Measuring Voice Quality Parameters After Speaker Pseudonymization

van Son, R.J.J.H.

DOI
10.21437/Interspeech.2021-26

Publication date
2021

Document Version
Final published version

Published in
Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH

INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Measuring Voice Quality Parameters after Speaker Pseudonymization

Rob J. J. H. van Son[1,2]

[1]Netherlands Cancer Institute, Amsterdam, The Netherlands
[2]ACLC, University of Amsterdam, The Netherlands
r.v.son@nki.nl

Abstract

Collecting and sharing speech resources is important for progress in speech science and technology. Often, speech resources cannot be shared because of concerns over the privacy of the speakers, e.g., minors or people with medical conditions. Current technologies for pseudonymizing speech have only been tested on "standard" speech for which pseudonymization methods are evaluated on speaker identification risk, intelligibility, and naturalness. For many applications, the important characteristics are para-linguistic aspects of the speech, e.g., voice quality, emotion, or disease progression. Little information is available about the extent to which speaker pseudonymization methods preserve such paralinguistic information. The current study investigates how well voice quality parameters are preserved by an example speech pseudonymization application. Correlations prove to be high between original and pseudonymized recordings for seven acoustic parameters and a composite measure of dysphonia, the *AVQI*. Root mean square errors for these parameters were reasonably small. A linear mixed effect model shows a link between the difference between source and target speaker and the size of the absolute difference in the *AVQI*. It is argued that new measures of quality are needed for pseudonymized non-standard speech before wide-spread application of pseudonymized speech can be considered in research and clinical practise.

Index Terms: voice privacy, speaker pseudonymization, voice quality, paralinguistics

1. Introduction

Collecting and sharing speech resources is important for progress in speech science and technology. A lot of progress in speech technology has been made possible by the availability of large speech corpora in combination with advanced statistical techniques [1, 2]. However, speech recordings also carry a privacy risks. This is especially true when the speakers have medical conditions, are minors, or the subject matter is sensitive. But these are also groups that might benefit from improvements in speech technology tailored to their needs. The privacy risks resulting from sharing speech recordings would be mitigated if the probability of speaker (re-)identification could be reduced while retaining useful linguistic and para-linguistic features. Successful de-identification of speech would shift the risk-benefit balance for sharing speech corpora towards more sharing.

De-identification of speech will always be a trade-off between risk of re-identification and usefulness. In general, pseudonymization is more realistic than attempting true anonymization. This trade-off suggests an approach to pseudonymization that is adjustable in the level of information removed from the speech while still preserving relevant features well enough to make the result useful. The goal is to develop a method in which the transformation of the speech can be tailored to the risk profile and features needed (c.f., [3]).

The *VoicePrivacy 2020* challenge [4] has been organized to stimulate development of pseudonymization and anonymization methods useful for speech technology deployment, balancing the risk of re-identification and speech quality. The first *VoicePrivacy 2020* challenge was directed at "normal" speech. However, research into paralinguistics, disordered and pathological speech, and the deployment of, e.g., cloud-based diagnostics on speech, would all benefit from being able to use pseudonymized speech to alleviate privacy risks. Non-standard speech has peculiarities that can make de-identification more or less difficult, depending on the task. In non-standard speech, every speaker has her or his own individual deviations from standard speech. These deviations are important for the task at hand, e.g., diagnosing pathologies, but are also powerful features to re-identify the speaker. However, these peculiarities can be inherently transient, or changing over time, making re-identifying speakers over time difficult even without pseudonymization. Non-standard speech can also be a challenge for automatic speech recognition or measures of naturalness.

These particulars of non-standard speech complicate the evaluation of pseudonymization methods on re-identification risk, intelligibility, or speech degradation. This is not always a problem as for many tasks, intelligibility or naturalness are irrelevant. One such task is the automatic evaluation of voice quality in speakers with larynx pathologies.

The current study makes a first step in the direction of evaluating the use of pseudonymization in research of paralinguistic phenomenons by investigating to what extent voice quality can be measured in pseudonymized speech using an algorithm that has been entered in the *VoicePrivacy 2020* challenge [5, 6, 7]. The algorithm performed reasonably on anonymization and WER in the *VoicePrivacy 2020* challenge and good on speech quality (entry `I1` in [6, 7]). In this study the performance of a range of voice quality parameters are compared between original speech and pseudonymized speech.

The questions addressed in the current study are:

1. Can original acoustic voice parameter values be determined from pseudonymized speech?
2. What is the error in these values when determined from pseudonymized speech?
3. Can the size of the error be predicted from the extent of the anonymization?

2. Methods

2.1. Speech recordings

Speech recordings from 43 patients (9F/34M) who had been treated for small laryngeal tumors with laser surgery or radiation therapy were obtained for analysis. Speech had been rou-

http://dx.doi.org/10.21437/Interspeech.2021-26

Table 1: *Average values (sd) of all parameters (Original) and pairwise differences with the Original recordings after pseudonymization (*P500*,* P550*,* P600*). Pseudonymization targets* P500*: ϕ=500 Hz, F_0=120 Hz;* P550*: ϕ=550 Hz, F_0=150 Hz;* P600*: ϕ=600 Hz, F_0=180 Hz; RMSE: Range of root mean square errors for (*P500*,* P550*,* P600*) compared to linear model from fit to* $Original \sim Pseudo$*. (see Figure 2); *: p< 0.001 pair-wise Student t-test against value for* Original*. See text.*

	AVQI		CPPs		HNR		Slope	
Original	4.469	(1.714)	10.887	(3.132)	14.108	(4.431)	-20.706	(3.748)
P500	-0.054	(0.516)	*0.400	(0.991)	*-0.882	(1.958)	*0.656	(1.554)
P550	-0.155	(0.526)	*0.635	(0.887)	-0.037	(1.964)	*1.338	(1.305)
P600	*-0.323	(0.623)	*0.810	(0.964)	0.408	(2.083)	*1.847	(1.515)
RMSE	[0.51 - 0.61]		[0.86 - 0.95]		[1.93 - 2.07]		[1.30 - 1.55]	

	Tilt		Jitter		Shimmer		ShdB	
Original	-10.646	(1.323)	1.380	(1.517)	8.149	(3.454)	0.733	(0.299)
P500	*-0.463	(0.607)	*-0.440	(0.905)	0.185	(2.298)	0.019	(0.197)
P550	-0.083	(0.485)	*-0.535	(1.04)	-0.292	(1.950)	-0.004	(0.159)
P600	-0.106	(0.634)	*-0.630	(0.982)	*-0.781	(1.784)	-0.049	(0.153)
RMSE	[0.46 - 0.57]		[1.27 - 1.46]		[1.88 - 3.21]		[0.16 - 0.28]	

tinely recorded during consultations, both before treatment and during 12 month follow-up. From each recording session, 3 seconds from a sustained vowel, preferably /a/, and 4 seconds of running speech, mainly from a neutral story read aloud, were selected. When no sustained vowel of at least 3 seconds was available, several sustained vowel realizations were concatenated to obtain a sample of 3 seconds. In total, 101 recorded sessions were available for analysis from these speakers, 42 recorded before the start of treatment (1 per speaker, missing for 1 speaker) and 59 recorded during follow up (0-3 per speaker). The distribution of recordings over speakers is uneven due to technical and administrative omissions.

The neutral first formant, ϕ, is a measure of the vocal tract length (VTL) [8] and is manipulated during pseudonymization together with F_0. The average ϕ for these speakers, calculated according to [5, 8, 9, 10] from all recorded sessions, is ϕ = 549 ($\pm$14) Hz and the mean F_0 = 142 ($\pm$56) Hz. Treatment for small laryngeal tumors is not expected to change the VTL, i.e., ϕ, but it must be noted that the pitch of these patients will have been affected before as well as after treatment.

Average values for vocally healthy speakers from the profiles in [10] are ϕ = 561 ($\pm$12) Hz and F_0 = 200 ($\pm$21) Hz for female speakers (N=116) and ϕ = 534 ($\pm$11) Hz and F_0 = 116 ($\pm$21) Hz for male speakers (N=107).

Table 2: *List of acoustic parameters tested. All are defined in [11] and can be calculated in* Praat *[12]. Note that* Jitter *is* not *part of the* AVQI*.*

AVQI	Acoustic Voice Quality Index
CPPs	Smoothed Cepstral Peak Prominence
HNR	Mean Harmonics to Noise Ratio (dB)
Slope	Slope of LTAS (dB)
Tilt	Tilt of trendline through LTAS (dB)
Jitter	Jitter (% period perturbation)
Shimmer	Shimmer (% amplitude perturbation)
ShdB	Shimmer (dB)

2.2. Voice quality parameters

Seven acoustic parameters related to voice quality and a composite measure of voice quality are used in this study (see Table 2). Overall voice quality is measured using the Acoustic

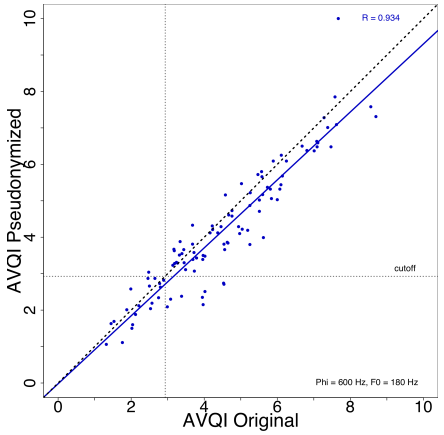

Figure 1: *Scatter plot and regression line of* AVQI *for* P600 *pseudonymized vs. original values. Diagonal dotted line: Ideal correlation. Horizontal and vertical dotted lines:* AVQI=*2.95 cutoff value between normal and dysphonic voice for Dutch. Spearman correlation coefficient R=0.934, p < 0.001*

Voice Quality Index (*AVQI*) [11, 13, 14, 15]. The *AVQI* falls on a scale from 0-10 (lower is better) with *AVQI*=2.95 being the demarcation point between normal and dysphonic voice for Dutch [11, 13]. The *AVQI* (version 2.03 [16, 17]) is calculated from primary acoustic parameters as (1):

$$\begin{aligned} AVQI = (3.295 - 0.111 \cdot CPPs - 0.073 \cdot HNR - \\ 0.213 \cdot Shimmer + 2.789 \cdot ShdB - \\ 0.032 \cdot Slope + 0.077 \cdot Tilt) \cdot 2.208 + 1.797 \end{aligned} \quad (1)$$

This differs slightly from the formula derived in [11, 15]. Acoustic parameters are determined on the voiced parts of the concatenation of 3 seconds of recorded sustained vowel, preferably /a/, and 4 seconds of recorded running speech according to [11]. All values were extracted from the *AVQI Praat* script version 2.03 [15, 16, 17], with *AVQI* values truncated between 0-10. No-voice conditions with undefined parameter values were scored as *AVQI*=10.

Of the 101 original speech samples, 81 (80%) are dysphonic (*AVQI*$\geq$2.95), 38/42 recorded before treatment (90%) and 43/59 after treatment (73%).

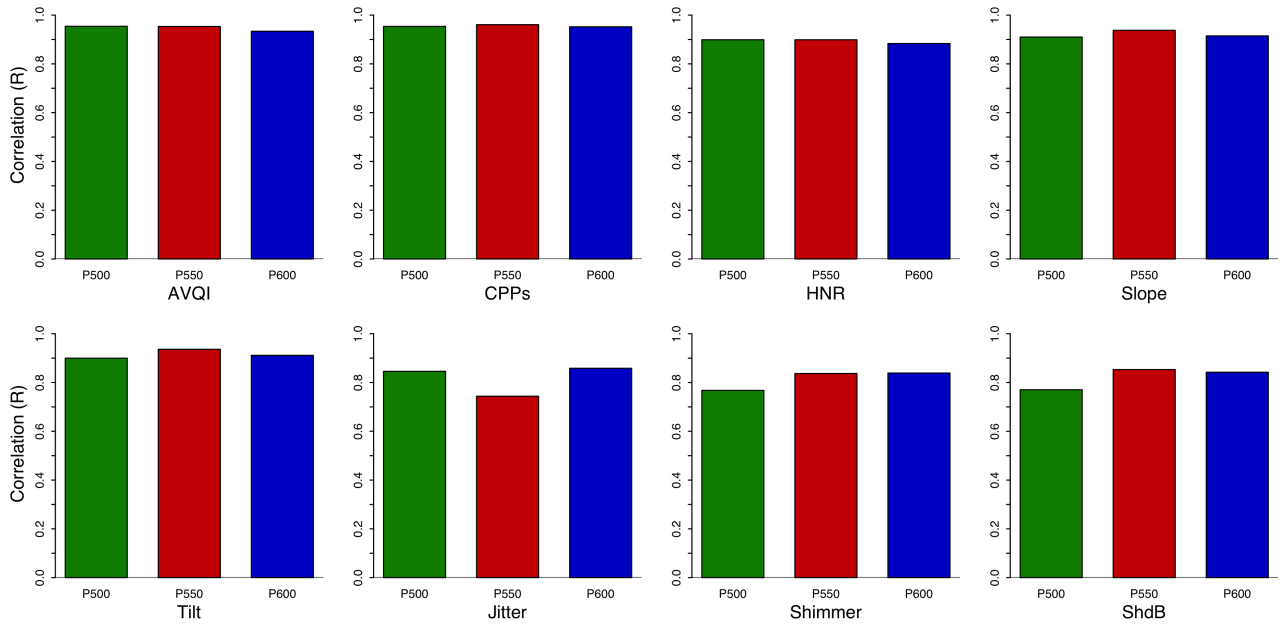

Figure 2: *Pearson's correlation coefficients between Pseudonymized and Original values of parameters. All R: $p < 0.001$.* P500 *(green): ϕ = 500 Hz, F_0 = 120 Hz;* P550 *(red): ϕ = 550 Hz, F_0 = 150 Hz;* P600 *(blue): ϕ = 600 Hz, F_0 = 180 Hz. See text.*

2.3. Speech pseudonymization

A signal processing approach is used in the current study for pseudonymization of speech. The method is based on changing the perceived acoustic length of the vocal tract and individual formant frequencies by changing the sampling rate ("playback speed") after which overlap-add [18] is applied to adjust the F_0 and the duration of the utterances. The standard *Praat* [12] command *Change Gender* is used to perform these changes. Details can be found in [5, 9, 10]. In *ABX* speaker identification experiments with pseudonymized normal speech, both expert and naive subjects identified speakers in $\leq$70% of trials [9]. This method has been entered into the *VoicePrivacy 2020* challenge [4] and performance results can be found in [5, 6, 7, 9]. See also the *Examples* folder in the media files.

In the current study, the procedure for the human listening experiments from [9] was used. The original recordings were pseudonymized to three fixed target "speakers" with random frequency and intensity shifts added to a low-frequency band around F_0 and the F_{3-5} bands. Speaking rate was fixed at 3 syllables/second for all targets. The three pseudonymization targets have neutral first formants, ϕ, of 500 Hz, 550 Hz, and 600 Hz and F_0 values of, respectively, 120 Hz, 150 Hz, and 180 Hz. They are indicated as, respectively, *P500*, *P550*, and *P600* in the remainder of this article. These target positions span the vocal track lengths and F_0 of male and female speakers.

It is expected that the extent of the changes needed to go from the characteristics of the source speaker to those of the target speaker will affect the voice quality of the resulting pseudonymized speech as measured with, e.g., a root-mean-square error (RMSE). The "distance" between the pseudonymized and original recordings is given by two parameters: The change in VTL, i.e., ϕ, and the change in median F_0. The procedure to change ϕ also changes the median F_0 and, therefore, will affect the additional changes needed to arrive at the target F_0. So, there will be an interaction between changes in ϕ and changes in median F_0. To determine how the sizes of these changes affect the voice quality outcomes, the absolute difference between the *AVQI* values from pseudonymized and original recordings, $|\Delta AVQI|$, are modelled from the absolute differences in ϕ, $|\Delta\phi|$, and median F_0, $|\Delta\overline{F_0}|$, between the pseudonymized and original recordings and a signed interaction term, $\Delta\phi{:}\Delta\overline{F_0}$, for all three targets pooled (*P500*, *P550*, and *P600*). A linear mixed effects model is constructed with $|\Delta\phi|$, $|\Delta\overline{F_0}|$, and $\Delta\phi{:}\Delta\overline{F_0}$ as fixed effects and *Speaker* as random effect according to (2):

$$|\Delta AVQI| \sim |\Delta\phi| + |\Delta\overline{F_0}| + \Delta\phi : \Delta\overline{F_0} + (1|Speaker) \quad (2)$$

3. Results

Statistics are done with R version 3.6.1 [19]. To compensate for the number of tests for 3 pseudonymization targets and 8 parameters, a Bonferroni correction is used and the level of significance is set at $p < 0.001$.

The average values for the 8 parameters for the original recordings and the differences with the pseudonymized recordings (pseudonymized - original value) are given in Table 1. The differences between pseudonymized and original speech seem to be systematic, with pseudonymized speech tending to have the "better" values, e.g., with pseudonymized speech having lower *AVQI* scores than the original recordings. As a result, the number of dysphonic speech samples (*AVQI*$\geq$2.95) is also reduced from 81 (80%) in original to 74-76 in pseudonymized speech (73-75%, not shown). This systematic difference is particularly noticeable for *CPPs*, *Slope*, and *Jitter*, where it is statistically significant for all three pseudonymization targets. For other parameters, i.e., *AVQI*, *HNR*, *Tilt*, and *Shimmer*, it is only significant for one of the extreme targets, *P500* or *P600*, which would be furthest from the average speaker cases.

Pairwise correlations between parameter values in original and pseudonymized speech are strong. Figure 1 gives

an example of a scatter plot for *AVQI* scores between *P600* pseudonymized speech and the original recordings. Correlation coefficients between pseudonymized speech and original for all parameters are presented in Figure 2. For most parameters $R \gtrsim 0.9$. *Jitter*, *Shimmer*, and *ShdB* correlate less well with $R > 0.7$.

Voice quality parameter values are strongly correlated between pseudonymized and original recordings for each pseudonymization target (see Figure 2). This implies that original parameter values can be estimated from the pseudonymized values using a linear model based on the target value (*P500*, *P550*, *P600*), i.e., lm($Original \sim Pseudonymized$).

A recipient of pseudonymized speech needs to know the expected measuring error before the data will be usable. A recipient only knows the (approximate) pseudonymization target. So it is important to estimate the average error for each pseudonymization target. The range of root mean square errors (RMSE) between the estimated and observed original parameter values for the three pseudonymization targets are presented in the last row of Table 1. For the parameters with lower correlations between original and pseudonymized values, *Jitter*, *Shimmer* and *ShdB*, these RMSEs are only marginally smaller than the standard deviation of the original values. For the other parameters, the RMSE values are a third of the standard deviation of the original values, e.g., for *AVQI* the RMSE $\sim$ [0.51-0.61], while the standard deviation of the original values is 1.7.

When preparing pseudonymized speech for distribution, it is important to select the pseudonymization parameters such that the measuring error will be within the intended range. The model of (2) gives a first approximation of the relation between absolute *AVQI* error and the pseudonymization parameters. The model fit for (2) results in contributions of *Intercept*, $|\Delta\overline{F_0}|$, and $\Delta\phi{:}\Delta\overline{F_0}$ that were statistically significant ($p<0.001$, [20]), while the contribution from $|\Delta\phi|$ was not ($p=0.0078$). No such effects were found for the other acoustic parameters that make up the *AVQI* ($p>0.001$, not shown).

The R^2 of the model fit in (2) is 0.12 for the fixed, and 0.37 for the random effects (*Speaker*) [21]. It can be inferred that changing the F_0 and the interaction, $\Delta\phi{:}\Delta\overline{F_0}$, had a significant effect on the general voice quality, *AVQI*, after pseudonymization. The fit of the model for fixed and random effects combined has R^2=0.49. Adding the original *AVQI* as a fixed effect to (2), increases this to total R^2=0.57 ($p<0.001$, ANOVA, ΔAIC=-9, not shown) with higher *AVQI* (=lower voice quality) associated with higher deviations in *AVQI* between pseudonymized and original speech.

4. Discussion

The results of the current study show that the questions posed in the *Introduction* can be answered in the affirmative. It is indeed possible to determine the original acoustic parameter values as well as the derived voice quality (*AVQI*) from pseudonymized speech within predictable RMSE tolerances. For most acoustic parameters, there is a strong correlation between the pseudonymized and original values and a bounded RMSE. For instance, for the overall voice quality, *AVQI*, the correlation coefficient $R>0.9$ and the RMSE is between 0.5-0.6 on a scale from 0-10. Although not ideal, such a system could already be used for some tasks, if care is taken to keep the RMSE within pre-determined limits.

It has also been found that, for the *AVQI*, the error sizes can be predicted from the pseudonymization parameters, within certain limits. The absolute difference between the *AVQI* of pseudonymized and original speech can be approximated with a linear mixed effect model using the speaker characteristics and the pseudonymization parameters. Such a relation has not (yet) been found for the composite parameters. As the signal processing approach to speech pseudonymization used in this study is tuneable, this relation helps to select settings that trade-off between anonymity and conservation of voice quality parameters, e.g., to lower RMSE.

The surprising fact that pseudonymized speech tends to sometimes score "better" on acoustic parameters than the original speech could be an artefact of the overlap-add procedure used to change the duration and F_0 of the pseudonymized speech. This procedure might "regularize" the pitch periods compared to the originals.

80% of the speech samples selected for this study are dysphonic (*AVQI* $\geq$ 2.95). This presents problems when evaluating the quality of the pseudonymization. First, the voice parameters which are to be preserved are also a good cue to speaker identity. Hence, it is difficult to evaluate the anonymization as both human listeners and Automatic Speaker Verification (ASV) applications will be able to identify the speakers to a large extent on their voice parameters. Second, the intelligibility of the speech is already lower due to their dysphonic nature. Both human listeners and Automatic Speech Recognition (ASR) applications will rate the speech with lower intelligibility. It will be difficult to tease out the additional deterioration of intelligibility due to the pseudonymization. And third, naturalness is difficult to evaluate in dysphonic speech, original or pseudonymized.

Before pseudonymized speech can be shared with other researchers or used for diagnostic purposes in a clinical setting, the anonymity and usefulness of the pseudonymized speech have to be vetted. To evaluate the quality of pseudonymization applications for non-standard speech, alternatives must be found for the customary ASV, ASR, and naturalness scores. Currently, a similarity/distance approach seems to be most promising for evaluating the quality of a pseudonymization application. Such an approach was already proposed for intelligibility in [5, 9].

5. Conclusions

Using a tuneable signal processing approach to speech pseudonymization, it is shown that voice quality parameters can still be measured after pseudonymization up to a predictable error. The method allows for implementing a trade-off between anonymity and conservation of voice quality parameters. For further advances in the use of speech pseudonymization methods in para-linguistic applications, it is argued that there is a need for evaluation standards for anonymity and speech quality in pseudonymized non-standard speech. Such new evaluation standards are needed before pseudonymization can be considered useful and safe in research and clinical practice.

6. Acknowledgements

This research has been approved by the Institutional Review Board at the Netherlands Cancer Institute (IRBd20-023). The Department of Head and Neck Oncology and Surgery of the Netherlands Cancer Institute receives a research grant from Atos Medical (Hörby, Sweden), which contributes to the existing infrastructure for quality of life research.

7. References

[1] Z. Zhang, N. Cummins, and B. Schuller, "Advanced data exploitation in speech analysis: An overview," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 107–129, 2017.

[2] Y. Ning, S. He, Z. Wu, C. Xing, and L.-J. Zhang, "A Review of Deep Learning Based Speech Synthesis," *Applied Sciences*, vol. 9, no. 19, p. 4050, Sep. 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/19/4050

[3] S. Kung, "A Compressive Privacy approach to Generalized Information Bottleneck and Privacy Funnel problems," *Journal of the Franklin Institute*, vol. 355, no. 4, pp. 1846–1872, Mar. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0016003217303162

[4] N. Tomashenko *et al.*, "The VoicePrivacy 2020 Challenge," VoicePrivacy, Feb. 2020. [Online]. Available: https://www.voiceprivacychallenge.org/

[5] S. P. Dubagunta, R. J. J. H. van Son, and M. Magimai-Doss, "Adjustable Deterministic Pseudonymisation of Speech: Idiap-NKI's submission to VoicePrivacy 2020 Challenge," 2020. [Online]. Available: https://www.voiceprivacychallenge.org/docs/Idiap-NKI.pdf

[6] N. Tomashenko *et al.*, "The voiceprivacy 2020 challenge: Challenge setup and results," https://www.voiceprivacychallenge.org/docs/1___VoicePrivacy_challenge_setup_and_results_N_Tomashenko.pdf, 2020, Online; accessed 2020-02-22.

[7] X. Wang *et al.*, "The voiceprivacy 2020 challenge subjective evaluation-1," 2020. [Online]. Available: https://www.voiceprivacychallenge.org/docs/6___Subjective_evaluation_1_naturalness_intelligibility_speaker_verifiability_X_Wang.pdf

[8] A. C. Lammert and S. S. Narayanan, "On Short-Time Estimation of Vocal Tract Length from Formant Frequencies," *PLOS ONE*, vol. 10, no. 7, p. e0132193, 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0132193

[9] S. P. Dubaguntaa, R. J. J. H. van Son, and M. Magimai-Doss, "Adjustable Deterministic Pseudonymization of Speech," submitted.

[10] R. J. J. H. van Son, "Pseudomymize speech," Netherlands Cancer Institute., 2019. [Online]. Available: https://robvanson.github.io/PseudonymizeSpeech/

[11] Y. Maryn, P. Corthals, P. Van Cauwenberge, N. Roy, and M. De Bodt, "Toward Improved Ecological Validity in the Acoustic Measurement of Overall Voice Quality: Combining Continuous Speech and Sustained Vowels," *Journal of Voice*, vol. 24, no. 5, pp. 540–555, Sep. 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0892199709000034

[12] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (computer program). version 6.1.06," 2019.

[13] Y. Maryn, M. De Bodt, and N. Roy, "The Acoustic Voice Quality Index: Toward improved treatment outcomes assessment in voice disorders," *Journal of Communication Disorders*, vol. 43, no. 3, pp. 161–174, 2010.

[14] B. Barsties and M. De Bodt, "Assessment of voice quality: current state-of-the-art," *Auris Nasus Larynx*, vol. 42, no. 3, pp. 183–188, 2015.

[15] G. Kishore Pebbili, S. Shabnam, M. Pushpavathi, J. Rashmi, R. Gopi Sankar, R. Nethra, S. Shreya, and G. Shashish, "Diagnostic Accuracy of Acoustic Voice Quality Index Version 02.03 in Discriminating across the Perceptual Degrees of Dysphonia Severity in Kannada Language," *Journal of Voice*, vol. 35, no. 1, pp. 159.e11–159.e18, Jan. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0892199719300736

[16] Vlaamse Vereniging voor Logopedisten, "praat-script-avqi-v0203," https://www.vvl.be/documenten-en-paginas/praat-script-avqi-v0203, accessed: 2021-02-22.

[17] B. Barsties and Y. Maryn, "The improvement of internal consistency of the Acoustic Voice Quality Index," *American Journal of Otolaryngology - Head and Neck Medicine and Surgery*, vol. 36, no. 5, pp. 647–656, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.amjoto.2015.04.012

[18] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech communication*, vol. 9, no. 5-6, pp. 453–467, 1990.

[19] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2019. [Online]. Available: http://www.R-project.org/

[20] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models," *Journal of Statistical Software*, vol. 82, no. 13, 2017. [Online]. Available: http://www.jstatsoft.org/v82/i13/

[21] S. Nakagawa and H. Schielzeth, "A general and simple method for obtaining R2 from generalized linear mixed-effects models," *Methods in Ecology and Evolution*, vol. 4, no. 2, pp. 133–142, Feb. 2013. [Online]. Available: http://doi.wiley.com/10.1111/j.2041-210x.2012.00261.x