



ISSN: 3060-4966



JIZPI XABARNOMASI

ILMIY-TEXNIK JURNAL

- 05.00.00 - Texnika fanlari
- 05.02.00 - Mashinasozlik va mashinashunoslik.
Mashinasozlikda materiallarga ishlov berish.
Metallurgiya. Aviasiya texnikasi
- 05.03.00 - Asbobsozlik, metrologiya va axborot-o'lchov
asboblari va tizimlari
- 05.04.00 - Radiotexnika va aloqa
- 05.05.00 - Energetika va elektrotexnika.
Qishloq xo'jaligi ishlab chiqarishini elektrlashtirish
texnologiyasi. Elektronika
- 05.06.00 - To'qimachilik va yengil sanoat materiallari
va buyumlari texnologiyasi
- 05.07.00 - Qishloq xo'jaligi ishlab chiqarishini
mexanizasiyalash texnologiyasi
- 05.08.00 - Transport
- 05.09.00 - Qurilish
- 05.10.00 - Inson faoliyati xavfsizligi

2026
maxsus son

 **JURNAL.JIZPI.UZ**

JIZPI XABARNOMASI

Muassis:

Jizzax politexnika instituti

O‘zR Oliy attestatsiya komissiyasi Rayosatining 2024-yil 25- dekabrda 365/4-son qarori bilan jurnal 05.00.00 – Texnika fanlari bo‘yicha OAK milliy nashrlari ro‘yxatiga kiritilgan

JizPI xabarnomasi

ILMIY-TEXNIK JURNAL

Vestnik DjizPI

НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Bulletin of JizPi

SCIENTIFIC AND TECHNICAL JOURNAL

Bir yilda to‘rt marta chop etiladi.

Jizzax viloyati Matbuot va axborot boshqarmasi tomonidan 2019 yil 14 martda 06-042 raqam bilan ro‘yxatga olingan.

Nashr uchun mas‘ul:

K.Karimova

Bosishga ruxsat etildi: 04.10.2026.

Qog‘oz bichimi: 60x84 1/8.

Bosma tabog‘i: 13,5.

Ofset bosma. Ofset qog‘ozi.

Adadi: 100 dona.

Bahosi kelishilgan narxda.

Buyurtma №14

“JIZPI TIPOGRAFIYASI” MChJ tomonidan tayyorlangan.

“JIZPI TIPOGRAFIYASI” MChJ bosmaxonasida chop etildi.

130100, Jizzax shahri, Islom Karimov shoh ko‘chasi, 4 - uy.

Telefon: (+99888) 140 00 74

BOSH MUHARRIR:

A.USMANKULOV, texnika fanlari doktori, professor

BOSH MUHARRIR O‘RINBOSARI:

J.ABDUNAZAROV, texnika fanlari doktori, professor

MAS‘UL MUHARRIR

K.KARIMOVA, texnika fanlari bo‘yicha falsafa doktori

TAHRIR HAY‘ATI:

S.M.Turobjonov, texnika fanlari doktori, akademik (TDTU)

S.S. Shaumarov, texnika fanlari doktori, professor (TDTrU)

U.Yu.Yuldashev, fizika-matematika fanlari doktori, professor (JizPI)

J.N.Abdunazarov, texnika fanlari doktori, professor (JizPI)

A.Parpiyev, texnika fanlari doktori, professor (JizPI)

I.S.Shukurov, texnika fanlari doktori, professor (SamGASU)
Q.J.Jumaniyozov, texnika fanlari doktori, professor (TEITI)
M.T.Xodjiyev, texnika fanlari doktori, professor (GulDU)
A.Z.Mamatov, texnika fanlari doktori, professor (TTYeSI)
U.B. Yusupov, texnika fanlari doktori, dotsent (TDTrU)
Sh. A.Qorabayev, texnika fanlari doktori, professor (NDTU)
S.L. Eshkobilov, PhD, professor (AQSH)
M.Mikusova, PhD, professor (Slovakia)
V.A.Kondratev, texnika fanlari doktori, professor (Belarus)
M.M. Ganiyev, texnika fanlari doktori, professor (Rossiya)
F.A. Egamberdiyev, texnika fanlari doktori, dotsent (JizPI)
I.Z.Abbazov, texnika fanlari bo'yicha falsafa doktori, dotsent (JizPI)
O.Zafarov, texnika fanlari doktori, dotsent (JizPI)
I.B.Asqarov, pedagogika fanlari doktori, professor (JizPI)
W. Pangwei, PhD, professor (Xitoy)
K. Kim, PhD, professor (Janubiy Koreya)
K. Széll, - PhD, Dotsent (Vengriya)
A. Cansiz – PhD, Professor (Turkiya)
M. Pogatsnik – PhD, Dotsent (Vengriya)
Sh. Madjidov – DSc, Dotsent (Xitoy)
V.I. Rimshin, texnika fanlari doktori, professor (Rossiya)
A.B. Zabiyeva, PhD, dotsent (Qozog'iston)
S.X.Baxriev, texnika fanlari nomzodi, dotsent (Tojikiston)
I.V.Yakimenko, texnika fanlari doktori, professor (Rossiya)
S.O. Eshbekova, fizika-matematika fanlari nomzodi, dotsent
O.K.Adilov, texnika fanlari nomzodi, professor (JizPI)
S.Tavboyev, Texnika fanlari nomzodi, dotsent (O'zMU JF)
Sh.M. Musayev, texnika fanlari bo'yicha falsafa doktori, dotsent (JizPI)
A.O. Sultonov, texnika fanlari bo'yicha falsafa doktori, dotsent (JizPI)
S.V.Djiyanbayev, texnika fanlari bo'yicha falsafa doktori, dotsent (JizPI)
G'I.Mamayev, texnika fanlari bo'yicha falsafa doktori, dotsent (JizPI)
O.B.Berdiyev, texnika fanlari nomzodi, professor (JizPI)
B.I.Matniyazov, texnika fanlari nomzodi, professor (JizPI)
S.A. Sattarov, fizika-matematika fanlari nomzodi, dotsent (JizPI)
G'.G'.Egamnazarov, texnika fanlari nomzodi, dotsent (JizPI)
B.B.Doniyorov, texnika fanlari bo'yicha falsafa doktori, dotsent (JizPI)
M.A.Doniyorova, texnika fanlari bo'yicha falsafa doktori, dotsent (JizPI)
T.A.Berdiyev, iqtisodiyot fanlari bo'yicha falsafa doktori, dotsent (JizPI)
O'.I.Jamolov, iqtisodiyot fanlari bo'yicha falsafa doktori, dotsent (JizPI)
S.A.Usmanov, pedagogika fanlari bo'yicha falsafa doktori, professor (JizPI)
A.A.Taylaqov, texnika fanlari bo'yicha falsafa doktori, dotsent (JizPI)
B.A.Tursunov, texnika fanlari bo'yicha falsafa doktori, dotsent (JizPI)
M.Anorboyev, texnika fanlari bo'yicha falsafa doktori, dotsent (JizPI)
E.Abdullayev, texnika fanlari bo'yicha falsafa doktori, dotsent (JizPI)
L.M.Qurbonova, kimyo fanlari falsafa doktori, PhD (JizPI)
B.U.Meliyev, texnika fanlari falsafa doktori, dotsent (JizPI)

TEXNIK MUHARRIRLAR:

M.Eshqulov - assistent o'qituvchi, (UZ, ENG, RUS)

X.Xamdamova – filologiya fanlari bo'yicha falsafa doktori, dotsent (UZ)

M.Ziyayeva - pedagogika fanlari bo'yicha falsafa doktori, dotsent (RU)

F.Xiloliddinova - v.b.dotsent (ENG)

Tahririyat manzili:

Jizzax shahri, Islom Karimov shoh ko'chasi, 4-uy

Faks: (0372) 226-45-47. Rasmiy sayt jurnal.jizpi.uz

Mundarija

1	Oybek Achilovich Kayumov, Nazokat Rashitovna Kayumova. A Safety-Constrained Speaker-Adaptive Algorithm for Foreign Language Pronunciation Assessment in Correctional Education	6
2	Хамиджонов Шермурод Хомиджон ўғли. Namangan viloyatida sanoat tarmoqlari joylashuvining oʻziga xos xususiyatlari va rivojlanish tendensiyalari	10
3	Fakhriddin A. Boboyev, Malokhat U. Ilkhamova, Shakhruzbe O. Ozodov. Optimizing Overlock Seam Parameters for Joining Footwear Uppers to Nonwoven Strobel Insoles	14
4	Nuriddin Tojiyev, Shohruh Safarov, Behzod Tahirov, Otabek Abdullayev. Thermo-Magneto-Electro-Elastic Nonlinear Bending of Anisotropic Plates of Complex Shape	18
5	Yusufov Khakim, Khamidov Ilyosbek. Investigation of the strength properties of polymer-modified asphalt concrete mixtures	24

A Safety-Constrained Speaker-Adaptive Algorithm for Foreign Language Pronunciation Assessment in Correctional Education

Oybek Achilovich Kayumov
Jizzakh Branch of the National University of
Uzbekistan
Jizzakh, Uzbekistan
oybekuzonlined3@gmail.com

Nazokat Rashitovna Kayumova
Jizzakh Branch of the National University of
Uzbekistan
Jizzakh, Uzbekistan
kayumovanazokat7@gmail.com

Abstract—This study presents a safety-constrained, speaker-adaptive algorithm for automatically assessing foreign language pronunciation in correctional education. The method separates stable acoustic speaker characteristics from target pronunciation competence so that pitch range, timbre, habitual speaking rate, and recording channel are not incorrectly treated as pronunciation errors. Two to six unlabeled enrollment utterances are summarized with robust statistics and passed to a small residual adapter placed above a frozen self-supervised speech encoder. The adapter is restricted to a nuisance subspace by orthogonal projection, a preservation loss, and a trust region. A safety gate combines enrollment quality, forced alignment reliability, predictive uncertainty, score agreement, and profile drift to accept the adapted estimate, return the global estimate, or request human review. A reproducible simulation used one hundred sixty virtual speaker profiles, three hundred twenty unlabeled enrollment records, and nine hundred sixty evaluation samples under group-separated cross-validation. The proposed adapter reduced mean absolute error from 9.04 to 8.31 points and increased quadratic weighted agreement from 0.586 to 0.668. Selective scoring covered 84.7 percent of samples with a mean absolute error of 7.98 points. The data are synthetic and do not represent incarcerated individuals. Real deployment requires ethical approval, independent expert annotation, external validation, subgroup auditing, and strict separation from legal or disciplinary decisions.

Keywords—automatic pronunciation assessment, speaker adaptation, residual adapter, forced alignment, prosody, uncertainty, fairness

INTRODUCTION

Automatic pronunciation assessment can support individualized foreign language learning when access to instructors is limited. In correctional education, local deployment is particularly relevant because Internet connectivity, device availability, and access to external services may be restricted. The assessment problem is nevertheless more difficult than speech recognition. A scoring system must estimate the quality of the learner's production, not merely recover the intended words, and it must do so without confusing stable voice characteristics with linguistic errors.

The same phoneme may be realized with different fundamental frequency ranges, vocal tract characteristics, habitual rates, and microphone responses. A global model can penalize an unfamiliar but stable acoustic profile, whereas unconstrained personalization can absorb a genuine pronunciation error into an individual norm. This tension is methodologically important for a vulnerable population: an apparently accurate adaptation mechanism may improve average error while weakening construct validity or producing unequal residual errors across acoustic subgroups.

Speaker adaptation in speech recognition includes feature normalization, speaker embeddings, parameter adaptation, and data augmentation [1]. Residual adapters can modify self-supervised representations without updating the full encoder [2], while conditioned speaker representations have been used in computer-assisted pronunciation training [3]. These techniques do not by themselves guarantee that phoneme accuracy, completeness, stress, or rhythm remain invariant under personalization.

This work proposes a safety-constrained algorithm with five contributions. First, unlabeled enrollment is used to estimate only nuisance characteristics. Second, a residual adapter is projected away from pronunciation-sensitive directions. Third, a preservation objective and trust region limit unwanted score changes. Fourth, a do-no-harm gate selects between an adapted score, the frozen global score, and human review. Fifth, profile drift and selective online updating are explicitly audited. The method is evaluated with a fully reproducible synthetic protocol; the resulting figures are engineering evidence about algorithmic behavior, not clinical, legal, or population claims.

RELATED WORK AND PROBLEM DEFINITION

Goodness of Pronunciation remains a standard phoneme-level indicator, but its reliability depends on the acoustic model and temporal boundaries. CTC segmentation provides scalable alignment [4], and alignment errors can propagate directly into automatic pronunciation evaluation [5]. Recent systems therefore combine phoneme, word, and utterance information [6], [7], [8], [9]. Self-supervised encoders such as wav2vec 2.0, HuBERT, and WavLM provide robust speech representations [10], [11], [12], and their transfer to pronunciation assessment has shown that limited labeled data can be used efficiently [13].

Pronunciation competence is both segmental and suprasegmental. Segmental evidence includes phoneme posterior quality, substitutions, deletions, insertions, and duration. Suprasegmental evidence includes pitch contour, lexical stress, energy, rhythm, pausing, rate, and fluency [14], [15]. Absolute pitch or words per minute cannot be interpreted as universal quality measures; they must be compared with language targets and with a speaker's robust baseline. At the same time, speech technologies have documented disparities across accents and demographic groups [16], [17]. Consequently, accuracy, calibration, coverage, abstention, and subgroup residual errors must be reported together.

Let x_{ut} denote utterance t of speaker u and let E denote a frozen encoder. The observed representation is decomposed into target competence, a stable nuisance component, and residual noise:

$$z_{ut} = E\theta_0(x_{ut}) = zc(c_{ut}) + Bn_u + \varepsilon_{ut} \quad (1)$$

Here c is the pronunciation construct, n is the speaker and channel nuisance vector, and B spans nuisance directions. Legal status, sentence information, identity, and disciplinary data are excluded from both n and the scorer. The technical objective is to reduce the effect of n while preserving sensitivity to c .

TABLE I.

ADAPTATION BOUNDARY AND SAFETY
RATIONALE

Component	Permitted adaptation	Protected evidence
Voice baseline	Median pitch, timbre, rate, channel	Pitch movement and target timing
Phonetic path	Context and alignment confidence	Phone identity, accuracy, completeness
Prosodic path	Within-speaker normalization	Stress position, rhythm, contour
Decision	Confidence-based selection	Global fallback and human review

PROPOSED METHOD

Unlabeled enrollment and robust profile

A new user reads two to six short, phonetically balanced prompts. Enrollment records are separated from scored test records and require no expert score. Quality control rejects clipping, severe truncation, insufficient phonetic coverage, and unstable repeated readings. For feature k , location and scale are estimated with the median and median absolute deviation:

$$m_{uk} = \text{median}_t h_k(x_{ut}), d_{uk} = 1.4826 \text{MAD}_t[h_k(x_{ut})] \quad (2)$$

The profile contains median pitch, pitch interquartile range, spectral tilt, formant scale, articulation rate, energy statistics, channel descriptors, and quality metadata. A bounded enrollment quality index combines duration sufficiency, phonetic coverage, signal-to-noise ratio, cross-recording stability, and clipping. The simulation weights are 0.30, 0.25, 0.20, 0.20, and -0.35, respectively. All inputs are normalized to the unit interval.

Projection-constrained residual adapter

The global WavLM type encoder and scoring heads remain frozen. A pronunciation-sensitive basis S is learned on the development corpus from gradients associated with segmental and prosodic targets. The orthogonal projector prevents the adapter update from moving directly along those directions:

$$P \perp = I - S(S^T S + \epsilon I)^{-1} S^T \quad (3)$$

A small bottleneck network maps the robust profile to an eight-dimensional adaptation code. A fixed learned lifting operator A maps this code to the 256-dimensional encoder representation. The adapted representation is

$$a_u = g\phi(e_u), z'_{ut} = z_{ut} + P \perp A a_u \quad (4)$$

The adapter norm is clipped to a trust region. A preservation loss penalizes changes in frozen phoneme logits and in stress and rhythm descriptors. Training minimizes score error, preservation error, adapter energy, and the variance of residual error across observed nuisance quartiles. The fairness term uses acoustic nuisance groups only; demographic or legal attributes are neither required nor inferred.

$$L = L_{\text{score}} + \lambda_1 L_{\text{pres}} + \lambda_2 \|a_u\|_2^2 + \lambda_3 L_{\text{fair}} \quad (5)$$

Safety gate and selective decision

The global and adapted scorers operate in parallel. Confidence C combines enrollment quality, forced alignment reliability, predictive certainty, and agreement between the two scores. In (6), δ denotes the absolute difference between adapted and global scores. The simulation uses weights 0.15, 0.55, 0.15, and 0.15 with an intercept of 0.06, making alignment reliability the dominant component. An adapted score is accepted only when confidence is at least 0.59, and its difference from the global score does not exceed 12 points. Otherwise, the system returns the global score; a detected profile drift or severe disagreement triggers reenrollment or expert review.

$$C_{ut} = \text{clip}[w^T(Q_u, R_{\text{align}}, 1 - U, 1 - \Delta/\tau) + b, 0, 1] \quad (6)$$

Adaptation is not designed to increase scores. Its purpose is to remove systematic nuisance-related bias, so it retains both positive and negative corrections in the audit trail. The final adapted score combines calibrated phonetic and prosodic

subscores with weight 0.55 for the phonetic path and a bounded difficulty correction that shrinks when enrollment quality is low.

Drift-controlled profile updating

Only records with high confidence, adequate audio quality, and small global-to-adapted disagreement may update the profile. An exponential moving average uses a maximum rate of 0.20 and decreases with the number of accepted updates. A diagonal covariance formed from the robust scales supports Mahalanobis drift detection. With eight profile dimensions, the simulation freezes adaptation when squared distance exceeds 20.1, corresponding to a conservative one percent tail threshold.

$$e_u(t+1) = (1 - \eta_t)e_u(t) + \eta_t h(x_{ut}), \quad \eta_t \leq 0.20 \quad (7)$$

Recommendations are produced only from reliable features. A low baseline pitch is never labeled an intonation error. Feedback is generated for an incorrect contour relative to the speaker's range, misplaced lexical stress, unstable rhythm, or a reliable phoneme-level deviation. Figure 1 summarizes the information flow and the three safety outcomes.

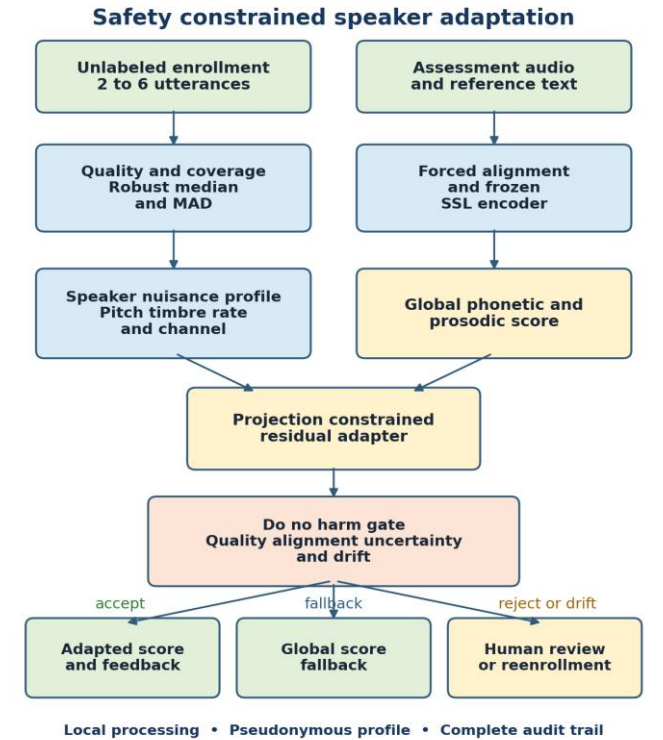


Fig. 1. Safety-constrained speaker adaptive assessment architecture.

EXPERIMENTAL DESIGN

No ethically approved corpus of incarcerated learners was available. The algorithm was therefore examined with a controlled synthetic generator. We created 160 virtual profiles. Each profile contained eight observations: two unlabeled enrollment records and six evaluation records. Four latent segmental and six latent prosodic competence features were generated from profile ability and task difficulty. Timbre and formant scale, baseline pitch, habitual rate, and energy or channel nuisance were mixed into the observed features through fixed loading matrices and measurement noise.

Six systems were compared. The global model ignored enrollment: full cepstral mean and variance normalization centered all features. The embedding baseline appended a speaker vector. The proposed adapter applied the projected residual update. The safety gate selected between adapted and global estimates. The oracle received the true nuisance vector and served as an upper bound. Eight-fold GroupKFold

separation ensured that a profile did not appear in both training and test partitions. We repeated the complete experiment with ten random seeds (1 to 10).

TABLE II. SYNTHETIC VALIDATION PROTOCOL

Item	Specification
Profiles and records	160 profiles; 320 enrollment; 960 evaluation
Latent features	4 segmental; 6 prosodic; 4 nuisance
Validation	8-fold profile-separated GroupKFold; 10 seeds
Metrics	Pearson r; Spearman rho; MAE; RMSE; QWK
Gate	Confidence at least 0.59; disagreement at most 12

Pearson correlation measures linear association, Spearman correlation measures rank consistency, mean absolute error and root mean square error are expressed on the 100-point scale, and quadratic weighted kappa uses strict 20-point score bands. Selective results are reported with coverage because an error calculated only on accepted cases can otherwise be misleading. Calibration size and acoustic nuisance quartiles are examined as ablations.

RESULTS

TABLE III. SYNTHETIC COMPARISON OF SPEAKER ADAPTATION METHODS

Method	Pearson r	MAE	QWK
Global	0.646	9.04	0.586
Full CMVN	0.233	11.66	0.111
Speaker embedding	0.688	8.57	0.636
Proposed adapter	0.712	8.31	0.668
Safety gate	0.700	8.45	0.650
Oracle nuisance	0.723	8.15	0.681

Table III and Fig. 2 show that unrestricted centering is harmful in this setting. Full CMVN removes nuisance variation but also suppresses stable proficiency information, producing the largest error and weakest agreement. A speaker embedding improves on the global system, but the proposed projection-constrained adapter performs better because it restricts the update to the intended nuisance space.

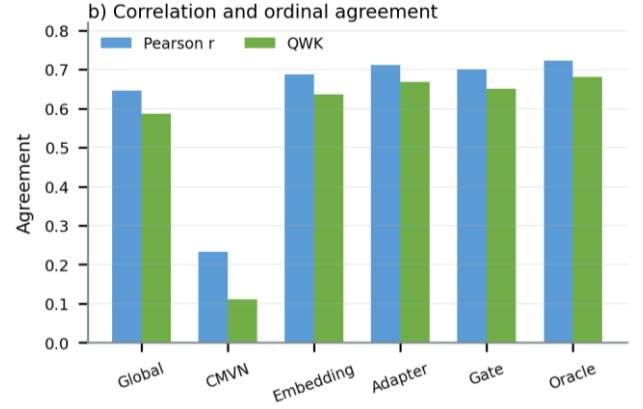
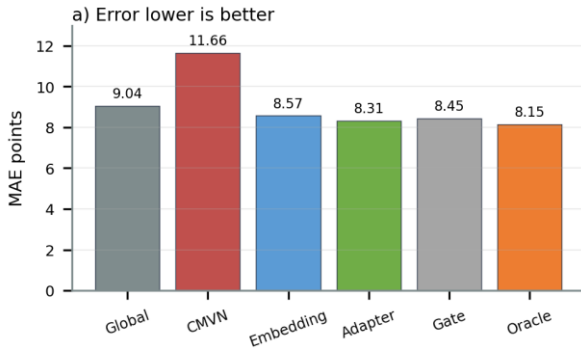


Fig. 2. Synthetic error and agreement across adaptation variants.

Relative to the global model, the proposed adapter reduces mean absolute error by 0.73 points, or approximately 8.1 percent, and increases quadratic weighted kappa from 0.586 to 0.668. Its mean absolute error is only 0.16 points above the oracle upper bound. The safety gate is intentionally conservative: its error is 0.14 points higher than the always adapted model because low-confidence changes are not accepted. For cases where the adapted score was released automatically, confidence selection achieved 84.7 percent coverage and a mean absolute error of 7.98 points.

TABLE IV. ENROLLMENT SIZE ABLATION

Enrollment records	MAE	Pearson r
0	9.04	0.646
1	8.41	0.705
2	8.31	0.712
4	8.24	0.716
6	8.25	0.716

The largest gain occurs between zero and two enrollment records. Four records provide a practical operating point, while six do not yield a meaningful additional improvement in this simulation. Error decreased in three of four nuisance quartile analyses, specifically timbre or formant, habitual rate, and energy or channel. The baseline pitch quartile difference was small. These acoustic quartiles are sensitivity tests and must not be interpreted as evidence of demographic fairness.

DISCUSSION

The main result is not simply that personalization lowers average error. The contribution is a mechanism that states what may be personalized, protects the linguistic construct, and exposes a controlled fallback when evidence is weak. Projection and preservation constraints formalize the separation between acoustic nuisance and target competence. The gate then turns model uncertainty into an operational decision rather than hiding it inside a single score.

The failure of full CMVN illustrates why unconstrained within-speaker normalization is unsuitable for assessment. A stable feature may encode nuisance, competence, or both. Removing every personal mean can therefore discard genuine proficiency. Speaker embeddings are less destructive but remain difficult to interpret because no intrinsic guarantee separates target and nuisance content. The proposed adapter approaches the oracle in the synthetic generator because the loading structure is known; this advantage will probably be smaller with real speech, where the two subspaces overlap, and alignment errors are not Gaussian.

A valid external study should include at least two trained phonetic raters, interrater intraclass correlation and weighted kappa, profile-separated testing, independent enrollment and

assessment prompts, unseen accents and microphones, calibration analysis, and profile-level bootstrap confidence intervals. Open corpora such as speechocean762 can support preliminary benchmarking [18], and joint assessment and mispronunciation detection can enrich diagnostic labels [19], but neither substitutes for an ethically approved target population study.

The synthetic protocol has four principal limitations. First, it assumes a low-dimensional stable nuisance vector. Second, the pronunciation-sensitive basis is learned from simulated target gradients. Third, expert score noise and rater severity are simplified. Fourth, subgroup analysis uses acoustic quartiles rather than protected attributes. Accordingly, the reported values demonstrate reproducibility and internal logic only; they do not estimate performance for any correctional institution or language group.

SECURITY AND ETHICAL SAFEGUARDS

Voice may constitute biometric data. Enrollment therefore requires a specific pedagogical purpose, freely given and documented consent, data minimization, local processing, encryption, role-based access, limited retention, and controls for deletion or reenrollment. The stored profile should contain compact nuisance statistics rather than raw audio whenever diagnostic traceability does not require the original signal. The system records the model version, adapter decision, confidence, score difference, drift event, and reviewer action in a tamper-evident audit log. The software modules should be tested against reliability, security, maintainability, and accountability requirements consistent with ISO IEC 25010 [20].

The system is intended only for language learning feedback. Incarceration status, sentence length, offense category, disciplinary history, or risk classification must never be model inputs. An automated score must not determine punishment, privilege, release, work assignment, security level, or access to services. Nonparticipation must have no adverse consequence, and every learner must have access to understandable feedback, human review, and a procedure for contesting an assessment.

CONCLUSION

We developed a speaker-adaptive pronunciation assessment algorithm around robust unlabeled enrollment, a frozen self-supervised encoder, a projection-constrained residual adapter, and a do-no-harm gate. The method personalizes pitch range, timbre, habitual rate, and channel effects while preserving phoneme identity, completeness, stress, rhythm, and contour evidence. Drift monitoring and selective profile updates prevent uncontrolled accumulation of adaptation error.

In the synthetic profile separated experiment, the proposed adapter achieved a Pearson correlation of 0.712, a Spearman correlation of 0.701, a mean absolute error of 8.31, a root mean square error of 10.51, and a quadratic weighted kappa of 0.668. Confidence-based release covered 84.7 percent of evaluation cases with a mean absolute error of 7.98. These values are not results from incarcerated individuals. The next research stage is an ethics-approved multilingual corpus study with independent phonetic raters, external validation, biometric security testing, and transparent subgroup auditing.

REFERENCES

1. P. Bell, J. Fainberg, O. Klejch, J. Li, S. Renals, and P. Swietojanski, "Adaptation algorithms for neural network-based speech recognition: An overview," *IEEE Open Journal of Signal Processing*, vol. 2, pp. 33–66, 2021, doi: 10.1109/OJSP.2020.3045349.
2. A. Bhatia, S. Sinha, S. Dingliwal, et al., "Do not stop self-supervision: Accent adaptation of speech representations via residual adapters," in *Proc. Interspeech*, 2023, pp. 3362–3366, doi: 10.21437/Interspeech.2023-2192.
3. Z. Zhang, Y. Wang, and J. Yang, "BiCAPT: Bidirectional computer-assisted pronunciation training with normalizing flows," in *Proc. Interspeech*, 2022, pp. 4511–4515, doi: 10.21437/Interspeech.2022-10937.
4. L. Kürzinger, D. Winkelbauer, L. Li, T. Watzel, and G. Rigoll, "CTC segmentation of large corpora for German end-to-end speech recognition," in *Speech and Computer*, LNCS 12335, 2020, pp. 267–278, doi: 10.1007/978-3-030-60276-5_27.
5. V. C. Mathad, T. J. Mahr, N. Scherer, et al., "The impact of forced alignment errors on automatic pronunciation evaluation," in *Proc. Interspeech*, 2021, pp. 1922–1926, doi: 10.21437/Interspeech.2021-1403.
6. B. Lin, L. Wang, X. Feng, and J. Zhang, "Automatic scoring at multi-granularity for second language pronunciation," in *Proc. Interspeech*, 2020, pp. 3022–3026, doi: 10.21437/Interspeech.2020-1282.
7. Y. Gong, Z. Chen, I. H. Chu, P. Chang, and J. Glass, "Transformer-based multi-aspect, multi-granularity non-native English speaker pronunciation assessment," in *Proc. ICASSP*, 2022, pp. 7262–7266, doi: 10.1109/ICASSP43922.2022.9746743.
8. F. A. Chao, T. H. Lo, T. I. Wu, Y. T. Sung, and B. Chen, "A hierarchical context-aware modeling approach for multi-aspect and multi-granular pronunciation assessment," in *Proc. Interspeech*, 2023, pp. 974–978, doi: 10.21437/Interspeech.2023-550.
9. H. C. Pei, H. Fang, X. Luo, and X. S. Xu, "Gradformer: A framework for multi-aspect, multi-granularity pronunciation assessment," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 554–563, 2024, doi: 10.1109/TASLP.2023.3335807.
10. A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 12449–12460.
11. W. N. Hsu, B. Bolte, Y. H. H. Tsai, et al., "HuBERT: Self-supervised speech representation learning by masked prediction of hidden units," *IEEE ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3451–3460, 2021, doi: 10.1109/TASLP.2021.3122291.
12. S. Chen, C. Wang, Z. Chen, et al., "WavLM: Large-scale self-supervised pre-training for full-stack speech processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 6, pp. 1505–1518, 2022, doi: 10.1109/JSTSP.2022.3188113.
13. E. Kim, J. J. Jeon, H. Seo, and H. Kim, "Automatic pronunciation assessment using self-supervised speech representation learning," in *Proc. Interspeech*, 2022, pp. 1411–1415, doi: 10.21437/Interspeech.2022-10245.
14. W. Dong, R. van Hout, C. Cucchiari, and H. Strik, "Second language prosody assessment by combining acoustic and neural model features," in *Proc. Speech Prosody*, 2024, pp. 165–169, doi: 10.21437/SpeechProsody.2024-34.
15. K. Fu, S. Gao, S. Shi, X. Tian, W. Li, and Z. Ma, "Phonetic and prosody-aware self-supervised learning approach for non native fluency scoring," in *Proc. Interspeech*, 2023, pp. 949–953, doi: 10.21437/Interspeech.2023-587.
16. N. Markl, "Language variation and algorithmic bias in British English automatic speech recognition," in *Proc. ACM FAccT*, 2022, pp. 521–534, doi: 10.1145/3531146.3533117.
17. A. Koenecke, A. Nam, E. Lake, et al., "Racial disparities in automated speech recognition," *Proceedings of the National Academy of Sciences*, vol. 117, no. 14, pp. 7684–7689, 2020, doi: 10.1073/pnas.1915768117.
18. J. Zhang, Z. Zhang, Y. Wang, et al., "speechocean762: An open source non native English speech corpus for pronunciation assessment," in *Proc. Interspeech*, 2021, pp. 3710–3714, doi: 10.21437/Interspeech.2021-1259.
19. H. Ryu, S. Kim, and M. Chung, "A joint model for pronunciation assessment and mispronunciation detection and diagnosis with multi-task learning," in *Proc. Interspeech*, 2023, pp. 959–963, doi: 10.21437/Interspeech.2023-337.
20. ISO IEC 25010:2023, *Systems and Software Engineering—Systems and Software Quality Requirements and Evaluation—Product Quality Model*. Geneva, Switzerland: International Organization for Standardization, 2023.