

Using YouTube to enhance English international communication abilities in China: multimodal learning strategies

Sijia Ma

Dalian University of Foreign Languages, Dalian, China

may052021@163.com

Abstract. This conceptual paper proposes a framework to address the gap between exam-oriented English education and the real-world communication needs of Chinese university students. It synthesizes existing research to introduce and theoretically justify two intervention strategies—Listening—Reading—Repetition—Memorization and YouTube-Based Structured Learning with Shadowing—proposed to support the development of listening comprehension and oral fluency. Grounded in relevant research, these methods are tailored to different proficiency levels and rely on freely available online resources to ensure accessibility for all learners. This approach not only has the potential to support pedagogical improvement but also may promote educational equity and communication-focused reform in China's EFL context.

Keywords: English as a Foreign Language (EFL), international communication, listening comprehension, oral proficiency, educational equity, multimodal learning, Chinese university students

1. Introduction

1.1. Background and problem statement

In today's increasingly globalized society, English has emerged as the predominant lingua franca, playing a critical role in international communication, cross-cultural interactions, and global collaboration [1-4]. Consequently, effective English communication skills have become increasingly important for Chinese students' academic success and future career opportunities in a globalized world. Nevertheless, the current approach to English education in Chinese high schools remains predominantly focused on grammar proficiency, vocabulary memorization, and exam-oriented preparation [5-8], often at the expense of practical communication skills needed for real-world interaction [9, 10].

As a direct consequence of this educational approach, many Chinese university students face significant and persistent challenges when they engage in real-life communicative contexts [11, 12]. In particular, students often struggle with comprehending spoken English, especially when confronted with different accents, rapid speech rates, and informal language [13, 14]. Additionally, they exhibit difficulties in spoken English, manifesting in poor pronunciation, insufficient fluency, limited vocabulary retrieval under pressure,

and a general lack of confidence in spontaneous oral interactions [13, 15]. These issues collectively indicate a deficiency in what can be referred to as "international communication abilities" [12, 13].

International communication abilities refer to learners' competencies in effectively understanding and using spoken language within authentic, culturally diverse communicative contexts [12, 13]. These skills include two primary components: listening comprehension, which involves accurately interpreting and processing speech characterized by diverse accents, speech speeds, and informal patterns [13, 14]; and oral proficiency, encompassing clear pronunciation, conversational fluency, appropriate vocabulary use, and the capability to engage confidently in spontaneous dialogues [12, 15].

Despite the acknowledged importance of practical communication skills, existing research remains limited—often focusing on either listening or speaking in isolation and favoring theoretical over practical approaches. Moreover, research addressing the challenges of Chinese university students is sparse, resulting in a significant gap in evaluating targeted interventions to improve listening comprehension and oral proficiency in this context. To address this gap, the present paper adopts a research-informed conceptual approach. Rather than reporting empirical findings, it proposes a pedagogical intervention framework informed by existing studies. The ultimate objective is to improve students' listening comprehension and oral fluency through structured multimodal learning activities.

1.2. International communication abilities

International communication abilities refer to a set of receptive and productive language competencies that enable learners to effectively engage in oral communication within linguistically and culturally diverse contexts. For Chinese university students learning English as a Foreign Language (EFL), these abilities are not limited to grammatical accuracy or vocabulary size but encompass the capacity to understand, interpret, and produce spoken English in real time, often under cognitively demanding or unfamiliar intercultural conditions.

A core component of international communication ability is listening comprehension, which involves the capacity to accurately interpret spoken English delivered at natural speed. Effective listening comprehension also entails familiarity with a range of English accents, including but not limited to American, British, and Australian varieties, as well as the ability to understand informal or reduced speech forms such as idioms, contractions, and elisions. Furthermore, learners must develop the resilience to comprehend speech in suboptimal conditions, including background noise, speaker variability, or lack of contextual support. Such capabilities require not only linguistic knowledge but also auditory discrimination and pragmatic inference skills.

The scope of international communication abilities therefore extends beyond linguistic competence to incorporate intercultural awareness and communicative flexibility. Communicative competence in English education must integrate social and intercultural understanding, as language use in real-world contexts is inherently shaped by cultural norms and expectations. Without this broader framework, language teaching risks producing technically proficient but communicatively ineffective speakers. Similarly, four essential dimensions of cross-cultural communication competence—linguistic ability, cultural literacy, intercultural adaptability, and communicative appropriateness—are as relevant to English learners as they are to language teachers. These dimensions highlight the importance of integrating language instruction with cultural content to prepare learners for global engagement.

Building on this perspective, Li applies Byram's model of Intercultural Communicative Competence (ICC) in the context of a college English curriculum and concludes that communicative success depends not only on mastery of language forms but also on learners' ability to interpret, relate, and respond appropriately across cultures [16]. This reinforces the view that international communication abilities are multifaceted, requiring

both language and cultural knowledge. Liu further elaborates on this by framing intercultural communication as a dynamic process involving language use, cultural adaptability, and a global mindset [17]. He underscores that learners must possess the cognitive and affective readiness to negotiate meaning across cultures, particularly in high-stakes academic and professional settings.

Zhou, Xu, and Bayley offer additional support for this integrated approach, arguing that international communication in the Chinese EFL context requires not only linguistic proficiency but also intercultural mediation skills [18]. Their research highlights the need for learners to function as cultural intermediaries who can interpret and reconcile differing communicative norms. This perspective is particularly salient in today's globalized environment, where English is used as a lingua franca across diverse sociolinguistic communities.

1.3. Aims and objectives

This paper aims to bridge the theory-practice gap by proposing a pedagogical framework informed by second language acquisition research. Its specific objectives are to:

1. Synthesize key research on multimodal input and accent exposure relevant to the Chinese EFL context.
2. Propose two structured intervention strategies (LRRM and YouTube-Based Shadowing) tailored to different proficiency levels.
3. Discuss the practical implications for educators and learners and outline the broader societal impact of implementing such strategies.

1.4. Theoretical underpinnings: multimodal and varied input for listening development

The following review is structured around the specific auditory and perceptual challenges faced by Chinese EFL learners, notably their reliance on orthographic rather than phonological processing and their limited exposure to authentic speech variety, and outlines the research-supported solutions that inform our proposed methods.

1.4.1. Multimodal input in listening and speaking

Chinese EFL learners often struggle to connect written English forms with their pronunciation, due in part to differences between the logographic Chinese script and the alphabetic English system. Consequently, phonological decoding practice—sounding out words—plays a crucial role in building phonemic awareness and improving pronunciation. Research shows that lower-proficiency learners benefit significantly from decoding strategies: for example, Zhao, Wei, and Yao found that beginners rely heavily on phonological decoding (whereas more advanced learners gradually shift to whole-word reading), and direct instruction in these skills led to measurable gains in word recognition [19, 20]. Theoretical models of reading and speech acquisition also underscore the importance of establishing strong sound-symbol connections. Ehri and Perfetti, Liu, and Tan argue that developing a robust phonological route facilitates vocabulary retention and accurate pronunciation, which are essential for listening comprehension and speaking fluency [21, 22]. Moreover, phonological decoding practice can heighten learners' phonemic awareness, strengthen the form-sound-meaning connection (thereby aiding vocabulary acquisition), and improve pronunciation through repeated exposure to English sound patterns. However, mastering decoding requires sustained practice. This practice demands a high level of consistency. Learners who have already achieved some reading fluency through sight-word memorization may see diminishing returns from additional phonics training [19]. Nevertheless, phonological decoding remains most beneficial for beginning to intermediate learners who have not yet internalized English phonemic rules or who struggle with pronunciation and word recognition, underscoring the need to scaffold it appropriately for those without prior phonics training. For novice learners,

instructors may integrate decoding exercises (such as phonics drills, minimal pair practice, or guided oral reading) into regular lessons to build these foundational skills.

At the same time, combining auditory and textual input through subtitled videos offers a multimodal scaffold to improve listening comprehension and vocabulary acquisition. Subtitles allow learners to match spoken English with written words in real time, easing the cognitive load of processing rapid speech. Empirical studies have demonstrated the benefits of subtitles: Li reported that Chinese university students retained new vocabulary best with bilingual subtitles, while Birulés-Muntané and Soto-Faraco observed improved comprehension of authentic speech when learners used second-language (L2) subtitles [23, 24]. The effectiveness of a given subtitle type depends on proficiency—beginners often need L1 (first language) support (as using L2-only subtitles too early can overwhelm novice learners), whereas more advanced learners profit from L2-only subtitles. Importantly, a review by Matielo, D'Ely, and Baretta emphasizes that subtitles are most effective when paired with active learning techniques (e.g., note-taking, transcription, or shadowing) rather than used alone [25]. Thus, a scaffolded approach is recommended: instructors and learners can begin with content featuring Chinese (L1) or bilingual subtitles and progressively transition to English-only subtitles as listening skills improve, all while engaging in active post-viewing exercises. For instance, after watching a video, a learner might transcribe key sentences or verbally summarize the content. Such follow-up activities reinforce comprehension and retention. This multimodal input strategy directly addresses Chinese EFL learners' difficulties with real-time listening and supports the development of both receptive and productive skills.

1.4.2. Exposure to multiple accents vs. multiple talkers

In an increasingly globalized world where English functions as a lingua franca across diverse sociolinguistic settings, exposure to a variety of English accents and speaking styles is critical for preparing learners to engage in authentic international communication. While most English language instruction in Chinese universities still relies on standardized native-speaker models (often American or British), this limited scope does not reflect the rich accentual diversity learners will encounter in real-world contexts. Recent research has therefore turned its focus toward the benefits of exposing second language (L2) learners to accentual and talker variability to improve listening comprehension and perceptual flexibility.

Two distinct but complementary strands of input variation are commonly studied: exposure to multiple talkers, which involves listening to different speakers of the same accent (e.g., multiple American English speakers with differing vocal characteristics), and exposure to multiple accents, which refers to encountering speech from speakers with different regional or national varieties of English (e.g., American, British, Indian, or Australian English). Each form of variability presents unique challenges and benefits for learners, particularly for those whose prior listening experience is limited to textbook recordings or monolingual instruction.

Potter and Saffran conducted a foundational study demonstrating the benefits of early and repeated exposure to multiple accents [26]. Although their work focused on infants, their findings have significant implications for adult L2 learners. The researchers found that exposure to a wide range of phonetic variability during early listening tasks enhanced participants' ability to generalize their comprehension to novel and unfamiliar accents. This generalization capacity—referred to as perceptual adaptability—is crucial for understanding accented speech in global communication contexts where English is spoken with diverse phonological patterns.

In a parallel line of inquiry, Bent and Holt examined how variation in talkers and foreign accents affects L2 spoken word identification [27]. Their results indicated that learners who were exposed to a variety of speakers, even within the same accent group, developed a heightened sensitivity to phonetic variability and

were better able to identify words spoken by unfamiliar voices. This exposure helped learners become more attuned to subtle differences in prosody, intonation, and articulation—features often overlooked in monologic or scripted speech recordings.

These findings are especially relevant for Chinese learners, many of whom report difficulties understanding non-standard or unfamiliar English accents due to limited exposure. Traditional listening comprehension exercises in the Chinese university context typically present a single "standard" voice, leading to an over-reliance on narrow phonological cues and poor adaptability in real-life communication scenarios. By contrast, exposure to a broader spectrum of accentual input encourages the development of bottom-up listening strategies, such as segmental decoding and contextual inference, while reducing overdependence on top-down strategies that may fail in unpredictable linguistic environments.

The advantages of incorporating both accent and talker variability into language instruction are well documented. First, it enhances learners' perceptual flexibility, allowing them to more readily process unexpected phonetic inputs. Second, it prepares students for real-world linguistic encounters—such as international travel, academic exchange, or global business contexts—where English is spoken in a range of accents. Third, such exposure helps normalize diversity in speech patterns, reducing the anxiety or resistance some learners experience when first encountering unfamiliar accents.

However, this approach is not without its limitations. For low-proficiency learners, initial exposure to unfamiliar accents or multiple speakers may result in reduced comprehension, increased cognitive load, and even listening fatigue. This can be discouraging and may lead to decreased motivation if not scaffolded appropriately. Furthermore, learners accustomed to highly standardized inputs may exhibit initial anxiety or frustration when confronted with less intelligible speech forms, particularly if such materials are presented without sufficient preparatory activities or follow-up support.

To mitigate these challenges, educators are advised to adopt a gradual and scaffolded approach. Instruction should begin with familiar and intelligible accents, such as General American or Received Pronunciation, using controlled listening materials like short, scripted dialogues. As learners gain confidence, instruction can expand to include unfamiliar or regional accents (e.g., Australian, Indian, or South African English) and authentic materials, such as interviews, vlogs, and spontaneous speech recordings. Supplementary activities—such as guided comprehension questions, transcription exercises, and discussion-based reflection—can also support learners in navigating the increased complexity. Thus, incorporating accent variability into training requires careful planning and scaffolding to yield effective outcomes.

2. Pedagogical framework

This section proposes two practical, research-informed intervention strategies tailored to learners at different proficiency levels to address the challenges outlined in previous sections. These strategies are intended for educators and self-directed learners aiming to improve international communication abilities, particularly in listening and speaking. Accordingly, they can be integrated into formal classroom curricula or adapted for self-paced learning as long as appropriate support is provided.

Conceptually, this pedagogical framework draws on several established theories. First, it aligns with multimedia learning theory by leveraging dual auditory-visual input to enhance comprehension. Second, it reflects Communicative Language Teaching (CLT) principles by prioritizing functional communication practice over rote memorization [28]. Third, it incorporates intercultural communicative competence (Byram's ICC model), emphasizing cultural awareness and appropriate intercultural interaction as part of language proficiency [16]. By adhering to these principles, the proposed methods aim to holistically improve learners'

listening and speaking skills in authentic contexts. Taken together, these frameworks provide a foundation for the combined use of multimedia input and communicative practice in the methods that follow.

2.1. Listening–reading–repetition–memorization method

This multimodal intervention is specifically designed for lower to intermediate-level Chinese university students who struggle with listening comprehension, pronunciation, and vocabulary retrieval in spontaneous English communication. These learners often face challenges decoding spoken input at natural speed, particularly when unfamiliar vocabulary, accents, or informal expressions are involved. Additionally, they may lack the confidence and automaticity required for fluent oral expression. The Listening–Reading–Repetition–Memorization method provides a structured, scaffolded approach that addresses these difficulties through a sequenced set of receptive and productive activities.

The method consists of four interconnected stages: listening with visual support, transcription and memorization, oral repetition, and productive output. These stages are grounded in the principles of multimodal learning and second language acquisition, with the aim of reinforcing phonological awareness, vocabulary retention, and spoken fluency.

The first stage involves listening to spoken English while reading along with the transcript. Learners engage with high-frequency vocabulary lists, short dialogues, or scripted passages using audio recordings paired with written text. This dual-input approach is particularly beneficial for learners at earlier stages of acquisition, as it allows them to match auditory forms with orthographic representations, improving decoding accuracy and word recognition. Empirical research has shown that reading-while-listening strategies significantly enhance comprehension and incidental vocabulary learning compared to reading-only approaches [29].

To maximize the effectiveness of this method, both teachers and learners need to select materials that are appropriately challenging and level-specific. One practical strategy is to begin with short articles, dialogues, or video clips and take a brief comprehension quiz afterward. If learners score around 70%, the material is considered suitably difficult—it provides enough challenge to promote learning without causing excessive frustration. Teachers can curate or create such materials in advance. At the same time, self-learners can find graded English listening resources online, including platforms like Elllo.org, BBC Learning English, or VOA Learning English, which offer audio with transcripts and quizzes.

For beginning-level learners, bilingual subtitles are especially helpful during the initial stages of listening practice. They provide immediate semantic support and help map spoken English to familiar concepts in the learner's first language. Teachers can support this process by creating customized video content with L1–English bilingual subtitles, particularly for frequently used sentence structures or thematic dialogues. As students progress to intermediate levels, educators should gradually shift from bilingual to English-only subtitles, in alignment with the scaffolding approach suggested by Li and Matielo et al. [23, 25]. This gradual transition helps learners develop independence in decoding English speech while receiving contextual support.

In the second stage, learners transcribe and memorize key phrases or collocations from the text. This may include dictation exercises or fill-in-the-blank activities designed to heighten attention to lexical form and usage. Memorization tasks focus on functional chunks of language, such as common sentence stems or collocational pairs, which contribute to fluency and syntactic awareness.

The third stage centers on oral repetition, in which learners imitate the model pronunciation of native speakers. By repeating the audio content aloud—paying attention to stress, intonation, and rhythm—learners develop more accurate and natural speech patterns. This shadowing technique supports phonological encoding

and oral fluency. Spring and Takeda emphasize that repeated speaking practice plays a crucial role in developing long-term retrieval fluency and pronunciation accuracy [30].

The final stage involves productive use of the learned material through retelling or paraphrasing, either orally or in writing. Learners reconstruct the original content using their own words, reinforcing internalization and promoting flexible language use in new communicative contexts. This output phase not only consolidates vocabulary and syntax but also aligns with Swain's Output Hypothesis, which highlights the cognitive benefits of being pushed to produce language [31].

This method offers several pedagogical advantages. It supports long-term vocabulary retention through active repetition and multimodal input, strengthens pronunciation and speech fluency via imitation of authentic audio models, and encourages the development of automaticity in both comprehension and production. Importantly, it provides a clear, repeatable structure that allows learners to track progress over time [30].

By offering level-appropriate materials and carefully managing subtitles based on learners' proficiency, educators can further enhance the accessibility and effectiveness of the Listening–Reading–Repetition–Memorization method.

Nevertheless, the method is time-consuming and requires a high level of learner motivation and consistency. Without sustained effort, learners may not experience the full range of benefits of this intensive practice. As such, its effectiveness depends on students' willingness to engage in daily or regular sessions and their ability to maintain focus during each stage of the cycle.

2.2. YouTube-based structured learning with shadowing

Designed for intermediate to advanced learners, the YouTube-Based Structured Learning method leverages the authentic, dynamic, and multimodal nature of YouTube videos to develop learners' listening comprehension, oral fluency, and accent adaptability. By combining real-world audiovisual materials with structured pre-task and post-task activities, this intervention aims to bridge the gap between textbook English and spontaneous, informal spoken discourse. Learners engaging in this method are typically seeking to enhance their natural fluency, expand their exposure to diverse English accents, and build confidence in using English for real-life communication.

This method unfolds through a sequenced set of tasks centered on repeated exposure, vocabulary focus, and oral imitation (shadowing). In the first stage, learners watch a selected YouTube video—such as an interview, vlog, or short lecture—three times with subtitles, focusing on understanding the general content. This repeated viewing with textual support allows learners to process authentic input at a manageable pace while building familiarity with key vocabulary, accents, and discourse features. As Yuyun and Simamora emphasize, YouTube videos provide rich, context-embedded input that facilitates both top-down comprehension and bottom-up decoding when accompanied by supportive strategies like repeated viewing and subtitles [32].

In the second stage, learners extract and analyze useful vocabulary and sentence structures from the video. Particular attention is paid to idiomatic expressions, discourse markers, and informal phrases, all of which contribute to more native-like speech production. Contextual learning of new words has been shown to improve vocabulary retention and usage accuracy, especially when learners encounter the same words in varied communicative contexts [33].

The third and central component of the method is shadowing practice, where learners play segments of the video and attempt to repeat the speech immediately after or simultaneously with the speaker. This oral imitation task focuses on pronunciation, stress, rhythm, and intonation. Shadowing not only improves

phonological awareness but also strengthens the learner's auditory-motor connection, which supports the development of fluent, natural speech. AlObaid highlights the value of shadowing combined with narrow listening from authentic YouTube content, noting that it fosters automaticity in oral production and builds learners' capacity to handle varied speech patterns [34].

An optional but beneficial extension is personal recording and self-feedback. Learners record their shadowing attempts and then compare their performance to the original speaker. This metacognitive process enhances self-monitoring, draws attention to areas of difficulty, and encourages learner autonomy.

This method presents several instructional advantages. First, it promotes exposure to a wide range of English accents, including British, American, Australian, and others, thereby enhancing learners' perceptual flexibility and global communicative readiness. Second, it simultaneously develops listening and speaking skills through tightly integrated input and output activities. Third, it increases learners' familiarity with informal, spontaneous speech, including fillers, contractions, and spoken grammar—elements often neglected in formal EFL instruction.

However, there are also challenges associated with this approach. Learners may initially struggle with understanding unfamiliar accents or highly informal discourse styles, especially if they have been primarily exposed to scripted or standardized classroom materials. Furthermore, the method requires teacher guidance or the use of self-regulated learning strategies, such as goal setting, time management, and self-assessment, to be implemented effectively.

3. Implications

3.1. Implications for educators

The proposed framework suggests several important implications for English language educators, particularly those working with Chinese university students. One of the most critical considerations is the need to differentiate instruction based on learner proficiency levels. Given the diverse linguistic backgrounds and skill profiles of students, educators must adopt intervention strategies aligned with students' current capabilities. For instance, lower-proficiency learners benefit more from the structured, scaffolded nature of the Listening–Reading–Repetition–Memorization method, which provides controlled input and repetitive practice. In contrast, intermediate to advanced learners can gain greater fluency and communicative confidence from the YouTube-Based Structured Learning with Shadowing method, which exposes them to authentic, spontaneous speech and diverse accents in real-world contexts.

Effective implementation of these interventions also requires careful task design and gradual progression in difficulty. Teachers should avoid overwhelming students, especially at the initial stages, by introducing manageable tasks that increase in complexity over time. For example, shadowing practice might begin with isolated words or short sentences before advancing to longer, unscripted monologues. Such pacing ensures that learners build confidence and skills incrementally, reducing the risk of frustration or disengagement.

Furthermore, it is essential that these strategies are not treated as supplemental or extracurricular. Rather, they should be integrated systematically into the core curriculum, such as listening courses, speaking workshops, or even content-based instruction. Embedding these multimodal approaches into the daily instructional cycle enhances their impact and signals to learners that practical communication skills are valued alongside grammatical knowledge and test performance.

Equally important is the role of teachers in fostering learner autonomy and self-regulation. Educators should explicitly model techniques such as shadowing, self-recording, and vocabulary logging in the classroom, and then guide students to continue these practices independently. This process not only builds

learner confidence but also cultivates long-term study habits that support ongoing language development beyond the classroom.

Lastly, educators must leverage technological tools to maximize the accessibility and effectiveness of these interventions. Teachers can introduce students to reliable online platforms and train them to use subtitle features, adjustable playback speeds, and audio recording software. By doing so, teachers help students harness digital resources to enrich their language input and create personalized, interactive learning environments.

3.2. Implications for learners

To fully benefit from the multimodal intervention strategies outlined in this paper, learners themselves must take an active and reflective role in their language development. A key first step is conducting self-assessment to identify specific weaknesses in listening and speaking skills.

Successful language development also hinges on establishing a consistent and realistic practice routine. Unlike short-term cramming for tests, building communicative competence requires long-term, incremental effort. Students are encouraged to set daily or weekly goals—for instance, engaging in 30 minutes of structured practice per day—and to integrate these methods into their regular study schedule. This approach not only supports gradual skill development but also fosters discipline and self-management, which are essential for sustained improvement.

Importantly, learners should understand that language improvement is not always immediately visible, nor does it follow a fixed timeline. It is not simply about practicing one or two hours daily but maximizing total exposure and use over time. The more time English is used actively or passively in real-life situations, the more improvement learners will see. Long-term immersion—both formal and informal—remains the most potent contributor to language development.

To this end, learners at all levels should seek to integrate English practice into their everyday lives. Beginning learners might start by watching movies or TV shows with bilingual subtitles, helping them understand content while familiarizing themselves with natural pronunciation and pacing. Intermediate learners can transition to reading English-language magazines, blogs, or simplified novels, gradually building their reading stamina and vocabulary. Advanced learners are encouraged to go a step further by using English as a tool for learning—not just learning the language itself.

A particularly valuable skill to cultivate is awareness of phonological features such as rhythm, stress, and intonation. In shadowing practice, students should go beyond surface-level repetition and actively attend to the prosodic elements of speech. Through repeated imitation, comparison with native models, and deliberate refinement, learners can develop more natural and intelligible pronunciation. Over time, this heightened awareness helps transform passive listening into active phonetic decoding and expressive articulation.

Another important consideration is selecting learning materials based on personal interests and authentic content. The YouTube-based approach emphasizes learner engagement through real-world input. Students should take initiative in choosing videos that align with their interests—such as technology, entertainment, or social issues—as this relevance increases motivation and facilitates deeper cognitive processing. Interest-driven content also mirrors the diversity of real-world communication, preparing learners for interactions beyond the classroom. Furthermore, students can benefit from seeking out real-life opportunities to use English—such as language exchange meetups or online communities—as a complement to these practices, thereby building confidence in genuine communication.

4. Discussion

Beyond their pedagogical value, implementing the proposed intervention strategies carries broader implications for addressing two persistent societal challenges in China. First, they offer a concrete response to a systemic weakness in the Chinese education system: the prioritization of exam performance over communicative competence. Despite years of English instruction, many students graduate with limited ability to engage in real-world communication. Second, these interventions offer a pathway toward greater educational equity, particularly in bridging the gap between students in urban and rural areas or from different economic backgrounds. Furthermore, these strategies align with national initiatives aimed at promoting educational equity and modernizing pedagogy, such as the 'Double First-Class' university plan which emphasizes developing globally competitive talent, and the 'Internet Plus Education' model which advocates for the integration of technology to bridge resource gaps between urban and rural areas.

By equipping teachers and students with structured, flexible, and cost-effective tools for enhancing listening and speaking skills, this research contributes to improving the overall quality of English education in China while promoting fairer access to language learning opportunities. As global communication becomes an increasingly essential competency, the ability to bridge educational divides and reform outdated teaching models becomes not only an academic concern but ultimately a social imperative.

The integration of YouTube-based multimodal learning also offers valuable pedagogical benefits. YouTube provides authentic, visually rich, and contextually grounded input, which can increase student engagement and attention compared to traditional learning materials. The ability to select content based on personal interest further enhances learner motivation, which is a key predictor of long-term language retention and overall proficiency. As learners repeatedly interact with audiovisual input—combining listening, reading, and speaking tasks—they build deeper phonological and lexical representations, improving both comprehension and production. Multimodal input supports cognitive reinforcement through multiple sensory channels, aligning with second language acquisition theories that emphasize the importance of meaningful, repeated exposure in varied formats.

4.1. Limitations and future research

As a conceptual paper, this study has several limitations. First, the proposed strategies require empirical validation to measure their efficacy on learner outcomes. Second, the effectiveness of these methods, particularly the YouTube-based approach, hinges on learner self-regulation and consistency. Third, the variable quality of online content and resources may impact the learning experience, as not all YouTube materials provide accurate subtitles or appropriate language. Future research should involve longitudinal empirical studies to track the sustained impact of these interventions across diverse learner profiles and institutions. For instance, without original data to demonstrate outcomes, any claims of effectiveness remain speculative. Relying on these untested methods in educational practice carries the risk that they may not yield the expected improvements or may require significant adaptations to different learning contexts. It is crucial for future studies to empirically evaluate the interventions – for example, through small-scale pilot implementations or quasi-experimental designs with pre- and post-test measurements of listening comprehension and speaking fluency. Researchers should also collect qualitative feedback from learners and instructors to identify practical challenges and refine the strategies accordingly. Ultimately, longitudinal research is needed not only to validate immediate effects but also to determine whether gains from these multimodal strategies are sustained over time and transferable to various educational settings.

5. Conclusion

This paper has proposed two differentiated, research-informed intervention strategies aimed at improving the international communication abilities of Chinese university students. Grounded in existing research and aligned with multimodal learning principles, these methods address key components of listening comprehension and oral proficiency, offering tailored support for learners at various proficiency levels.

By integrating multimodal learning through YouTube, learners have the opportunity to enhance their communicative competence, and this approach has the potential to contribute to broader social goals, such as improving engagement in global dialogue and fostering culturally competent graduates. Nonetheless, these approaches should be implemented with care, as their effectiveness hinges on factors such as learner motivation, resource quality, and instructor support. In this sense, strengthening students' international communication skills is not merely an educational objective, but potentially a step toward China's greater participation in the global community. As outlined, future empirical work is necessary to validate these strategies through longitudinal studies and to explore their adaptation for different learning environments.

References

- [1] Nickerson, C. (2005). English as a lingua franca in international business contexts. *English for Specific Purposes*, 24(4), 367–380. <https://doi.org/10.1016/j.esp.2005.02.001>
- [2] Seidlhofer, B. (2012). Understanding English as a Lingua Franca. <https://doi.org/10.1111/j.1473-4192.2011.00305.x>
- [3] Michaud, M., & Colpitts, B. D. F. (2015). English as a Lingua Franca: Globalization, ownership, and the diversification of English. *Kwansei Gakuin University Humanities Review*, 20, 125–131.
- [4] Tan, J. (2024). English as a Lingua Franca in Global Business: Balancing Efficiency and Cultural Sensitivity. *Research Studies in English Language Teaching and Learning*. <https://doi.org/10.62583/rseltl.v2i2.42>
- [5] Liao, J. (2022). A Systematic Review of the Backwash Effect of the English Education in China. *Overseas English Testing: Pedagogy and Research*. <https://doi.org/10.12677/OETPR.2022.41001>
- [6] Zhao, H. (2022). The Influence of Test-oriented Teaching on Chinese Students' Long-term use of English. *International Journal of Education and Humanities*. 6, 123-128. <https://doi.org/10.54097/ijeh.v6i2.3658>
- [7] Lou, M. (2021). Classroom Activities in Chinese Secondary EFL Classes: Teachers' Uptake of the Learner-centered Curriculum Reform. *International Journal of TESOL Studies*. Vol. 3 (2) , 168-186. <https://doi.org/10.46451/ijts.2021.07.04>
- [8] Lucenta, L. (2011). China's Higher Education Lacks Higher Learning. *Phi Delta Kappan Magazine*, 93(4), 76–77. <https://doi.org/10.1177/003172171109300421>
- [9] Zhang, X. (2024). Exploring the Improvement of Chinese Junior High School Students' English Capacities in Oral Speaking. *Lecture Notes in Education Psychology and Public Media*. <https://doi.org/10.54254/2753-7048/45/20230234>
- [10] Ren, L. (2023). Ways of Getting Students Practicing English Speaking in Chinese High School. *English Language Teaching and Linguistics Studies*. <https://doi.org/10.22158/eltls.v5n3p66>
- [11] Yang, X., & Duan, Y. (2016). Difficulties of Chinese Students with Their Academic English: Evidence from a China-United States University Program.
- [12] Han, F. (2013). Pronunciation problems of Chinese learners of English. *ORTESOL Journal*, 30, 26-30.<https://www.proquest.com/scholarly-journals/pronunciation-problems-chinese-learners-english/docview/1940921406/se-2>
- [13] Amoah, S., & Yeboah, J. (2021). The speaking difficulties of Chinese EFL learners and their motivation towards speaking the English language. *Journal of Language and Linguistic Studies*, 17(1), 56-69. Doi: [https://doi.org/10.54254/2753-7048/17\(1\)/56-69](https://doi.org/10.54254/2753-7048/17(1)/56-69)

10.52462/jlls.4

- [14] Zhang, W., & Ding, L. (2011). A preliminary study on Chinese EFL learners' attitudes towards their accent.
- [15] Ma, L. (1994). English learning: An analysis of Chinese students' problems in pronunciation.
- [16] Li, Y. (2019). Collaborative Online International Learning: A Pilot Study in Intercultural Communicative Competence in Chinese Higher Education Background. *US-China Education Review A*. <https://doi.org/10.17265/2161-623x/2019.07.004>.
- [17] Liu, W. (2022). A Study on the Integration of Business English Teaching and Intercultural Communication Skills Cultivation Model Based on Intelligent Algorithm. *Security and Communication Networks*. <https://doi.org/10.1155/2022/2873802>.
- [18] Zhou, Y., Xu, S., & Bayley, J. (2011). *Intercultural Competence and EFL Teaching in China*., 138-166.https://doi.org/10.1057/9780230299481_7.
- [19] Zhao, R., Wei, H., & Yao, Z. (2025). The effects of L2 proficiency on L2 word reading strategies: Evidence from Chinese English bilinguals. *Applied Psycholinguistics*. <https://doi.org/10.1017/s0142716424000432>
- [20] Ng, S. (2006). The effects of direct instruction in phonological skills on L2 reading performance of Chinese learners of English.
- [21] Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4
- [22] Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review*, 112(1), 43–59. DOI: 10.1037/0033-295X.112.1.43
- [23] Li M. (2016). Investigation into the differential effects of subtitles (first language, second language, and bilingual) on second language vocabulary acquisition. <http://hdl.handle.net/1842/22013>
- [24] Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PLoS ONE*, 11(6), e0158409. <https://doi.org/10.1371/journal.pone.0158409>
- [25] Matielo, R., D'Ely, R., & Baretta, L. (2015). The effects of interlingual and intralingual subtitles on second language learning/acquisition: A state-of-the-art review. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, 54, 161–182. <https://doi.org/10.1590/0103-18134456147091>
- [26] Potter, C. E., & Saffran, J. (2017). Exposure to multiple accents supports infants' understanding of novel accents. *Cognition*, 166, 67–72. <https://doi.org/10.1016/j.cognition.2017.05.031>
- [27] Bent, T., & Holt, R. (2013). The influence of talker and foreign-accent variability on spoken word identification. *The Journal of the Acoustical Society of America*, 133(3), 1677–1686. <https://doi.org/10.1121/1.4776212>
- [28] Luo, X. (2024). Communicative Language Teaching and Contextual Teaching and Learning: Challenges and Prospects in Chinese High-School English Education. *Science, Technology and Social Development Proceedings Series*.<https://doi.org/10.70088/qdqnck28>
- [29] Zhang, P., & Peng, Y. (2024). Multimodal reading in reading-only versus reading while-listening modes: evidence from Chinese language learners. *Chinese as a Second Language Research*. <https://doi.org/10.1515/caslar-2024-2003>.
- [30] Spring, R., & Takeda, J. (2024). Teaching Phrasal Verbs and Idiomatic Expressions Through Multimodal Flashcards. *STEM Journal*. <https://doi.org/10.16875/stem.2024.25.2.40>.
- [31] Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235–256). Newbury House.
- [32] Yuyun, I., & Simamora, F. (2021). THE USE OF YOUTUBE TO SUPPORT EFL STUDENTS' LISTENING SKILLS. *ELLTER Journal*. <https://doi.org/10.22236/ellter.v2i2.7512>.
- [33] Khairat, N. (2024). The Use of YouTube to Enhance Learners' English Listening Skills. *Surakarta English and Literature Journal*. <https://doi.org/10.52429/selju.v7i2.265>.

[34] Alobaid, A. (2022). ICT Virtual Multimedia Learning Tools/Affordances: The Case of Narrow Listening to YouTube Multimedia-Based Comprehensible Input for the Development of ESL Learners' Oral Fluency. Technologies, *Artificial Intelligence and the Future of Learning Post-COVID-19*. https://doi.org/10.1007/978-3-030-93921-2_28.