Meitinger, Katharina, Sabien van der Sluis, and Matthias Schonlau. 2024. "Keep the Noise down: On the Performance of Automatic Speech Recognition of Voice-Recordings in Web Surveys." *Survey Practice* 17 (February). https://doi.org/10.29115/SP-2023-0022.

ARTICLES

Keep the noise down: On the performance of automatic speech recognition of voice-recordings in web surveys

Katharina Meitinger[1], Sabien van der Sluis[1], Matthias Schonlau[2]

[1] Methods & Statistics, Utrecht University, [2] University of Waterloo

Keywords: voice-recording, automatic speech recognition, validity, accuracy, web survey, open-ended question

https://doi.org/10.29115/SP-2023-0022

Survey Practice

Vol. 17, 2024

Voice-recordings are increasingly implemented in web surveys, but the resulting audio data need to be transcribed before analysis. Since manual coding is too time- and work-intensive, researchers often rely on automatic speech recognition (ASR) systems for the transcription of the voice-recordings. However, ASR tools might create partly incorrect transcriptions and potentially change the content of responses. If the ASR performance (i.e., accuracy and validity) differs by subgroup and contextual factors, a bias is introduced in the analysis of open-ended questions. We assessed the impact of sociodemographic and contextual factors on the accuracy and validity of ASR transcriptions with data from the Longitudinal Internet Studies for the Social Sciences (LISS) panel collected in December 2020. We find that background noise reduces the accuracy and validity of ASR transcriptions. In addition, validity improved when the respondent was alone during the survey. Fortunately, we did not find any evidence of systematic differences across subgroups (age, sex, education), devices or respondent location.

Introduction

Due to technical advancements, the amount of audio data vastly increased over the past decades (Fallgren, Malisz, and Edlund 2018). Automatic speech recognition (ASR) automatically transcribes audio into text data which then can be used for analysis (Ghai and Singh 2012). A recent area of application is social science surveys that provide respondents the possibility to answer parts of the survey with voice-recordings (Gavras and Höhne 2022; Gavras et al. 2022; Höhne 2021; Lenzner and Höhne 2022; Revilla, Couper, and Ochoa 2018; Revilla et al. 2020; Revilla and Couper 2021; Ziman et al. 2018). Voice-recordings potentially reduce survey completion time (Revilla et al. 2020) and improve criterion validity (Gavras and Höhne 2022). These studies rely on ASR because it drastically reduces transcription time and cost (Revilla and Couper 2021; Ziman et al. 2018).

However, the transcription and the voice-recording might differ. The accuracy of ASR transcriptions might be low due to longer, shorter, missing, added text, or compound words (Errattahi et al. 2019; Ghannay, Estève, and Camelin 2020). A common measure of ASR *accuracy* is the word error rate (WER) which is the number of transcription errors divided by the answer

length (Kim et al. 2019; Tancoigne et al. 2022). The higher the WER, the worse the transcription. Additionally, the ASR transcription could change the meaning of transcribed words (i.e., *validity*). Accuracy and validity might differ by context factors and subgroups (e.g., sex and age), which potentially introduces bias in the analysis of open-ended answers. We assume that low accuracy might also deteriorate the validity of the transcription.

This study has two research questions:

> RQ1: Does the accuracy of ASR transcriptions differ by subgroups and context factors?
>
> RQ2: Does the validity of ASR transcriptions differ by subgroups and context factors?

Previous research reported on a gender bias in ASR performance due to an overrepresentation of men in the ASR training database (Garnerin, Rossato, and Besacier 2003). Most ASR models are also trained with young people, which potentially reduces ASR performance for the elderly (Pellegrini et al. 2012).

> *H_1: Accuracy and validity are lower for women than for men.*
>
> *H_2: Accuracy and validity are higher for younger than for older respondents.*

Context factors such as background noise can reduce ASR performance (Morgan, Wegmann, and Cohen 2013; Rodrigues et al. 2019).

> *H_3: Background noise reduces accuracy and validity.*

Research on voice-recordings in surveys is often done in mobile surveys (Revilla et al. 2020; Revilla, Couper, and Ochoa 2018). Voice-recordings can be implemented in web surveys where respondents also answer on PCs, laptops, or tablets. Mobile respondents are more used to voice-recordings than PC or tablet users, which improves ASR performance (Revilla and Couper 2021; Rodrigues et al. 2019). Previous research also indicates that ASR performance decreases with increasing distance between the microphone and the speaker (Renals and Swietojanski 2017; Kim et al. 2019). PC and laptops are therefore potentially more challenging for ASR systems than mobiles due to the need for distant speech recognition.

> *H_4: Smartphone users have a higher accuracy and validity than respondents using other devices.*

Social context might affect ASR performance, too. Nonresponse increased in a voice-recording study when respondents felt they were being overheard (Revilla and Couper 2021). Respondents might whisper answers to avoid bystanders' eavesdropping. At home, respondents may dare to speak louder

and can more easily control the volume of background noises. Background noises might also be less diverse than in public places. ASR systems are trained in very controlled conditions and perform worse with diverse background noises or low volume responses (Kalinli et al. 2010; Morgan, Wegmann, and Cohen 2013).

> *H_5: Accuracy and validity is higher for respondents answering the survey from home than for respondents who are not at home.*
>
> *H_6: Accuracy and validity is higher for respondents who are alone during the survey than for respondents who are not alone.*

Methods and data

Data

Data were collected in December 2020 by the Longitudinal Internet Studies for the Social Sciences (LISS) panel, a probability online panel of the Dutch population (see https://www.lissdata.nl/about-panel). The goal of the study was to assess whether voice-recording studies are feasible in a probability based mixed-device web panel of the general population. Based on a between-subjects design, respondents were randomly assigned to one of the experimental conditions at the beginning of the survey: written response only, voice-recording only, or choice between written and voice-recording. Since the focus of our analysis was on voice-recordings, we allotted more respondents to the voice-recording and choice conditions than to the written response condition. In total, 661 respondents (19.35%) were assigned to the written response condition, 1,306 respondents (38.26%) were assigned to the voice-recording condition, and 1,449 respondents (42.39%) to the choice condition. In the current study, we focus only on respondents that used the voice-recording option.

Average completion time was 5.0 minutes (median 3.7 minutes). Voice-recordings were collected with the browser-based questfox tool (https://questfox.online/en/questmanagement) that uses Google speech-to-text as the ASR system. Once the respondents submitted their voice-recordings, they could not view or edit their transcribed responses.

The initial data set had 3,416 respondents of which 358 received a voice-recording option as the only answering option or selected voice-recording when given the choice and met technical and legal requirements for answering with voice-recordings. In this study, we only included respondents who completed the survey, provided permission to use audio data, and had a voice-recording as well as the transcribed text. Sample size was reduced further because some responses were unintelligible or it was impossible to distinguish between words in the transcription (>20 words not or incorrectly transcribed, 25 cases) or because respondents had included more than one entry (5 cases).

The final sample had 212 respondents of which 100 were female and 112 male. The average age was 48.79 (SD = 17.43), 54.7% of the respondents had a higher education (higher vocational education or university), and 22.2% of respondents answered the survey on a smartphone. Around a quarter of the respondents were alone when responding to the survey (23.1%) and had background noise in their recordings (25.9%). The majority of respondents filled out the survey at home (97.2%).

It is important to note that the respondents included in the final sample are on average higher educated, younger, and more often male than the respondents who were assigned to the choice condition and opted for a written response instead of a voice-recording (higher education: 37.0%, average age: M =53.2 [SD=18.5], women: 54.0%).

Tested Question

For our analysis, we used voice-recordings to a category-selection probe, a special kind of open-ended question (Willis 2005). The closed question was "In general how would you rate the current state of the economy in the Netherlands?" (English translation). The category-selection probe then asked the respondent to "Please explain why you selected [answer at question mentioned before]" (English translation).

Coding Procedure

We developed a coding schema based on previous research on common ASR error types and sources (Errattahi et al. 2019; Ghannay, Estève, and Camelin 2020). It distinguishes between error types (e.g., error in verbs, nouns) and error sources (e.g., background noise, pronunciation errors) and assessed whether the content/meaning of the response changed due to the transcription.

Responses were available as voice-recordings and automatically transcribed text. Both response types were coded for discrepancies by a human coder and partly (19 respondents) recoded by a second coder. The interrater agreement was 86.4%, which is good given the complexity of the coding schema and task (Gisev, Bell, and Chen 2013). All coding disagreements were resolved and integrated into the final data set. We analyzed the data with SPSS 25.

Measures

DEPENDENT VARIABLES

Our dependent variable for the *accuracy* analysis is WER, a count variable of ASR transcription errors by response length in words. A transcription with a WER of .25 means that 25% of the words would need to be altered (via substitutions, deletions, or insertions) to perfectly match the reference transcript. WER can exceed 1.00 because words can appear in the transcription but not in the voice-recording (e.g., voice-recording: 'rentes';

transcription: 'rente is'). WER is a good measure for speech analytics if the WER is below 25% (Park et al. 2008). For the *validity* analysis, we used a binary variable that indicates whether the meaning of the response was changed due to a transcription error or not.

INDEPENDENT VARIABLES

We used sex (reference category: male), age, background noise (reference: no background noise), self-reported mobile usage (reference: no mobile), location of the respondent during the study (reference: at home), and social context (reference: alone) as independent variables. We included education as a binary control variable (reference: high education).

We assessed all assumptions of our regression models. We removed an extreme outlier with a WER exceeding 1000%. We tested for multicollinearity and found very low VIF values (1.04 to 1.14). In addition, non-linear effects of age were tested and found to be non-significant for both models.

Analytical Strategy

For the accuracy analysis, we regressed a count variable with log of the answer length (number of words) as an offset variable in a negative binomial regression model to obtain the WER. Negative binomial regression accounted for the overdispersion present in our data (Hilbe 2014). For the validity analysis, we calculated a logistic regression with change of meaning (yes/no) as the dependent variable.

Results

Accuracy

WER ranged from 0 to 3.33 and the average WER was 0.20 (SD=0.36), which means that it is an appropriate measure of speech analytics for our study. The histogram (see Figure 1) shows the frequencies of word error rates.

When compared to the Poisson model (AIC = 1184.19, BIC = 1211.04), a negative binomial model is indeed more appropriate for our data (AIC = 936.23, BIC = 966.43). The estimate of the dispersion coefficient B negative binomial was 0.951 (SD = 0.15), indicating that there is a small amount of overdispersion present (see Table 1, Model Accuracy). The negative binomial regression model was statistically significant ($\chi^2(7)=21.26$, $p=.003$). Only background noise significantly predicted WER ($p<.001$). Responses with background noise had a 2.08-times higher WER than responses without background noise. Neither sex, age, mobile usage, location, nor social context significantly predicted WER. Education was also not significant.

Validity

In 60.8% of the analyzed responses, the meaning of at least one word changed due to the ASR transcription (see Figure 2).

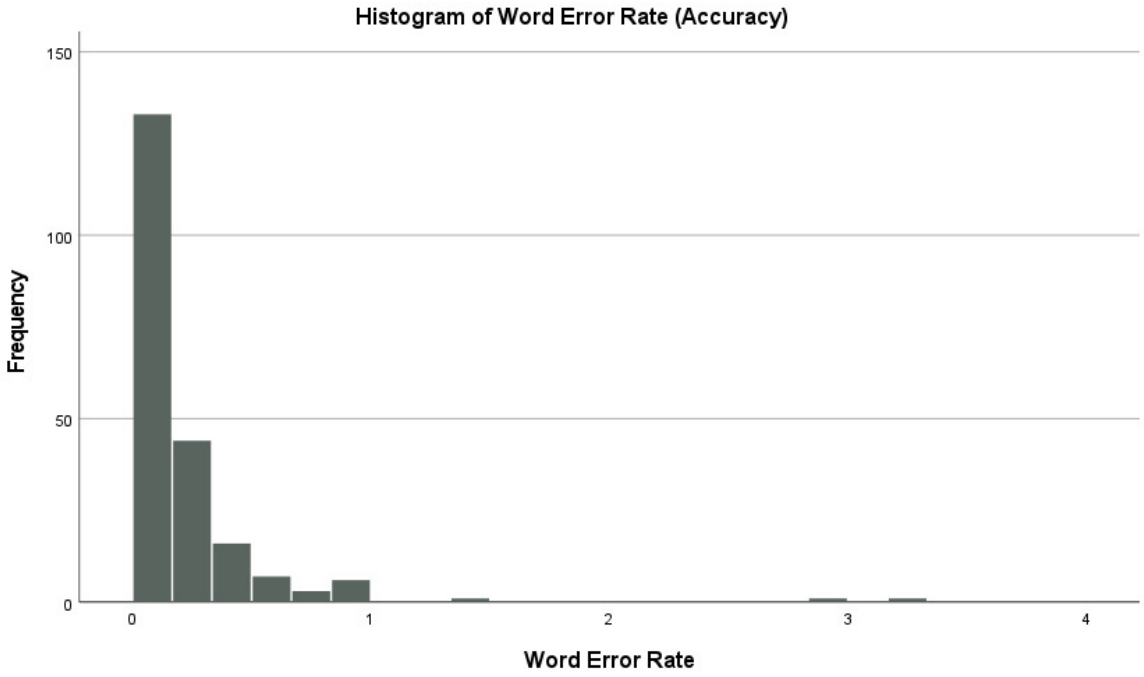

Figure 1. Histogram of word error rate

Table 1. Negative binomial regression of word error rate and logistic regression of changed meaning.

	Accuracy: WER N=212			Validity: Changed Meaning N=212		
	B	(95% Wald CI)	IRR	B	OR	(95% CI OR)
(Intercept)	-1.56***	[-2.15, -0.97]	0.21	1.13*	3.09	
Sex (Ref. category: Male)						
Female	-0.19	[-0.53, 0.16]	0.83	0.47	1.60	[0.88, 2.92]
Age	-0.01	[-0.02, 0.00]	0.99	-0.02	0.99	[0.97, 1.00]
Background noise (Ref.: No)						
Yes	0.76***	[0.34, 1.12]	2.08	0.80*	2.21	[1.08, 4.53]
Mobile usage (Ref.: No)						
Yes	-0.26	[-0.69, 0.16]	0.77	-0.60	0.55	[0.27, 1.12]
Location (Ref.: At home)						
Not at home	-0.66	[-1.58, 0.27]	0.52	-0.17	0.84	[0.14, 5.21]
Social context (Ref.: Not alone)						
Alone	-0.14	[-0.54, 0.27]	0.87	-0.84*	0.43	[0.22, 0.86]
Education (Ref.: High)						
Low	0.29	[-0.04, 0.62]	1.34	-0.11	0.90	[0.50, 1.61]
(Negative binomial)	0.95	[0.70, 1.29]				

NOTE: IRR: Incidence Rate Ratio. OR: Odds Ratio *p < .05. **p < .01. ***p < .001.

The logistic regression model (see Table 1, Model Validity) was statistically significant ($\chi^2(7)=16.70$, $p=.019$) and 66.5% of respondents were correctly classified (Nagelkerke R^2 =.103). Responses with background noise had 2.21-times higher odds that the meaning of the response changed than

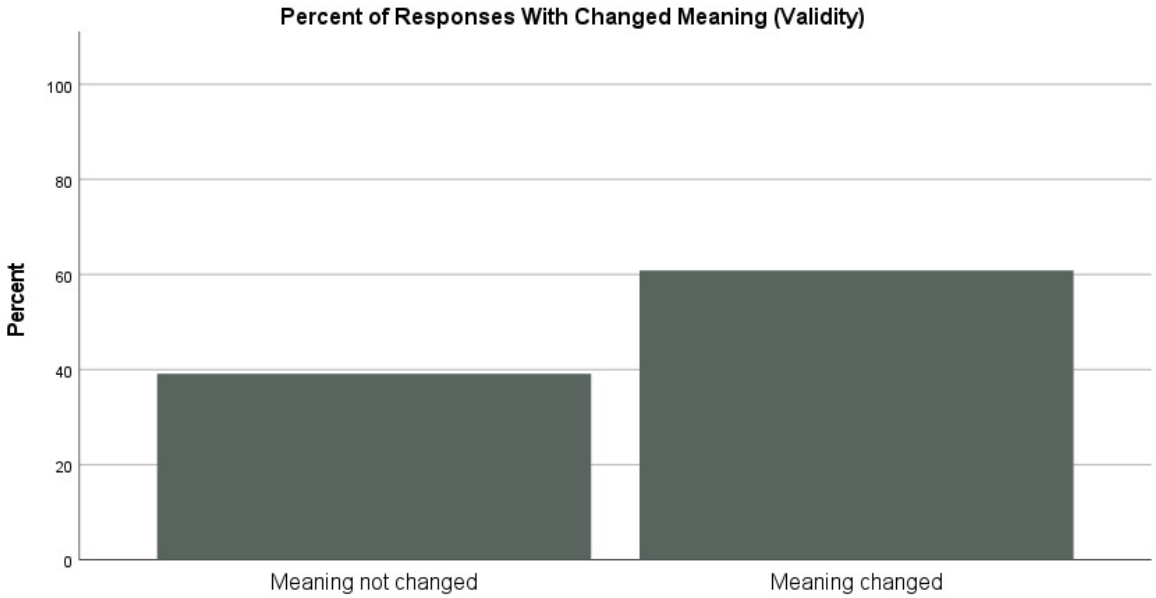

Figure 2. Percentage of responses with a changed meaning

responses without background noise (p=.030). Additionally, respondents who were alone had lower odds of a changed meaning than respondents who were not alone (OR: 0.43, p=.017). None of the other predictors were significant. The sample had enough (80%) power to distinguish between an 0.188 difference in proportion, e.g., 0.5 versus 0.688.

Discussion

We have evaluated the impact of group characteristics and context factors on the accuracy and validity of ASR transcriptions of voice-recordings in social science web surveys.

We were concerned that performance differences introduce a bias in the analysis of open-ended questions. Accuracy and validity did not significantly differ by sex and age (H_1 and H_2 not confirmed). Although this is contrary to previous research (Garnerin, Rossato, and Besacier 2003; Pellegrini et al. 2012), this is good news because we did not find evidence for bias in the ASR transcriptions for these subgroups. The likely explanation for our null findings is that the performance of ASR systems improved since previous research has been conducted (Kim et al. 2019).

Additionally, we assessed different context factors that potentially deteriorate ASR performance. Contrary to our hypotheses, we found no significant effect for smartphone usage (H_4 not confirmed), which is good news for the implementation of voice-recordings in mixed device surveys. Location (i.e., not at home) was also not a significant predictor in both regression models (H_5). The most likely explanation for this null finding is that the sample did

not contain sufficient respondents that filled out the survey outside of their home. Whereas accuracy was not significantly influenced by social context (i.e., not alone), being alone significantly improved the validity of ASR transcriptions (H_6 partly confirmed). The optimistic interpretation regarding accuracy would be that ASR performance is nowadays performing well in different locations and social contexts. This is in line with current ASR research that reports sinking WER rates in various conditions and languages due to training effects (Kim et al. 2019; Proksch, Wratil, and Wäckerle 2019; Tancoigne et al. 2022). The more pessimistic interpretation would be that situations might be more complex than assumed. For example, being at home could mean a calm surrounding but could also be quite noisy in the case of a family with young children. While social context did not matter regarding accuracy, validity was affected, indicating that ASR systems might perform differently depending on whether respondents are alone or not. One explanation would be that respondents dare to speak louder when they are alone, which improves ASR performance (Morgan, Wegmann, and Cohen 2013). This also underlines that an assessment of ASR performance should not only rely on WER but should be supplemented by some content analysis (see also Proksch, Wratil, and Wäckerle 2019). In line with previous research (Morgan, Wegmann, and Cohen 2013; Rodrigues et al. 2019), background noise significantly reduced accuracy and validity (H3 confirmed). When coding the voice-recordings, we also coded what type of background noise appeared and found that background noises are rarely due to social settings (e.g., café, workplace sounds) or outside locations (e.g., traffic noise) but due to, for example, television sounds or white noise.

Several of these background noises can potentially be reduced by a proactive survey design. For example, respondents could be informed about the voice-recording in the invitation letter and provided with some advice regarding optimal recording conditions to fill out the survey (e.g., switch off TV, optimal set-up of the microphone). During the survey, precise feedback on microphone performance could be given with targeted advice on how to resolve potential issues. It is possible that improved noise cancellation technology might reduce the impact of the background noise in the future.

Even when the remaining issues with ASR performance are resolved, several challenges for the implementation of voice-recordings in web surveys remain. In our study, only a few respondents show a preference to record their response instead of typing it, and the respondents who are willing to record or who opt for this option are slightly younger, higher educated and more often men than women. This difference might even be more pronounced in a cross-sectional survey design. Our respondents were members of the LISS panel and are therefore used to answering a web survey, sometimes also in nontraditional response formats and tasks (e.g., collection of biomarkers, Avendano 2010). Previous research revealed that voice-recordings potentially reduce survey completion time (Revilla et al. 2020) and can improve criterion

validity (Gavras and Höhne 2022). However, to reach its full potential the willingness to do voice-recordings needs to be increased and potential biasing effects of its implementations need to be addressed.

Limitations and Future Research

We assessed the performance of one ASR system (Google speech-to-text) with a question on the economy. Further research could investigate different topics (e.g., sensitive topics) and performance differences of ASR systems (e.g., YouTube, Microsoft Azure) across groups and context factors.

Lead author's contact details

Katharina Meitinger, Department of Methods & Statistics, Utrecht University, Padualaan 14, 3584 CH Utrecht, The Netherlands. Email: k.m.meitinger@uu.nl

Acknowledgments

This research was supported by the Social Sciences and Humanities Research Council of Canada under grant (SSHRC # 435-2013-0128) and (SSHRC #435-2021-0287). We are thankful for the feedback and helpful suggestions made by the Survey Practice reviewers and editors.

Submitted: June 26, 2023 EST, Accepted: October 21, 2023 EST

REFERENCES

Avendano, Mauricio. 2010. "Can Biomarkers Be Collected in an Internet Survey? A Pilot Study in the LISS Panel 1." In *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*, edited by M. Das, P. Ester, and L. Kaczmirek, 1st ed., 371–412. New York/East Sussex: Routledge. https://doi.org/10.4324/9780203844922-15.

Errattahi, Rahhal, Asmaa EL Hannani, Thomas Hain, and Hassan Ouahmane. 2019. "System-Independent Asr Error Detection and Classification Using Recurrent Neural Network." *Computer Speech & Language* 55 (May):187–99. https://doi.org/10.1016/j.csl.2018.12.007.

Fallgren, P., Z. Malisz, and J. Edlund. 2018. "Bringing Order to Chaos: A Non-Sequential Approach for Browsing Large Sets of Found Audio Data." In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Garnerin, M., S. Rossato, and L. Besacier. 2003. "Gender Representation in Open-Source Speech Resources." *arXiv* Preprint:2003.08132.

Gavras, Konstantin, and Jan Karem Höhne. 2022. "Evaluating Political Parties: Criterion Validity of Open Questions with Requests for Text and Voice Answers." *International Journal of Social Research Methodology* 25 (1): 135–41. https://doi.org/10.1080/13645579.2020.1860279.

Gavras, Konstantin, Jan Karem Höhne, Annelies G. Blom, and Harald Schoen. 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: Statistics in Society* 185 (3): 872–90. https://doi.org/10.1111/rssa.12807.

Ghai, Wiqas, and Navdeep Singh. 2012. "Literature Review on Automatic Speech Recognition." *International Journal of Computer Applications* 41 (8): 42–50. https://doi.org/10.5120/5565-7646.

Ghannay, Sahar, Yannick Estève, and Nathalie Camelin. 2020. "A Study of Continuous Space Word and Sentence Representations Applied to ASR Error Detection." *Speech Communication* 120 (June):31–41. https://doi.org/10.1016/j.specom.2020.03.002.

Gisev, Natasa, J. Simon Bell, and Timothy F. Chen. 2013. "Interrater Agreement and Interrater Reliability: Key Concepts, Approaches, and Applications." *Research in Social and Administrative Pharmacy* 9 (3): 330–38. https://doi.org/10.1016/j.sapharm.2012.04.004.

Hilbe, Joseph M. 2014. *Modeling Count Data*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781139236065.

Höhne, Jan Karem. 2021. "Are Respondents Ready for Audio and Voice Communication Channels in Online Surveys?" *International Journal of Social Research Methodology* 26 (3): 335–42. https://doi.org/10.1080/13645579.2021.1987121.

Kalinli, O., M. L. Seltzer, J. Droppo, and A. Acero. 2010. "Noise Adaptive Training for Robust Automatic Speech Recognition." *IEEE Transactions on Audio, Speech, and Language Processing* 18 (8): 1889–1901. https://doi.org/10.1109/tasl.2010.2040522.

Kim, J.Y., C. Liu, R.A. Calvo, K. McCabe, S.C. Taylor, B.W. Schuller, and K. Wu. 2019. "A Comparison of Online Automatic Speech Recognition Systems and the Nonverbal Responses to Unintelligible Speech." *arXiv* preprint:1904.12403.

Lenzner, Timo, and Jan Karem Höhne. 2022. "Who Is Willing to Use Audio and Voice Inputs in Smartphone Surveys, and Why?" *International Journal of Market Research* 64 (5): 594–610. https://doi.org/10.1177/14707853221084213.

Morgan, N., S. Wegmann, and J. Cohen. 2013. "What's Wrong With Automatic Speech Recognition (ASR) and How Can We Fix It?" Berkeley, CA: International Computer Science Institute.

Park, Youngja, Siddharth Patwardhan, Karthik Visweswariah, and Stephen C. Gates. 2008. "An Empirical Analysis of Word Error Rate and Keyword Error Rate." In *Proc. Interspeech*, 2070–73. ISCA. https://doi.org/10.21437/interspeech.2008-537.

Pellegrini, Thomas, Isabel Trancoso, Annika Hämäläinen, António Calado, Miguel Sales Dias, and Daniela Braga. 2012. "Impact of Age in ASR for the Elderly: Preliminary Experiments in European Portuguese." In *Advances in Speech and Language Technologies for Iberian Languages*, 139–47. Berlin: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-35292-8_15.

Proksch, Sven-Oliver, Christopher Wratil, and Jens Wäckerle. 2019. "Testing the Validity of Automatic Speech Recognition for Political Text Analysis." *Political Analysis* 27 (3): 339–59. https://doi.org/10.1017/pan.2018.62.

Renals, Steve, and Pawel Swietojanski. 2017. "Distant Speech Recognition Experiments Using the AMI Corpus." In *New Era for Robust Speech Recognition: Exploiting Deep Learning*, 355–68. Cham: Springer. https://doi.org/10.1007/978-3-319-64680-0_16.

Revilla, Melanie, and Mick P. Couper. 2021. "Improving the Use of Voice Recording in a Smartphone Survey." *Social Science Computer Review* 39 (6): 1159–78. https://doi.org/10.1177/0894439319888708.

Revilla, Melanie, Mick P. Couper, Oriol J. Bosch, and Marc Asensio. 2020. "Testing the Use of Voice Input in a Smartphone Web Survey." *Social Science Computer Review* 38 (2): 207–24. https://doi.org/10.1177/0894439318810715.

Revilla, Melanie, Mick P. Couper, and Carlos Ochoa. 2018. "Giving Respondents Voice? The Feasibility of Voice Input for Mobile Web Surveys." *Survey Practice* 11 (2): 2713. https://doi.org/10.29115/sp-2018-0007.

Rodrigues, Ana, Rita Santos, Jorge Abreu, Pedro Beça, Pedro Almeida, and Sílvia Fernandes. 2019. "Analyzing the Performance of ASR Systems: The Effects of Noise. Distance to the Device. Age and Gender." In *Proceedings of the XX International Conference on Human Computer Interaction (Interacción'19), June 25-28, 2019, Donostia, Gipuzkoa, Spain*. New York, USA: ACM. https://doi.org/10.1145/3335595.3335635.

Tancoigne, Elise, Jean Philippe Corbellini, Gaëlle Deletraz, Laure Gayraud, Sandrine Ollinger, and Daniel Valero. 2022. "Un mot pour un autre? Analyse et comparaison de huit plateformes de transcription automatique." *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 155 (1): 45–81. https://doi.org/10.1177/07591063221088322.

Willis, G. 2005. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage Publications.

Ziman, Kirsten, Andrew C. Heusser, Paxton C. Fitzpatrick, Campbell E. Field, and Jeremy R. Manning. 2018. "Is Automatic Speech-to-Text Transcription Ready for Use in Psychological Experiments?" *Behavior Research Methods* 50 (6): 2597–2605. https://doi.org/10.3758/s13428-018-1037-4.

SUPPLEMENTARY MATERIALS

Coding scheme

Download: https://www.surveypractice.org/article/90388-keep-the-noise-down-on-the-performance-of-automatic-speech-recognition-of-voice-recordings-in-web-surveys/attachment/187482.docx