions and answer options and type in answers to open-ended questions, which can be burdensome, particularly for respondents with lower literacy or language skills (e.g., illiterates or second language learners). For instance, Grotlüschen, Buddeberg, Dutz, Heilmann, & Stammer (2019) estimated that in 2018, about 12% (or six million) of the adult population in Germany could not read and write adequately. Another 21% (or 11 million) had spelling errors even for commonly used words. In addition, the UNESCO Institute for Statistics (2017) estimates that about 14% (or 750 million) of the world population in 2016 could not read and write. Thus, people with low literacy and language skills may either not be able to participate in text-based web surveys or may provide data of lower quality (e.g., leave questions unanswered, give uninterpretable answers to open-ended questions, or give contradictory answers to a set of conceptually related items).

However, text-based communication in web surveys does not only pose a challenge for those with limited literacy and language skills. It is known from previous research that especially open-ended survey questions place a considerable burden on respondents, and thus, they are often associated with higher item nonresponse rates (Bradburn, 1978; Lambert & Miller, 2015) or very short answers (Lambert & Miller, 2015; Revilla & Ochoa, 2016). In particular, this is the case when respondents answer open-ended questions on smartphones with a virtual on-screen keyboard shrinking the viewing space of the web survey content (Lambert & Miller, 2015; Lugtig & Toepoel, 2016; Peytchev & Hill, 2010; Struminskaya et al., 2015). For instance, Lambert and Miller (2015) reported that smartphone respondents were less likely to answer open-ended questions, and when they did, they provided much shorter answers than PC respondents (i.e., the median answer length for PC respondents was almost twice as high as for smartphone respondents).

Audio and voice communication channels in web surveys are expected to be less burdensome for respondents than text-based communication channels (Tourangeau et al., 2000, chapter 10). Respondents only need to press a play button to have the questions read aloud to them (audio channel) or a recording button to provide their answers (voice channel). This methodology may be well suited for respondents with lower literacy levels or with lower competence in the language of the web survey. Particularly, the voice communication channel allows to collect "qualitative" answers in standardized web surveys. This combination of qualitative and quantitative methodologies enables researchers to gather unprecedented rich and nuanced information on respondents' attitudes, behaviors, and beliefs by eliciting open narrations (Gavras & Höhne, 2020; Gavras et al., 2022). It also makes the communication process between researchers and respondents more natural by simulating daily conversation. Because audio and voice communication channels gain importance in everyday life (e.g., in the form of virtual assistant technology; see Struminskaya, Lugtig, Keusch, & Höhne, 2020), it seems worthwhile to investigate their usefulness and usability for web survey research as well. Following Conrad et al. (2017), giving respondents the choice to select the communication

channel that is most suitable for them may enhance survey data quality (see also Miller (2017) for a comprehensive discussion of the future of surveys).

Although the use of audio and voice communication channels in (mobile) web surveys might be associated with several benefits, little research has been conducted on whether respondents are willing to take part in such web surveys. To the best of our knowledge, only two studies have been published on this subject to date (see Höhne, 2021; Revilla et al., 2018). In the first study, Revilla et al. (2018) asked respondents of the Netquest online access panel in Spain whether they would be willing to use the voice input function of their smartphone to provide voice answers to open-ended questions in web surveys. More than 50% of the participating respondents indicated that they would be willing to give voice answers to open-ended questions. In addition, the authors examined what respondent characteristics are associated with the willingness to use the voice input function. Willingness was higher among respondents who already used voice input in everyday life, used their smartphone more frequently, were native Spanish speakers, and reported no difficulties in answering the current survey. Interestingly, these two latter findings suggest that the voice input is not necessarily perceived as helpful by respondents who are supposed to benefit most from this methodology (e.g., second language learners and respondents experiencing difficulties in survey responding). Instead, voice input in smartphone surveys seems to attract frequent and competent smartphone users.

Extending the research by Revilla et al. (2018), Höhne (2021) also examined respondents' willingness to use audio communication channels in web surveys. Specifically, the author first asked about respondents' willingness to participate in future surveys via their smartphone to have the questions read aloud to them and then about respondents' willingness to participate in future surveys via their smartphone to provide voice answers. However, in contrast to Revilla et al. (2018), he asked whether respondents would use both communication channels for answering survey questions in general and not specifically for answering open-ended questions. Using data from the probability-based German Internet Panel, he found a considerably lower willingness for voice input than Revilla et al. (2018). About 25% of the respondents indicated that they would agree with an audio channel, and only about 16% indicated that they would agree with a voice channel. Similar to Revilla et al. (2018), Höhne (2021) found that respondents' willingness to engage in web surveys with audio and voice communication channels, respectively, was higher among respondents using a smartphone for web survey completion and more frequent and competent Internet users. Moreover, younger and male respondents, as well as respondents with higher extraversion and agreeableness, were more willing. According to the author, these findings indicate that most respondents seem not yet ready for audio and voice communication channels. However, both channels might be appealing to some respondent groups, such as millennials (born between 1982 and 2003; Strauss & Howe, 1991), who are more reluctant to participate in traditional web surveys (see also Revilla & Höhne, 2020).

While both of these earlier studies provide insights on the extent to which respondents drawn from different samples in different countries are willing to engage in web surveys with audio and voice communication channels and on contributing respondent characteristics, the reasons for respondents' (non)willingness remain completely unclear. With the current study, we aim at filling this research gap by replicating and extending the earlier studies by Revilla et al. (2018) and Höhne (2021). For this purpose, we conducted a cross-sectional web survey (no device limitation) in a nonprobability online panel in Germany ($N = 2146$). We address the following three research questions:

1. What is the share of respondents who are willing to participate in web surveys with a smartphone to have the survey questions read aloud to them and to give voice answers?
2. What are the reasons for respondents' (non)willingness to participate in web surveys with a smartphone to have the survey questions read aloud to them and to give voice answers?

3. What respondent characteristics are associated with respondents' (non)willingness to participate in web surveys with a smartphone to have the survey questions read aloud to them and give voice answers?

Method

Data Collection and Sample

We conducted a web survey in the Respondi online access panel (respondi.com) in Germany in November and December 2020. The current study was part of a larger study with several unrelated experiments that were independently randomized to reduce the possibility of systematic carryover effects. Overall, participation in the web survey took about 15 minutes. The questions analyzed in this study were placed in the fourth quarter of the web survey.

To draw a sample that is representative of the German adult online population, we used cross-quotas for gender (female and male) and age (18–29 years, 30–39 years, 40–49 years, 50–59 years, and 60 years and older). Of the 4994 respondents starting the web survey, 301 were screened out because of full quotas. Another 2248 respondents were randomly routed to a different web survey that did not include our questions, and 299 broke off before receiving the questions analyzed in this study. This leaves us with 2146 respondents for statistical analysis.

These respondents were between 18 and 83 years old with a mean age of 44.6 ($SD = 14.8$), and 49.3% were female.[2] In terms of education, 16.8% had graduated from a lower secondary school (low education level), 27.1% from an intermediate secondary school (medium education level), and 55.0% from a college preparatory secondary school (high education level). Furthermore, 0.8% still attended school or had finished without a diploma, and 0.3% did not report their highest level of education.

Questions

We used the same two questions as Höhne (2021) on respondents' willingness to participate in future web surveys with a smartphone to use audio and voice communication channels. Each question (Q1 and Q2) was followed by an open-ended probing question (P1 and P2) asking respondents to provide reasons for their (non)willingness. The questions and probes were formulated as follows:

Q1. In general, are you willing to participate in upcoming surveys with your smartphone to have the questions read out loud to you?

P1. The previous question asked whether you would generally be willing to participate in future surveys using your smartphone to have the questions read out loud to you.

Your answer was: "[respondent's answer]"
Please elaborate on your answer. Why did you choose this answer?

Q2. In general, are you willing to participate in upcoming surveys with your smartphone to give oral answers via voice input?

P2. The previous question asked whether you would generally be willing to participate in future surveys using your smartphone to give oral answers via voice input.

Your answer was: "[respondent's answer]"
Please elaborate on your answer. Why did you choose this answer?

The questions and probes were presented on separate web survey pages (single question presentation). Q1 and Q2 offered the following vertically aligned answer options: "Definitely yes," "Probably yes," "Probably no," "Definitely no," "Don't use a smartphone," and "Don't know how it works." A space visually separated the last two options from the remaining options (see Appendix A for screenshots of the questions and probes) and respondents who selected one of these options did not receive the open-ended probes.

Analytical Strategies

Regarding our first and third research questions, we follow the analytical strategies used by Höhne (2021) and Revilla et al. (2018). Thereby, we aim at increasing the comparability of the results. Data preparations and analyses were conducted with SPSS version 24 (RQ1 and RQ3) and Microsoft Excel 365 (RQ2). To examine our first research question, we look at the answer distributions (in percentages) of the two questions on respondents' willingness to participate in web surveys with a smartphone to have the survey questions read aloud to them and to give voice answers.

To analyze the reasons for respondents' (non)willingness (second research question), respondents' answers to the two open-ended probing questions were coded by the first author based on two initial coding schemes, including four willingness and seven nonwillingness categories, respectively. Both coding schemes were gradually expanded during the coding and finally consisted of 10 (audio willingness) and 12 (audio nonwillingness) categories as well as 10 (voice willingness) and 14 (voice nonwillingness) categories, respectively. Using these final coding schemes, a student assistant independently coded a randomly selected subset of 10% ($n = 407$) of the answers to the probes so that we could estimate inter-coder reliability. The subset size of probe answers and number of judges were initially chosen for time and budget constraints. An agreement of 81% among the two coders and a Cohen's kappa value of 0.80 indicate that the number of coders was adequate (see Nunnally (1978) and Rust and Cooil (1994) for a comprehensive discussion on minimally acceptable levels of inter-rater agreement and appropriate numbers of judges). Discrepancies between the two ratings were discussed between the two coders until they reached a consensus. The reasons for respondents' (non)willingness are reported in percentages.

To answer our third research question, we examine independent variables associated with respondents' willingness for audio and voice channels (coding of dependent variables: 1 "definitely no," 2 "probably no," 3 "probably yes," and 4 "definitely yes") by running two separate ordinary least squares (OLS) regressions. The (non-substantive) answer options "don't use a smartphone" and "don't know how it works" are excluded from the regression analyses. We use the following independent variables that were suggested by prior studies to be associated with respondent behavior (Höhne, 2021; Revilla et al., 2018; van Vaerenbergh & Thomas, 2013): smartphone use in this survey (1 = "yes"), smartphone competence (1 = "low" to 5 = "high"), general Internet use (1 = "low" to 6 = "high"), Internet use via smartphone (1 = "yes"), belief that Internet facilitates communication (1 = "low" to 5 = "high"), affinity for technology (1 = "low" to 5 = "high"), survey evaluation: difficulty (1 = "low" to 5 = "high"), survey evaluation: enjoyment (1 = "low" to 5 = "high"), survey evaluation: length (1 = "low" to 5 = "high"), survey evaluation: topic sensitivity (1 = "low" to 5 = "high"), and need for cognition (1 = "low" to 7 = "high").[3] In addition, we control for the following socio-demographic variables: age (in years), female (1 = "yes"), and education with low as reference: medium (1 = "yes") and high (1 = "yes"). English translations of the independent and control variables are documented in Appendix B.

Results

Research Question 1

As shown in Table 1, more than half of the respondents (52.3%) indicated that they definitely or probably would be willing to participate in future web surveys using audio channels. The willingness of respondents to give answers via voice input was somewhat lower. Nevertheless, more than one-third of the respondents (39.7%) indicated that they definitely or probably would be willing to use this communication channel in future surveys. About 5% of the respondents indicated that they do not use a smartphone or do not know how the audio play and/or voice recording functions work.

Research Question 2

We next examined respondents' reasons for their (non)willingness to use audio and voice communication channels. Tables 2 and 3 show the results. The most frequently mentioned reason for willingness to use the audio channel was a general openness to new technologies and procedures (41.8%). This is followed by the impression that this communication channel (13.8%) and the use of a smartphone (10.7%) facilitate answering web surveys. Another 5% of the respondents indicated their willingness without providing a substantial reason (e.g., "Why not" and "I think that's good"). Several reasons were also mentioned by only a few respondents (less than 5%): willing, but preferring reading on a PC, enjoying completing surveys irrespective of the device, overcoming vision impairments, and using audio functions regularly. About 20% of the respondents provided other reasons (e.g., "Depends on the survey" and "Depends on the incentive") or noninformative reasons or nonsense (e.g., "Because it is so" and "Hkhjkj").

Similar to the audio channel, the most frequently mentioned reasons for willingness to use the voice channel were a general openness to new technologies and procedures (29.7%) and the impression that this communication channel facilitates answering web surveys (27.2%). About 8% of the respondents indicated their willingness without providing a substantial reason (e.g., "No problem" and "Would be ok"), and about 7% indicated that they were willing but prefer answering on a PC. Reasons mentioned by less than 5% of the respondents included regular use of voice functions, enjoying completing surveys irrespective of the device, the impression that a smartphone facilitates answering surveys, and overcoming vision impairments. Finally, about one quarter (24.3%) provided other reasons (e.g., "I would like to contribute to research" and "I can be bought,

Table 1. Respondents' Willingness to Participate With Their Smartphone in Surveys to Have the Survey Questions Read Aloud to Them and to Give Voice Answers Via Voice Input.

Answer Options	Audio Channel (%)	Voice Channel (%)
Definitely no	14.7	24.5
Probably no	27.5	30.9
Probably yes	28.7	25.0
Definitely yes	23.6	14.7
Don't use a smartphone	1.4	1.2
Don't know how it works	4.1	3.7

Note. N = 2146. The last two options were visually separated by a space from the remaining options (see Appendix A for screenshots of the two questions).

Table 2. Reasons for Willingness to Use the Audio and Voice Channels in Upcoming Surveys.

Reasons for Willingness	Audio Channel (%)	Voice Channel (%)
General openness to new technologies and procedures	41.8	29.7
Communication channel facilitates answering	13.8	27.2
Smartphone facilitates answering	10.7	0.5
Willingness (without providing a substantial reason)	5.3	7.7
Willing, but respondents prefer reading and/or PC completion	3.1	7.2
Enjoying completing surveys irrespective of device	2.6	0.8
Overcoming vision impairments	1.2	0.4
Regular use of audio and voice functions	0.9	2.2
Other	4.9	7.2
Noninformative reasons and nonsense	15.7	17.1

Note. n = 1123 (audio) and *n* = 852 (voice).

Table 3. Reasons for Nonwillingness to Use the Audio and Voice Channels in Upcoming Surveys.

Reasons for Nonwillingness	Audio Channel (%)	Voice Channel (%)
Preferring reading or writing instead of listening (audio) or speaking (voice)	20.9	15.1
Preferring completing surveys via PCs	17.0	7.7
Nonwillingness (without providing a substantial reason)	11.3	18.5
No added value (in comparison to reading or writing)	9.1	2.1
Requiring too much effort	7.2	6.4
Third parties around	5.9	7.7
Low and infrequent smartphone use	5.0	2.1
Technical obstacles	2.0	1.2
Privacy and data security concerns	1.9	7.7
Hearing (audio) or speaking (voice) difficulties	1.2	0.4
Disliking voice function	—	5.6
Lack of ability to control (altering or deleting recordings)	—	1.9
Other	7.8	9.4
Noninformative reasons and nonsense	10.8	14.1

Note. n = 905 (audio) and *n* = 1189 (voice). Because of rounding, percentages do not add up to 100%.

for money I will also speak into the microphone") or noninformative reasons and nonsense (e.g., "Self-explanatory" and "hmmm").

With respect to respondents who indicated that they were not willing to use the audio channel, more than one fifth (20.9%) preferred reading instead of listening and almost one fifth (17.0%) preferred completing surveys on a PC. Another 11% indicated their nonwillingness without providing a substantial reason (e.g., "I just don't want to" and "Not interested"), and 9% questioned the added value of the audio channel in comparison to traditional communication channels. Some respondents indicated too much effort (about 7%), concerns about third parties being around or being bothered by completing the survey in this way (about 6%), and low or infrequent smartphone use (5%). Again, several reasons were mentioned by less than 5% of the respondents: technical obstacles, privacy and data security concerns, and hearing difficulties. Finally, about 20% provided

Table 4. Standardized OLS Regression Coefficients of Independent Variables on Respondents' Willingness to Participate in Surveys With Audio and Voice Communication Channels.

	Audio Channel	Voice Channel
Independent Variables	Coefficients (Standard Errors)	Coefficients (Standard Errors)
Smartphone use in this survey	0.107*** (0.047)	0.071** (0.050)
Smartphone competence	0.055* (0.032)	0.047 (0.034)
General Internet use	−0.014 (0.026)	−0.027 (0.028)
Internet use via smartphone	0.091*** (0.072)	0.056* (0.077)
Belief that Internet facilitates communication	0.038 (0.025)	−0.020 (0.027)
Affinity for technology	0.140*** (0.028)	0.124*** (0.030)
Survey evaluation: difficulty	0.002 (0.033)	−0.007 (0.035)
Survey evaluation: enjoyment	0.140*** (0.032)	0.152*** (0.034)
Survey evaluation: length	−0.037 (0.024)	0.007 (0.026)
Survey evaluation: topic sensitivity	−0.040 (0.026)	−0.064* (0.028)
Need for cognition	0.010 (0.022)	0.031 (0.024)
Age	−0.278*** (0.002)	−0.125*** (0.002)
Female	−0.055** (0.042)	−0.017 (0.044)
Education with low as reference		
Medium	−0.016 (0.065)	0.009 (0.069)
High	−0.064* (0.063)	−0.039 (0.067)

Note. *p < .05, **p < .01, ***p < .001. Intercepts are statistically significant. Coding of the dependent variables: 1–"definitely no," 2–"probably no," 3–"probably yes," and 4–"definitely yes," Answers to the options "don't use a smartphone" and "don't know how it works" are excluded from the analyses. We used listwise deletion of missing values. See "Analytical Strategies" for the coding of the independent variables. We excluded respondents selecting the answer option "other" when asked for their gender because the share was very low ($n = 4$).

other reasons (e.g., "My smartphone is a company phone" and "In order to be able to carry out the survey, an app is probably required, which I do not want to download") or noninformative reasons and nonsense (e.g., "xxx" and "Because").

Among respondents who indicated that they were not willing to use the voice channel, the largest share did not provide a substantial reason for their nonwillingness (18.5%). This is followed by a preference for writing instead of speaking (15.1%). Preferring completing surveys via PC, privacy and data security concerns, and concerns about third parties were mentioned by about 8% of the respondents, respectively. The next most frequently cited reasons were too much effort (about 6%) and disliking the voice recording function (about 6%). Reasons mentioned by less than 5% of respondents included low and infrequent smartphone use, the impression that the voice channel would not add any value compared to traditional communication channels, and lack of ability to control the voice answers. Again, about one quarter (23.5%) provided other reasons (e.g., "I'd feel stupid doing that" and "That sounds strange") or noninformative reasons and nonsense (e.g., "Don't know" and "Several reasons").

Research Question 3

Finally, we investigated variables associated with respondents' willingness for audio and voice channels (coding of the dependent variables: 1 "definitely no," 2 "probably no," 3 "probably yes," and 4 "definitely yes," respectively) by running two separate OLS regressions. We used

several independent variables suggested by previous research to impact respondent behavior (Höhne, 2021; Revilla et al., 2018; van Vaerenbergh & Thomas, 2013). We did not exclude respondents who provided noninformative or nonsensical answers to the open-ended probes about their (non)willingness to use both channels, as we found no obvious signs that they engaged little with the survey other than providing low-quality answers to these open-ended questions. As a sensitivity check, we re-ran the regressions excluding these respondents. The results were similar to the ones we obtained using the full sample and none of our main conclusions changed. Hence, we decided to report the results for the full sample.

Table 4 shows the results. The regression models on audio [F(15,1961) = 36.80, $p < .001$, adjusted-$R^2 = 0.21$] and voice channels [F(15,1973) = 16.02, $p < .001$, adjusted-$R^2 = 0.10$] are statistically significant. Interestingly, the coefficients in both models show almost identical patterns. Respondents participating with a smartphone, using the Internet via the smartphone more frequently, and having a higher affinity to technology show a higher willingness. A higher smartphone competence only increases willingness for audio channels, but not for voice channels. Enjoying the participation in the web survey increases willingness (audio and voice) and finding the topics of the web survey sensitive decreases willingness (voice only). Younger respondents are more willing to participate in web surveys using audio and voice channels. Finally, female respondents and higher educated respondents show lower willingness (audio only).

Discussion and Conclusion

Replicating and extending earlier research by Höhne (2021) and Revilla et al. (2018), this study examined respondents' (non)willingness to engage in web surveys using audio and voice communication channels. In addition, we looked at respondents' reasons for (non)willingness and contributing respondent characteristics. Our results showed that more than 50% of respondents would be willing to have the questions read aloud (audio channel) and about 40% of respondents would also be willing to give answers via voice input (voice channel). These figures are considerably higher than those reported by Höhne (2021), who found a willingness for the audio channel of 25% and a willingness for the voice channel of 16%. However, our figures for the voice channel are similar to those reported by Revilla et al. (2018), who found a willingness of 54%.

In his article, Höhne (2021) argued that the differences between his results and the results reported by Revilla et al. (2018) might be due to the following reasons. First, the surveys were conducted in different countries (Germany vs. Spain). Second, the surveys were conducted in different panels (probability vs. nonprobability panel). Third, the two samples were characterized by different age structures; respondents in the study by Revilla et al. (2018) were younger than those in the study by Höhne (2021). Because the current study and the study by Höhne (2021) were conducted in Germany, national or cultural differences can be ruled out as an explanation for the differences in the magnitude of willingness. Also, taking a closer look at the age structure in our study and in the study by Höhne (2021), we only find minor support for the third explanation. The median age category in our study is "41 to 45" years, compared to "51 to 55" years in the study by Höhne (2021). Instead, our results rather support the second explanation, namely, that willingness to use audio and voice channels is higher among members of nonprobability online panels than among members of probability-based panels. Members of nonprobability online panels, most frequently, voluntarily sign up online for the panels themselves, while members of probability-based online panels are usually recruited offline (see Blom et al., 2015). Thus, it is not surprising that the first are more open to Internet-based technological innovations than the latter.

With respect to the reasons for respondents' (non)willingness, we found relatively similar patterns for both communication channels. The most frequently mentioned reasons for willingness to use this technology were a general openness to new technologies and the impression that these communication channels facilitate answering web surveys. These two reasons accounted for more than 55% of the reasons provided by respondents. Interestingly, the second reason (i.e., communication channel facilitates answering) was mentioned about twice as often for voice than for audio channels. This indicates that respondents perceive the voice channel as much more facilitating than the audio channel.

Respondents who expressed their nonwillingness to use one or both communication channels mostly argued that they prefer written communication (i.e., reading and writing instead of listening and speaking) and completing surveys on PCs instead of smartphones. Some respondents also questioned the added value of this new technology in comparison to traditional web survey interaction. This particularly applies to having the questions read aloud to them (audio channel). Potentially, these respondents could be convinced to participate in web surveys with audio and voice channels if they are provided with a sound explanation of the usefulness. In contrast, conversion attempts might be less successful among respondents concerned about third parties and privacy and data security. However, these reasons were mentioned comparatively rarely and only in relation to the voice channel with a relevant frequency.

Regarding respondent characteristics associated with respondents' willingness to participate in web surveys with audio and voice channels, our findings largely overlap with those reported by Höhne (2021) and Revilla et al. (2018). Similar to these earlier studies, we found higher willingness for those participating in the current study with a smartphone and more frequent and competent Internet users. Moreover, an affinity for technology and enjoying participation in the current survey correlated positively with willingness, while age was negatively associated with willingness to use audio and voice channels. Taken together, these findings suggest that audio and voice channels in smartphone surveys appeal primarily to frequent and competent smartphone users as well as younger and tech-savvy respondents. Finally, perceiving the topics of the web survey sensitive decreased respondents' willingness to use the voice channel. This finding can be explained by the fact that the voice channel (partially) mimics interviewer-administered survey modes, in which respondents are less willing to answer sensitive questions than in self-administered modes (see Tourangeau and Yan (2007) for a comprehensive review of research on sensitive questions in surveys). Employing sensitive open-ended questions with requests for voice answers might be critical because of social desirability. However, this is only an assumption that needs to be investigated empirically. We therefore recommend that future studies examine the proneness of voice answers to socially desirable answer behavior when it comes to sensitive question topics.

Three limitations to this study provide avenues for future research. First, we only asked respondents for their hypothetical willingness and did not examine their actual compliance when presented with audio and voice communication channels. Future research should investigate whether the response rates of smartphone surveys with audio and voice channels are indeed as high as suggested by respondents' alleged willingness. For instance, these studies could build on existing theories, such as the Technology Acceptance Model (Davis, 1989), to draw more theoretically driven conclusions. Second, we conducted our survey among members of an online access panel, and thus, we cannot infer to the general population. For instance, as shown by Höhne (2021), willingness among the general population might be considerably lower than reported for our sample. If members of probability-based online panels are indeed more reluctant to use this technology than members of nonprobability online access panels, then researchers and practitioners would have to weigh the (possible) benefits of collecting richer and more nuanced data (by employing the audio and voice channels in nonprobability web surveys) against the possibility of

inferring from the sample to the general population. Therefore, it is important that future research examines differences in the willingness of respondents from nonprobability and probability-based online panels. Third, we argued that audio and voice channels in smartphone surveys may be particularly well suited for respondents with lower literacy levels and with lower competence in the language of the web survey. However, we were not able to include measures of these competencies in the web survey. Therefore, we encourage future research to examine this issue more thoroughly.

As more and more respondents participate in web surveys via smartphones and audio and voice communication channels become increasingly important in everyday life, it seems to be only a matter of time before these communication channels also find their way into web surveys. Some age cohorts, such as millennials, are already quite familiar with these means of communication options (see Bosch, Revilla, & Paura, 2019). In addition, it is likely that other age cohorts follow and adapt their behavior to the new technologies and communication options. From a data analysis perspective, powerful data processing and analysis capabilities, such as Automatic Speech Recognition, Natural Language Processing, and Text-as-Data Methods, enable proper handling of audio and voice data (even from large-scale web surveys). Thus, it seems wise to begin examining the benefits and limitations of these communication channels in web surveys and deriving best practices for their implementation.

Acknowledgements

Declaration of Conflicting Interest

Funding

Data Availability

Replication data and documentation are available at https://doi.org/10.7910/DVN/0T6MNG

ORCID iD

Timo Lenzner ⓘ https://orcid.org/0000-0002-0177-2227

Notes

1. Audio and voice communication channels in mobile web surveys differ from ACASI in interviewer-administered surveys and IVR in telephone-administered surveys in terms of the survey mode. ACASI is mainly used to deliver sensitive questions, such as questions on sexual behavior and diseases, via headphones without the knowledge of the interviewer (Caspar, 2008). The answering process is nonverbal (e.g., entering a value via a keyboard), precluding open narrations. IVR consists of an audio delivery of survey questions but it is frequently touchtone-administered (i.e., respondents enter values via the telephone keypad; Currivan, 2008).

Even if answers are verbally provided, they are usually restricted to a minimum (e.g., answer categories, such as "agree"), which also precludes open narrations. IVR is most commonly used to engage customers (e.g., in company hotlines) and to extend business hours (e.g., after closing time and on the weekend). Compared to other survey administration methodologies, IVR is quite rare in survey research.

2. About 0.2% ($n = 4$) of the respondents selected the answer option "other." The remaining respondents selected the answer option "male."

3. To measure need for cognition (NFC), we used the short scale developed by Beißert et al. (2015) that consists of four questions. The questions are rated on a response scale from (1) "agree not at all" to (4) "neither/nor" to (7) "agree strongly". In line with the authors, NFC scores were calculated as follows: initially, we recoded the first and fourth questions so that low/high values indicate a low/high NFC. We then added up respondents' answers to the four NFC questions and divided the total score by the number of questions answered. Hence, the NFC scores range from 1 to 7.

References

Antoun, C., Couper, M. P., & Conrad, F. G. (2017). Effects of mobile versus PC web on survey response quality: A crossover experiment in a probability web panel. *Public Opinion Quarterly*, *81*(S1), 280–306. doi:10.1093/poq/nfw088

Antoun, C., Katz, J., Argueta, J., & Wang, L. (2018). Design heuristics for effective smartphone questionnaires. *Social Science Computer Review*, *36*(5), 557–574. doi:10.1177/0894439317727072

Beißert, H., Köhler, M., Rempel, M., & Beierlein, C. (2015). Kurzskala need for cognition NFC-K. *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. doi:10.6102/zis230.

Blom, A. G., Bosnjak, M., Cornilleau, A., Cousteaux, A.-S., Das, M., Douhou, S., & Krieger, U. (2015). A comparison of four probability-based online an mixed-mode panels in Europe. *Social Science Computer Review*, *34*(1), 8–25. doi:10.1177/0894439315574825

Bosch, O. J., Revilla, M., & Paura, E. (2019). Answering mobile surveys with images: An exploration using a computer vision API. *Social Science Computer Review*, *37*(5), 669–683. doi:10.1177/0894439318791515

Bradburn, N. (1978). Respondent burden. In *Health survey research methods*. DHEW publication No. 79 (pp. 49–55). Washington, DC: U.S. Department of Health, Education, and Welfare.

Callegaro, M. (2010). Do you know which device your respondent has used to take your online survey? *Survey Practice*, *3*(6), 1–12.

Caspar, R (2008). Audio computer-assisted self-interviewing(ACASI). In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods – first volume* (pp. 44–45). New York, NY: SAGE.

Conrad, F. G., Schober, M. F., Antoun, C., Yan, H. Y., Hupp, A. L., Johnston, M., & Zhang, C. (2017). Respondent mode choice in a smartphone survey. *Public Opinion Quarterly*, *81*(S1), 307–337. doi:10.1093/poq/nfw097

Couper, M. P., & Peterson, G. J. (2017). Why do web surveys take longer on smartphones? *Social Science Computer Review*, *35*(3), 357–377. doi:10.1177/0894439316629932

Currivan, D. B. (2008). Interactive voice response(IVR). In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods – first volume* (pp. 342–344). New York, NY: SAGE.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *The Milbank Quarterly*, *13*(3), 319–340. doi:10.2307/249008

de Bruijne, M., & Wijnant, A. (2013). Comparing survey results obtained via mobile devices and computers: An experiment with a mobile web survey on a heterogeneous group of mobile devices versus a computer-assisted web survey. *Social Science Computer Review*, *31*(4), 482–504. doi:10.1177/0894439313483976

de Bruijne, M., & Wijnant, A. (2014). Improving response rates and questionnaire design for mobile web surveys. *Public Opinion Quarterly*, *78*(4), 951–962. doi:10.1093/poq/nfu046

Gavras, K., & Höhne, J. K. (2020). Evaluating political parties: Criterion validity of open questions with requests for text and voice answers. *International Journal of Social Research Methodology*, *25*(1), 135–141. doi:10.1080/13645579.2020.1860279

Gavras, K., Höhne, J. K., Blom, A., & Schoen, H (2022). Innovating the collection of open-ended answers: The linguistic and content characteristics of written and oral answers to political attitude questions. *Journal of the Royal Statistical Society (Series A)*. Advance online publication. doi:10.1111/rssa.12807

Grotlüschen, A., Buddeberg, K., Dutz, G., Heilmann, L. & Stammer, C. (2019). LEO 2018 – Living with low literacy. Retrieved September 16, 2020 from https://leo.blogs.uni-hamburg.de/wp-content/uploads/2019/07/LEO_2018_Living_with_Low_Literacy.pdf

Gummer, T., Quoß, F., & Roßmann, J. (2019). Does increasing mobile device coverage reduce heterogeneity in completing web surveys on smartphones? *Social Science Computer Review*, *37*(3), 371–384. doi:10.1177/0894439318766836

Gummer, T., & Roßmann, J. (2015). Explaining interview duration in web surveys: A multilevel approach. *Social Science Computer Review*, *33*(2), 217–234. doi:10.1177/0894439314533479

Höhne, J. K. (2021). Are respondents ready for audio and voice communication channels in online surveys? *International Journal of Social Research Methodology*. Advance online publication. doi:10.1080/13645579.2021.1987121

Krebs, D., & Höhne, J. K. (2021). Exploring scale direction effects and response behavior across PC and smartphone surveys. *Journal of Survey Statistics and Methodology*, *9*(3), 477–495. doi:10.1093/jssam/smz058

Lambert, A. D., & Miller, A. L. (2015). Living with smartphones: Does completion device affect survey responses? *Research in Higher Education*, *56*(2), 166–177. doi:10.1007/s11162-014-9354-7

Lugtig, P., & Toepoel, V. (2016). The use of PCs, smartphones, and tablets in a probability-based panel survey: Effects on survey measurement error. *Social Science Computer Review*, *34*(1), 78–94. doi:10.1177/0894439315574248

Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Review*, *31*(6), 725–743. doi:10.1177/0894439313485201

Miller, P. V. (2017). Is there a future for surveys? *Public Opinion Quarterly*, *81*(S1), 205–212. doi:10.1093/poq/nfx008

Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Peterson, G., Griffin, J., LaFrance, J., & Li, J. (2017). Smartphone participation in web surveys. In P. P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg & B. T. West (Eds.), *Total survey error in practice* (pp. 203–233). Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781119041702.ch10

Peytchev, A., & Hill, C. A. (2010). Experiments in mobile web survey design similarities to other modes and unique considerations. *Social Science Computer Review*, *28*(3), 319–335. doi:10.1177/0894439309353037

Poggio, T., Bosnjak, M., & Weyandt, K. (2015). Survey participation via mobile devices in a probability-based online-panel: Prevalence, determinants, and implications for nonresponse. *Survey Practice*, *8*(1), 1–7. doi:10.29115/sp-2015-0002.

Revilla, M., Couper, M. P., & Ochoa, C. (2018). Giving respondents voice? The feasibility of voice input for mobile web surveys. *Survey Practice*, *11*(2), 1–11. doi:10.29115/sp-2018-0007.

Revilla, M., & Höhne, J. K. (2020). Comparing the participation of millennials and older age cohorts in the CROss-National online survey panel and the german internet panel. *Survey Research Methods*, *14*(5), 499–513. doi:10.18148/srm/2020.v14i5.7619

Revilla, M., & Ochoa, C. (2016). Open narrative questions in PC and smartphones: Is the device playing a role? *Quality & Quantity*, *50*(6), 2495–2513. doi:10.1007/s11135-015-0273-2

Revilla, M., Toninelli, D., Ochoa, C., & Loewe, G. (2016). Do online access panels really need to allow and adapt surveys to mobile devices? *Internet Research*, *26*(5), 1209–1227. doi:10.1108/intr-02-2015-0032.

Rust, R. T., & Cooil, B. (1994). Reliability measures for qualitative data: Theory and implications. *Journal of Marketing Research*, *31*(1), 1–14. doi:10.2307/3151942

Strauss, W., & Howe, N. (1991). *Generations: The history of America's future, 1584 to 2069*. New York, NY: William Morrow & Company.

Struminskaya, B., Lugtig, P., Keusch, F., & Höhne, J. K. (2020). Augmenting surveys with data from sensors and apps: Opportunities and challenges. *Social Science Computer Review*. Advance online publication. doi:10.1177/0894439320979951

Struminskaya, B., Weyandt, K., & Bosnjak, M. (2015). The effects of questionnaire completion using mobile devices on data quality. Evidence from a probability-based general population panel. *Methods, Data, Analyses*, *9*(2), 261–292.

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

Tourangeau, R., & Yan, T (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883. doi: 10.1037/0033-2909.133.5.859.

UNESCO Institute for Statistics (2017). Literacy rates continue to rise from one generation to the next. Retrieved January 18, 2022, from http://uis.unesco.org/sites/default/files/documents/fs45-literacy-rates-continue-rise-generation-to-next-en-2017_0.pdf

van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217. doi:10.1093/ijpor/eds021

Wells, T., Bailey, J. T., & Link, M. W. (2014). Comparison of smartphone and online computer survey administration. *Social Science Computer Review*, *32*(2), 238–255. doi:10.1177/0894439313505829

Appendix A

Screenshots of the Questions and Probes on Reasons for (Non)willingness

Figure A1. Screenshot of the question on audio channel (presentation on a PC).

Bei der vorherigen Frage ging es darum, ob Sie grundsätzlich bereit wären, an zukünftigen Befragungen mit Ihrem Smartphone teilzunehmen, um die Fragen vorgelesen zu bekommen.

Ihre Antwort lautete: „**Eher ja**"

Bitte erläutern Sie Ihre Antwort noch etwas näher. Warum haben Sie sich für diese Antwort entschieden?

WEITER

Figure A2. Screenshot of the probe on reasons for (non)willingness to use the audio channel (presentation on a PC).

Wären Sie grundsätzlich bereit, an zukünftigen Befragungen mit Ihrem Smartphone teilzunehmen, um Ihre Antworten mündlich über die Spracheingabe abzugeben?

○ Definitiv ja

○ Eher ja

○ Eher nein

○ Definitiv nein

○ Nutze kein Smartphone

○ Weiß nicht wie das funktioniert

WEITER

Figure A3. Screenshot of the question on voice channel (presentation on a PC).

Bei der vorherigen Frage ging es darum, ob Sie grundsätzlich bereit wären, an zukünftigen Befragungen mit Ihrem Smartphone teilzunehmen, um Ihre Antworten mündlich über die Spracheingabe abzugeben.

Ihre Antwort lautete: „**Definitiv nein**"

Bitte erläutern Sie Ihre Antwort noch etwas näher. Warum haben Sie sich für diese Antwort entschieden?

WEITER

Figure A4. Screenshot of the probe on reasons for (non)willingness to use the voice channel (presentation on a PC).

Appendix B

English Translations of the Independent and Control Variables

Smartphone Use in This Survey: Determined via Paradata.

Smartphone competence: In general, how would you rate your ability to use smartphones? Answer options: (1) very bad, (2) bad, (3) moderate, (4) good, and (5) very good.

General Internet use: How often do you use the Internet, the World Wide Web, or email for private purposes whether at home, at work, or elsewhere? Answer options: (1) more than 2 hours a day, (2) one to 2 hours a day, (3) less than 1 hour a day, (4) several times a week, (5) once a week, and (6) once a month or less.

Internet use via smartphone: Did you use a smartphone to access the Internet during the last three months? Answer options: (1) yes and (0) no.

Belief that Internet facilitates communication: To what extent do you agree or disagree with the following statement? The Internet simplifies communication between people. Answer options: (1) disagree strongly, (2) disagree somewhat, (3) neither/nor, (4) agree somewhat, and (5) agree strongly.

Affinity for technology: To what extent do you agree or disagree with the following statement? It is exciting to try out new technologies or equipment. Answer options: (1) disagree strongly, (2) disagree somewhat, (3) neither/nor, (4) agree somewhat, and (5) agree strongly.

Survey evaluation: difficulty: How difficult did you find the survey? Answer options: (1) not at all, (2) hardly, (3) partly, (4) quite, and (5) very.

Survey evaluation: enjoyment: How did you like the survey? Answer options: (1) not at all, (2) hardly, (3) partly, (4) quite, and (5) very.

Survey evaluation: length: How lengthy did you find the survey? Answer options: (1) not at all, (2) hardly, (3) partly, (4) quite, and (5) very.

Survey evaluation: topic sensitivity: How personal did you find the survey? Answer options: (1) not at all, (2) hardly, (3) partly, (4) quite, and (5) very.

Need for cognition: a. Simply knowing the answer rather than understanding the reasons for the answer to a problem is fine with me. b. I prefer my life to be filled with puzzles that I must solve. c. I would prefer complex to simple problems. d. I think primarily because I have to. Answer options: (1) agree not at all – (4) neither/nor – (7) agree strongly.

Age: In what year were you born? Answer field: [Please enter your year of birth].

Gender: What is your gender? Answer options: (1) male, (2) female, and (3) other.

Educational attainment: What is your highest school degree? Answer options: (1) still attending school, (2) left school without a degree, (3) "Polytechnische Oberschule GDR," completion of eighth or ninth grade, (4) "Polytechnische Oberschule GDR," completion of 10th grade, (5) lower secondary school degree, (6) intermediate secondary school degree, (7) UAS-entrance degree, and (8) university-entrance degree.

Note. The German wordings of all questions including answer options are available from the first author upon request.