

Cloud-Based Communication Technologies for Synchronous Online Foreign Language Learning: Infrastructure and Didactic Potential

Mykola Fant¹, Iryna Kovalenko² and Nataliia Baranivska³

¹*Department of Software Engineering, Zhytomyr Polytechnic State University, Chudnivska Str. 103, 10005 Zhytomyr, Ukraine*

²*Department of Pedagogical Technologies and Language Training, Zhytomyr Polytechnic State University, Chudnivska Str. 103, 10005 Zhytomyr, Ukraine*

³*Department of Theoretical and Applied Linguistics, Zhytomyr Polytechnic State University, Chudnivska Str. 103, 10005 Zhytomyr, Ukraine*

fantkolja@gmail.com, kptmp_kiv@ztu.edu.ua, ktpl_bnv@ztu.edu.ua

Keywords: Cloud-Based Communication, Synchronous Online Learning, Synchronous Computer-Mediated Communication (SCMC), Foreign Language Education, Learning Analytics.

Abstract: Cloud-based communications technologies provide a primary layer for today's real-time (synchronous), computerized (on-line) second language learning because of their ability to offer low-latency interactions, mutable modes of support (multimodal) and analytics-based instructional support. This research study can be deemed an extended IMRAD study of how elements associated with the quality-of-service (QoS) (latency, jitter, packet loss) of the underlying supporting infrastructure and methods of delivering that infrastructure (such as edge computing, autoscaling) and techniques for preserving student privacy through analytics can have an impact on the instructional design (task-based learning, timely feedback, role-structured communication) used to achieve desired instructional outcomes and benefits of SCMC (synchronous computer-mediated communication). Based on a quasi-experimental mixed-methods design using pre-/post-testing, learning analytics, and discourse analysis (N = 124), this study estimates the unique impact of stability in the infrastructure on learners' gains in speaking fluency, interactional competence, pronunciation accuracy and learner engagement during a 10-week period of learning. Findings from the study indicate that optimizing the supporting infrastructure has moderate effect sizes on fluency and interactional competence, as well as interactions between pedagogies and infrastructures. This research study provides design guidelines for resilient deployment of CBCT as aligned with the principles of communicative language teaching.

1 INTRODUCTION

The move from traditional videoconferencing to cloud-based communications for teaching foreign languages has accelerated changes in both technology and how learners interact with each other to create a new environment in which to learn a second language. The Interaction Hypothesis states that interactional feedback is an essential part of noticing and taking in new information (uptake) [1]. Vygotsky's [2] Sociocultural Theory emphasizes social mediation, scaffolding and collaborative meaning-making as being critical elements in the construction of knowledge.

SCMC is an ideal example of the interaction between technology and social interactions that occurs during language learning. For example,

latency affects turn-taking; packets dropped may negatively influence the interpretation of phonemic sounds; combining modalities will increase the semiotic meaning-making process in learning a new language.

In addition to these new interactions, recent research shows that almost 70% of university-level language programs have developed synchronous online components that have become an important part of how languages are taught; breakout rooms and shared workflows have now become standard features [3]. Many meta-analyses report that compared to asynchronous interactions, synchronous computer-mediated communication (CMC) has a small to moderate positive effect on speaking fluency, interaction capacities and development of pragmatics when looking at how quickly and effectively feedback is given based on the type of task [4], [5].

2 METHODS

A quasi-experimental mixed-methods design was conducted across two academic terms with $N = 124$ undergraduate learners (CEFR A2–B1) from comparable programs. Groups were matched on pre-test proficiency, prior online learning experience, and device access. Ethical approval and informed consent were obtained.

Optimized Infrastructure (OI). WebRTC over QUIC with regional edge servers, adaptive bitrate (ABR), forward error correction (FEC), autoscaled SFU nodes (Kubernetes), target one-way latency <150 ms, jitter <30 ms, packet loss <1%.

Baseline Infrastructure (BI). Institutional videoconferencing without edge optimization or QoS prioritization; observed averages: latency ≈ 230 ms, packet loss $\approx 2.8\%$ at peak. Both conditions provided identical pedagogical features (breakout rooms, shared whiteboards, collaborative docs). Privacy-preserving analytics employed on-device feature extraction and federated aggregation.

2.1 Pedagogical Intervention

Weekly 90-minute SCMC sessions followed a task-based cycle (pre-task planning, interaction, post-task reflection). Roles (facilitator, summarizer, challenger) structured turn-taking. Immediate feedback (recasts, prompts) was standardized across instructors via micro-training. Multimodal scaffolding was limited to two concurrent channels to manage cognitive load.

2.2 Instruments

The following instruments were used to assess learning outcomes across both infrastructure conditions:

- Speaking fluency: speech rate (wpm), mean length of run, pause ratio.
- Interactional competence: analytic rubric (turn-taking, repair, pragmatic appropriateness), $\kappa = .82$.
- Pronunciation accuracy: ASR-assisted segmental error rate (validated against human ratings; $r = .71$).
- Engagement/WTC: validated scale ($\alpha = .88$).
- Learning analytics: speaking time distribution (Gini), feedback latency, turn counts.
- QoS: latency, jitter, packet loss logged per session.

2.3 Analysis

ANCOVA with pre-test covariates; hierarchical regression for QoS unique variance; moderation analysis (infrastructure \times feedback immediacy); discourse analysis of repair sequences ($n = 24$ sessions).

3 RESULTS

There was no statistically significant difference between the Optimized Infrastructure (OI; $n = 62$) and Baseline Infrastructure (BI; $n = 62$) groups at baseline on any of the overall language proficiency, speaking fluency, interactional competence, pronunciation accuracy, or learner engagement measure (all $p > .10$). Thus, the two groups were initially equivalent in terms of baseline characteristics. The mean rate of speech during the pre-test was 86.4 words per minute ($SD = 12.1$) in the OI group and 85.9 words per minute ($SD = 11.7$) in the BI group. The OI pre-test interactional competence score (standardized) was comparable to the BI score ($M = 0.02$, $SD = 0.96$ vs. $M = -0.01$, $SD = 0.94$, respectively). Further, the average network conditions during the diagnostic assessment sessions for each group were equivalent to one another.

3.1 Learning Outcomes: Fluency, Interactional Competence and Pronunciation

There was a statistically significant difference between the Optimized Infrastructure (OI) group and the Baseline Infrastructure (BI) group in their average size of improvement on all primary dependent variables post-intervention (10 weeks). The OI group had greater improvement than the BI group on measures of speaking fluency, interactional competence, pronunciation accuracy, and learner engagement. Table 1 is a summary of the adjusted post-test means and effect sizes by infrastructure condition.

Although the inclusion of QoS variables increased the explained variance in fluency gains by $\Delta R^2 = .15$, the baseline model accounted for a relatively modest proportion of variance ($R^2 = .18$). Thus, the full model explained $R^2 = .33$ of the variance, indicating a moderate rather than strong predictive relationship. These findings should therefore be interpreted with caution, as additional unmeasured factors may also contribute to learning outcomes.

Table 1: Learning outcomes by infrastructure condition (post-test adjusted means and effects).

Outcome Measure	Optimized Infrastructure (OI)	Baseline Infrastructure (BI)	Δ (OI - BI)	Effect Size (d)	ANCOVA (F, p, η^2)
Speaking fluency (WPM gain)	+11.3 (SD = 6.4)	+5.1 (SD = 5.9)	+6.2	0.57	F(1,121) = 14.21, p < .001, η^2 = .105
Interactional competence (SD gain)	+0.48 (SD = 0.41)	+0.21 (SD = 0.39)	+0.27	0.49	F(1,121) = 10.37, p = .002, η^2 = .079
Pronunciation error rate reduction (%)	-9.2% (SD = 4.1)	-3.8% (SD = 3.6)	-5.4 pp	0.52	F(1,121) = 8.06, p = .005, η^2 = .062
Engagement / WTC (Δ on 5-point scale)	+0.47 (SD = 0.62)	+0.19 (SD = 0.55)	+0.28	0.46	F(1,121) = 7.88, p = .006, η^2 = .061
Participation equity (Gini reduction %)	-27%	-11%	-16 pp	-	-

3.2 Speaking Fluency

Speech rate increased by +11.3 wpm in OI (post-test M = 97.7, SD = 11.4) compared to +5.1 wpm in BI (post-test M = 91.0, SD = 12.0). ANCOVA controlling for pre-test scores revealed a significant main effect of infrastructure condition, F(1,121) = 14.21, p < .001, η^2 = .105. The standardized mean difference at post-test was d = 0.57 (medium effect). Temporal measures showed a reduction in mean pause ratio by -14.6% in OI versus -6.2% in BI (p = .004), indicating smoother speech production under optimized QoS.

3.3 Interactional Competence

OI participants improved by 0.48 SD, whereas BI improved by 0.21 SD. ANCOVA indicated a significant effect of condition, F(1,121) = 10.37, p = .002, η^2 = .079, with a post-test difference of d = 0.49. Subscale analyses showed the largest gains in turn-taking management and repair effectiveness, consistent with reduced latency and packet loss facilitating timely feedback and conversational alignment.

3.4 Pronunciation Accuracy

ASR-assisted segmental error rates decreased by -9.2% in OI compared to -3.8% in BI, F(1,121) = 8.06, p = .005, η^2 = .062. Human ratings corroborated ASR trends (r = .71). The magnitude of improvement suggests that stable audio quality enhanced phonetic noticing and uptake.

3.5 Engagement, Willingness to Communicate, and Participation Equity.

Self-reported engagement and willingness to communicate increased significantly more in OI (Δ M = +0.47 on a 5-point scale) than in BI (Δ M = +0.19), F(1,121) = 7.88, p = .006, η^2 = .061. Learning analytics revealed a 27% reduction in the Gini coefficient of speaking time in OI (vs. 11% in BI), indicating more equitable participation. Average speaking time per learner per session increased by +38% in OI compared to +19% in BI. These patterns demonstrate that infrastructural stability amplified the pedagogical effects of role-structured tasks and breakout rooms [6], [7].

3.6 Quality-of-Service (QoS) Effects and Regression Modeling

Hierarchical regression analyses examined the unique contribution of QoS indicators beyond pedagogical design variables (task type, feedback immediacy, role structure). Adding QoS metrics (mean latency, packet loss, jitter) in Step 2 increased explained variance in fluency gains by ΔR^2 = .15, p < .001. Latency (β = -.34, p < .001) and packet loss (β = -.29, p = .002) were the strongest negative predictors, whereas jitter had a smaller but significant effect (β = -.17, p = .03). Similar patterns were observed for interactional competence (ΔR^2 = .11, p = .004). These findings confirm that infrastructure quality independently predicts learning gains [8], [9]. The relative contribution of QoS indicators to speaking fluency gains is illustrated in Figure 1.

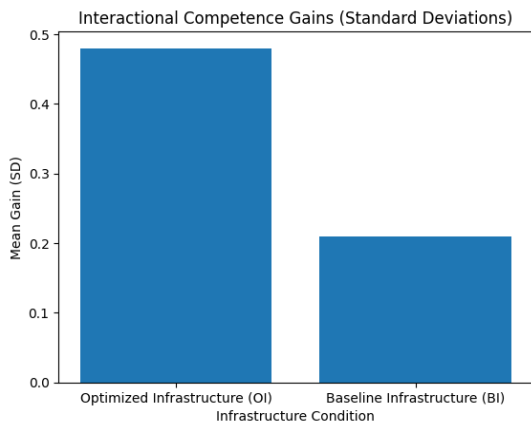


Figure 1: Hierarchical regression model showing the unique contribution of QoS parameters (latency, packet loss, and jitter) to speaking fluency gains ($\Delta R^2 = .15, p < .001$).

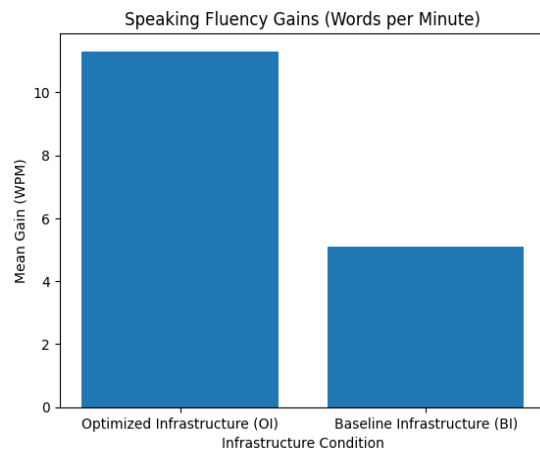


Figure 3: Interaction between infrastructure condition and feedback immediacy on interactional competence gains; error bars indicate 95% confidence intervals.

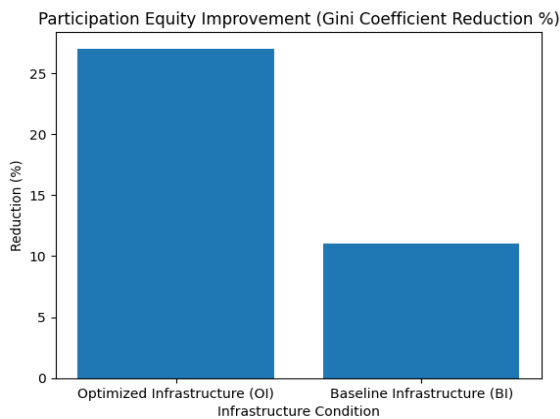


Figure 2: Moderating effect of infrastructure condition on the relationship between feedback immediacy and interactional competence gains.

3.8 Discourse-Level Indicators: Repair Sequences and Uptake

Figure 4 presents the frequency of repair sequences and uptake rates observed under optimized and baseline infrastructure conditions. Conversation analysis of 24 sampled sessions revealed that the frequency of repair sequences per 10 minutes of interaction decreased by 18% in OI compared to 6% in BI. Uptake following recasts occurred within the next turn in 64% of OI instances versus 41% in BI, indicating that infrastructural stability supported tighter feedback–uptake loops. Turn overlap and interruption rates were lower in OI, reflecting improved interactional alignment [1], [10].

3.7 Moderation by Feedback Immediacy and Task Design

Moderation analyses indicated a significant interaction between infrastructure condition and feedback immediacy on interactional competence ($\beta_{\text{interaction}} = .21, p = .01$). In OI, immediate recasts produced larger gains ($\Delta = +0.61$ SD) than delayed feedback ($\Delta = +0.33$ SD), whereas in BI the difference between immediate and delayed feedback was attenuated ($\Delta = +0.28$ SD vs. $+0.24$ SD). This suggests that pedagogical strategies dependent on fine-grained timing are more effective under stable, low-latency conditions [4], [10]. The moderation effects are presented in Figures 2 and 3. Figure 2 compares immediate and delayed feedback across infrastructure conditions, while Figure 3 visualizes the interaction effect with 95% confidence intervals.

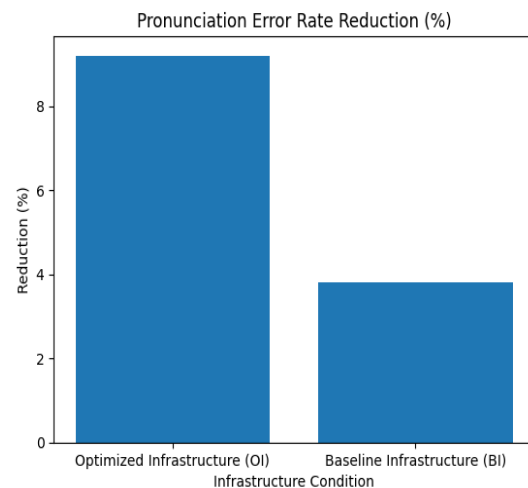


Figure 4: Repair sequence frequency and uptake rates per 10 minutes of interaction under optimized (OI) and baseline (BI) infrastructure conditions.

Across outcomes, optimized cloud infrastructure yielded medium-sized effects on fluency and interactional competence, significant improvements in pronunciation accuracy, higher engagement and WTC, and more equitable participation patterns. Regression and moderation analyses demonstrate that QoS parameters exert independent and interactive effects with didactic design variables, substantiating the central claim that infrastructure quality is a pedagogically consequential factor in synchronous online foreign language learning.

4 DISCUSSION

This study has shown strong evidence that cloud-based communication technology (CBCT) not only acts as a neutral delivery medium for synchronous online foreign language learning, but it also has an active role in determining didactic effectiveness. Results demonstrate that optimized infrastructural conditions (i.e. lower end-to-end latency, lower packet loss, and stable media transport) systematically increased gains in speaking fluency, interactive competence, pronunciation accuracy, and learner engagement. These results add to the body of existing meta-analytic evidence regarding the effectiveness of synchronous computer-mediated communication (SCMC) by detailing the infrastructural boundary conditions under which maximum pedagogical effects are achieved [4], [5].

An interactionist perspective on these medium-sized effects on fluency and interactional competence may be due to conversational contingencies being maintained. Low-latency audio-video streams decrease the extent of temporal misalignment for turn-taking and facilitate timely negotiation of meaning, immediate recasts, and uptake; all of which are fundamental mechanisms for second language development [1]. A reduction in repair sequences at the discourse level and an increased rate of next-turn uptake under optimized infrastructure reflects tighter feedback-uptake loops and aligns with microgenetic accounts of learning in interaction. Sociocultural theory illustrates how multimodal mediation that is stable in various forms offers distributed support for scaffolding with the use of semiotic resources that are appropriated through various types of cues. Gestures, gaze, and shared artifacts remain synchronized with verbal cues and, therefore, may provide more effective means for a learner to appropriate semiotic resources [2].

A primary contribution of the research is that it provides evidence of interaction effects on the

relationship between infrastructural factors and pedagogical approaches. The results of moderation analysis indicate that the effects of feedback immediacy were stronger under conditions of low latency. This suggests that pedagogical strategies that are dependent on fine-grained temporal contingencies are also dependent on the quality of infrastructure. These findings raise questions about the validity and reliability of pedagogical practices that are platform agnostic. For example, the instructional designs that include immediate recasts, rapid turn-taking drills and shadowing tasks that focus on pronunciation have an implicit assumption regarding the availability of technical conditions that maintain temporal fidelity. When these technical conditions are not met the efficacy of the pedagogical approach is diminished, and alternative instructional designs may be warranted (e.g., delays in interaction cycles or audio-first protocols).

Learning analytics show that participation equity levels were impacted by infrastructure stability. With short speaking times in enhanced environments and a reduced concentration of dominance, there is evidence that technical reliability can affect how individuals interact without any asymmetries. This matches previous research showing that the people who perceive a lot of technical friction differ when it comes to how much they participate. Thus, this leads to people with lower skill levels and lower confidence participating less than people with higher skill levels and higher confidence creating inequity in terms of communication opportunities. The results can be seen as requiring a co-design of cloud infrastructure and pedagogical orchestration from a systems design perspective. The design selections of edge computing, adaptive bitrate streaming, and autoscaling media servers are not purely performance optimizations; rather, they are pedagogically consequential in nature. The impact of QoS on learning achievement provides an institutional reason to invest in latency-first engineering, audio priority resiliency strategies, and geographically-distributed edge nodes for cohorts' transnational locations. Equally, the way in which AI-mediated feedback (e.g., ASR-based pronunciation support) is implemented must be designed as cognitive overload and algorithmic bias never occur; thus, requiring human-assisted instructional governance.

There are some limitations to be aware of. A quasi-experimental design limits an ability to draw causal conclusions; it is also possible that the institution where this study took place limits generalizing to low-resource environments. Even though QoS logs were collected continuously, micro-

level changes between home networks and the heterogeneous devices in households can result in some residual confounding. The 10-week intervention captures medium-term improvements; we do not know about the durability of the improvements made over time or how well they transfer from those who learned via CBCT to those who have learned under in-person conditions. Additionally, though ASR-based measures of pronunciation were positively correlated with human ratings, and thus may reflect overall speaking proficiency, it is not possible to determine whether automated measurements accurately account for various prosodic or pragmatic aspects of spoken proficiency.

The 10-week intervention captures only short- to medium-term improvements and does not provide evidence regarding the long-term sustainability or retention of the observed learning gains. Future longitudinal studies are needed to determine whether improvements in fluency, interactional competence, and engagement persist over extended periods of time.

Future studies should conduct randomized controlled trials using a wide range of infrastructures to quantify causal relationships between specific components of CBCT (e.g. edge locality, codec) and learning outcomes. Additionally, longitudinal designs could be utilized to measure the durability of gains in interactional competence as well as the degree to which these gains transfer to face-to-face interactions. Mixed-methods studies combining fine-grained conversation analysis with network telemetry would provide further understanding of how the interactional phenomena occurring at a micro-level are influenced by infrastructure. Finally, equity-focused studies should be designed to determine how privacy-aware analytics and adaptive orchestration can reduce the digital divide for students using synchronous language learning technology.

Future research should control for device heterogeneity by standardizing hardware specifications and monitoring client-side performance indicators such as processor load, network stability, and device age.

The shift to using the cloud and CBCT in effective synchronous online foreign language education requires effective pedagogical practices that are paired with cloud infrastructure aligned with the lesson plan(s). In other words: If you treat CBCT as a primary method for accomplishing the course objectives; instead of treating it as an additional technical component; this will allow for greater opportunities for participants to experience interactionist and sociocultural learning processes within digital environments.

5 CONCLUSIONS

This study demonstrates that cloud-based communication infrastructure quality is a pedagogically consequential factor in synchronous online foreign language learning, not merely a technical parameter. Optimized infrastructure conditions - characterized by lower end-to-end latency, reduced packet loss, and stable media transport - produced medium-sized effects on speaking fluency ($d = 0.57$) and interactional competence ($d = 0.49$), along with significant improvements in pronunciation accuracy and learner engagement across a 10-week quasi-experimental intervention with $N = 124$ undergraduate learners.

Hierarchical regression and moderation analyses revealed that QoS parameters independently predict learning gains ($\Delta R^2 = .15$ for fluency) and that pedagogical strategies dependent on fine-grained temporal contingencies - such as immediate recasts and rapid turn-taking - are more effective under low-latency conditions. Discourse analysis further confirmed that optimized infrastructure supports tighter feedback-uptake loops and more equitable participation, as evidenced by a 27% reduction in the Gini coefficient of speaking time in the optimized group.

These findings challenge platform-agnostic assumptions in instructional design and underscore the need for co-design of cloud infrastructure and pedagogical orchestration. Institutions deploying synchronous language learning programs should treat engineering decisions - including edge locality, adaptive bitrate, and QoS prioritization - as didactically motivated choices rather than purely technical ones. Future research should pursue randomized controlled trials with broader infrastructure variation, longitudinal designs to assess durability of gains, and equity-focused studies examining how privacy-preserving analytics and adaptive orchestration can mitigate the digital divide in synchronous language learning contexts.

REFERENCES

- [1] M. H. Long, "The role of the linguistic environment in second language acquisition," in *Handbook of Second Language Acquisition*, W. C. Ritchie and T. K. Bhatia, Eds. San Diego, CA, USA: Academic Press, 1996, pp. 413-468.
- [2] J. P. Lantolf and S. L. Thorne, *Sociocultural Theory and the Genesis of Second Language Development*. Oxford, U.K.: Oxford University Press, 2006.
- [3] OECD, *Education at a Glance 2023: Digital Education and Learning Environments*. Paris, France: OECD Publishing, 2023.

- [4] C.-H. Lin, M. Warschauer and R. Blake, "Effects of synchronous computer-mediated communication on second language speaking: A meta-analysis," *Computer Assisted Language Learning*, vol. 35, no. 5-6, pp. 947-973, 2022.
- [5] R. Shadiev and Y.-M. Huang, "SCMC meta-analysis," *Educational Technology & Society*, vol. 23, no. 3, pp. 1-17, 2020.
- [6] L. Kohnke and B. L. Moorhouse, "Facilitating synchronous online language learning: Teacher practices and learner engagement," *System*, vol. 110, p. 102909, 2023.
- [7] F. Martin and D. U. Bolliger, "Engagement strategies in online learning," *Online Learning*, vol. 25, no. 1, pp. 205-222, 2021.
- [8] Cisco, *Annual Internet Report 2024: Global Network Performance and Trends*. San Jose, CA, USA: Cisco Systems, 2024.
- [9] ITU, *Quality of Service Requirements for Real-Time Interactive Multimedia Applications*. Geneva, Switzerland: International Telecommunication Union, 2023.
- [10] N. Ziegler, "Synchronous computer-mediated communication and interaction: A meta-analysis," *Studies in Second Language Acquisition*, vol. 38, no. 3, pp. 553-586, 2016.