Leschanowsky, Anna; Das, Sneha; Bäckström, Tom; Perez Zarazaga, Pablo

Perception of Privacy Measured in the Crowd−Paired Comparison on the Effect of Background Noises

Published in:
Proceedings of Interspeech

DOI:
10.21437/Interspeech.2020-2299

Published: 01/01/2020

Document Version
Publisher's PDF, also known as Version of record

INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Perception of Privacy Measured in the Crowd – Paired Comparison on the Effect of Background Noises

Anna Leschanowsky, Sneha Das, Tom Bäckström, Pablo Pérez Zarazaga

Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland
(annakatharina.leschanowsky, sneha.das, tom.backstrom, pablo.perezzarazaga)@aalto.fi

Abstract

Voice based devices and virtual assistants are widely integrated into our daily life, but the growing popularity has also raised concerns about data privacy in processing and storage. While improvements in technology and data protection regulations have been made to provide users a more secure experience, the concept of privacy continues to be subject to enormous challenges. We can observe that people intuitively adjust their way of talking in a human-to-human conversation, an intuition that devices could benefit from to increase their level of privacy. In order to enable devices to quantify privacy in an acoustic scenario, this paper focuses on how people perceive privacy with respect to environmental noise. We measured privacy scores on a crowdsourcing platform with a paired comparison listening test and obtained reliable and consistent results. Our measurements show that the experience of privacy varies depending on the acoustic features of the ambient noise. Furthermore, multiple probabilistic choice models were fitted to the data to obtain a meaningful ordering of noise scenarios conveying listeners' preferences. A preference tree model was found to fit best, indicating that subjects change their decision strategy depending on the scenarios under test.

Index Terms: privacy in speech communication, crowdsourcing, perception of privacy, acoustic environment

1. Introduction

The popularity of voice based devices and virtual assistants has grown and they have been widely integrated into our daily life. Individual actions, such as performing tasks and using services, are fundamentally influenced by those devices, and humans can benefit emotionally from interacting with chatbots in a similar way as when interacting with other humans [1, 2]. However, privacy concerns around smart devices, specifically conversational agents, have been widely raised [3, 4], whereby the access to sensitive speech information needs to be addressed on multiple levels. Action was taken by the EU with the general data protection regulations (GDPR) [5], which has also led to an increased awareness among users regarding personal data and their distribution. Nevertheless, a recent survey investigating the awareness on storage behaviour of voice recordings among users of smart speakers in the US showed that users lack an understanding of how their personal data can be processed and stored [6]. This is crucial as users' perception of risks and benefits influences the willingness to adopt a technology. Therefore, privacy has been addressed in connection with technology acceptance models and concepts of trust in smart personal assistants [7, 8].

In human-to-human interaction we can observe that people intuitively modify their way of talking depending on 1. the content, 2. the level of trust they have in their conversation partners and 3. the environment. Communication Privacy Management Theory (CPM) provides a systematic approach to how humans make decisions to disclose or protect their private information by using the concept of privacy boundaries and rules for third party disclosures when interacting with others [9]. Sannon et al. [3] extended that theory towards human-agent interactions to understand humans' privacy expectations of agents and possible violations of privacy during interaction. Moreover, in multimedia communication, especially video conferencing systems, three major factors for users' perception of privacy have been identified: Information Sensitivity, Receiver and Usage [10].

Environmental influence on revealing private information has so far been mainly addressed in the field of room acoustics, where acoustic properties of the environment are adjusted to limit the propagation of voice [11]. This is especially important when designing open offices as productivity can be increased whenever people are not able to understand background speech [12]. Although privacy may not be the major intent in this context, similar measures, such as articulation index and speech transmission index, can be used to quantify the level of privacy. However, those measures do require controlled conditions and specialized equipment [11]. If our devices were able to easily quantify the level of privacy with respect to surroundings, they could adapt their strategy of sharing information between devices and the communication environment in a similar way as humans do. Consequently, this could increase the level of trust between users and technology.

In previous work a conversational speech corpus was presented to quantify the experience of privacy within different real-life scenarios [13]. While the focus of the work was the compilation of the corpus, a first analysis showed that different environments influence human perception on privacy. Building on those findings, in the current study we carried out a crowdsourced listening test to enable in-depth analysis on a sufficiently large sample size and to analyse features that people use to quantify how much information they are willing to reveal in the specific surrounding. Our purpose was to investigate the perception of privacy with respect to ambient noise; the experimental setup and listening test design are described in section 2 and 3, respectively. Section 4 focuses on the consistency of the obtained results. We analyse them, in section 5 by a preference choice analysis of different environmental background noises as well as a comparison of different elimination-by-aspects (EBA) models showing how different acoustic environmental features influence peoples' perception of privacy.

2. Listening Test Material

For our listening test, we used 50 speech files from the TIMIT database [14] and 3 noise files from the QUT database [15] to produce the noisy stimuli. As we wanted the participants to focus on the ambient noise rather than the content, speech files were chosen to less likely be interpreted contextually during the

http://dx.doi.org/10.21437/Interspeech.2020-2299

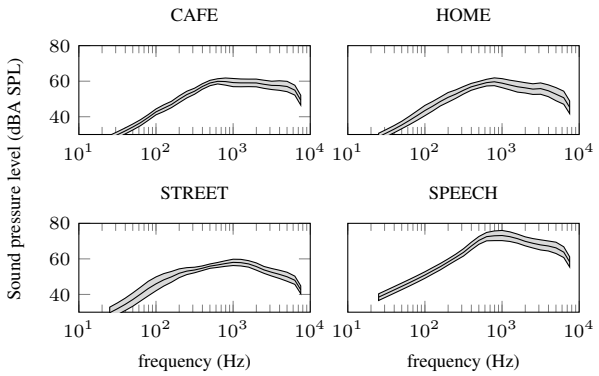

Figure 1: *Sound pressure levels (solid lines) over third octave bands for the equalized noise files and speech files and their standard deviations (shaded area).*

test. For example, sentences expressing facts were chosen over sentences containing a persons's name or opinion. Moreover, samples were balanced in gender of speaker and dialects. The three noise files, coffee shop, living room and street noise were chosen to be clearly distinguishable and close to real-life scenarios. Each speech file was mixed with an individual noise sample of each noise scenario. Therefore, 50 files were sampled randomly from the coffee shop and street noise file but had to be chosen manually for the living room noise. In this case, acoustic characteristics varied considerably over time so the manual choice insured that all participants would listen to the same acoustic features.

Since differences in loudness can influence the participants' decision, we applied normalization on the individual speech and noise files before creating the noisy stimuli. As we wanted the stimuli of the test to be as close as possible to conversational real-life scenarios we based our normalization on measured broadband SPL of speech recorded for different noise levels [16]. The average A-weighted sound pressure level was calculated for each of the three noise scenarios. The difference in sound pressure level was up to 10 dBA so each individual noise file was equalized and some attenuation was applied. After normalization, the average SPL was 55 dBA $\pm$ 2 dBA for the street and 55 dBA $\pm$ 2.4 dBA for the coffee and living room scenarios. Similarly, the speech signals were equalized but sound pressure level was increased so that the average SPL was 62 dBA $\pm$ 1.6 dBA.

Fig. 1 shows the average root mean square level in third-octave bands for the three noise environments and the standard deviation. We will refer to them as "coffee", "home" and "street" condition. The noisy mix signal was achieved by simply adding the pre-processed speech and noise signals. The averaged SNRs for the coffee, home and street noise resulted in 7.9 dB, 5.9 dB and 4.8 dB.

3. Test Design and Procedure

To evaluate the influence of acoustic features on the perception of privacy, we chose a forced-choice paired comparison listening test without reference signal, also known as preference test. By design, participants have to focus on fewer stimuli simultaneously compared to multi-stimulus tests and reliability can be measured more easily [17]. This is especially important when recruiting from crowdsourcing platforms. Participants were recruited using Prolific [18], a platform producing high-quality

Table 1: *Test questions to evaluate the perception of privacy.*

Q1:	Are you more likely to share a secret, in normal voice, in the acoustic environment A or B?
Q2:	If you share a secret in these acoustic environments, will you share it with a louder voice in environment A or B?
Q3:	If an eavesdropper is present, is it more likely for them to hear your normal voice in the acoustic environment A or B?

data with more diverse participants compared to Amazon Mechanical Turk and CrowdFlower (since March 2018 rebranded as Figure Eight) [19]. The combination of the three different scenarios and the original speech signal resulted in 6 pairs in total which were presented to each listener in random order. Tests were performed using an extension of the browser-based listening-test framework webMUSHRA with three questions (see Table 1) displayed on the same page [20]. Those questions, set up to quantify the perception of privacy were taken from previous work [13]. Participants could use the button "Play" to listen to either scenario "A" or "B" and switch seamlessly between them. Moreover, listeners were allowed to listen as often as they liked and rate the questions in any order they liked. Nevertheless, they were forced to listen at least once to the scenarios and to rate each question before continuing to the next page. Following the guidelines of [21] we included two Gold Standard questions where we asked participants to either choose "A" or "B" for all of the questions on that page. In case of those verification questions, the same clean speech signal was presented as stimuli "A" and "B". Again following [21], a questionnaire asking for demographic data was presented at the end of the test. In total, 100 participants were recruited with a completion time of 5 to 10 minutes and an average reward of 5 £ per hour. None of the listeners had participated in one of our tests related to the same subject before.

4. Consistency Checks

When conducting crowdsourced experiments the experimenter has less control over the experimental execution than in lab-based experiments. Therefore we will first investigate the reliability of our participants before carrying out a preference choice analysis and fitting different probabilistic choice models.

A first attempt to remove unreliable listeners from our test results is based on the two verification questions that were added to the listening test and on the completion code each listener had to submit on the Prolific platform. The correct completion code was only available after finishing the whole test. We excluded 24 listeners who failed one or both of the verification questions or submitted the wrong completion code to our study. While the majority of the 76 remaining listeners fall in the age group of 18 to 24, gender bias was almost entirely avoided (55% male, 45% female).

Reliability of participants can furthermore be assessed by checking for consistent ratings. The property of transitivity of preferences states that whenever a participant prefers scenario A to B and B to C then they furthermore prefer A to C. Otherwise we call the triad intransitive or a circular triad. Kendall [22] refers to the property of transitivity in a more general way as consistency in preferences. In the following, consistency is checked for each participant individually before transitivity violations and agreement amongst listeners are examined.

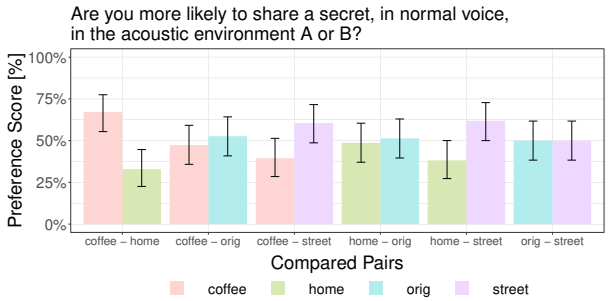

Figure 2: *Preference Score given in percentage with 95% CI for each compared pair for test question 1.*

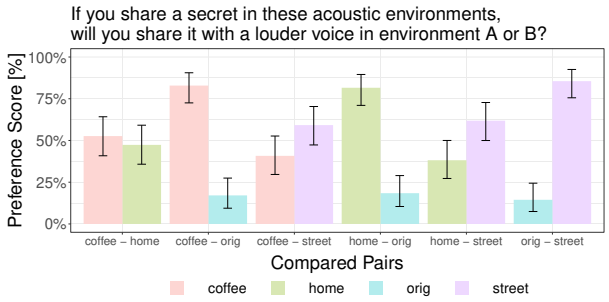

Figure 3: *Preference Score given in percentage with 95% CI for each compared pair for test question 2.*

Individual consistency checks were performed based on Kendall's coefficient of consistence [22]. We computed Kendall's ζ for each participant and question resulting in $(0.83, 0.875, 0.89)$ in average for question 1 to 3. While the average value suggested individual consistent results, we noticed a large variance in ζ values. Individual inconsistency is widely accepted as a measure to identify unreliable listeners. However, not only would the unreliability of the listeners influence the coefficient but also the stimuli under test [22]. Inconsistency can be a sign that the objects are indistinguishable for the listener, or that the dependent variable may not have linear property. Therefore comparisons may not reduce to simple rank ordering. Further analysis showed that the number of participants with a low coefficient of inconsistence resulted in 28 out of 76. This suggests that we might have asked the participants to perform a difficult task rather than that they are not able to express their preferences. Especially, expressing opinion on how likely one would be sharing a secret in a specific environment seemed to be rather difficult to answer for our participants, as most of the circular triads (N=18) were found for the first question. Moreover, circular triads were distributed over participants and questions so that none of our listeners had a coefficient of consistence close to zero throughout all questions. Due to these reasons, we did not exclude results for further analysis and continue with an overall consistency check.

Global consistency checks are conducted by looking at the stochastic transitivity and computing Kendall's coefficient of agreement. Stochastic transitivity measures take into account variable behaviour of the decision-maker and therefore consider that intransitive choices can occur with a certain probability. The results were summarized in a (4 x 4) preference matrix in which for each scenario i that had been preferred over the scenario j the entry t_{ij} was increased by 1. Based on this preference matrix we checked for stochastic transitivity using the

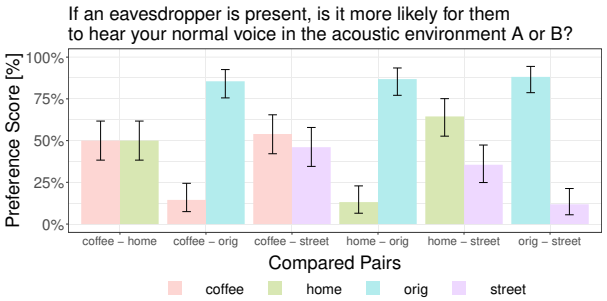

Figure 4: *Preference Score given in percentage with 95% CI for each compared pair for test question 3.*

R package 'eba' [23, 24]. We computed weak (WST), moderate (MST) and strong (SST) stochastic transitivity. Let P_{ij} be the empirical probability that scenario i is chosen over scenario j. When $P_{ij} \geq 0.5$ and $P_{jk} \geq 0.5$ for three scenarios S_i, S_j and S_k, then WST is satisfied if $P_{ik} \geq 0.5$, MST is satisfied if $P_{ik} \geq min\{P_{ij}, P_{jk}\}$ and SST is satisfied if $P_{ik} \geq max\{P_{ij}, P_{jk}\}$.

While violations of MST or SST are less severe, systematic violations of WST may hinder the representation of the paired comparison results on a global preference scale. The number of violations was counted for each question in our test. Ratings for the first question showed 1 violation of MST and 2 violations of SST whereas question 2 and 3 showed no violations.

Moreover, we computed Kendall's u-coefficient using the preference matrix for each question. In case that all participants agree, the coefficient results in a maximum of 1. With the number of participants $M = 76$ we can compute the minimal u-coefficient as $u_{min} = -1/(M-1) = -0.013$. The computed u-coefficients resulted in (0.024, 0.228, 0.277) for question 1 to 3 respectively. To test if the agreement was caused by chance, we carried out a χ^2 test. It suggested that our u-value differs significantly ($p < 0.05$) from the one obtained when ratings were caused by chance for all three questions. Due to the relatively low u-value obtained for the results of the first question, we can assume that the likeliness of sharing a secret was answered rather subjectively compared to the results for the second and third question.

Based on those results we conclude that all our participants after postscreening were capable of making judgements and that preferences were not assigned randomly. Moreover, inconsistency of individual judgements can be seen as an expression of individual preferences possibly influenced by our test design.

5. Results

Our overall hypothesis stated that environmental noise influences humans' perception of privacy. Figures 2, 3 and 4 show the preference scores in percentage for each tested comparison and question. A χ^2 test with Bonferroni correction for multiple testing showed a significant effect for how likely people are to share a secret for scenarios "coffee" and "home" ($p < 0.05$). Effect size was computed based on Cohen's w [25] and showed a medium effect ($w > 0.3$). When it comes to the questions about loudness and eavesdropping, comparisons between each noisy scenario and the original clean speech signal showed a significant effect ($p < 0.01$). Again effect size was computed for the pairs that showed a significant difference resulting in strong effects ($w > 0.5$). Paired comparison tests come with some disadvantages compared to rank order test designs. While

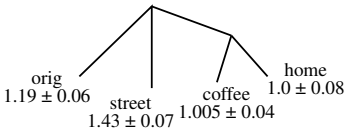

Figure 5: *Schematic Preference Tree Representation. Estimated ratio scale values are normalized with respect to the home scenario and shown with their 95% CI.*

the participant has a relatively simple task of making a preference choice, it is on the experimenter to construct a subjective scale and model decision strategies [26]. One of the widely used probabilistic choice models in psychoacoustic and QoE evaluation ([17],[27]) is the Bradley-Terry-Luce (BTL) model ([28], [29]). It was shown that under certain assumptions the result of a paired comparison can be established on a ratio-scale. However, the rather simple relationship between preference probabilities and scale values of the final model has one major drawback. BTL requires "context independence" of the paired comparison judgement which means that listeners judgement is based on the same auditory feature independent of the stimuli under test. However, if participants rate the pair "coffee - street" based on the auditory feature "background speech present/not present", they might need to adapt their decision strategy for "coffee - home" comparison, as background speech is present in both of those scenarios. A less restrictive model, the Elimination-by-aspects (EBA) was first introduced by Tversky [30]. Considering that attributes of a certain stimulus define the subjects' preferences, EBA's use a modified version of BTL so that the obtained scale values reflect only the influence of features that are present within one stimulus and absent in the compared one. Preference Trees [31] were later introduced as a special case of EBA's, where features follow a hierarchical order. Moreover, it turns out that BTL can be seen as a special case of EBA if only one unique attribute characterizes each stimulus [30]. We computed a BTL model as well as a Pretree model for the results of each question in our test using [24].

The likelihood-ratio test for the three BTL models resulted in $\chi^2_1(3) = 5.95, p_1 = 0.11, \chi^2_2(3) = 0.19, p_2 = 0.98$ and $\chi^2_3(3) = 1.61, p_3 = 0.66$. Moreover, stimulus equality was tested based on the comparison with a model where all stimuli are perceived equally. The hypothesis that all scenarios are equal in terms of preference can be clearly rejected ($p < 0.05$). Based on [24] we would reject the model if the p-value is less than 10%. Accordingly, we could accept all the BTL models, but following the idea that our stimuli might not be context independent we will fit a Pretree Model to our results. The best one that was found for the first question ratings is pictured in Fig. 5. The path lengths starting from the root node are proportional to the scale values, while the lengths of the branches show the degree of similarity between scenarios connected to the node. This model was fitted to the ratings obtained for each question with Likelihood-ratio test results of $\chi^2_1(2) = 1.38, p_1 = 0.50, \chi^2_2(2) = 0.19, p_2 = 0.91$ and $\chi^2_3(2) = 6.36, p_3 = 0.04$. While this model seems to fit better for the first question, it shows no difference for the second and performs even worse for the last question. Because of the fact that our two models are nested, we perform a likelihood ratio test for comparison [24]. The results show that we can reject the BTL model in favor of the Pretree for the first question ($p < 0.05$) but not for the second and the third. That matches our goal to fit the model to our first question test results without considering the other preference matrices. In addition, Akaike's information criterion [32] confirmed this observation with lower scores stating a better model fit (Table 2).

Table 2: *Akaike's information criterion for the BTL and Pretree model applied to test results of each question.*

Question/Model	BTL	Pretree
Question 1	40.5	37.9
Question 2	33.1	35.1
Question 3	33.9	40.7

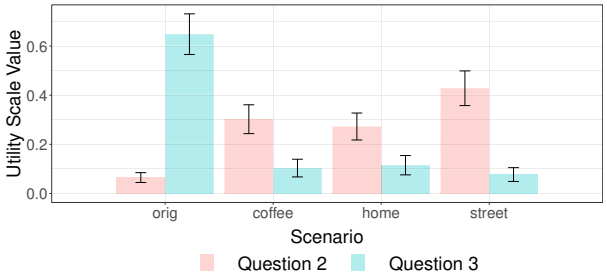

Figure 6: *Ratio scale preference estimated by the BTL model with errorbars showing 95% CI for test question 2 and 3.*

The preference tree model resulted in a better fit for the first question, but the BTL model outperformed any other tested preference model for the second and the third question. We conclude that listeners use the same perceptional feature to compare stimuli when they judge loudness or how likely an eavesdropper can hear their voice. Instead, this linear ranking does not hold with respect to the likeliness of sharing a secret. Here, dependent on the pair of stimuli, listeners adapt their decision strategy accordingly.

Given the preference tree structure, the likeliness of sharing a secret can be estimated on a ratio-scale. The scale values for each scenario are given in Fig. 5 together with their 95% confidence intervals. To improve interpretation, we normalize with respect to the "home" scenario. The BTL model coefficients for the second and third question are shown in Fig. 6 where utility scale values for each model sum up to unity.

6. Conclusion

In a crowdsourced paired comparison listening test setup we measured perception of privacy with respect to three different environmental noise scenarios. The crowdsourced test was found to provide reliable results and that acoustic information does affect the listeners' perception of privacy. When asked for the likeliness to share a secret, participants rated rather subjectively. However, results showed significant differences when comparing coffee shop and home noise, scenarios that varied in perceivable background speech. Furthermore, a BTL probabilistic choice model could be successfully fitted to the ratings of the questions addressing loudness and the possibility of eavesdropping. This indicates that listeners use the same auditory feature to evaluate the different noise scenarios. However, when asking directly for the likeliness to share a secret in a certain scenario, the preetree model showed a better fitting. That indicates that people adapt their decision strategy based on the scenarios under comparison and their acoustic features. In our setup, the scenarios could be grouped according to their acoustic feature of having background speech present.

7. References

[1] G. McLean and K. Osei-Frimpong, "Hey Alexa ... examine the variables influencing the use of artificial intelligent in-home voice assistants," *Computers in Human Behavior*, vol. 99, pp. 28 – 37, 2019.

[2] A. Ho, J. Hancock, and A. S. Miner, "Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot," *Journal of Communication*, vol. 68, no. 4, pp. 712–733, 2018.

[3] S. Sannon, B. Stoll, D. DiFranzo, M. F. Jung, and N. N. Bazarova, ""I just shared your responses": Extending communication privacy management theory to interactions with conversational agents," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. GROUP, 2020.

[4] B. van der Sloot, *The Handbook of Privacy Studies: An Interdisciplinary Introduction*. Amsterdam University Press, 2018.

[5] G. D. P. Regulation, "Regulation (EU) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation)," *Official Journal of the European Union (OJ)*, vol. 59, no. 1-88, p. 294, 2016.

[6] Y. Javed, S. Sethi, and A. Jadoun, "Alexa's voice recording behavior: A survey of user understanding and awareness," in *Proceedings of the 14th International Conference on Availability, Reliability and Security*, ser. ARES '19. Association for Computing Machinery, 2019.

[7] W. Wilkowska and M. Ziefle, "Perception of privacy and security for acceptance of e-health technologies: Exploratory analysis for diverse user groups," in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, 2011, pp. 593–600.

[8] N. Zierau, C. Engel, M. Söllner, and J. M. Leimeister, "Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda," in *WI2020 Zentrale Tracks*. GITO Verlag, Mar. 2020, pp. 99–114.

[9] S. Petronio and I. Altman, *Boundaries of Privacy: Dialectics of Disclosure*, ser. Boundaries of Privacy: Dialectics of Disclosure. State University of New York Press, 2002.

[10] A. Adams, "Users' perception of privacy in multimedia communication," in *CHI '99 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '99. Association for Computing Machinery, 1999.

[11] F. Dunn, W. M. Hartmann, D. M. Campbell, and N. H. Fletcher, *Springer handbook of acoustics (2nd edition)*. Springer Publishing Company, Incorporated, 2015.

[12] S. Utami, J. Sarwono, N. Rochmadi, and N. Suheri, "Speech privacy and intelligibility in open-offices as an impact of sound-field diffuseness," in *Inter-noise*, 2014.

[13] P. Pérez Zarazaga, S. Das, T. Bäckström, V. Vegesna, and A. Vuppala, "Sound privacy: A conversational speech corpus for quantifying the experience of privacy," in *Interspeech 2019*, 2019, pp. 3720–3724.

[14] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue, "Timit acoustic-phonetic continuous speech corpus," *Linguistic Data Consortium*, 1992.

[15] D. Dean, S. Sridharan, R. Vogt, and M. Mason, "The QUT-NOISE-TIMIT corpus for evaluation of voice activity detection algorithms," in *Proceedings of the 11th Annual Conference of the International Speech Communication Association*, K. Hirose, S. Nakamura, and T. Kaboyashi, Eds. International Speech Communication Association, 2010, pp. 3110–3113.

[16] A. Weisser and J. M. Buchholz, "Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions," *The Journal of the Acoustical Society of America*, vol. 145, no. 1, pp. 349–360, 2019.

[17] K.-T. Chen, C.-C. Wu, Y.-C. Chang, and C.-L. Lei, "A crowdsourceable QoE evaluation framework for multimedia content," in *Proceedings of the seventeen ACM international conference on Multimedia - MM '09*. ACM Press, 2009, pp. 491–500.

[18] Prolific, Oxford, UK, 2020. [Online]. Available: https://www.prolific.co

[19] E. Peer, L. Brandimarte, S. Samat, and A. Acquisti, "Beyond the Turk: Alternative platforms for crowdsourcing behavioral research," *Journal of Experimental Social Psychology*, vol. 70, pp. 153 – 163, 2017.

[20] M. Schoeffler, S. Bartoschek, F.-R. Ströter, M. Roess, S.Westphal, B. Edler, and J. Herre, "webMUSHRA – a comprehensive framework for web-based listening tests," *Journal of Open Research Software*, vol. 6, no. 1, 2018.

[21] I.-T. R. P.808, "Subjective evaluation of speech quality with a crowdsourcing approach," 2018.

[22] M. G. Kendall, *Rank correlation methods*, 4th ed. London Griffin, 1970.

[23] F. Wickelmaier and C. Schmid, "Probabilistic choice models for psychological scaling," *The joint CFA/DAGA 2004 congress*, 2004.

[24] ——, "A Matlab function to estimate choice model parameters from paired-comparison data," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 1, p. 29–40, 2004.

[25] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 1988.

[26] K. Zimmer, W. Ellermeier, and C. Schmid, "Using probabilistic choice models to investigate auditory unpleasantness," *Acustica United with Acta Acustica*, vol. 90, no. 6, pp. 1019–1028, 2004.

[27] L. Fernández Gallardo, "A paired-comparison listening test for collecting voice likability scores," *Speech Communication*, p. 5, 2016.

[28] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, 1952.

[29] R. D. Luce, *Individual Choice Behavior: A Theoretical analysis*. Wiley, 1959.

[30] A. Tversky, "Elimination by aspects: A theory of choice," in *Psychological Review*, vol. 79, no. 4, 1972, pp. 281–299.

[31] A. Tversky and S. Sattath, "Preference trees," *Psychological Review*, vol. 86, no. 6, pp. 542–573, 1997.

[32] H. Akaike, "On entropy maximization principle," in *Applications of Statistics*, 1977.