NISTIR 8328

Mission Critical Voice Quality of Experience Access Time Measurement Method Addendum

Chelsea Greene
Jesse Frey
Zainab Soetan
Jaden Pieper
Tim Thompson

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8328

National Institute of
Standards and Technology
U.S. Department of Commerce

NISTIR 8328

Mission Critical Voice Quality of Experience Access Time Measurement Method Addendum

Chelsea Greene
Jesse Frey
Zainab Soetan
Jaden Pieper
Tim Thompson
Communications Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8328

December 2020

U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Undersecretary of Commerce for Standards and Technology

National Institute of Standards and Technology
Interagency or Internal Report 8328
Natl. Inst. Stand. Technol. Interag. Intern. Rep. 8328, 18 pages (December 2020)

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8328

Acknowledgments

The authors would like to thank Steve Voran from the National Telecommunications and Information Administration's Institute for Telecommunications Sciences for his superb guidance and willingness to share his in-depth knowledge in developing measurement methods for voice communications systems. The authors would also like to thank Don Bradshaw of the National Institute of Standards and Technology's (NIST) Public Safety Communications Research (PSCR) Division for his leadership and direction in developing measurement systems and properly communicating test results to the public. Finally, the authors would like to acknowledge the mathematical contributions and helpful revisions provided by Cara O'Malley and William Magrogan of NIST PSCR.

Abstract

Access time generally describes the time associated with the establishment of a talk path upon user request to speak and has been identified as a key component of quality of experience (QoE) in voice communications. The National Institute of Standards and Technology's (NIST) Public Safety Communications Research (PSCR) Division has previously developed an access time measurement method in NISTIR 8275 [1]. Improvements were implemented in the access delay measurement method to better handle the idiosyncrasies of the various push-to-talk (PTT) technologies tested. These improvements include a more robust audio alignment procedure, and graceful communications failure detection and handling. This paper covers those improvements to the access delay measurement system, results of testing Project 25 (P25) technologies with and without encryption turned on, as well as some initial Long Term Evolution (LTE) measurements.

Key words

Access delay; Articulation Band Correlation Modified Rhyme Test (ABC-MRT); A-weight; Encryption; Key performance indicator (KPI); Land mobile radio (LMR); Latency; Mission Critical Push-to-Talk (MCPTT); Modified Rhyme Test (MRT); Mouth-to-ear (M2E); Packetized; Project 25 (P25); Public Safety; Push-to-talk (PTT); Quality of experience (QoE); Receive; Streaming; Transmit; Vocoder; Voice.

Table of Contents

List of Tables

List of Figures

Acronyms

ABC-MRT Articulation Band Correlation Modified Rhyme Test. i, 1, 2, 6

ADB Android Debug Bridge. 4, 5

AW A-weighting. 4, 5

dBA A-weighted decibel. 4

E2E end-to-end. 1, 5

KPI key performance indicator. i, 1, 5

LMR land mobile radio. i, 4

LTE Long Term Evolution. i, 1, 4, 6, 8–10

M2E mouth-to-ear. i, 1, 5, 6

MCPTT Mission Critical Push-to-Talk. i, 1, 4, 8, 9

MCV mission critical voice. 1, 10

MRT Modified Rhyme Test. i, 1, 2, 9

P25 Project 25. i, 1, 2, 5, 6, 8

PSCR Public Safety Communications Research. i, 1

PTT push-to-talk. i, 2, 3, 8

QoE quality of experience. i, 1

Interested users can use this example as a template for developing their own restart script. When a restart script is used, the AW check is run as before except now when the number of retries threshold is exceeded, the ADB script is called. The script is given the number of times that it has been called for a given trial and returns whether or not the test should continue. At this stage, a user-created script could attempt various recovery strategies and, when all possibilities are exhausted, pause for the user to fix the problem.

3. Example Measurements of Encrypted P25 Communications

Encryption is important for security and widely used by public safety in sensitive situations. The original example access delay measurements performed in Ref. [1] were conducted without encryption, but it is imperative that our measurement systems can successfully measure encrypted technologies, given their widespread use. It is also informative to study the effect that encryption has on access delay and, in turn, the E2E access time measurements.

The process to enter encrypted communication mode varies by the radio brand; the radios tested have a two-position concentric switch that was programmed to turn on encryption. Once in encrypted communication mode, the test proceeded as usual. Table 1 summarizes the difference encrypted communication mode has on various P25 technologies used during testing. The values in Table 1 represent the access delay required to achieve a raw intelligibility of 85%.

Table 1. Example measurements of encrypted and unencrypted E2E access time results. 2500 ms audio clip access delay values for a raw intelligibility of 85%. Uncertainties are reported as 95% confidence intervals

	Access Delay [ms]	M2E Latency [ms]	E2E Access Time [ms]
P25 Direct Encrypted	174.1 ± 9.7	296.8 ± 1.7	470.9 ± 9.9
P25 Direct Unencrypted	126.9 ± 8.8	243.3 ± 0.3	370.2 ± 8.8
P25 Trunked Phase 1 Encrypted	695.8 ± 7.1	437.4 ± 6.3	1133.2 ± 9.5
P25 Trunked Phase 1 Unencrypted	643.3 ± 6.6	379.6 ± 2.1	1022.8 ± 6.9
P25 Trunked Phase 2 Encrypted	701.2 ± 6.5	636.1 ± 27.3	1337.3 ± 28.0
P25 Trunked Phase 2 Unencrypted	655.1 ± 6.3	601.9 ± 16.1	1257.0 ± 17.3

Because the encryption process effectively inserts additional fixed time processes into the communication chain, it was expected that M2E latency would be the KPI most impacted by the presence of encryption. Our initial expectations were for access delay and intelligibility to be similar regardless of encryption, due to equivalent controlled, low-noise laboratory testing environments. As reflected in the results in Table 1, access delay and M2E latency were both impacted by the transition from clear to encrypted transmissions. For P25 technologies where encryption is turned on, we hypothesize that the delay is due to the loading time of the AES 256 cipher block.

The curve parameters for each test, both encrypted and unencrypted, are shown in Table 2. The column labeled I_0 describes the asymptotic intelligibility for each technology. As expected, encryption did not significantly impact the asymptotic intelligibility for any given technology. The curve parameter related to the steepness of the intelligibility transition, λ, also remained relatively consistent, except for P25 trunked Phase 1. Table 2 shows that encryption has the greatest impact on the intelligibility curve midpoint, t_0, which increases. This, combined with the other curve fit parameters remaining unchanged, demonstrates that access delay is increased by a relatively consistent offset when encryption is used. For most P25 technologies, encryption added approximately 30-50 ms of delay.

This offset can be seen for most α levels. α is an intelligibility scaling factor used in the definition of access delay [1]. Figure 1 shows the visual comparison for the P25 trunked Phase 1 system. For a given α level, the access delay is larger when encrypted communication mode is on. Figure 2 shows the same information represented in Fig. 1, except access delay is shown as a function of the raw intelligibility of the clip from ABC-MRT, rather than relative intelligibility based on α values.

Table 2. Curve parameter results with 95% confidence intervals included in parentheses. I_0 describes the asymptotic intelligibility of the system for the words under test. t_0 describes the time at which intelligibility is 50% of its asymptotic level. And λ describes the inverse of the slope of the intelligibility curve at time t_0.

	I_0	t_0 [s]	λ [s^{-1}]
P25 Direct Encrypted	0.916, (0.914, 0.917)	0.069, (0.065, 0.073)	−0.041, (−0.045, −0.037)
P25 Direct Unencrypted	0.915, (0.914, 0.916)	0.025, (0.021, 0.029)	−0.040, (−0.044, −0.036)
P25 Trunked Phase 1 Encrypted	0.917, (0.916, 0.918)	0.569, (0.566, 0.572)	−0.050, (−0.053, −0.047)
P25 Trunked Phase 1 Unencrypted	0.916, (0.914, 0.917)	0.528, (0.525, 0.531)	−0.045, (−0.048, −0.043)
P25 Trunked Phase 2 Encrypted	0.903, (0.901, 0.905)	0.573, (0.570, 0.575)	−0.046, (−0.049, −0.044)
P25 Trunked Phase 2 Unencrypted	0.901, (0.899, 0.903)	0.525, (0.523, 0.527)	−0.046, (−0.049, −0.044)
LTE	0.996, (0.995, 0.997)	0.033, (0.031, 0.036)	−0.040, (−0.042, −0.038)

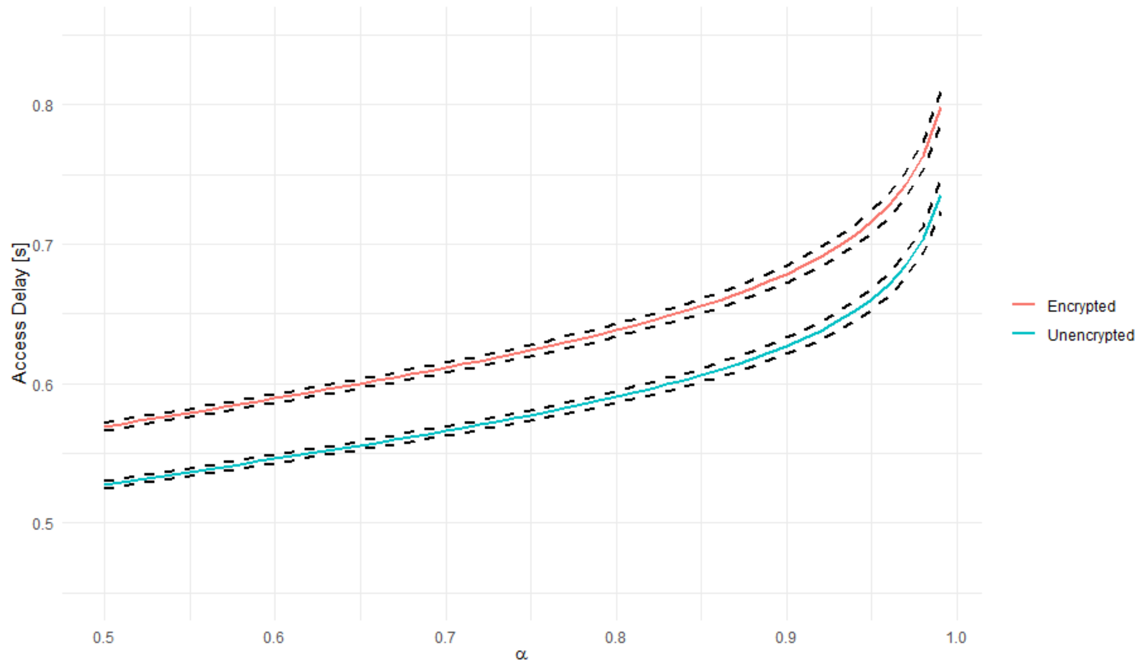

Fig. 1. P25 trunked Phase 1 access delay as a function of α. The dashed black lines represent the 95% confidence intervals.

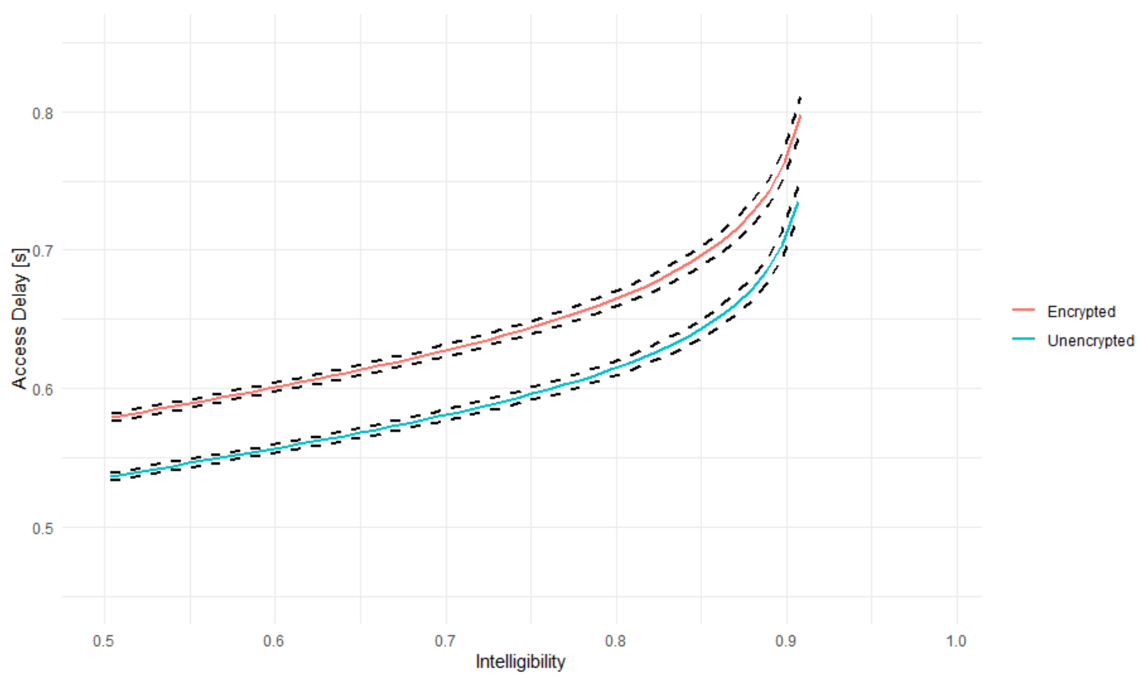

Fig. 2. P25 trunked Phase 1 access delay as a function of raw intelligibility. The dashed black lines represent the 95% confidence intervals.

4. Example Measurements of LTE Communications

LTE MCPTT technology is beginning to emerge on the market. Preliminary measurements were taken using early stages of MCPTT applications on Android devices. All measurements were performed using an RF enclosure. As previously discussed, Table 2 shows the curve parameters for LTE testing; asymptotic intelligibility was very high. In Fig. 3, access delay is shown as a function of α and in Fig. 4 as a function of intelligibility. Note that because the asymptotic intelligibility was nearly 1, these two plots are almost identical. The results for t_0 and λ were most similar to those of unencrypted P25 Direct. For a raw intelligibility of 85%, access delay was measured as 103.9 ± 4.1 ms with a level of confidence of 95%.

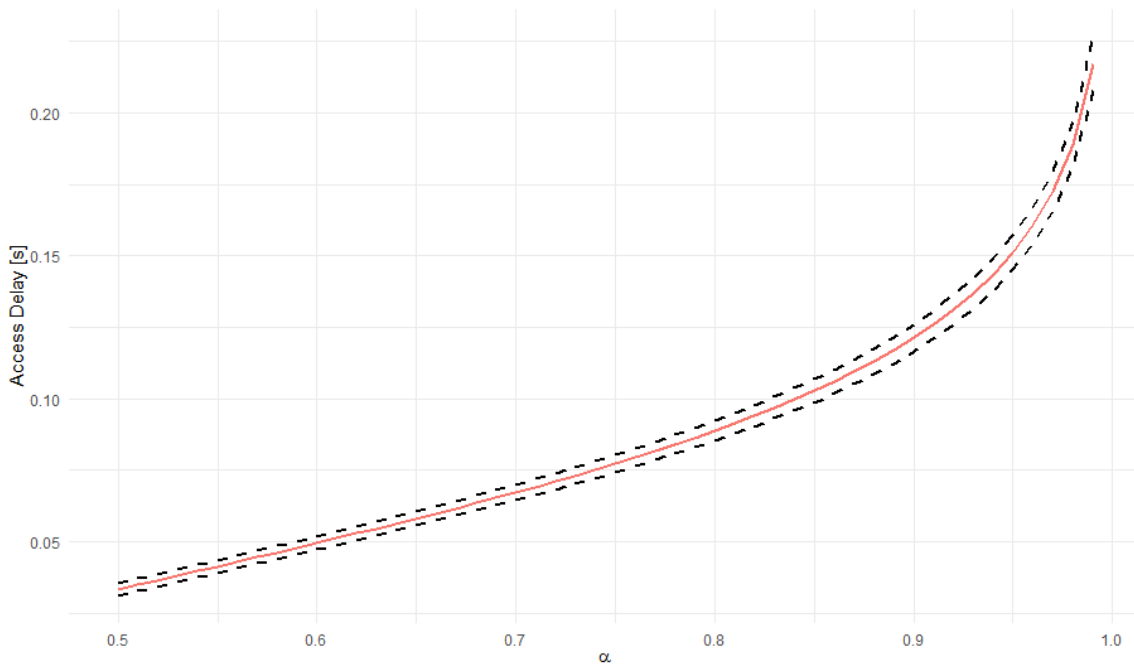

Fig. 3. LTE access delay curve as a function of α. The dashed black lines represent the 95% confidence intervals.

5. Conclusion

This paper describes several improvements made to the access time measurement system [1]. In particular, the audio alignment stage of the measurement process was improved to be more robust to specific impairments observed with low-rate vocoders. Further, the measurement system was made more robust and user-friendly for testing of prototype and in-development PTT applications, which are more prone to application failure in their early development stages. This paper also demonstrates the measurement system can successfully measure encrypted P25 and LTE communication systems. It is shown that encryption

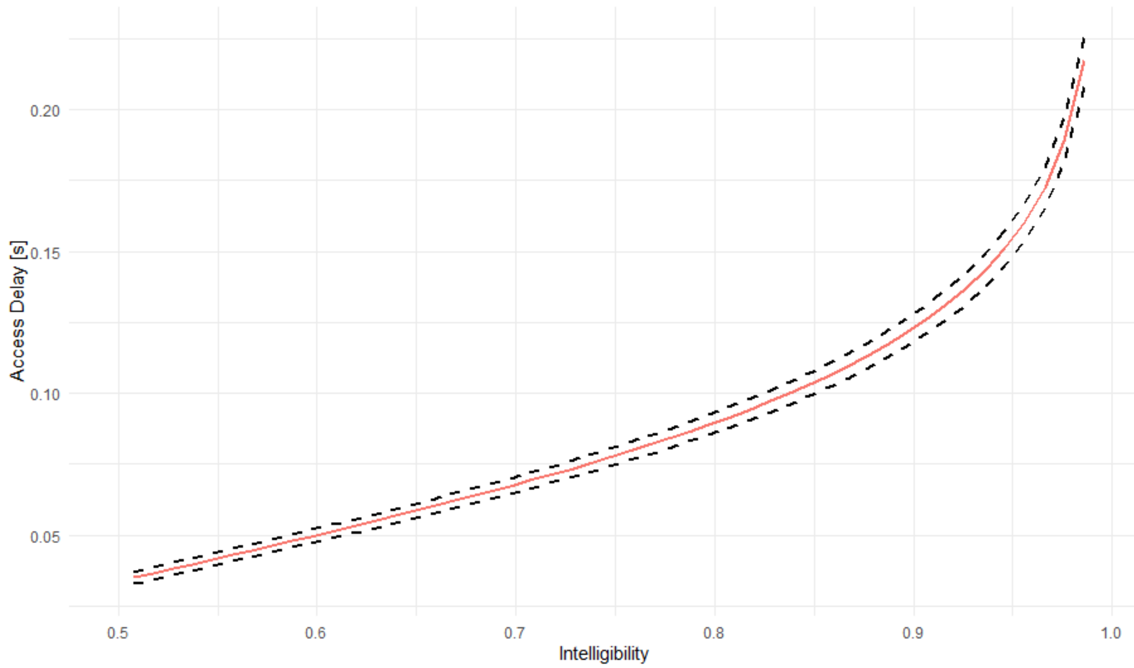

Fig. 4. LTE access delay as a function of raw intelligibility. The dashed black lines represent the 95% confidence intervals.

adds an almost constant amount of access delay to a given communication system. Finally, the measurement system was also able to measure access delay of LTE devices, despite the early stages of development of prototype LTE MCPTT applications.

5.1 Future Work

5.1.1 Curve Fitting Improvements

Limitations were identified with the current procedure for fitting curves to measured intelligibility data and we are currently working on a new and improved curve fitting regime. In particular, we are making the access delay measurements less dependent on characteristics specific to audio clip structure rather than MRT keywords themselves. Access delay currently depends on a subjective determination of when an MRT keyword is spoken within a full MRT phrase. This can potentially add bias to an access delay measurement and cause underestimates of true system access delay. This work is currently ongoing.

5.1.2 Access Focused Intelligibility Research

We also have recognized the need for truth data on the sort of impairments seen in access delay testing. In particular, we are interested in collecting data describing the subjective intelligibility response to partially muted words. Work is currently underway to characterize this response with new MRT intelligibility testing.

5.1.3 Packetized Voice Impairments

Loss, pause, and jump impairments are common in packetized voice and are known to impact speech quality measurements, as described in Ref. [8]. While analyzing LTE device data, a handful of trials with low P_2 intelligibility were noted. The waveforms of such trial recordings contained a gap of varying length in the middle of P_2. Figure 5 shows examples of these impairments: the top waveform shows a trial exhibiting typical behavior, the middle figure contains a pause impairment in a filler word, and the bottom figure shows an example of a pause impairment with P_2 which causes a low estimated intelligibility score. At the time of writing, detecting these pause impairments relies on detecting trials with suboptimal P_2 intelligibility values and reviewing the associated waveform. Further investigation is required to detect, identify, and handle this impairment within the MCV measurement system.

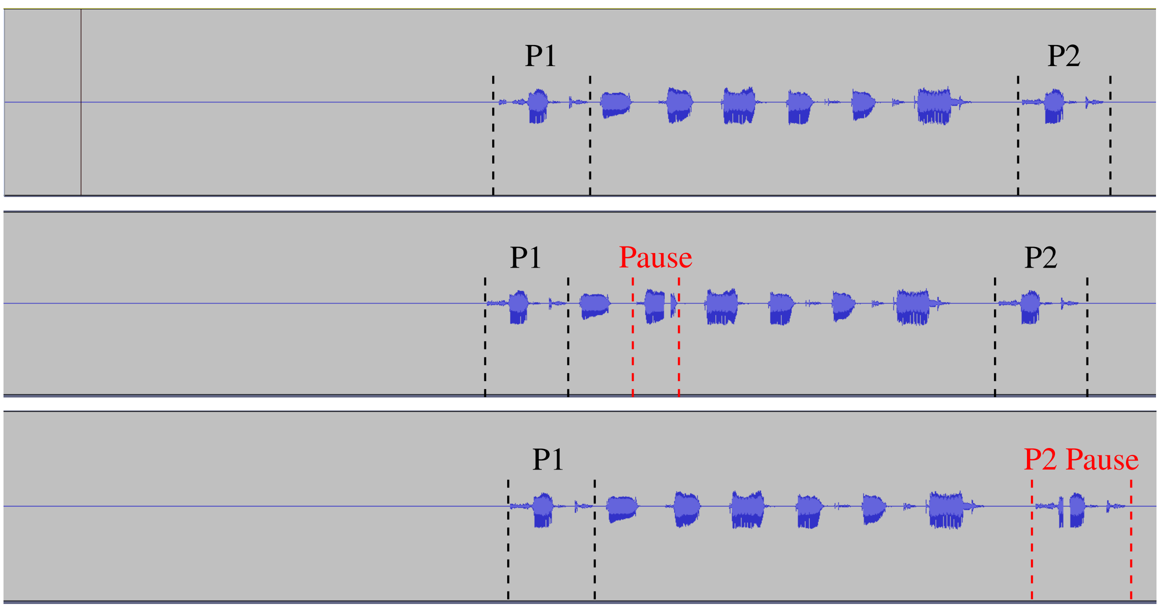

Fig. 5. LTE trial recording waveforms showing pause impairments

References

[1] Pieper J, Frey J, Greene C, Soetan Z, Thompson T, Voran S, Bradshaw D (2019) Mission Critical Voice QoE Access Time Measurement Methods (NIST), IR-8275. doi: 10.6028/NIST.IR.8275

[2] Frey J, Pieper J, Thompson T (2018) Mission Critical Voice QoE Mouth-to-Ear Latency Measurement Methods (NIST), IR-8206. doi: 10.6028/NIST.IR.8206

[3] ANSI/ASA (2009) ANSI/ASA S3.2-2009 Method for Measuring the Intelligibility of Speech over Communication Systems.

[4] Voran SD (2017) A multiple bandwidth objective speech intelligibility estimator based on articulation index band correlations and attention. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp 5100–5104. doi: 10.1109/ICASSP.2017.7953128

[5] Voran SD (2004) A bottom-up algorithm for estimating time-varying delays in coded speech. *Proceedings of the 3rd International Conference on Measurement of Speech and Audio Quality in Networks* .

[6] Fletcher H, Munson WA (1933) Loudness, its definition, measurement and calculation. *The Bell System Technical Journal* 12(4):377–430. doi: 10.1002/j.1538-7305.1933.tb00403.x

[7] (2020) Android Debug Bridge (ADB). URL https://developer.android.com/studio/command-line/adb.

[8] Voran SD (2003) Perception of temporal discontinuity impairments in coded speech – a proposal for objective estimators and some subjective test results. *Proceedings of the 2nd International Conference on Measurement of Speech and Audio Quality in Networks, Prague, Czech Republic, May 2003* .