Journal of **Linguistics and Communication** Studies

Volume 5, Number 1 **March 2026**

CONTENTS

Journal of Linguistics and
Communication Studies
ISSN 2958-0412
www.paradigmpress.org/JLCS
Volume 5 Number 1 March 2026

Affixal Processes in Ogba and English Languages

Fashion Giobari Zabbey[1] & Chinedum Isaac[2]

[1] Department of English and Literary Studies, Faculty of Humanities, Rivers State University, Nigeria
[2] Department of Linguistics, Faculty of Humanities, Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Nigeria
Correspondence: Fashion Giobari Zabbey, Department of English and Literary Studies, Faculty of Humanities, Rivers State University, Nigeria.

doi:10.63593/JLCS.2026.03.01

Abstract

This paper contrasted the affixal processes in in Ogba and English languages. The aim of the study is to identify the areas of similarities and dissimilarities between the affixal processes in both languages. The study adopted the Contrastive Analysis Theory (CA). The data used in this study were elicited through the use of unstructured oral interview of L1 and L2 users of Ogba and English languages respectively. The data for the study were analyzed through the use of descriptive and contrastive methods of data analysis. The data were presented using the Leipzig glossing pattern. It was observed that prefixes, suffixes and suprafixes exist in both languages. This study also found out that prefixes are segmental phonemes in both languages. However, it was noticed that whereas the prefixes in Ogba have either V or N (where N represents syllabic nasal) syllabic structure, the prefixes in English language have irregular syllabic structures which include: VC, CV, VCCV, CVCV and CVC. This study further observed that it is only the verb that host prefixes in Ogba while nouns, adjectives and verb can host the prefixes in English language. Again, it was affirmed that English language has more suffixes than the Ogba language. Only four suffixes were identified in the Ogba while the English language has several suffixes. Additionally, whereas suffixation is inflectional in Ogba, it is both inflectional and derivational in the English language. It was noticed that the suprafixes in Ogba are tone (variation in the pitch of the voice) and nasalization while the suprafix in the English language is stress. In view of that, this study recommended that English language teachers within Ogba speech communities should focus on other forms of nominalization and stress in the English language.

Keywords: affix, prefix, suffix, supra fix, root word

1. Introduction

Language is a key tool for human interaction or communication. In other words, language is an important instrument in the lives of all humans in a given society. It is observed that every natural or human language behaves like other living organisms in the society. Like other living organisms, every natural language has the ability to either grow or die. This means that no language remains stagnant. In other words, languages change from time to time (Kari, 2015). Also, linguistic researches have shown that natural languages do not permit the stringing or

lacing of morphemes arbitrarily. Every natural language has its system of word structure. There is no language without rules that govern its morphological, syntactic, phonological, semantic and pragmatic operations (Yul-Ifode, 2001). The level of linguistic analysis that examines how sounds combine to form morphemes (the smallest meaningful unit in a word) and words is morphology. Morphology deals with the internal structure of words. Hence, some lexical items in a natural language can be segmented or separated into diverse morphological units or components. That is, a lexical item in a natural language may consist of a root word and an affix. It is important to note that even though affixes are usually bound morphemes in all natural languages, the parameter for their attachment is usually language specific. For that reason, this study contrasts the affixal processes in Ogba and English languages with the aim of identifying the areas of similarities and dissimilarities between the affixal processes in both languages.

2. The Ogba Language and Its Speakers

The nomenclature "Ogba" is commonly used to designate both the language and its speakers. However, the morpheme "ndé" 'people of' is sometimes used to differentiate the language and its speakers. For instance, one may say ndé Ogba 'people of Ogba'. The language is linguistically classified or grouped as an Igboid (Igbo related) language. It belongs to the Benue-Congo family of the Niger-Congo phylum. Nwokoji and Isaac (2024) affirm that Ogba is spoken in over forty-one communities in Ogba/Egbema/Ndoni Local Government Area of Rivers State, Nigeria. The language is closely related to Ekpeye, Igbo, Ikwere and Echie which are spoken in different local government areas of Rivers State, Nigeria.

3. Theoretical Framework

This study was anchored on the Contrastive Analysis Theory (CA). According to Melefa (2015) and Nwala (2015), the proponent of the CA theory is Robert Lado in 1957. The theory is based on the notion that all natural languages have their areas of relatedness and variations. Thus, the theory assists in the identification of resemblances and differences between two or more languages. Richard (1992) and Yang (1992), the CA theory can be applied in the different levels of grammatical study such as phonetics and phonology, morphology, semantics and syntax. Also, they admit that the CA is relevant because it enhances language teaching and translation. Richard (1992) and Yang (1992) further note that the CA identifies the areas of difficulties for language learners that could be handled with correct exercise. Thus, this paper posits that the CA theory can support the researchers in the documentation of the similarities and discrepancies between the affixal processes in the Ogba and the English languages.

4. Review of Related Literature

Isaac (2018), and Eze and Isaac (2023) unanimously affirmed that the affixal processes that result in nominalization Ogba include: prefixation, suffixation, and interfixation alone, or in combination with other nominal derivation processes. It is deduced from the assertions of Isaac (2018), and Eze and Isaac (2023) that the term affixation has to do with the process of attaching an affix to a word or root either for inflectional or derivational purposes. The term "affix" refers to an element in the morphological structure of a word which is not part of the root or stem (Mathews, 2007). According to Kari (2015), an affix is a sub-class of morpheme that is added to a root. He noted that the relationship between the root or stem of a word and affixes can be compared to the structure of a tree. Like a tree, a word can consist of different components. Kari (2015) also admitted that in morphology, the term "root" has to do with the part of a lexical item to which other word-building elements (affixes) can be added. Trask (1993) and Katamba (1993) unanimously describe an affix as a bound morpheme which can only be attached to a root or stem. On the other hand, Katamba says that a root is the irreducible core of a word with absolutely nothing attached to it. Onumajuru (2023) asserted that affixes in human languages can perform inflectional and derivational functions. She stated that a derivational affix is an affix which serves to derive new words. She also noticed that derivational affixes sometimes change the word class of the base to which it is attached. She further observed that when an affix changes the word class of the base, it is called class changing derivational affix and when it retains the word class and does not change it, it is called class maintaining derivational affix. This study affirms through literature reviewed that affixation may be inflectional or derivational morphological process which involves the attachment of an affix to the base of a word.

5. Methodology

The data used in this study were elicited through

the use of unstructured oral interviews. The unstructured oral interview was used in obtaining in-depth data through face-to-face verbal communication. The use of oral interviews in particular yielded a lot of information, which enriched our study. The data were collected from two sources which include the primary data and secondary data. The primary data were drawn from proficient L1 and L2 users of Ogba and English languages respectively, while the secondary data were drawn from related literatures, especially Isaac (2018). The method of data analysis adopted in this study is the descriptive and contrastive methods. The data were presented using Leipzig glossing pattern.

6. Discussion and Result

Affixation is the process of attaching a bound morpheme (affix) to a word. This implies that affixes in Ogba and English languages mandatorily have a host which is usually a word (root or stem). There are four different criteria commonly employed in the classification of affixes in both languages. The criteria are position, function, meaning, and productivity. However, for the sake of precision, the present study presents affixes in both language based on position. This study observed that on the basis of their position in relation to the root words, affixes in Ogba and English languages are prefixes, suffixes and superfixes.

6.1 Prefixation in Ogba and English Languages

Prefixation is the process of attaching a prefix to a host (root word). The term prefix refers to an affix, a bound morpheme which is always attached at the beginning of a root word that is before the base or host. The data gathered for this study proved that prefixes in Ogba and English languages are unit of syllables. In Ogba, prefixes are generally tone bearing units. It is also observed that all the vowels and syllabic consonants in Ogba can morphologically function as prefixes:

Table 1. Examples of Prefixation in Ogba language

S/N	Prefixes	Roots	Gloss	Derived Word	Gloss
1a.	á-	ríò	'beg'	áríò	'(act of) begging'
b.	ẹ́-	rnà	'suck'	èẹ́rnà	'breast'
c.	é-	mú	'laugh (v)'	émú	'laughter (n)'
d.	í-	wè	'be angry'	íwè	'anger'
e.	ị́-	shí	'smell (v)'	ị́shí	'smell (n)'
f.	ó-	rnú	'work (v)'	órnú	'work (n)'
g.	ò-	hnụ̀	'see'	òhnụ̀	'vision'
h.	ú-	kó	'scarce'	úkó	'scarcity'
i.	ụ́-	hú	'grow'	ụ́hú	'growth'
j.	m-	má	'be beautiful'	mmá	'beauty'
k.	n-	vná	'scold (v)'	nvná	'scolding'

The examples in the Table 1 indicate that the syllabic structures of the prefixes in Ogba are V or N, where V is a vowel and N a syllabic nasal (semi-vowel). It is also noticed that all the host (root or stem) of the prefixes in examples 1a-k are verbs while the derived forms are nominals (nouns). This denotes that in Ogba, verbs generally host prefixes. In other words, verbs are the only grammatical category or word class that host prefixes in Ogba. Simply put, prefixes cannot be attached to other words in the language. It is further observed that prefixes in the language do not have any semantic quality in isolation. That is, without the verb which serves as the host, prefixes in Ogba are grammatically considered as segmental phonemes. The structural configuration of prefixation in Ogba is Pref. + V (where pref. is a prefix which is either a vowel or a syllabic nasal and V a verb root) which is one of the most productive morphological operations in the language.

Table 2. Examples of Prefixation in English language

S/N	Prefixes	Roots	Derived Word
2a.	ac-	knowledge	acknowledge
b.	anti-	malarial	antimalarial
c.	co-	occur	co-occur
d.	de-	merit	demerit
e.	dis-	play	display
f.	ex-	militant	ex-militant
g.	ir-	regular	irregular
h.	im-	moral	immoral
i.	mal-	practice	malpractice
j.	mono-	lingual	monolingual
k.	re-	view	review

The examples in the Table 2 reveal that the prefixes in English have irregular syllabic structures which include: VC, CV, VCCV, CVCV and CVC. It is also noticed that the hosts (roots or stems) of the prefixes in examples 2a-k belong to different word classes. For instance, the roots of 2a, 2b, 2d, 2f and 2j are nouns; 2c, 2e, 2i and 2k are verbs while 2g is an adjective. This implies that prefixes can be attached to nouns, verbs and adjectives in English language. With the exception of the negative prefix "un-" which can sometime have an adverb as its host as in un-fairly = unfairly, this study did not find any example of a prefix that can be attached to an adverb. It is further of observed that the words that are derived through prefixation in the language belong to different classes. In 2a, it is noticed that the noun knowledge changes to a verb (acknowledge) while the verb play remains a verb as in display in 2e. This proves that prefixation in English language is both class-maintaining and class-change morphological operation in the language. The data on Table 2 indicate that some of the prefixes in English language have inherent meanings (semantic qualities). For example, the prefixes anti- as in 2b, ex- as in 2f and mono- as in 2j have the meanings: against, former, and one respectively.

6.2 Suffixation in Ogba and English Languages

The process of attaching a suffix to a host is known as suffixation. A suffix is an affix that is attached at the end of a root word or host. Suffixation in Ogba and English languages may result in either derivation or inflection. It is observed that there are few suffixes in Ogba.

Table 3. Examples of Suffixation in Ogba language

S/N	Roots	Gloss	Suffixes	Derived Form	Gloss
3a.	zú̩	'train'	-má	zú̩má	'trained'
b.	kpó̩-	'call'	-mé/-mè̩	kpó̩mè	'called'
c.	biá	'come'	-lí	biá-lí	'come + ability'
d.	gbá	'write'	-shí̩	gbá-shí̩	'write + into'

The examples in 3a-d on Table 3 show that the suffixes in Ogba have the CV syllabic structure. This suggests that suffixes in the language comprise a consonant and a vowel which constitutes segmental phoneme. Like the prefixes, the suffixes in Ogba are tone bearing units. The examples also prove that suffixat+ion is inflectional and derivational in the language. For instance, the attached of the suffix "-má" to the verb root "zú̩" 'train' results in the perfective form "zú̩má" 'trained'. Similarly, the suffix "-lí" is commonly attached to a lexical verb to mark ability. Consider the following constructions:

4a. Wó̩ biá áhiá.

3PL come-PST market.
'They came to the market.'

b. Wọ́ biá-mè áhiá.
3PL come-PERF. market.
'They have come to the market.'

c. Wọ́ ká biá-lí áhiá.
3PL NEG. come-ABIL market.
'They could not come to the market.'

d. Wọ́ biá-shí yè íjè.
3PL come-APPL 1PL visit
'They visited us.'

The examples in sentences 4a-d show that the verb root (biá 'come') which originally expresses simple past can take different suffixes. They suffixes perform divergent grammatical roles. The suffix "-mè" in 4b changes the construction in 4a to a past perfect construction. The suffix "-lí" expresses ability as in 4c while the suffix "-shí" in 4d is an applicative marker (a morpheme which allows a verb to take additional argument or object). In view of the foregoing, this paper posits that suffixation is inflectional in Ogba. The examples 5a-g on Table 4 illustrate suffixation in English language.

Table 4. Examples of Suffixation in English language

S/N	Roots	Suffixes	Derived Word
5a.	boy	-s	boys
b.	baby	-ies	babies
c.	assign	-ment	assignment
d.	good	-ness	goodness
e.	supervise	-ee	supervisee
f.	kill	-er	killer
g.	Act	-ion	action

The examples in 5a-g reveal that suffixation is both inflectional and derivational morphological operation in English language. For instance, the attachment of the suffixes -s and -ies to the roots boy and baby which are singular in 5a and 5b respectively result in their plural forms. On the contrary, the attachment of the suffixes -ment, -ness, -ee, -er and ion to the roots assign, supervise, kill and act result in the realization of the nouns assignment, goodness, supervisee, killer and action respectively. These examples also show that suffixation is both a class-maintaining and class-changing morphological operation in the English language. Whereas the derived words in 5a-b maintain the same class with their roots, the derived words in 5c and 5e-g changed from verb to nouns while 5d changes from adjective to noun. It is important that suffixation is one of the productive inflectional and derivational morphological operation in the language. There are several suffixes in the language.

6.3 Suprafixation in Ogba and English Languages

Suprafixation refers to the morphological operation which involves the use of a suprasegmental feature to create difference in the meaning of a lexical item or grammatical construction. The suprasegmental features in natural languages are stress, tone, nasalization, intonation, etc. It is noticed that tone and nasalization are the suprafixes in Ogba. Tone and nasalization are as essential as the phonemic sounds (consonants and vowels) in the language. This is because like the segmental phonemes (consonants and vowels), they (tone and nasalization) can alter the meaning of words, phrases, clauses and sentences which otherwise are the same in terms of their segmentals. Consider the examples on Table 5:

Table 5. Suprafixation in Ogba language

S/N	Word	Gloss	Suprafixes
6a.	égbé	'hawk'	HHT
	ègbè	'trap'	LLT
	ègbé	'gun'	LHT
b.	ēkwé	'drum'	MHT
	èkwè	'yam barn'	LLT
c.	ché	'wait'	HT
	chè	'think'	LT
d.	rá	'drink'	
	rná	'suck'	Nasalization
e.	tá	'chew'	
	tná	'entice'	Nasalization
f.	tù	'measure'	
	tnù	'select'	Nasalization

It is noticed that the dissimilarities between the words in examples 6a-f is the variation in the pitch of the voice (tone) and nasalization. For instance, a change from high tone to low tone changes the meaning of the word "ché" 'wait' to 'think' as in 6c while the nasalization of the vowel in 6d-f result in change in meaning. Nasalization is orthographically represented in Ogba through the insertion of the letter "n" in between the preceding sound and the nasalized vowel.

Table 6. Suprafixation in English language

S/N	Word	Suprafix
7a.	REcord	Stress
	reCORD	
b.	EXport	Stress
	exPORT	
c.	SUBject	Stress
	subJECT	
d.	INsult	Stress
	inSULT	
e.	PROject	Stress
	ProJECT	

The examples in 7a-e on Table 6 show that the suprafix in the English language is stress. It is observed that the differences between the words in each of the examples are the shift in stress which results in difference in meaning and the class of the words. It is also noticed that when the stress is placed on the first syllable of the bi-syllabic words in 7a-e, they function as nouns but when placed on the second syllable, they function as verbs. This indicates that a shifting stress is one of the affixal processes in the language.

7. The Similarities and Differences Between the Affixal Processes in Ogba and English Languages

This study has been able to ascertain that the affixal processes in the Ogba and English languages have some similarities and differences. This corroborates the Contrastive Analysis Theory (CA) by Robert Lado that natural languages have their areas of similarities and differences. In terms of similarities, it was observed that prefixes, suffixes and suprafixes exist in both languages. This study also found out that prefixes are segmental phonemes in both languages. However, it was noticed that whereas the prefixes in Ogba have either V or N (where N represents syllabic nasal) syllabic structure, the prefixes in English language have irregular syllabic structures which include: VC, CV, VCCV, CVCV and CVC. This study further observed that it is only the verb that host prefixes in Ogba while nouns, adjectives and verb can host the prefixes in English language. Again, it was affirmed that the words that are derived through prefixation in Ogba are usually the nominals (nouns). In other words, prefixation strictly result in nominalization in the language but the words that are derived through prefixation in the English language belong to different classes. This indicates that that whereas prefixation is firmly a class-changing morphological operation in the Ogba language, it is both class-changing and class-maintaining morphological operation in the English language.

This study revealed that English language has more suffixes than the Ogba language. Only four suffixes were identified in the Ogba while the English language has several suffixes. The syllabic structure of all the suffixes in Ogba is CV whereas the suffixes English language has irregular syllabic structure such as CVC, CVCC, etc. Additionally, whereas suffixation is inflectional in Ogba, it is both inflectional and derivational in the English language.

It is noticed that the suprafixes in Ogba are tone (variation in the pitch of the voice) and nasalization while the suprafix in the English language is stress. In view of that, this study posits that tones are like the segmental phonemes (consonants and vowels) in Ogba. That is, tones can alter the meaning of words, phrases, clauses and sentences which otherwise are the same in terms of their segmentals. This is because Ogba is a tone language. On the other hand, English is a stress sensitive language that is wrong placement of stress can result in misinterpretation.

8. Conclusion

This paper contrasted the affixal processes in Ogba and English languages with the aim of identifying the areas of similarities and dissimilarities between the affixal processes in both languages. The study was anchored on the Contrastive Analysis Theory (CA). The data used in this study were elicited through the use of unstructured oral interview. The data were collect from two sources which include the primary data and secondary data. The primary data were drawn from proficient L1 and L2 users

of Ogba and English languages respectively, while the secondary data were drawn from related literatures. The method of data analysis adopted in this study was the descriptive and contrastive methods. The data were presented using Leipzig glossing pattern.

This study corroborated the Contrastive Analysis Theory (CA) which posits that natural languages have their areas of similarities and differences. In terms of similarities, it was observed that prefixes, suffixes and suprafixes exist in both languages. This study also found out that prefixes are segmental phonemes in both languages. However, it was noticed that whereas the prefixes in Ogba have either V or N (where N represents syllabic nasal) syllabic structure, the prefixes in English language have irregular syllabic structures which include: VC, CV, VCCV, CVCV and CVC. This study further observed that it is only the verb that host prefixes in Ogba while nouns, adjectives and verb can host the prefixes in English language. Again, it was affirmed that the words that are derived through prefixation in Ogba are usually the nominals (nouns). This study also revealed that English language has more suffixes than the Ogba language. Only four suffixes were identified in the Ogba while the English language has several suffixes. The syllabic structure of all the suffixes in Ogba is CV whereas the suffixes English language has irregular syllabic structure such as CVC, CVCC, etc. Additionally, whereas suffixation is inflectional in Ogba, it is both inflectional and derivational in the English language. It was noticed that the suprafixes in Ogba are tone (variation in the pitch of the voice) and nasalization while the suprafix in the English language is stress. In view of that, this study posits that tones are like the segmental phonemes (consonants and vowels) in Ogba. That is, tones can alter the meaning of words, phrases, clauses and sentences which otherwise are the same in terms of their segmentals. This is because Ogba is a tone language. On the other hand, English is a stress sensitive language that is wrong placement of stress can result in misinterpretation.

9. Recommendation

Based on the assumption of the Contrastive Analysis Theory (CA) that the areas of dissimilarities between two or more languages can present challenges for the L2 learner, this paper recommends that English language teachers within Ogba speech communities should focus on other forms of nominalization and stress in the English language.

References

Eze, A. and Isaac, C. (2023). Nominalization processes in Ogba. In *Zien Journal of Social Sciences and Humanities*. Online.

Isaac, C. (2018). An analysis of Ogba noun phrase. An unpublished M.A. dissertation. University of Port Harcourt.

Kari, E. E. (2015). *Morphology: An introduction to the study of word structure.* University of Port Harcourt Press Ltd.

Katamba, F. (1993). *Morphology*. Macmillan Press.

Mathews, P. H. (2007). *Oxford concise dictionary of linguistics*. Oxford University Press.

Melefa, M. O. (2015). Contrastive analysis of tense and aspect in English and Abinu. In O.M. Ndimele (ed.). *Language endangerment: Globalisation & the fate of minority languages in Nigeria: A festschrift for Appolonia Uzoaku Okwudishu*. M & J Grand Orbit Communications Ltd.

Nwala, M. A. (2015*). Introduction to linguistics: A first course*. Obisco Nig. Enterprises.

Nwokoji, C. and Isaac, C. (2024). Personal pronouns in Ogba and English languages. *NIU Journal of Humanities-Nexus International University*, 257-264.

Onumajuru, V. C. (2023). Morphology. In Christie U. Omego (ed.), *Linguistics, language and the media*. Gabby Sambros Enterprises.

Richard. J. C. (1992). *Longman Dictionary of Language Teaching and Applied Linguistics.* Longman.

Trask, R. L. (1993). *A dictionary of grammatical terms in linguistic*s. Routledge.

Yang, V. J. (1992). *Linguistics and second language acquisition*. Macmillan.

Yul-Ifode, S. (2001). *An introduction to language in history and society*. University of Port Harcourt Press.

Journal of Linguistics and Communication Studies
ISSN 2958-0412
www.paradigmpress.org/JLCS
Volume 5 Number 1 March 2026

The Interactional Function of "*Hǎo Lei*" in the Sequence Organization of WeChat Interactions

Qingjiao Fu[1]

[1] College of Foreign Languages, Ocean University of China, China
Correspondence: Qingjiao Fu, College of Foreign Languages, Ocean University of China, China.

doi:10.63593/JLCS.2026.03.02

Abstract

As a common conversational practice in Mandarin Chinese, "*hǎo lei*" is often employed to express the speaker's acceptance or agreement with its prior turn. Based on data of WeChat interactions and using Conversation Analysis as the research methodology, this study analyzes the sequence organization and interactional function of "*hǎo lei*" in WeChat interaction. Observation reveals that "*hǎo lei*" typically occurs in the responding position and the sequence-closing position, which express the speaker's positive attitude or commitment to the action implemented by the previous turn. "*Hǎo lei*" performs a range of interactional functions, depending on different sequential situations and contexts, including expressing acceptance, giving receipt tokens, and granting the request.

Keywords: "*hǎo lei*", WeChat interaction, interactional function, sequence organization, Conversation Analysis

1. Introduction

With the widespread use of smartphones and high-speed internet connections, applications that afford synchronous text-mediated online interaction, such as WeChat and WhatsApp, have become increasingly prevalent. As a result, synchronous text-mediated online interaction has emerged as a prominent topic in the fields of Conversation Analysis (CA), sociolinguistics, and other social interaction studies. The rising use of CA for online data analysis provides researchers with new insights into the dynamics of online interactions (Meredith, 2019) and supports the analysis of multimodal practices in synchronous text-mediated online interactions, particularly in WeChat.

"*Hǎo lei*", a commonly used linguistic practice in Chinese interactions, is widely examined in sociological and linguistic studies. Although previous research has mainly focused on the discourse and pragmatic functions of "*hǎo lei*," its interactional role in synchronous text-based online interactions remains primarily understudied. A few studies have briefly addressed its discourse function; for instance, Guo (2000) compared its function as a positive response in Chinese and Korean imperative sentences, while Gong (2018) explored its active answering function in online interactions. However, the interactional function of "*hǎo lei*" is largely unexplored.

Drawing on data from naturally occurring WeChat interactions and applying CA as the research methodology, this thesis seeks to develop a deeper understanding of the sequence

organization and interactional function of "*hǎo lei*" in online contexts. The research aims to address the following questions:

(1) What are the characteristics of the position and composition of "*hǎo lei*" in WeChat interactions?

(2) What are the interactional functions of "*hǎo lei*" in WeChat interactions?

2. Research Methodology and Data Collection

Emerging from the ethnomethodological tradition in sociology, CA is an action-oriented approach dedicated to describing human behavior through the meticulous observation of everyday interactional practices (Drew, 2013). By the late 1960s, CA had been established as an independent field focused on understanding the organization of interaction. It was initially applied to analyze technologically-mediated interactions through pioneering research on telephone calls by Schegloff and Sacks. In the early 2000s, this approach was extended to the analysis of synchronous text-mediated online interactions (Giles *et al.*, 2015). Online interactions, which are neither random nor unstructured, are sequentially organized and are considered quintessential forms of naturally occurring data (Meredith, 2019).

Guided by CA, this study employs micro-analytic qualitative methods to analyze the underlying mechanics of "*hǎo lei*" in WeChat interactions, aiming to elucidate its general properties in specific contexts. Therefore, this study will annotate and catalog the characteristics of "*hǎo lei*", including its composition and position in WeChat interactions. Furthermore, by conducting a case-by-case examination of each instance at its specific moment of production, the study will explore the relationship between the particularities of turn design and their sequential positions.

The dataset for this research comprises 200 screenshots, representing 207 instances of "*hǎo lei*" occurring naturally within WeChat interactions among classmates, friends, family, and colleagues communicating in Mandarin Chinese. Utilizing this self-built corpus, the study will qualitatively analyze the 207 instances to identify the generic, context-independent properties of its sequence organization and interactional function. To protect the providers' privacy, names, phone numbers, and other personal information in examples have been carefully altered.

3. The Characteristics of "*Hǎo Lei*" in WeChat Interactions

Humans form and maintain social relationships through interactions with others. They tend to cooperate and support each other to build harmonious relationships. In the context of constructing and maintaining social relationships, "*hǎo lei*" can express agreement with the previous turn, but at times it may also express token acceptance or disagreement. The study of the normative relationships and practices in conversation is a key objective of conversation analysis. Observation reveals that the interactional functions of "*hǎo lei*" is highly dependent on its position within the conversation sequence.

3.1 Turn Design Characteristics

In Mandarin Chinese, "*hǎo lei*" is a combination of the functional character "*hǎo*" and the modal particle "*lei*". The functional character "*hǎo*" literally means "*okay*" and is often used as a responding token to express agreement, recognition, evaluation, and suggestion in interaction (Shao & Zhu, 2005). The addition of the modal particle "*lei*" imbues "*hǎo lei*" with the speaker's emotion and attitude, differentiating it from the simple functional character "*hǎo*".

According to the concept of affordances, the particular technological platform WeChat, may "have an impact on how a user interacts with it" (Meredith, 2019: 243). Compared with face-to-face interaction, online interaction could not provide us with the speaker's intonation and emotion, but the analysis of the punctuations, emojis, and other transactional practices employed in synchronous text-mediated online interaction, allowing for an in-depth understanding of the WeChat interaction.

3.1.1 Configuring with Punctuations and Modal Particles to Express the Speaker's Positive Attitude or Commitment

In WeChat interactions, the contextual arrangement of "*hǎo lei*" alongside various conversational practices can result in different interactional functions. Analysis of the data indicates that "*hǎo lei*" frequently appears with punctuations, emojis, and modal particles. These elements enhance interactional success by expressing the speaker's positive attitude or commitment to the action suggested by the preceding turn. In WeChat interactions, punctuations and emojis effectively substitute for the intonation and facial expressions present in

face-to-face interactions. These multimodal practices are crucial for conveying emotion and tone, features uniquely adapted to online communication scenarios.

Out of the 207 instances in the data, six include tildes (~) and ten include exclamation marks (!), which convey an intimate stance toward the recipient and express willingness and strong commitment to the action anticipated by the previous speaker. In Excerpt 1, speaker Li uses a tilde after "*hǎo lei*" rather than a comma or another punctuation mark to create a less formal and more relaxed tone. The tilde adds a lighthearted, lively element to the conversation, contributing to a positive and cheerful atmosphere that the recipient perceives as comforting and agreeable.

Excerpt 1:

01 孙：（（文件）2023 孙晓晓.docx）

02 班长，这是俺的国际法论文

03 [玫瑰]

04 李：好嘞~

Emojis or kaomojis are also frequently employed after "*hǎo lei*" to enhance the expressiveness and maintain a relaxed interaction between speakers. Common emojis like "rose", "grin", and "heart" effectively convey the speaker's emotional stance, extending presence within the interaction. Emojis thereby help recipients discern stances and attitudes, enhance understanding, and occasionally indicate a reluctance to continue the interaction. The quantitative analysis of the data shows 17 instances of "*hǎo lei*" accompanied by modal particles, such as "*óhǒu* (*哦吼*)", "*áoáo* (*嗷嗷*)", "*hahaha* (*哈哈哈*)", "*òò* (*哦哦*)", "*ngng* (*嗯嗯*)", and "*āihei* (*哎嘿*)".

Excerpt 2:

01 晓：嗨~咱这节课论文有啥要求嘛 (是中文 5000 英文 3000 还是必须用英文写 5000 来着) [可怜]

02 丽：我记得是最好用英文 3000 词来着，中文多少字我没记住[囧]不知道记得准不准[捂脸]

03 晓：哈哈哈好嘞(就是说可以用中文的写是吧[让我看看]

04 丽：应该是吧哈哈我有些不确定

05 晓：哈哈哈啊哈好嘞好嘞

06 丽：嗯嗯

07 晓：[谢谢]

In Excerpt 2, "*hǎo lei*" is used twice with the modal particle "*hahaha*". Xiao initiates the exchange with a greeting, followed by a question and two possible answers in brackets, simplifying Li's task to choose one. However, Li expresses uncertainty and uses two emojis to mitigate the awkwardness, while Xiao subsequently adjusts her question, seeking less detailed information. Li still fails to give a definitive answer, leading Xiao to use "*hahaha*" before "*hǎo lei*". This strategy alleviates any discomfort and signals Li that Xiao intends to seek additional information elsewhere, which aligns with Li's expectation.

3.1.2 Configuring with Supplementary Information to Facilitate Successful Interactions

In WeChat interactions, "*hǎo lei*" is typically accompanied by supplementary elements such as expressions of gratitude, information receipt signs, and additional accounts, all of which contribute to reinforcing agreement and communication clarity. Among the 89 instances where "*hǎo lei*" is used with supplementary messages, three distinct contexts are identified.

a. Followed by Expressions of Gratitude

"*Hǎo lei*" frequently serves as a receipt token in the sequence-closing position of question-answer exchanges. When used this way, it is often accompanied by practices like emojis, gratitude expressions, or additional explanations to convey enhanced appreciation. In these interactions, the questioner initiates the sequence with a query, the responder provides an answer, and "*hǎo lei*" followed by gratitude completes the sequence. An illustrative example is shown below, where "*hǎo lei*" in line 03 acknowledges the answer, subsequently followed by gratitude.

Excerpt 3:

01 威：内个，你知道校医院什么时候开门吗，我眼睛有点不舒服想拿眼药水

02 浩：这个我不太清楚呀，但我室友说那边应该一直有值班的

03 威：好嘞谢谢

According to Levinson's concept of conversational economy, information exchange inherently involves a sense of indebtedness (Li, F. & Li, Z., 2022). This social expectation explains the frequent use of gratitude expressions following "*hǎo lei*".

b. Used with Information Receipt Signs

"*Hǎo lei*" commonly precedes information receipt tokens, such as "*shōu dào* (*收到*)", in WeChat interactions. In this Excerpt 4, Ping and Qian are

members of the same group for a class presentation, with Ping responsible for collecting presentation materials. After reminding Qian to revise their contribution in the first two lines, Qian responds by sending the corrected document. Upon receipt, Ping uses "*hǎo lei*" followed by "*shōu dào*" to acknowledge receipt, reinforcing the acknowledgment function of "*hǎo lei*".

Excerpt 4:

01 萍：@倩

02 uu 你的部分记得更正一下哈

03 倩：（（文件）文本分析.docx）

04 萍：好嘞收到

05 辛苦[玫瑰]

06 倩：[玫瑰]

c. Followed by Accounts to Prior Questions

When "*hǎo lei*" is used in the sequence-closing position, it often serves to acknowledge a prior answer while providing an explanation for the initial query. In Excerpt 5, Zhang poses a question in lines 01-02. Following Liu's response, Zhang replies with "*hǎo lei*" and offers an explanation in line 05, clarifying the rationale behind her inquiry. The conversation suggests that Zhang was inquiring about an order, possibly for flowers, thus engaging with "*hǎo lei*" as a transition between receiving an answer and expressing her reasoning.

Excerpt 5:

01 张：什么时候会发货呀？

02 今天能嘛

03 刘：嗯 都是今天发货

04 张：好嘞

05 坐等

3.2 The Position of "Hǎo Lei" in WeChat Interactions

Based on the fact that conversation is organized sequentially, CA posits that the positioning of an utterance within ongoing interaction is crucial for understanding its meaning and significance as an action (Sidnell & Stivers, 2013). The foundational sequence structure of conversation is the adjacency pair (Sacks *et al.*, 1974), consisting of the initiating turn and the responding turn, referred to as the first pair part (FPP) and the second pair part (SPP) (Schegloff, 2007; Yu, 2022). Following the adjacency pair, a post-expansion may occur, which can be minimal or non-expansive. The minimal post-expansion typically adds a turn after the adjacency pair to conclude the sequence, known as the sequence-closing third (Schegloff, 2007).

Adjacency pairs are regarded as social norms in conversation (Yu & Guo, 2020). Once the interaction context is set, the FPP and SPP establish themselves, illustrating the conditional relevance of the sequence (Wu, 2022). In this study, "*hǎo lei*" is identified as occurring either in a responding position, addressing the preceding turn, or in the sequence-closing position, providing a response to a turn where the recipient has been addressing the prior participant's turn.

Specially speaking, the practice "*hǎo lei*" may appear in the responding position as a reply to an act of informing, reminding, offering, inviting, encouraging, suggesting, or requesting. There is a total of 88 instances of "*hǎo lei*" occurring in the responding position, which constitutes nearly half of the instances. Additionally, "*hǎo lei*" also frequently appears in the sequence-third or sequence-closing position within question-answer sequences, request-acceptance/refusal sequences, suggestion-accordance/discordance sequences, and reminder-response sequences. There is a total of 119 instances of "*hǎo lei*" in the sequence-closing position, representing more than half of the observed occurrences.

4. The Interactional Functions of "*Hǎo Lei*" in WeChat Interactions

In online interactions, the expression "*hǎo lei*" is often used to signal agreement with a preceding turn, analogous to the function of the character "*hǎo*". However, its interactional function is affected by the addition of the modal particle "*lei*". To fully comprehend the function of "*hǎo lei*", it is necessary to consider the situational context and contextual factors. This section explores the interactional functions of "*hǎo lei*" across various contexts.

4.1 Expressing Acceptance

As a common conversational practice in online interactions, the primary interactional function of "*hǎo lei*" is to express the speaker's alignment with the stance conveyed by another participant, akin to the active affirmation function of "*hǎo*" (Shao & Zhu, 2005). As a direct acceptance of the preceding turn, "*hǎo lei*" is often employed to indicate concordance between the action suggested by the previous speaker and the expectations of the recipient. This acceptance may manifest in the responding position in

response to a suggestion, proposal, invitation, offer, or encouragement.

Firstly, when responding to a suggestion or proposal, "*hǎo lei*" may be used to accept and express agreement with the initial participant. In Excerpt 6, colleagues Ping and Qing are exchanging contact details. Ping requests details from Qing in line 01, and Qing complies, offering a suggestion for efficiency in line 05. Ping's "*hǎo lei*" signifies acceptance and closes the sequence.

Excerpt 6:

01 萍：晴姐，名扬王蕾的微信或者 qq 发我下吧，我加上她

02 晴：好

03 ((个人名片王蕾))

04 QQ 123456789

05 都加一下吧，跟他们业务往来频繁，QQ 传文件方便

06 萍：好嘞

Secondly, "*hǎo lei*" can also express acceptance of invitations or offers, embodying agreement between the recipient's expectations and the offer made by another participant, which can be illustrated in Excerpt 7. Initially, Yao requests to borrow Mei's book, which is agreed upon. Then, Yao gives an account for the request and expresses gratitude to Mei, which incurs Mei's invitation to sit together with Yao while sharing the book. The offer demonstrates that while granting Yao's request, Mei has understood the presupposition and would like to sit next to Yao. What's more, it could be predicted that if Mei doesn't offer the invitation, Yao would make another request in the following interaction to achieve her final goal. Finally, Yao accepts this offer with "*hǎo lei*", ensuring alignment with her initial request motives.

Excerpt 7:

01 瑶：姐妹，你买了实用英语写作的书吗

02 美：买了

03 瑶：我这周能先和你看一本嘛

04 美：可以的

05 瑶：我想着买电子版，结果没有[苦涩][苦涩][苦涩]我刚下单实体书

06 [谢谢]

07 美：嗷嗷，那今天我们坐一起吧

08 瑶：好嘞，我去找你

09 谢谢!!!!

10 美：不客气！

11 瑶：[爱心]

Thirdly, acknowledging encouragement is crucial, and "*hǎo lei*" is a common way to express appreciation and commitment to meeting expectations. This response conveys that the recipient appreciates the encouragement and will do their best to meet the expectations that come with it, making the encourager feel valued and supported. In this exchange, Wang and Liu are colleagues. When Liu receives encouragement from Wang, he replies with "*hǎo lei*", emphasizing readiness to meet expectations, bolstered by a "strong" emoji.

Excerpt 8:

01 王：刘老师好，因为明天调试人比较多，创新团队尽量 10:40 以后过来哈。我让突出贡献那边尽量 10:00-10:40 完成调试哈

02 刘：好的，谢谢[玫瑰]

03 王：客气啦，明天加油

04 刘：好嘞![强壮]PPT 明天 9:00 前发给你哈[抱拳][爱心]

05 王：好哒

4.2 Giving Receipt Tokens

Another significant function of "*hǎo lei*" is to signal receipt of information or assistance, akin to the function of "*okay*" (Beach, 2020). When used as a receipt token, "*hǎo lei*" serves as an informing reply. In simpler terms, when "*hǎo lei*" is used in a responding position, it signals that the recipient has received and acknowledged the information or reminder from the speaker.

Firstly, "*hǎo lei*" serves as a response token when acknowledging received information. In such cases, the speaker provides the recipient with important information that they need to know or collect. The recipient then uses "*hǎo lei*" to acknowledge that they have received the information from the speaker and that they have understood the information and that they are ready to take any necessary action.

Excerpt 9:

01 生：老师我做完核酸了

02 师：好嘞

This interaction involves a situation where a student is required to inform the teacher once his covid test has been taken. The exchange only consists of two lines. In the first line, the student informs the teacher about the covid test result, which the teacher is obligated to know. Then, the

teacher responds with "*hǎo lei*" to acknowledge receipt of the information.

Secondly, "*hǎo lei*" can also occur in a remind-reply sequence-closing position. In this sequence, the first part is a reminder, and the second part is a response to the reminder. "*Hǎo lei*" not only serves as a response to the previous turn, but is used to indicate the end of the sequence. In the following excerpt of WeChat group interaction, the teacher reminds the student to change his alias in the group (line 04). After receiving confirmation from the teacher, the student uses "*hǎo lei*" to acknowledge receipt and conclude the interaction.

Excerpt 10:

04 师：@生 顺便这位同学改一下群备注哈

05 生：好哒，已修改，谢谢

06 师：好嘞

Thirdly, in sequence-closing spots, "*hǎo lei*" marks receipt of acceptance to suggestions, which informs its prior participant that he/she has received the acceptance. This interaction occurred between two classmates, where Yao was asked to prepare some materials and provide them to Xue within a specific time frame. In line 07, Yao begins a new turn constructional unit (TCU) by asking a question and giving an account. However, Xue responds negatively and makes a proposal to Yao, asking her to deliver it that night. Then, Yao accepts the proposal. Xue then uses the phrase "*hǎo lei*" to indicate that she has received Yao's acceptance and that they will meet later that night.

Excerpt 11:

06 瑶：学姐，你在宿舍吗？我的材料整理差不多了，一会给你送上去

07 雪：我现在不在呢~

08 晚上吧可以吗

09 瑶：可以可以

10 雪：好嘞

4.3 Granting the Request

Requests are one of the widely studied speech acts in the fields of pragmatics, cross-cultural communication, and conversation analysis. The act of requesting is the imposition of the requester's will on the requested person, a social act that requires the requested person to give and the requester to benefit (Wu & Liu, 2020). When the FPP is a request, or implied request, "*hǎo lei*" in the responding position can be used to grant the request.

Excerpt 12:

01 瑶：婷婷

02 帮我带个皮筋和黑夹子好吗

03 谢谢

04 瑶 ((瑶拍了拍婷))

05 婷：好嘞

06 婷 ((婷拍了拍瑶))

07 瑶：[玫瑰]

Request is a very common social behavior, through which social members can directly obtain the help of other social members (Yu, 2019). In this excerpt, Yao and Ting are roommates who attended the same class on that day. In the first four lines, Yao asks Ting for help and uses the WeChat feature "tickle" to get Ting's attention. In line 05, Ting grants Yao's request and also employs the transactional practice of "tickle" to interact with Yao, which helps to develop their relationship in the interaction (Wu, 2021). Sometimes, participants may not make a request directly and explicitly. Instead, they may use a pre-expansion sequence that lays the groundwork for the base first-pair part of the request in some way.

Excerpt 13:

01 晓：在宿舍吗

02 @鹿

03 鹿：在

04 晓：婷婷手机没带

05 你看看

06 鹿：是的

07 好嘞

08 晓：[感谢]

The above WeChat interaction took place on a weekday morning. Xiao, Lu, and Ting are roommates who were attending the same course that morning. Xiao and Ting left earlier than Lu, but Ting forgot her smartphone at the dormitory. In line 01, Xiao initiates the interaction through a yes/no question, which receives a positive answer from Lu. Xiao then gives an account as to why the question was asked, which avoids possible request sequence. The interaction could have ended here, but Lu immediately responds with "*hǎo lei*" indicating that she has understood Xiao's request implied in the pre-sequence question, that is, if Lu can locate Ting's smart phone, she should bring it with her to the

classroom for her. Therefore, Lu offers to help before Xiao makes her final request. Additionally, Xiao's gratitude in line 08 confirms Lu's prediction of Xiao's ultimate goal.

5. Conclusion

With the aim of addressing previous research gaps, this study systematically examines the sequence organization and interactional function of "*hǎo lei*" in WeChat interactions, using CA as the research methodology. Observation reveals that "*hǎo lei*" performs various interactional functions, highly dependent on its position within the conversation sequence. It emerges either as a response to the preceding turn in the responding position or in the sequence-closing position, addressing the turn where the recipient has previously engaged with a participant's action.

What's more, in WeChat interactions, "*hǎo lei*" exhibits two primary characteristics across different sequence organizations. On the one hand, it frequently combines with various punctuations, emojis, and modal particles, enhancing the speaker's positive attitude or commitment to the action suggested by the previous turn. On the other hand, "*hǎo lei*" is typically followed by supplementary information, such as expressions of gratitude, information receipt tokens, and additional explanations, all of which help strengthen agreement and understanding within the interaction.

Finally, "*hǎo lei*" is a prevalent conversational practice in online exchanges, predominantly used to align the speaker's stance with that of another participant. Firstly, it appears in the responding position to signify acceptance of suggestions, proposals, invitations, or encouragements. Secondly, "*hǎo lei*" serves as a crucial receipt token, signaling that the recipient has acknowledged the information or assistance provided by the speaker. This function manifests in the responding or sequence-closing positions within informing-reply or reminding-reply sequences. Thirdly, "*hǎo lei*" can also be employed in the responding position to grant requests or implied requests. The conventional use of this conversational practice allows both parties to express agreement, acceptance, receipt of information, and transition between topics while maintaining harmonious social relationships.

References

Beach W A. (2020). Using prosodically marked "Okays" to display epistemic stances and incongruous actions. *Journal of Pragmatics*, (169), 151-164.

Drew P. (2013). Conversation analysis and social action. *Journal of Foreign Languages, 36*(3), 2-20.

Giles D, Stommel W, Paulus T, Lester J, Reed D. (2015). Microanalysis of online data: The methodological development of "digital CA". *Discourse, Context and Media*, (7), 45-51.

Gong Xiaohui. (2018). The variation phenomenon and its pragmatic factors in the network words: Take "good + words" as an example. *Journal of Qiqihar Teachers College*, (03), 59-61.

Guo Yuling. (2000). Analysis of imperative sentences in Chinese and Korean and their affirmative responses. *Journal of Capital Normal University*, (3), 106-111.

Li Feng, Li Zhen. (2022). Research on "epistemology" from the perspective of conversational analysis. *Studies in Philosophy of Science and Technology*, *39*(03), 20-26.

Meredith J. (2019). Conversation analysis and online interaction. *Research on Language and Social Interaction*, (52), 241-256.

Sacks H, Schegloff E M, Jefferson G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696-735.

Schegloff E A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.

Shao Jingmin, Zhu Xiaoya. (2005). The discourse functions of hao and its evolution toward functional usage. *Studies of the Chinese Language*, (5), 399-407+479.

Sidnell J, Stivers T. (2013). *The Handbook of Conversation Analysis*. New Jersey: Wiley-Blackwell.

Wu Yaxin, Liu Shu. (2020). Approaching delicacy in requesting from the perspective of sequence organization. *Modern Foreign Languages*, *43*(01), 32-43.

Wu Yaxin. (2021). A conversation analytic approach to identity. *Journal of Foreign Languages*, *44*(03), 49-59.

Wu Yaxin. (2022). The conversation analytic approach to particles in Mandarin Chinese.

Journal of Foreign Languages, 45(06), 21-31.

Yu Guodong, Guo Hui. (2020). Adjacency pair as a social norm. *Modern Foreign Languages, 43*(01), 18-31.

Yu Guodong. (2019). Conversational practices as evidence for taking request as a face threatening act. *Journal of Beijing International Studies University, 41*(04), 3-19.

Yu Guodong. (2022). *What Is Conversation Analysis?*. Shanghai: Shanghai Foreign Language Education Press.

Journal of Linguistics and Communication Studies
ISSN 2958-0412
www.paradigmpress.org/JLCS
Volume 5 Number 1 March 2026

The Effects of Noise on English Vowel Perception by Chinese EFL Learners of Different Proficiency Levels

Bingyi Wu[1]

[1] Ocean University of China, Shandong, China
Correspondence: Bingyi Wu, Ocean University of China, Shandong, China.

doi:10.63593/JLCS.2026.03.03

Abstract

Daily communication usually occurs in various kinds of noise backgrounds. Noise interferes with the process of recognizing the target sounds, thus affecting the accuracy of speech perception. In the present study, three speech perception environments were set up: quiet, English babble noise and Chinese babble noise. 60 undergraduate English majors were selected as subjects and divided equally into a higher proficiency level group and a lower proficiency level group. By measuring and calculating their correct identification rates of twenty English vowels in three different listening conditions, the study investigated the effects of second language proficiency and noise type on Chinese English learners' vowel perception. The results showed that learners at the higher level showed a significant advantage over those in the lower level group in identifying English vowels in both quiet and noisy environments, indicating a positive correlation between L2 proficiency and English vowel perception. In contrast, the results of the higher and lower level groups showed different trends in English and Chinese noise. The accuracy of the higher proficiency level group in English babble and Chinese babble are almost the same, while the lower proficiency level group performed significantly better in Chinese babble than in English babble. This suggests that both L2 proficiency and the types of noise have an effect on English vowel perception, and that there is an interactive effect between the two factors.

Keywords: vowel perception, noise, informational masking

1. Introduction

1.1 Research Background

Speech perception is of vital importance in language communication. It not only plays an important role in conveying messages between the speaker and listener, but also contributes to helping L2 learners to develop their second language listening ability and pronunciation. Speech sounds can be divided into vowels and consonants, of which vowels are the nucleus of syllables as well as the direct bearer of major suprasegmental speech flow features such as stress and intonation (Ladefoged, 2006).

Most speech communication in daily life occurs in noisy background. In real life and work environments, various noises interfere with the tracking and recognition of target speech signals at various stages of sound processing, which is called auditory masking of noise (Watson, 2005). The impact of noise on second language communication is particularly obvious. In some noisy places, the perception and understanding of a second language become extremely

challenging (Shimizu *et al.*, 2002; Yang & Zhao, 2014; Xu *et al.* 2018). Generally speaking, noise can produce energetic masking and informational masking to interfere with listeners' speech perception. There has been enormous research on the speech perception of L2 learners. However, most of them concentrate on their performance in a quiet environment and ignore the influence of noise, which is much ubiquitous in real life.

1.2 Purpose and Significance of the Study

The present study aims to investigate the English vowel perception in babble noise by Chinese English learners. The first objective of this study is to consider whether there is a difference in L2 learners' vowel perception as their English proficiency improves. The second objective is to explore the effect of informational masking with different masker types on the performance of Chinese English learners of different proficiency levels. Finally, the study attempts to explore whether there is an interactive effect between the two factors, English proficiency and masker intelligibility, in influencing English vowel perception.

This study is significant because speech perception is essential in the process of language communication. First of all, the accuracy of listeners' speech perception not only influences their communication with others, but also plays an important role in L2 speakers' acquisition of the language, especially the acquisition of pronunciation. Besides, among all the speech sounds, vowel perception directly affects the accuracy of pronunciation acquisition and the effect of language communication. Moreover, noise can be everywhere in daily life, so the interference caused by noise should be paid more attention to. In an occupational environment such as airports, speech perception is vital. Workers are required to perceive speech sounds accurately in noisy environments to guarantee safety (Shimizu *et al.*, 2002; Yang & Zhao, 2014; Xu *et al.*, 2018). Finally, this study is meaningful in the comprehension of Chinese EFL learners' speech perception performance in noise. In this way, teachers can be inspired to improve their teaching methods and students can adjust their way to study in order to improve their speech perception ability and further improve their communication ability. Further, the results of this study can be applied in the area of language assessment to enhance the validity and reliability of language listening tests.

2. Literature Review

2.1 Informational Masking of Babble Noise

Informational masking is believed to be caused by the competition between the masker and the target sound in the auditory center for processing resources (Yang *et al.*, 2014). In the exploration of second language speech perception, researchers often use white noise, pink noise, brown noise, speech-shaped noise, babble noise and other noise backgrounds with different signal-to-noise ratios (SNRs) to obtain different levels of informational masking effects. Some studies also incorporate background noise into speech signals to avoid the ceiling effect, which has become a standard method (Winters & O'Brien, 2013; Yang & Zhao, 2014).

Research has been conducted to study the possible factors that may influence the amount of informational masking on speech perception. Mi *et al.* (2003) measured the perception of L2 English vowels in quiet, long-term speech-shaped noise (LTSSN) and multi-talker babble (MTB) by English-native (EN) listeners and Chinese-native listeners in the U.S. (CNU) and China (CNC). In MTB, CNU listeners performed significantly better than CNC, which not only suggested that informational masking played a different role in speech perception by listeners with different backgrounds, but also indicated that exposure to native English may reduce informational masking of MTB. Different types of noise cause a range of various degrees of informational masking, which differ in their masking effects for listeners with different language backgrounds. Jin and Liu (2012) investigated English sentence perception in quiet and two types of maskers, MTB and LTSSN, by English-, Chinese-, and Korean-native listeners. The results found that non-native listeners were more affected by background noise than native listeners. Besides, the masking effects of MTB were greater for Chinese listeners, which indicated that there might be interaction effects between the listeners' native language and competing noise.

2.2 Native and Non-Native Speakers' Speech Perception in Noise

The effect of noise on speech perception is much greater for non-native than for native listeners (Mayo *et al.*, 1997; Rogers *et al.*, 2006; Bradlow & Alexander, 2007; Cooke *et al.*, 2008). The question was first addressed in the laboratory by Black and Hast (1962), nearly half a century ago. They

compared word perception scores in quiet and white noise at SNRs of +4, 0, and -4dB. A modest disadvantage for non-native relative to native listeners in quiet conditions of about 11% grew to 16%, 25% and 29% with increasing noise. In Cooke's (2008) study, English and Spanish native speakers were required to recognize keywords in English sentences presented in quiet and in noise, including stationary noise and competing utterances. The study suggested the native advantage in all the listening conditions. Even for bilinguals who started learning the second language at an early age, they were still not comparable with native speakers. The experiments of Rogers *et al.* (2006) confirmed that although the perceptual performance of bilinguals who began to learn a second language at an early age was similar to that of native speakers under quiet conditions, there was still a considerable difference in noise.

Some studies compared native and non-native speakers' performance in quiet, low noise and high noise conditions and found that the size of the native advantage varied in different conditions. While Bradlow and Bent (2002) found that the native advantage remained constant as noise level increased, Garcia Lecumberri and Cooke (2006) demonstrated that native advantage increased in the high noise condition. In the study conducted by Shimizu *et al.* (2002), the researchers found that in the task of English speech perception under quiet conditions, white noise, pink noise and aircraft noise, the speech perception ability of Japanese college students decreased with the decrease of the signal-to-noise ratio of various noise backgrounds.

Moreover, the effects also vary with the noise types (Brungart, 2001; Shimizu *et al.*, 2002; Cutler *et al.*, 2004; Lecumberri & Cooke, 2006; Cooke *et al.*, 2008; Cutler *et al.*, 2008; Broersma & Scharenborg, 2010; Jin & Liu, 2012; Calandruccio *et al.*, 2014). For example, Van Engen (2010) investigated the recognition performance of English sentences by native English speakers and Chinese English learners in the background of two-talker noise (including both English and Chinese conditions) and pointed out that both groups were more affected under English noise conditions.

2.3 Individual Differences in Speech Perception

The performance of L2 speakers in noise is not only different from that of native speakers, but also exhibits a significant individual difference. In most linguistic studies, researchers have focused on two major factors: the age at the onset of second language learning (e.g., Mayo *et al.*, 1997) and second language experience (e.g., Flege *et al.*, 1999; Bradlow & Bent, 2002).

Flege *et al.* (1999) reported that experienced non-native listeners showed better English vowel production and perception than inexperienced non-native listeners. Exposure to a second language environment improves non-native speakers' L2 experience. In Mi's (2003) study, Chinese-native speakers in the U.S. performed better than those in China when they perceived English vowels in multi-talker babble, suggesting the possible effects of second language experience. Studies suggest that non-native speakers who started learning a second language earlier performed better than those who were later, but in most cases, the age of onset and L2 experience have shared effects. Mackay *et al.* (2001) found that when perceiving the first consonant of an English word under noise, Italian native speakers who began to learn English earlier performed better than those who started to learn English later, but there was no difference in their performance when perceiving the last consonant of a word. At the same time, even if the subjects started learning English early, their perception was also affected by the use of their mother language.

Except for the age of onset and L2 experience, researchers are discovering other factors that have potential effects on speech perception in noise. For example, McGowan (2015) measured native listeners' perception and transcription of Chinese-accented English in noise and found there were social factors such as social expectation that caused individual differences in participants' performance. All the participants transcribed more accurately when they were presented with a Chinese face than a Caucasian face. The individual differences in non-native speakers' proficiency have also been considered. Zhou *et al.* (2010) studied the speech perception of twenty RP English vowels by 88 English majors. Each vowel occurred three times in different carrier words in the recording. Participants were required to transcribe all the phonetic symbols they had heard, including consonants and vowels. After the experiment, their answers were rated and only the vowel transcription was scored. Results showed that the mean identification accuracy was 75% among English majors, and their speech perception

ability was significantly related to the overall ability of English, oral English proficiency level and gender.

The degree of individual differences varies with subjects and listening conditions. In the study of Shimizu *et al.* (2002), the individual difference in Japanese college students' English speech perception performance was small under quiet conditions, but it became larger as the signal-to-noise ratio increased in noisy background. The subjects with a high L2 proficiency continuum were expected to perform better in noisy conditions, whereas those with low proficiency performed worse.

2.4 Summary

Based on the above review of previous research, it is clear that compared with native listeners, non-native listeners suffer more difficulties from the informational masking effects produced by noise when perceiving English speech sounds. The decline in noise varies depending on the noise types and listeners' individual differences.

However, research concerning second language speech perception in noise focused more on sentence comprehension (Van Engen, 2010; Jin & Liu, 2012), tone recognition (Mao & Xu, 2016) and perception of individual phonetic symbols (Mi *et al.*, 2013). Although most of the previous research investigated sentence recognition and consonant identification in noise, few have studied vowel identification in noise by non-native listeners, in which there are no suprasegmental cues. Moreover, as to the individual difference, researchers mainly focused on the age of arrival and learning experience, and ignored the second language proficiency of Chinese learners. Besides, few researchers have explored the effect of different languages of noise, or combined it with the effect of L2 proficiency to analyze.

Therefore, this study adopted English vowel identification tasks for Chinese EFL learners at higher and lower proficiency levels to test their vowel perception performance in quiet, Chinese babble noise and English babble noise, aiming to investigate the relationship between noise types and listeners' L2 proficiency levels.

3. Methodology

3.1 Research Questions

The study aims to answer the following three questions:

1) What are the effects of second language proficiency in influencing Chinese learners' English vowel perception in noise?
2) What are the effects of masker types in influencing Chinese learners' English vowel perception?
3) What are the interactive effects between L2 proficiency and masker types in influencing Chinese learners' English vowel perception?

3.2 Participants

All the participants were freshmen of English major at the Ocean University of China. The learners have received 9-year compulsory education and thus were of comparable educational background. A total of 67 students (18-20 years) were recruited and asked to take the *Oxford Quick Placement Test* at first. According to their test grades, 60 of them were selected to participate in the following experiment. To ensure the L2 proficiency discrepancy, 7 students were excluded because their grades were at the medium level. Then the 60 participants were averagely divided into two groups, the higher proficiency group (HG) and the lower proficiency group (LG), according to the test results and grades of their latest English exams. The mean test scores of HG and LG were respectively 92.70 and 78.57 out of a maximum 120 points. Before the formal experiment, all the participants were required to get familiar with the 20 English vowels and make sure they were able to recognize the vowels in a short time and make a choice.

Table 1. Participant Background Data Across HP and LP Groups

	HG (N=30)		LG (N=30)	
	Mean	SD	Mean	SD
Age	18.80	0.76	18.61	0.74
Years of English learning	10.00	0.79	9.76	0.70
OQPT grades	92.70	6.09	78.57	4.94

Age is expressed in years; OQPT=*Oxford Quick Placement Test*, scores out of a maximum 120.

3.3 Materials

The test was conducted under three different situations: quiet, four-talker English babble noise (EN) and four-talker Chinese babble noise (CN). Two kinds of noise backgrounds were set. English monolingual and Chinese monolingual recordings of texts were respectively selected

from *New Concept English* and HSK textbooks. They respectively included two male voices and two female voices. The speakers were all native speakers of English or Chinese. The content of recordings consisted of both daily conversations and articles concerning various subjects, such as nature, technology and art.

20 Received Pronunciation (RP) English vowels served as speech stimuli. The speech materials used for this study consisted of 60 monosyllabic CNC (Consonant - Nucleus - Consonant) words (eg., *hard*, *fit*, *love)*, in which each vowel occurred three times. British pronunciation of these sixty carrier words was downloaded from the *Longman Dictionary of Contemporary English Online* (LDOCE), produced by a male voice. There was an interval of three seconds between two neighboring words to ensure there was enough time for participants to react and answer. The 60 words were presented in isolation with different sequences to produce three pieces of speech materials. Two of them were respectively mixed with EN and CN through an audio processing software. The duration of each recording was four to five minutes.

Table 2. Twenty English Vowels and Carrier Words Used in the Tests

Vowel	Word 1	Word 2	Word 3
/iː/	leave	heed	feed
/ɪ/	fit	hid	lit
/ɔː/	fort	maul	caught
/ɒ/	toss	lot	hod
/uː/	booth	loose	fool
/ʊ/	hook	book	full
/əː/	birth	hurt	perch
/ə/	was	rubber	ago
/ɑː/	cart	bark	laugh
/ʌ/	cut	love	bus
/e/	check	head	fetch
/æ/	pad	fat	had
/eɪ/	cape	late	bake
/aɪ/	live	fight	hide
/ɔɪ/	boil	join	voice
/ɪə/	beard	dear	fear
/ɛə/	chair	fair	bear
/ʊə/	tour	lure	poor
/aʊ/	house	loud	foul
/əʊ/	quote	rose	load

3.4 Data Collection

The research first selected 60 students after they took the *Oxford Quick Placement Test* and divided them into two proficiency groups based on their test results and the latest English exam scores. Each participant took the test three times under different conditions: quiet, four-talker English babble noise and four-talker Chinese babble noise. Listeners were seated in front of a computer, which presented 20 English vowels as options on the screen. After reading the instructions, they practiced with other words in order to get familiar with the experimental process. In the formal experiment, each of them listened to the three speech materials in random order. After they heard a target word, they were expected to choose which vowel they heard in the word. Each test was expected to cost about 15 to 20 minutes. Between every two words, each of the participants was given enough time to react, make a choice and move on to prepare for the next word. The answers were scored after the tests and the accuracy of vowel identification was calculated.

A pretest was conducted to verify the feasibility of the experiment. To avoid ceiling effects, the signal-to-noise ratios were selected through pilot testing to endure similar identification scores under each condition (in the 50-75% range).

3.5 Data Analysis

After the procedure of the identification experiment, the data collected were analyzed in the following way. First, the identification score was determined by a strict correct answer count. Answers choosing more than one vowel were regarded as false except that for those words which contain two vowels such as *rubber*, participants were required to choose both vowels they had heard. However, only the target vowel was graded. Then the scores were converted to percentages to represent identification rates. Finally, all analyses were carried out using SPSS Version 25.0. A one-way repeated measures ANOVA with L2 proficiency as a between-subject factor and noise type as a within-subject factor was conducted.

4. Results and Discussion

4.1 Vowel Perception by Listeners of Different L2 Proficiency Levels

Table 3. Descriptive Statistics of Mean Identification Rates of HG and LG

	N	Minimum	Maximum	Mean	SD
HG	90	15.00	100.00	67.00	20.48
LG	90	10.00	80.00	53.00	16.73

The mean identification scores by LG and HG groups under three conditions were calculated and listed in Table 3. The results indicated that learners of higher proficiency levels generally performed better than learners of lower proficiency levels in quiet and noise. Overall, listeners with higher L2 proficiency (67% correct) outperformed listeners with lower proficiency (53% correct) significantly (p<0.01). Therefore, the vowel identification scores had a strong correlation with L2 proficiency.

Previous studies mainly focused on the comparison of native and non-native speakers. They found that non-native listeners had more difficulty in recognizing speech than native listeners (Shimizu *et al.*, 2002; Bradlow & Bent, 2002; Lecumberri & Cooke, 2006; Pinet *et al.*, 2011), and the disadvantage was much larger in noise than in quiet. Few research has studied the non-native speakers with different proficiency levels, especially Chinese English learners. The present results indicated that L2 proficiency might be one influence factor for non-native speakers' perception. However, whether it makes a difference in each of the three conditions in the present study is still to be analyzed.

Besides, as shown in the table, the SD of both groups were large, then the individual differences within groups were considerable. According to former research, individual differences among non-native speakers were mainly the age of onset and second language experience, which were already controlled in the experiment. All the participants started English learning as a second language at 7-9 years old, and they all did not have ever been exposed to English environment for a long time, for example, living abroad in an English-speaking country. The only criterion for grouping was their grades in OQPT. Their grades were a continuum and to divide the participants into two proficiency groups at one point along the continuum was not as precise as possible. Besides, more various grouping criteria should be adopted to ensure the proficiency discrepancy between the two groups.

Table 4. Descriptive Statistics of Mean Identification Rates in Quiet

	N	Minimum	Maximum	Mean	SD
HG	30	35.00	100.00	73.00	19.23
LG	30	20.00	75.00	54.33	16.49

In quiet, the mean accuracy of the HG and LG group was 73% and 54.33%. Under both conditions, listeners with higher proficiency levels outperformed listeners with low proficiency levels significantly (p<0.01). This is consistent with the overall situation above. However, the overall identification accuracy was not as high as expected. Although the percentages were slightly higher in quiet, the situation was not optimistic. In previous research, the mean identification accuracy in a quiet environment by L2 learners was at least over 75% (Van Engen, 2010; Zhou *et al.*, 2010; Xu *et al.*, 2018), while in the present study, the accuracy was much lower with 73% by higher proficiency level learners as the highest. There may be two reasons to explain that. First, the participants are all freshmen in the university and their English competence does not achieve a relatively high level. Second, although all the participants are guaranteed to have known all the English vowels and be able to recognize them before the experiment, some students were still not very familiar with them at the time of data collection. Therefore, under the pressure of taking tests, it would probably be challenging for them to make the best choice during the experiment process.

4.2 Vowel Perception Under Different Types of Noise

Table 5. Descriptive Statistics of Mean Identification Rates in Quiet and Noise

	N	Minimum	Maximum	Mean	SD
Quiet	60	20.00	100.00	63.67	20.10
EN	60	10.00	100.00	56.25	18.99
CN	60	15.00	100.00	60.08	20.30

The mean identification rates by participants in quiet, EN, and CN were respectively 63.67%, 56.25%, and 60.08%. Data showed that due to the masking effects of noise, participants performed worse in two types of noisy environments than in quiet. What was interesting about the data in the table was that the mean identification rate in

Chinese noise was slightly higher than that in English noise, though the difference was not significant (p=0.288). It meant that there may be effects of noise types, i.e., the languages of maskers, in influencing participants' different performances, but whether the situation is the same for the higher or lower proficiency group is still to be analyzed. The results were partly in agreement with Van Engen's (2010) study, in which 20 native English speakers and 20 non-native English speakers whose native language was Chinese were recruited to do second-language recognition test in both English and Chinese 2-talker babble noise. Van Engen found that both groups of participants experienced greater difficulty in English noise than in Chinese noise, and this was consistent with the present study. The greater interference from English noise for both groups suggested that acoustic and/or linguistic similarity between the speech signal and the noise might be the most essential factor in driving noise language effects.

However, Garcia Lecumberri and Cooke (2006) did not observe the different effects of L1 and L2 noise types. They investigated the performance of Spanish speakers on English consonant perception respectively in Spanish and English competing speech and found that there was no significant differences in the masking effects of the two languages. The disagreement in findings may be accounted by that their study differed from the present study and Van Engen's study in that the former used single talker's speech as the noise while the latter respectively used 4-talker babble and 2-talker babble. Since the signal density increased, greater energetic masking could be induced, and then renders informational masking effects more observable (Van Engen, 2010).

4.3 Interplay Between L2 Proficiency and Masker Types

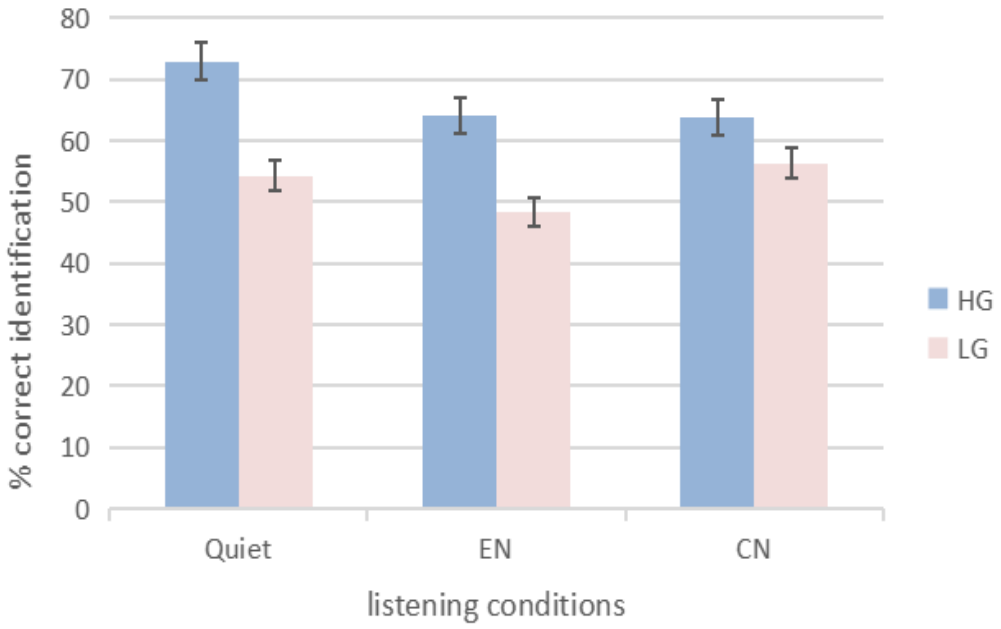

Figure 1. Mean Identification Rates by HG and LG in Three Conditions

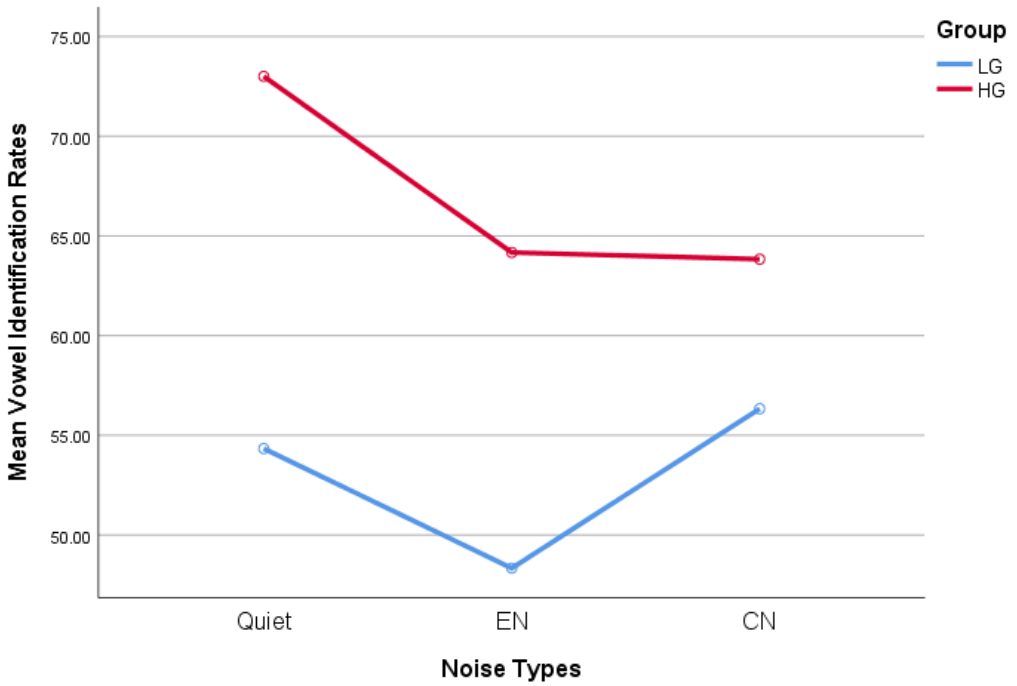

Figure 2. Changes of Mean Identification Rates Across Three Conditions

A repeated-measures ANOVA with L2 proficiency as a between-subject factor and noise type as a within-subject factor revealed that the effect of noise type was significant ($F(2,57)=5.411$, $p<0.01$, $\eta^2=0.16$). There was also an interaction effect between noise type and L2 proficiency ($F(2,57)=3.168$, $p\leq0.05$, $\eta^2=0.1$). The main effect of L2 proficiency was significant ($F(1,58)=12.362$, $p<0.01$, $\eta^2=0.176$).

Since there was an interaction effect between L2 proficiency levels and noise types, a simple effect analysis was performed after correction by the Bonferroni method. Results showed that in quiet and in EN, the simple effects of L2 proficiency were significant ($F(1,58)=16.284$, $p<0.01$, $\eta^2=0.219$; $F(1,58)=12.448$, $p<0.01$, $\eta^2=0.177$), while in CN, it was not significant ($F(1,58)=2.084$, $p=0.154$, $\eta^2=0.035$). Pairwise comparisons indicated that among HG learners, their mean identification score in quiet was the highest and showed strong advantages over CN and EN environments ($p<0.05$). Though they performed slightly better in EN than in CN, data showed no significant difference between the two noise types ($p>0.05$). However, among LG learners, their identification score in CN was significantly higher than that in EN ($p<0.05$), and there was no significant difference between that in quiet and in EN, or in CN.

4.3.1 The Effects of L2 Proficiency

Table 6. Descriptive Statistics of Mean Identification Rates in Noise

	N	Minimum	Maximum	Mean	SD
HG	60	15.00	100.00	64.00	20.58
LG	60	10.00	80.00	52.33	16.96

In the two types of noisy environments, the mean vowel identification rates by HG and LG were respectively 64% and 52.33%. Compared with those in quiet, there was an apparent decrease. Likewise, participants in HG outperformed the others significantly, which meant that in noise, second language proficiency still had an impact on participants' vowel perception. However, on account that two different types of maskers were employed in the study, it should be analyzed separately whether L2 proficiency played a role in each noisy environment.

Table 7. Descriptive Statistics of Mean Identification Rates in EN

	N	Minimum	Maximum	Mean	SD
HG	30	35.00	100.00	64.16	18.15
LG	30	10.00	65.00	48.33	16.57

In English noise, the mean identification rates of the HG and LG group were 64.17% and 48.33%.

Participants of higher proficiency levels generally got significantly higher scores than those of lower proficiency levels. Thus, it could be concluded that both in quiet and English noise, L2 proficiency would influence the English vowel perception of Chinese learners. The more proficient participants were at the second language, the better they performed in speech perception.

Table 8. Descriptive Statistics of Mean Identification Rates in CN

	N	Minimum	Maximum	Mean	SD
HG	30	15.00	100.00	63.83	23.06
LG	30	30.00	80.00	56.33	16.65

On the contrary, in Chinese noise, the mean vowel identification rates of the HG and LG groups were respectively 63.83% and 56.33%. There was no significant difference between the two groups (p=0.154). This indicated that though L2 proficiency played an important role in influencing listeners' performance in quiet and in English noise, it seemed to have no impact on English vowel perception by Chinese EFL learners when there was Chinese babble noise.

4.3.2 The Effects of Noise Types

Table 9. Descriptive Statistics of Mean Identification Rates of LG

	N	Minimum	Maximum	Mean	SD
Quiet	30	20.00	75.00	54.33	16.49
EN	30	10.00	65.00	48.33	16.57
CN	30	30.00	80.00	56.33	16.65

The vowel identification rates by the LG participants in quiet, English noise, and Chinese noise were respectively 54.33%, 48.33%, and 56.33%. Performance in quiet, which was similar to performance in CN, did not have apparent advantages over the two other conditions. It was due to their poor performance in quiet environment, which was probably caused by participants' unfamiliarity with the English vowels. There was a significant correlation between vowel identification accuracy and noise types among listeners with lower proficiency levels (p<0.01). This indicated that LG learners were more likely to be affected by environment and noise types. As noted in the previous chapter, Van Engen (2010) investigated native English speakers and Chinese native speakers and found that both groups suffered more from the interference by English noise than Chinese noise. It was worth noting that Chinese speakers experienced a smaller masking release in Chinese noise relative to English noise. In other words, compared with English native speakers, non-native speakers did not show relatively significant advantages in L2 perception in Chinese noise. Lecumberri *et al.* (2010) have tried to explain this by proposing the multiple effects of informational masking. On the one hand, due to the intelligibility of masking sounds, listeners faced a great challenge when processing the target sounds and masking sounds simultaneously. On the other hand, if listeners were familiar with the masking sounds, they would suffer greater interference. As Chinese speakers were more familiar with their mother tongue than L2 (English), it might be more challenging for them when the masking sounds were Chinese.

In the present study, participants were all EFL learners, so they were assumed to be more familiar with Chinese than with English. Therefore, they were expected to suffer greater interference in Chinese noise rather than in English noise. However, the current results suggested that participants showed strong advantages in Chinese noise compared with English noise. As in English babble noise, the speech materials and maskers were the same languages. Identifying the target sounds required more cognitive resources when there was a similarity between target sounds and masking sounds, so English noise would cause more interference for listeners when perceiving English vowels. In conclusion, for participants of lower L2 proficiency levels in this study, the effects of typological similarity between the target sounds and masking sounds played a relatively dominant role over listeners' degree of familiarity with noise.

Table 10. Descriptive Statistics of Mean Identification Rates of HG

	N	Minimum	Maximum	Mean	SD
Quiet	30	35.00	100.00	73.00	19.23
EN	30	35.00	100.00	64.16	18.15

CN	30	15.00	100.00	63.83	23.06

The vowel identification rates by the HG participants in quiet, English noise, and Chinese noise were respectively 73%, 64.17%, and 63.83%. Participants performed best in quiet, and significantly better than in noise (p<0.01), while in contrast to the results of participants at lower proficiency levels, there was no significant difference between EN and CN environments (p=0.951). The results suggested that although the overall data showed a relative disadvantage in EN, learners of HG seemed to be less susceptible to the influence of different masker types, or languages of noise. The findings were in agreement with the study of Garcia Lecumberri and Cooke (2006), who found no apparent differences in Spanish speakers' perception of English consonant in Spanish and English noise. Besides, Van Engen and Bradlow (2007) used 6-talker babble and also found no effects of languages of noises. On the contrary, many other studies have reported that there were effects of noise types. Van Engen (2010), for example, claimed that listeners' language experience and signal similarity would modulate the interference they suffered when they perceived speech in different noises and further contribute to different performances. One possible reason that may account for the different outcomes was the different speech perception tasks adopted. Van Engen (2010) investigated keyword recognition in sentences, while the present study and Garcia Lecumberri and Cooke (2006) measured identification of individual phones, vowels and consonants respectively. Then it seemed that, when participants listened to sentence-length materials, they were more likely to be interfered by noise in a specific language. As participants tried to identify keywords in sentences rather than consonants, more linguistic structures needed to be processed, so processing inefficiencies across levels of linguistic processing (Cutler *et al.*, 2004) would accumulate for non-native listeners. Moreover, in the perception of individual phones in isolation, it did not involve much cerebral activity to process and comprehend the target sounds and masking sounds.

In conclusion, L2 proficiency and noise types respectively played a dominant role in certain conditions. Results showed that for the non-native listeners at lower proficiency levels, interference from a 4-talker masker in the target language (English) was greater than interference from the listeners' native language (Chinese). This finding suggests that signal similarity (a match between target and noise languages) is at least as important as second language proficiency in driving noise language effects in general. For Chinese EFL learners of higher proficiency levels, their familiarity with their native language would offset the effects of signal similarity so that no significant difference in English versus Chinese noise was observed. These results supplemented and expanded the findings of Van Engen (2010) and Lecumberri *et al.* (2006) to a certain degree, in that similarity and familiarity would differ in their degree of influences on non-native speakers of different proficiency levels.

5. Conclusion

5.1 Major Findings of the Study

The primary goal of this study is to examine the effects of noise types and second language proficiency in influencing Chinese EFL learners' perception of English vowels. Subjects' English abilities were measured at first. Then they participated in three vowel identification tests, respectively in quiet, English babble noise and Chinese babble noise. After the experiment, participants' answers were scored and analyzed. The experimental results prove that noise does have masking effects on English vowel perception by Chinese EFL learners. The effect varies in degree across different kinds of noise types and different L2 proficiency levels. Based on the three research questions, the research findings of the present study are as follows:

Listeners at higher L2 proficiency levels perceive English vowels better than listeners at lower L2 proficiency levels. In quiet, English babble noise and Chinese babble noise, the higher proficiency learners all possess the advantages.

The effects of noise are different for listeners at higher proficiency and lower proficiency levels. For the former, the presence of noise does have an impact on vowel perception, but changing the noise types will not cause much difference to the interference. However, for the latter, English noise is proven to be more challenging for them to resist than Chinese noise.

There are interactive effects between the two factors, L2 proficiency, and noise types. If the listeners are at a high proficiency level, noise types will have little effect because as previous research suggests, advanced English learners are

less likely to be influenced by noisy environments. However, if the listeners are at a lower proficiency level, the change in noise types may cause greater influence. As it has been difficult for them to identify the vowels compared with advanced learners, the interference caused by English noise, which is similar to the target sounds, requires much more cognitive resources to process the information.

5.2 Implications of the Study

According to the results and discussion above, some practical implications of the present study are addressed briefly. Firstly, this study has shed light on some notable problems in non-natives' English vowel perception, especially in unfavorable listening conditions such as noise. Speech perception is the premise of speech acquisition and speech production. It determines learners' L2 pronunciation acquisition. In the current study, the accuracy of speech perception is proven to be influenced by L2 proficiency. This further improves that individual differences, including L2 experience, age of arrival and L2 proficiency, are internal factors that will influence learners' speech perception.

Then, the ability to deal with different noisy environments when perceiving English vowels also differs across L2 proficiency groups. Thus, in a real teaching situation, learners with different proficiency levels should adopt relevant measures, such as controlling the listening conditions and using group instruction, to guarantee teaching quality and results.

Furthermore, from a more clinical perspective, the findings in this study can be extended to some research on subjects with hearing disorders, including patients with hearing degradation or impairment and so on.

5.3 Limitations and Suggestions for Future Studies

Due to the restriction of actual conditions, the present study has limitations and needs to be improved.

First, the total number of participants is sixty. They are divided into two groups and thus, each group consists of thirty subjects. This is a relatively small number for such an experiment which requires participants to listen and judge, so subjectivity is a potential factor to influence the results. Future studies should increase the number of subjects in order to minimize the effect of individual differences.

Second, as mentioned above, the overall accuracy is not ideal. This is owing to the generally low second language proficiency level of participants recruited in the present study. Besides, the only criterion for grouping is the grades of OQPT and their latest English exam. This may lead to irrationality and one-sidedness to some extent. Future studies should consider strictly selecting subjects and taking various measures to determine the differences among their proficiency levels.

Third, the present study focuses on second language learners' vowel perception in multi-talker babble noise. In real studying and working conditions, there may be various kinds of noise. Thus, more noise types should be considered in future studies. Whether there is a difference in identification accuracy across the 20 vowels also should be further investigated.

References

Bent, T., & Atagi, E. (2015). Children' s perception of nonnative-accented sentences in noise and quiet. *The Journal of the Acoustical Society of America, 138*(6), 3985–3993.

Black, J. W., & Hast, M. H. (1962). Speech Reception with Altering Signal. *Journal of Speech Language and Hearing Research, 5*(1), 70–75.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America, 121*(4), 2339–2349.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America, 112*(1), 272–284.

Broersma, M., & Scharenborg, O. (2010). Native and non-native listeners' perception of English consonants in different types of noise. *Speech Communication, 52*(11-12), 980–995.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101–1109.

Calandruccio, L., Bradlow, A. R., & Dhar, S. (2014). Speech-on-speech Masking with Variable Access to the Linguistic Content of the Masker Speech for Native and Nonnative English Speakers. *Journal of the American*

Academy of Audiology, 25(4), 355–366.

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America, 123*(1), 414–427.

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *The Journal of the Acoustical Society of America, 124*(2), 1264–1268.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America, 116*(6), 3668–3678.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America, 106*(5), 2973–2987.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America, 108*(6), 3013–3022.

Ji, X., Zhang, H., Li, A. & Gong, J. (2018). An empirical study of English intonation perception among learners at different proficiency levels. *Foreign Language Teaching and Research, 50*(03), 393–406+480–481.

Jin, S.-H., & Liu, C. (2012). English sentence recognition in speech-shaped noise and multi-talker babble for English-, Chinese-, and Korean-native listeners. *The Journal of the Acoustical Society of America, 132*(5), EL391–EL397.

Ladefoged, P. (2006). *A Course in Phonetics (5th Ed.)*. Boston: Thomson Wadsworth.

Lecumberri, M. L. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America, 119*(4), 2445–2454.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication, 52*(11-12), 864–886.

Lewis, D., Hoover, B., Choi, S., & Stelmachowicz, P. (2010). Relationship Between Speech Perception in Noise and Phonological Awareness Skills for Children with Normal Hearing. *Ear and Hearing, 31*(6), 761–768.

Lin, S. (2011). The Effect of Age on the English Speech Perception and Production. *Journal of Tibet University, 26*(03), 175–178.

Mackay, I. R. A., Meador, D., & Flege, J. E. (2001). The Identification of English Consonants by Native Speakers of Italian. *Phonetica, 58*(1-2), 103–125.

Mao, Y., & Xu, L. (2016). Lexical tone recognition in noise in normal-hearing children and prelingually deafened children with cochlear implants. *International Journal of Audiology, 56*(2), 23–30.

Masuda, H. (2016). Misperception patterns of American English consonants by Japanese listeners in reverberant and noisy environments. *Speech Communication, 79*, 74–87.

Mayo, L. H., Florentine, M., & Buus, S. (1997). *A*ge of Second-Language Acquisition and Perception of Speech in Noise. *Journal of Speech Language and Hearing Research, 40*(3), 686–693.

McGowan, K. B. (2015). Social Expectation Improves Speech Perception in Noise. *Language and Speech, 58*(4), 502–521.

Mi, L., Tao, S., Wang, W., Dong, Q., Jin, S.-H., & Liu, C. (2013). English vowel identification in long-term speech-shaped noise and multi-talker babble for English and Chinese listeners. *The Journal of the Acoustical Society of America, 133*(5), EL391–EL397.

Pinet, M., Iverson, P., & Huckvale, M. (2011). Second-language experience and speech-in-noise recognition: Effects of talker-listener accent similarity. *The Journal of the Acoustical Society of America, 130*(3), 1653–1662.

Ren, H. (2022). L2 Speech perception and its Correlation with Production：Evidence from Speech Acquisition of Japanese Gemination by Chinese Learners. *Foreign Language Learning Theory and Practice*, (03), 115–127+162.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27*(03), 465-485.

Shimizu, T., Makishima, K., Yoshida, M., &

Yamagishi, H. (2002). Effect of background noise on perception of English speech for Japanese listeners. *Auris Nasus Larynx, 29*(2), 121–125.

Strori, D., Bradlow, A. R., & Souza, P. E. (2020). Recognition of foreign-accented speech in noise: The interplay between talker intelligibility and linguistic structure. *The Journal of the Acoustical Society of America, 147*(6), 3765–3782.

Van Dommelen, W. A., & Hazan, V. (2010). Perception of English consonants in noise by native and Norwegian listeners. *Speech Communication, 52*, 968–979.

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication, 52*(11-12), 943–953.

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America, 121*(1), 519–526.

Van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America, 111*(4), 1906–1916.

Watson, C. S. (2005). Some comments on informational masking. *Acta Acustica United with Acustica, 91*(3), 502–512.

Weiss, W., & Dempsey, J. J. (2008). Performance of bilingual speakers on the English and Spanish versions of the hearing in noise test (HINT). *Journal of the American Academy of Audiology, 19*, 5–17.

Winters, S., & O'Brien, M. G. (2013). Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Speech Communication, 55*(3), 486–507.

Xu, C., Yang, X., Wang, Y., Zhang, H., Ding, H. & Liu, C. (2018). Informational Masking of Six-talker Babble on Mandarin Chinese Vowel and Tone Identification: Comparison between Native Chinese and Korean Listeners. *Studies of Psychology and Behavior, 16*(01), 22–30.

Yang, X. & Zhao, Y. (2014). The effects of background noise on second-language speech perception. *Advances in Psychological Science, 22*(06), 934–942.

Yang, X., Zhao, Y. & Wang, Y. (2018). Differences in duration processing in the production and perception of English /e/–/æ/ by Chinese college students. *Foreign Languages and Their Teaching*, (02), 79–89+149.

Yang, X., Zhao, Y., Xu, L., Tang, W., Zhu, X., Bao, A. & Zhao, Z. (2017). Effects of Noise on English Listening Comprehension among Chinese College Students. *Contemporary Foreign Languages Studies*, (01), 12–17+109.

Yang, Z., Zhang, T., Song, Y. & Li, L. (2014). Subcomponents of auditory informational masking: Evidence from behavioral and neuroimaging studies. *Advances in Psychological Science, 22*(03), 400–408.

Zhou, W., Shao, P. & Chen, H. (2010). An empirical study of RP English vowel perception among English-major university students. *Foreign Languages Bimonthly, 33*(06), 45–49+128.

Journal of Linguistics and Communication Studies
ISSN 2958-0412
www.paradigmpress.org/JLCS
Volume 5 Number 1 March 2026

The Construction of Virtual Intimacy: A Multimodal Discourse Analysis on the Interactive Mechanism in Virtual Livestreaming

Yuxuan Liu[1]

[1] School of Foreign Studies, University of Science and Technology Beijing, Beijing, China
Correspondence: Yuxuan Liu, School of Foreign Studies, University of Science and Technology Beijing, Beijing, China.

doi:10.63593/JLCS.2026.03.04

Abstract

The emerging phenomenon of virtual livestreaming has garnered academic attention. However, there remains limited multimodal analysis addressing its interactive mechanisms. Grounded in the interpersonal metafunction of systemic functional linguistics and the interactive meaning framework of visual grammar, the present study employs a corpus-based approach to investigate how virtual streamers on YouTube construct a sense of virtual intimacy through verbal, visual, and their integrated multimodal resources. Results show that: (1) verbally, virtual streamers primarily use declarative sentences, medium-value modality, and probability to demonstrate a highly formulaic pattern of emotional expression; (2) visually, they foster an immersive, face-to-face-like interactive atmosphere through frontal, eye-level close-up shots combined with frequent level demands and smiles; (3) correlation analysis reveals a positive correlation between the offer contact and declarative sentence, also between level demand and low-value modality, but a negative correlation is found between smiles and declarative sentences, downward demands and interrogative sentences; (4) the multimodal coordination enables virtual streamers to effectively transform emotional interactions into sustained consumption engagement with virtual symbols. This study offers a novel perspective for research on the virtual streaming industry and digital consumption culture.

Keywords: virtual streamers, multimodal, interactive mechanism

1. Introduction

Driven by the convergence of ACG (anime, comics, and games) subculture, livestreaming culture, and online tipping practices, virtual streamers have rapidly emerged as a new form of cultural consumption. They are typically operated by human performers, with computer-generated animated avatars representing their on-screen presence, and interact with audiences via livestreaming platforms (Qin, 2022). Their revenue streams usually come from audience tipping, platform commissions, sales of related merchandise, etc. This novel form of virtual entertainment can provide the audience with a new interactive and consumption experience while eliminating the physical constraints of traditional streamers.

The origins of the virtual streaming industry can

be traced back to Japan. The performers were initially active on YouTube so they became widely known as "Vtubers". Distinct from traditional virtual idols produced through audio database synthesis, virtual streamers rely on real human performers (called *Nakanohito*) who animate their avatars in real time through motion capture and voice acting technology. This way allows for more realistic and dynamic livestreamed performances. Their virtual appearances (called "avatars") are typically designed by production companies, which can be two-dimensional, three-dimensional, or even animal forms with human characteristics, such as the popular "shark girl" avatar (Tan & Greene, 2025). This diversified character design enhances the audience's visual experience while offering performers greater creative space. It can be seen that the rapid growth of this industry is largely attributed to its unique interactivity and high levels of audience engagement. The audience is attracted by the creative content, interactivity, and the aesthetic appeal of virtual avatars (Peng & Chen, 2024). Besides, the anonymity afforded by these virtual avatars grants performers greater freedom of expression, enabling them to establish a close relationship with audiences without being limited by real-world identity.

In addition to attracting a large number of viewers, the rapid development of the virtual livestreaming industry has also drawn increasing scholarly attention. Existing research mainly centers on the characteristics of virtual streamers and the mechanisms of their interaction with the audience. Studies on their characteristics have examined how factors such as attractiveness (Liu & Zhao, 2025), coolness (Gao et al., 2025), emotionally expressive language styles (Gong & Sun, 2025), and anthropomorphism (Chen et al., 2024; Peng & Chen, 2024) influence the audience's consumption behavior. For instance, Liu and Zhao (2025) measured attractiveness of virtual streamers through dimensions of similarity, familiarity, and likability, as well as the two complementary aspects of voice and speech, showing that these factors can positively affect the audience's tipping behavior and consumption intentions. Moreover, higher textual similarity and familiarity can enhance the sense of closeness between virtual streamers and their audience, thus playing a potential role in prompting viewers' initiative to give financial support. Through a series of experimental studies, Gong and Sun (2025) found that the emotional language used by virtual streamers leads to a higher consumer intention to follow the advice (CIFA). Emotional language, by fostering emotional resonance, is considered more effective in stimulating audiences' purchase intentions than rational language. It further promotes consumption behavior by enhancing the audience's perceived agency and perceived experience. What's more, virtual streamers, due to their unique motion-capture technology, can create a hyperreal experience through highly realistic visual avatars and actions, thus strengthening viewers' immersion and making it easier for them to accept recommended products and services (Chen et al., 2024; Peng & Chen, 2024). However, the semiotic production process of these streamers has also brought an array of critical issues. For example, the subjectivity of the "person behind the avatar" (*Nakanohito*) is often erased, with their real physical body hidden. As a result, the virtual persona consumed by the audience has actually become a highly patterned digital body (Peng & Chen, 2024).

On the other hand, scholars have also explored the interaction mechanisms between virtual streamers and their audience. The following three key factors play a crucial role in the establishment of the interaction process. First, the symbolic interaction motivated by virtual bodies. From the perspective of communication semiotics, Qin (2022) points out that virtual streamers, based on the inherent symbolic characteristics of their virtual bodies, enable their audience to achieve a certain degree of cultural identity through self-projection. The audience, as the recipient of symbols, may also psychologically separate themselves from the real world and invest considerable emotional efforts to maintain the virtual world jointly established by virtual streamers and their fan groups (Peng & Chen, 2024). Second, the live content itself can enhance audience participation (Chen et al., 2025; Lu et al., 2021). Based on Consistency Theory and Dramaturgical Theory, Chen et al. (2025) found that the consistency between the viewer's interest-live content congruence (IC) and viewer's value-streamer's value congruence (VE) contributes to the audience's immersion, which in turn affects their attitude and behavior intentions. Third, the role-playing ability and performance of virtual streamers can also facilitate interaction with the audience. Their ability to convincingly present specific persona settings during livestreaming

tends to positively moderate the relationship between IC and immersion, while negatively moderating the relationship between streamer's persona-live content congruence (PC) and immersion (Chen et al., 2025). The result suggests that the role-playing ability of virtual streamers can bolster audience involvement in interaction. In addition, they are better able to capture viewers' attention and evoke emotional resonance through the performance of their virtual personas (Lu et al., 2021). To sum up, existing research on virtual streamers has showed their interest in individual characteristics and the single dimension of interaction with the audience. Yet, these studies tend to overlook the importance of multiple semiotic resources collaboratively constructing meaning. Given that these streamers' interaction mechanisms involve the coordination of multiple modalities such as language, images and embodied performance, how these modalities work together to enhance the intimacy between virtual streamers and their audience remains an area that needs further exploration.

Modality refers to the channels and media of communication, including systems of signs such as language, technology, images, color, and music (Zhu, 2007). Multimodality, in turn, concerns how meaning is constructed through the interplay of various semiotic resources, which is the use of multiple modes of communication in the design of symbolic artifacts or events and how these modes are combined in specific, structured ways (Kress & van Leeuwen, 2006). Current research on multimodality mainly draws from three different perspectives. First of all, the social semiotic approach advocated by Kress and van Leeuwen (2006) emphasizes how different modalities interact to produce complex meanings based on systemic functional linguistics. This framework attends to how social and cultural factors shape the use of semiotic resources and provides a systematic approach for analyzing the design of multimodal combinations. For instance, the "visual grammar" they proposed provides a structured framework for interpreting how visual elements convey meaning. The second perspective, derived from O'Toole (1994), applies the systemic functional grammar framework to visual arts by breaking down works like paintings into hierarchical compositional levels. This was later extended by O'Halloran (2005) to describe the grammatical systems of various semiotic resources and their metafunctions. The third approach is Norris's (2004) multimodal interaction analysis. It builds on interactional sociolinguistics and mediated discourse analysis to investigate the collaborative effects of language, gesture, movement, and other modalities in human communication, with a particular focus on how meaning is co-constructed in natural interactions.

To date, the scope of multimodal discourse analysis has expanded beyond static artifacts such as picture books (Qi, 2022), comics (Zhao, 2022), literary texts (Gu & Catalano, 2022), image-text advertisements (Kenalemang-Palm, 2023), and AI-generated images (Putland et al., 2025), to increasingly include dynamic audiovisual media, especially live streaming. For example, Wang and Pan (2022) employed conversation analysis to demonstrate that the multimodal linguistic interaction in e-commerce live streaming constitutes an effective persuasive strategy. Through frequent use of interactive symbols and emotionally personal expressions, streamers guide audiences toward purchasing behaviors, turning linguistic interaction into a form of economic vitality. Huang et al. (2020) also found that the success of Chinese livestreamer Li Jiaqi illustrates how gender identity can be mobilized as a resource to attract audiences, challenging conventional gender norms and expectations and offering new directions for the development of livestream e-commerce. Besides, in constructing the interactive significance of e-commerce livestreams, verbal modality serves to provide information and demonstrate objectivity, while visual modality functions to capture consumer attention and reduce interpersonal distance (Sun, 2024). However, it is vital to note that existing multimodal analyses of livestreaming interactions have largely focused on human streamers.

Based on previous studies, it is apparent that current research gaps exist in two main areas. On the one hand, research on virtual streamers has mainly converged on macro-level, singular characteristics such as coolness and attractiveness, or examined audience interaction mechanisms from a single perspective. But the collaborative effects of multimodal features such as language, facial expressions, and other resources remain underexplored. On the other hand, the application of multimodal discourse analysis within the livestreaming context has largely centered on human streamers, with comparatively little attention paid to virtual

streamers. In response to these gaps, the present study adopts the frameworks of interpersonal metafunction from systemic functional linguistics and interactive meaning from visual grammar to investigate how the multimodal features of verbal and visual modalities, as well as their interplay, construct the intimacy between virtual streamers and their audience. This study not only helps to broaden the research perspectives on virtual streamers but also offers more practical insights for the livestreaming industry and digital consumer culture. The following research questions will be addressed:

RQ1: What are the characteristics of verbal and visual modalities in virtual livestreaming?

RQ2: How do virtual streamers jointly construct a sense of intimacy with their audience through the collaboration of verbal and visual modalities?

2. Data and Methods

2.1 Data Description

Although virtual streamers currently broadcast on platforms such as YouTube, Twitch, and Bilibili, Vtubers hold a dominant position within the livestreaming market. According to data, Vtubers account for 80% of the top 10 streamers (Playboard, 2022b). In view of this, the present study draws its corpus from the real-time livestreams of English-speaking virtual streamers on the YouTube platform. Considering the wide range of livestreaming genres, including performances, chats, and games, etc., and their differences in interaction density, this study mainly focuses on English chatting livestreams, which feature a high proportion of verbal communication and frequent interaction. These characteristics make them particularly suitable for examining the construction of virtual intimacy.

In order to capture the interactive characteristics of streamers with different identity attributes, a total of five livestream videos were selected from Nijisanji EN's company-affiliated Vtubers and independent Vtubers unaffiliated with any management company. During data collection, about 10 minutes of high-interaction clips were randomly extracted from each video to ensure that the clips contain more concentrated audience comments and Super Chat (SC) triggers. After collecting data, this study built a multimodal corpus, which included video and text data. The total duration of video data is 3,073.488 seconds, and the text data totals 257,710 tokens, covering the streamers' spoken English output.

To ensure the data quality and the validity of subsequent analysis, the following steps were carried out in the corpus preprocessing. First, on the basis of YouTube's auto-caption, the subtitle text was transcribed through manual proofreading, with sentence-by-sentence comparison against the original video audio to correct errors related to tone words, ellipses, automatic recognition errors, etc. Next, the obtained text data were cleaned by removing meaningless filler words (e.g., *uh*, *hmm*), redundant characters (e.g., *wwww*), and background noise. Then, for multimodal interactive meaning analysis, video clips were converted into frame-by-frame visualization formats, and a video annotation table was created and synchronized with the corresponding textual timeline. The final multimodal corpus thus covers verbal, visual, and verbal-visual collaborative modal features.

2.2 Analytical Framework

2.2.1 Interpersonal Metafunction in Systemic Functional Linguistics

The current study integrates two multimodal analytical approaches, social semiotics and systemic functional grammar, and adopts the frameworks of interpersonal metafunction in systemic functional linguistics and interactive meaning in visual grammar as its analytical basis. Systemic functional linguistics (SFL), proposed by Halliday, emphasizes the social semiotic functions of language and categorizes its metafunctions into three types: ideational, interpersonal, and textual. Among these, the interpersonal metafunction focuses on how language constructs relationships between speakers and listeners, expresses attitudes, and negotiates interactional strategies in communicative processes. Based on Halliday (2004), this study analyzes the verbal modality of virtual livestreaming discourse by examining two subsystems in the interpersonal meaning framework: mood and modality.

The mood system reflects the fundamental interactive structure between speakers and listeners, indicating the type of speech function being enacted in communication. It primarily involves three types: declaratives, interrogatives, and imperatives. Declaratives are used to convey information or express opinions, interrogatives serve to elicit responses or prompt interaction, while imperatives typically issue requests or commands. The modality system, on the other

hand, encodes the speaker's subjective attitude and degree of commitment toward the proposition, expressing meanings that fall between affirmation and negation. Halliday (2004) divides modality into three levels: high, median, and low, which correspond to the speaker's strong, moderate, and weak attitudes towards probability, usuality, obligation, or inclination. Among them, probability and usuality belong to modalization, while obligation and inclination fall under modulation. For example, modal adverbs such as *definitely*, *probably*, and *maybe* or modal verbs like *can*, *must*, and *should* can convey the speaker's varying degrees of judgment towards the topic. Grounded in the interpersonal meaning system of SFL, this study selects mood and modality as core analytical indicators to examine the interpersonal interaction strategies in the verbal modality of virtual livestreaming and to explore how these verbal forms affect the interactive order and atmosphere in the virtual environment.

2.2.2 Interactive Meaning in Visual Grammar

Drawing on Halliday's theory of metafunctions, Kress and van Leeuwen (2006) proposed the visual grammar framework, which systematically explicates how images realize representational, interactive, and compositional meanings. This study focuses on the interactive meaning system within this framework, which include three key elements: contact, social distance, and attitude. These elements are employed to analyze the interactive strategies used by virtual streamers through visual resources to construct a simulated sense of intimacy during livestreams.

Contact refers to whether the character in the image engages in eye contact with the audience. Kress and van Leeuwen categorize this into two types: *demand* and *offer*. In *demand* images, the character directly gazes at the viewer, making an emotional or cognitive demand to enhance interactivity. By contrast, *offer* images position viewers as observers, as the character does not look directly at them. Social distance determined by shot scale is divided into three categories: *close-up*, *medium shot*, and *long shot*. Close-up shots convey a strong sense of intimacy by enlarging to the face or shoulders; medium shots present the part above the knees to create a moderate social distance; and long shots capturing the full body and surrounding environment tend to weaken closeness. Attitude concerns the angle of image shooting, which implies the participant's stance, including *horizontal angle* and *vertical angle*. The horizontal angle includes *frontal* and *oblique* angles, while the vertical angle involves *high*, *eye-level*, and *low* angles.

According to the original framework, it is worth noting that the vertical angle refers to the position of the camera rather than changes in eye gaze direction. However, given virtual streamers are limited by motion-capture technology in virtual livestreaming, camera angles remain basically fixed. As such, this study refines the original classification of contact and further divides *demand* into three categories based on gaze direction: *level demand*, *upward demand*, and *downward demand*. Besides, this study also supplements facial expressions like smiles and surprises as indicators of affective engagement. The detailed analytical framework is presented in Table 1.

Table 1. Analytical framework of the present study

Modality type	Dimension	Subcategory	Source
Verbal modality	Mood	declarative, interrogative, imperative	Halliday (2004)
	Modality	high, medium, low	
Visual modality	Contact	demand (level angle, downward angle, upward angle), offer	Kress and van Leeuwen (2006), self-defined
	Social distance	close shot, medium shot, long shot	
	Attitude	horizontal angle, vertical angle	
	Facial expression	smile, surprise, etc.	Self-defined

2.3 Procedure

Grounded in the multimodal discourse analysis framework above, this study conducts a systematic analysis of the collected corpus. First, the annotation process was carried out on the self-built multimodal corpus, covering both verbal and visual modalities. For the verbal modality, mood and modality features in the Vtubers' spoken discourse were identified by combining the UAM Corpus Tool 6 and manual annotation, following the interpersonal metafunction framework of systemic functional linguistics. Mood is classified into three types, declarative, interrogative, and imperative, while modality is categorized into high, medium, and low levels according to Halliday's interpersonal system.

For the visual modality, the study annotated features of contact, social distance, attitude, and facial expressions through the ELAN 6.7 multimodal annotation tool based on the interactional meaning system in Kress & van Leeuwen's (2006) visual grammar framework. In terms of contact, this study refined the original classification and divided *demand* into *level demand*, *downward demand*, and *upward demand* to capture whether streamers make direct eye contact with the audience and the direction of the gaze. Social distance is categorized according to shot scale into *close-up*, *medium shot*, and *long shot*. Attitude includes both *horizontal angle* and *vertical angle*. The facial expression category covers smiles, surprises and other common emotional expressions. All annotations align frame by frame with the corresponding video and text data to ensure precise correspondence between the verbal and visual modalities. To guarantee reliability, each annotation result was reviewed and proofread three times.

After annotation, the frequency and proportion of each verbal and visual feature were calculated through the data analysis function of UAM Corpus Tool 6 and ELAN 6.7 respectively. Finally, a Spearman correlation analysis was conducted in SPSS 27 to examine the relationships between features across the two modalities. The study also generated the heatmap via Python to reveal the distribution patterns and collaborative mechanisms of multimodal features in virtual livestreaming.

3. Results

3.1 Distribution of Verbal Features

Table 2. Distribution of mood system

Mood Clauses	N	%
Declarative clause	942	88.42
Interrogative clause	60	3.89
Imperative clause	82	7.69
Total	1080	100

This study conducts frequency statistics on the verbal modality features in virtual livestreams. In terms of the mood system, as seen in Table 2, virtual streamers primarily use declarative sentences (88.42%) during their livestream interactions. This result suggests that virtual streamers tend to communicate with their audience through direct and explicit statements. By contrast, imperative sentences account for 7.69%, while interrogative sentences are the least used, accounting for only 3.89%. This distribution feature may reflect virtual streamers' preference for maintaining the rhythm of conversations and the dominance of topics through declarative language.

Table 3. Distribution of modality system

Values of Modality	Probability	Usuality	Obligation	Inclination	Total	%
High	32	19	9	23	83	40.49
Medium	60	1	1	44	106	51.71
Low	14	2	0	0	16	7.80

Total	106	22	10	67	205	100
%	51.70	10.73	4.89	32.68	100	

Regarding the modality system, the data in Table 3 indicate that the use frequency of medium-value modality is the highest (51.71%), followed by high-value modality (40.49%), while low-value modality appears the least (7.80%). In addition, among the four modality types, probability is employed most often (51.70%). For instance, the highly frequent use of expressions of medium to high modality such as *I'm sure* and *probably* suggests that virtual streamers tend to avoid absolute or overly assertive statements during livestreams. In contrast, obligation modality is used the least (4.89%), which indicates that virtual streamers generally refrain from making compulsory requirements for their audience. In a nutshell, the verbal modality in virtual livestream interactions is characterized by a highly declarative discourse style, with a preference for medium to high-value modality expressions that carry a degree of uncertainty.

3.2 Distribution of Visual Features

Table 4. Distribution of visual features

Interactive Meanings	**Realizations**	**N**	**Time (s)**	**%**
Contact	level demand	159	2069.217	47.60
	downward demand	31	180.754	9.28
	upward demand	5	10.685	1.50
	offer	139	812.832	41.62
Social distance	close shot	5	3073.488	100
Horizontal angle	frontal angle	5	3073.488	100
Vertical angle	eye-level angle	5	3073.488	100
Facial expression	smile	105	235.793	87.50
	surprise	15	32.572	12.50

According to Table 4, the interaction of virtual streamers at the visual level is presented in a fixed close-up shot, frontal angle, and eye-level angle. In most cases, the streamer occupies the central position on screen and looks directly at the audience. The analysis of contact shows that virtual streamers primarily interact with viewers through demand gaze, with level demand accounting for the highest proportion (47.60%) and the longest duration (2069.217). This indicates that virtual streamers mainly simulate eye contact with the audience through direct, level gazes. But it is found that the proportion of downward demand (9.28%) and upward demand (1.50%) is obviously low, suggesting that their visual interaction strategies tend to emphasize equality in social relations. For facial expressions, smiles dominate at 87.50%, which constitutes the main emotional expression for virtual streamers, while expressions such as surprises (12.50%) are typically used to enhance the performative and dramatic aspects of the livestream through dynamic facial changes.

3.3 Spearman Correlation Between Verbal and Visual Features

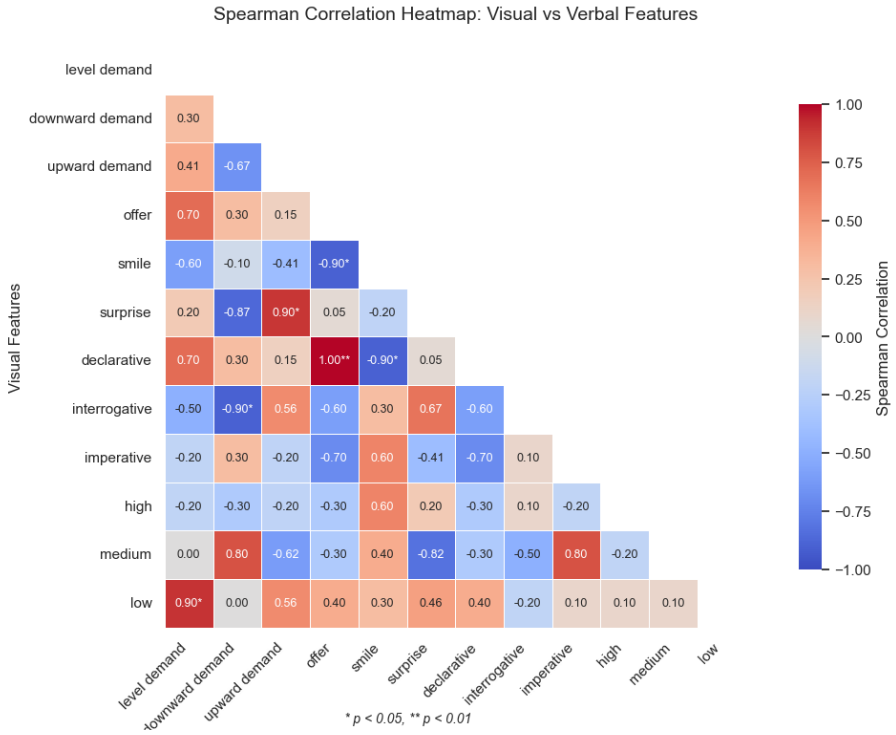

Figure 1. The heatmap for spearman correlations between verbal and visual features

In the analysis of the relationship between the verbal and visual layers, this study employs Spearman's rank correlation coefficient test to examine the correlations between multimodal features in virtual livestreams. As shown in Figure 1, several significant correlations are identified between verbal and visual features. There is a strong positive correlation between the streamer's offer gaze and the use of declarative sentences (r = 1.000, p = 0.000), indicating that virtual streamers often cooperate with an averted gaze to create a relaxed and casual interactive atmosphere when making information statements. In addition, a significant positive correlation is observed between level demand and low-value modality expressions (r = 0.900, p = 0.037). In other words, virtual streamers tend to maintain a level gaze when expressing uncertainty or adopting a more tentative attitude.

On the contrary, smiles show a significant negative correlation with declarative sentence structures (r = -0.900, p = 0.037). This implies that virtual streamers may smile less at the same time when making declarative statements possibly for the sake of maintaining seriousness or authority while delivering information. Similarly, downward demand is also significantly negatively correlated with the use of interrogative sentences (r = -0.900, p = 0.037), indicating that virtual streamers rarely use a downward gaze when posing questions to the audience, likely in order to avoid a sense of dominance. These findings reveal the coordinated distribution features of verbal and visual modalities in virtual livestreams.

4. Discussion

4.1 Constructing Intimacy Through Language: Mood and Modality in VTubers' Performances

The intimacy constructed by virtual streamers through the verbal modality centers on the strategic coordination of the mood system and modality system. From the perspective of systemic functional linguistics, the verbal features of virtual streamers exhibit the coexistence of highly formulaic and emotional patterns, and this contradictory unity forms the very foundation of their sense of intimacy.

In terms of the mood system, virtual streamers predominantly employ declarative sentences, which is consistent with Shi (2024)'s findings. These declaratives not only serve to transmit information, express opinions, and respond to live comments but also act as carriers of emotional projection. For example, in this study's corpus, expressions like *I remember I read the first two books* and *I want to go back to university* focus on the streamer's personal experiences and subjective feelings, avoiding the language pressure that imperative statements might

impose on the audience. It is evidenced that streamers have established an interactional framework centering on sharing rather than informing. Such intimate narrative behavior invites the audience into streamers' emotional world through self-sharing to bring the psychological distance closer to each other. It also indicates that they attempt to use declarative sentences to control the rhythm of livestreams and refocus attention on themselves in a timely manner (Qin, 2022). It is worth noting that many declaratives can be regarded as stylized templates of intimate speech even if they seem to be personal habitual expressions. For instance, virtual streamers often repeatedly use *I'm so proud of you*, *I'm really proud of you*, and *I'm glad that you enjoy hearing my voice* when responding to Super Chats. These expressions evidently form a standardized model of one-way emotional output through the fixed collocation of the subject *I* with affective adjectives like *proud* or *glad*. Such declaratives are frequently accompanied by the accumulation of adverbs of degree (e.g., *so*, *really*, *absolutely*), reflecting how streamers compensate for the affective limitations of their virtual personas in emotional communication via quantitative marking of language intensity.

The use of the modality system in virtual livestreams also displays clear intimacy-oriented features. Medium-value modality appears most frequently, with probability being the most common. This type of modality is often used to convey speculation, assumptions, wishes and other subjective judgments, which help avoid absolute assertions and create space for the audience to agree, supplement, or refute. It can also mitigate the potential hierarchy of discursive power between the streamer and audience. For instance, when discussing the relationship between game IPs, a streamer's vague statement like *most likely that it is a spiritual successor* conveys professional knowledge while avoiding the risks of assertive claims. Virtual streamers also frequently use expressions such as *I think* and *I figure* as typical markers of probability that attribute the language responsibility to personal experience, so as to avoid invasive judgments on the audience's own knowledge. This modality construction mechanism can effectively maintain the equal dialogue relationship and encourage the audience's response and sympathy based on shared perspectives, thereby realizing the dynamic construction of interactive intimacy.

Besides, the most innovative feature of virtual streamer discourse lies in the nested structure of modality expressions. In response to wishes about a medical exam, for example, the streamer built a complex modality chain: *You're going to pass the exam, or you are going to be diagnosed as completely healthy. And if you are not, you will heal and everything will be fine*. This turn of speech contains four consecutive expressions of modality, forming an emotional reinforcement network that ensures the speech itself possesses emotional effectiveness regardless of the actual result. This confirms the view of Gong and Sun (2025) that the use of affective language can enhance the audience's emotional resonance.

4.2 Creating Equality Through Visual Modality: Visual Affection in Virtual Livestreaming

The visual presentation of virtual livestreaming is highly standardized and patterned. It mainly realizes continuous visual engagement with the audience and the imitation of the interactive situation through the use of fixed close-up shots, frontal angles, and eye-level angles, which conforms to the results of Shi (2024). Such visual design effectively dispels the sense of distance between the streamer's virtual identity and real-time interaction, offering viewers an immersive experience of face-to-face-like communication during watching.

Figure 2. The screenshot of the VTuber's livestreaming

The consistent use of fixed close-up shots, frontal angles, and eye-level angles ensures that facial expressions and gaze direction of virtual streamers always occupy the center of the picture through deliberately imitating the perspective of video calls, which visually frames the avatar within the visual range of intimate distance. In practice, the streamer image usually occupies about 60-70% of the area in the center of the picture and the background matches the avatar's image setting (see Figure 2). Compared with traditional livestreams, this visual arrangement not only enhances the subjectivity of streamers but also serves a prominent emotional focus function that can amplify micro-expressions and emotional changes for the audience to detect during the interaction. However, considering the realistic technical factors of virtual livestreaming, such a fixed visual design is closely related to the actual space of the human performer in front of the desktop computer. Due to the limitations of motion capture technology, it is rare for virtual streamers to fully display whole-body movements or the interactions with the physical world (Lu et al., 2021).

In livestreaming, the demand gaze constitutes the most fundamental visual pattern in virtual streamer interactions, with level demand accounting for as high as 47.60% and sustaining the longest duration. The level demand is commonly associated with equality, friendliness, and closeness. Virtual streamers' eye movements can be controlled by the capture device to look directly at the audience in a straight-facing way, forming an "eye-to-eye" interaction. Moreover, in interactive moments such as replying bullet comments, the system often triggers specific eye animations, such as the effect of pupil dilation or blinking, to simulate nonverbal reactions when receiving positive feedback. Through this eye contact with the audience, virtual streamers construct a silent yet emotional interaction atmosphere, which makes viewers feel the emotional value of "being looked at" and "being listened to". Therefore, viewers' psychological connection and trust in the streamer can be established while also dissociating the implicit hints of power.

Figure 3. The screenshot of a smiling VTuber

In addition, smile is the most frequently employed facial expression in virtual livestreams. As a nonverbal symbol generally interpreted in a positive way, smile serves to alleviate social tension and foster intimacy. Virtual streamers use continuous smiles to convey a friendly and welcoming attitude to viewers. The high frequency of smiles is likely to elevate the emotional atmosphere and also effectively mitigates the interaction barriers inherent in virtual identity, so that viewers are allowed to experience simulated real-time conversations. For example, when receiving rewards or comments from the audience, virtual streamers are typically accompanied by smiling expressions and a frontal and eye-level angle, which can instantly narrow the emotional distance between themselves and the audience (see Figure 3). However, as Lu et al. (2021) point out, virtual livestreams is considered a hybrid form of integration of the virtual and real world, where technical glitches such as avatar clipping or persistent smiling due to model design may result in moments where the avatar appears to be smiling, but the performer may in fact be looking down, or the avatar might be designed to maintain a constant smile to perpetuate a sense of warmth and approachability.

4.3 Multimodal Coordination: Virtual Intimacy and Affective Consumer Mobilization

Through the results of the Spearman correlation analysis, it can be found that there is a significant coordination pattern between the language

selection and visual presentation of virtual streamers. And the interactive atmosphere created by different combination strategies can directly affect the audience's emotional experience and perceptions of intimacy.

First of all, a strong positive correlation is found between the virtual streamer's offer gaze and the use of declarative sentences. This indicates that when virtual streamers avert their gaze from direct eye contact, they are more inclined to use declarative sentences to share personal experiences, express opinions, or respond to comments. This combination strategy reduces the pressure and directness of speech, which makes the interaction appear more casual to facilitate a relaxed and intimate atmosphere. For instance, in terms of some scenes like where the streamer say *I wanted to watch cartoons with a plot with excitement you know what I mean*, the casual words and non-direct gaze place the audience within an informal and friendly communication environment where intimacy is more readily generated.

There is also a significant positive correlation observed between level demand and low-value modality. When engaging in direct level gaze with viewers, virtual streamers are customary to use expressions implying possibility or speculation which lower the compulsiveness and evaluation of language. This strategy weakening verbal pressure may further promote the audience's willingness to participate in the interactive process, and even stimulate tipping behavior (Qin, 2022). It is crucial that this verbal-visual coordination is especially evident when Super Chats or tipping notifications appear. Whenever a viewer sends a tip or gift and the system prompt appear, virtual streamers often immediately switch to a level demand, combine it with a smile, and use highly emotive thank-you words and nicknames, such as *thank you so much for the five gifted membership thank you so much thank you*. Although the modality value remains fixed, the intensity of emotional expression is dramatically enhanced, the smile is emphasized, and the gaze shifts instantly from offer to level demand. This instantaneous multimodal linkage produces an effect of "you are paid special attention", which effectively activates viewers' emotional identification mechanism. This mimic intimacy assumes an important consumption-driven function under the commercial logic of virtual livestreaming, that is, the audience is prone to continuous consumption behavior under the induction of virtual closeness. This suggests that the multimodal intimacy strategy employed by virtual streamers function not only as tools for emotional interaction but also as integral components of the consumption logic supported by emotional labor in the digital livestream economy. In this process, viewers seem to be autonomous consumers, yet in fact they remain entranced by the illusion of subjectivity constructed through images (Yang, 2011).

In contrast, there exists a significant negative correlation between smiles and declarative sentences. It means that when virtual streamers smile, they are more likely to cooperate more with highly interactive utterances such as questions, exclamations, or emotional calls to mobilize the audience's response and maintain a lively interaction rhythm. For example, a smiling streamer asks, "what did you do and what do you want to do if you have a time machine to go back to high school life," which motivates the audience to participate through real-time facial expressions and interactive speech. While strengthening intimacy, it also reinforces the audience's sense of identity as a consumer subject in the consumption-oriented context. What's more, downward demand negatively correlates with the use of interrogative sentences because a downward gaze conveys authority and distance, which is not conducive to stimulating viewers' response. Yet, as highly interactive utterances, interrogative sentences need to rely on a relaxed and equal communication environment. In brief, by leveraging the multimodal intimacy mechanism, virtual streamers shape their personas into objects that can be emotionally projected and psychologically identified by the audience, and continue to stabilize the subject-object structure of the livestreaming economy.

5. Conclusion

Based on the interpersonal metafunction of systemic functional linguistics and the interactive meaning in visual grammar, this study using a corpus-based approach conducts a multimodal discourse analysis of virtual livestreams to explore how virtual streamers construct intimacy through the verbal, visual modality, and the coordinated interplay of both. The main findings are as follows: (1) at the verbal level, virtual streamers foster a friendly and equal interactive atmosphere by frequently using declarative sentences, medium-value modality, and probability mood speech; (2) at the visual level, streamers consistently employ frontal, eye-level

close-up shots, level demand and smiles, effectively creating an intimate experience like face-to-face communication; (3) at the level of multimodal coordination, there is a strong positive correlation between offer gaze and declarative sentences, and between level demand and low-value modality; in contrast, a significant negative correlation is found between smiling expressions and declarative sentences, also between downward demand and interrogative sentences; (4) virtual streamers release emotional intimacy signals through multimodal means, which bolsters the audience's awareness participation and implicitly mobilizes emotional engagement to encourage continuous consumption. It thus becomes evident that intimacy in virtual livestreaming is not merely a form of emotional interaction, but a kind of simulated emotional labor accompanied with consumption mobilization.

This study, however, has certain limitations. On the one hand, the sample data mainly derives from English-speaking virtual livestreams. Cross-cultural differences in language and norms may affect the universality of multimodal interaction strategies, which can be extended to the livestream content of virtual streamers in different language and cultural contexts. On the other hand, the size of samples in this study is relatively limited, which may result in insufficient data coverage. However, this study enriches the understanding of the multimodal interaction mechanisms in the virtual livestreaming industry, and also provides a new perspective for the phenomenon of emotional economy in digital consumption culture.

References

Chen, H., Shao, B., Yang, X., Kang, W., & Fan, W. (2024). Avatars in live streaming commerce: The influence of anthropomorphism on consumers' willingness to accept virtual live streamers. *Computers in Human Behavior*, *156*, 108216. https://doi.org/10.1016/j.chb.2024.108216.

Chen, Y., Li, L., & Zhou, W. (2025). Impact of viewer-streamer-content congruence on users' behavioral intention in virtual streaming: The moderating effect of role-playing. *Electronic Commerce Research and Applications*, *70*, 101492. https://doi.org/10.1016/j.elerap.2025.101492.

Gao, W., Jiang, N., & Guo, Q. (2025). How cool virtual streamer influences customer in live-streaming commerce? An explanation of stereotype content model. *Journal of Retailing and Consumer Services*, *82*, 104139. https://doi.org/10.1016/j.jretconser.2024.104139.

Gong, X., & Sun, P. (2025). Can virtual streamers express emotions? Understanding the language style of virtual streamers in livestreaming e-commerce. *Journal of Retailing and Consumer Services*, *82*, 104148. https://doi.org/10.1016/j.jretconser.2024.104148.

Gu, X., & Catalano, T. (2022). Representing transition experiences: A multimodal critical discourse analysis of young immigrants in children's literature. *Linguistics and Education*, *71*, 101083. https://doi.org/10.1016/ j.linged.2022.101083.

Halliday, M. A. K. (2004). *An introduction to functional grammar*. Routledge.

Huang, H., Blommaert, J., & Van Praet, E. (2020). "OH MY GOD! BUY IT!" a multimodal discourse analysis of the discursive strategies used by Chinese ecommerce live-streamer Austin Li. In C. Stephanidis, G. Salvendy, J. Wei, S. Yamamoto, H. Mori, G. Meiselwitz, F. F.-H. Nah, & K. Siau (Eds.), *HCI International 2020 – Late Breaking Papers: Interaction, Knowledge and Social Media* (pp. 305–327). Springer International Publishing.

Kenalemang-Palm, L. M. (2023). The beautification of men within skincare advertisements: A multimodal critical discourse analysis. *Journal of Aging Studies*, *66*, 101153. https://doi.org/10.1016/j.jaging.2023.101153.

Kress, G., and van Leeuwen, T. (2006). *Reading images: The grammar of visual design*. (2nd ed). Routledge.

Liu, H., & Zhao, J. (2025). VTuber attractiveness and its effect on viewer gifting. *International Journal of Human–Computer Interaction*, 1–16. https://doi.org/ 10.1080/10447318.2025.2465866.

Lu, Z., Shen, C., Li, J., Shen, H., & Wigdor, D. (2021). More kawaii than a real-person live streamer: Understanding how the otaku community engages with and perceives virtual YouTubers. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/ 3411764.3445660.

Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. Routledge.

O'Halloran, K. L. (2005). *Mathematical discourse: Language, symbolism and visual images*. Continuum.

O'Toole, M. (1994). *The language of displayed art*. Leicester University Press.

Peng, J., & Chen, J. (2024). Research on the body aesthetics of virtual hosts and the mechanisms of symbolic consumption: A Baudrillardian perspective. *Journal of Southwest Minzu University (Humanities and Social Sciences Edition)*, *45*(11), 144–153. https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFDLAST2025&filename=XNZS202411016.

Playboard. (2022. Dec 16). Most super chatted. *Playboard*. https://playboard.co/en/ youtube-ranking/most-superchatted-all-channels-in-worldwide-total.

Putland, E., Chikodzore-Paterson, C., & Brookes, G. (2025). Artificial intelligence and visual discourse: A multimodal critical discourse analysis of AI-generated images of "dementia". *Social Semiotics*, *35*(2), 228–253. https://doi.org/10.1080/10350330.2023.2290555.

Qi, F. (2022). A study of traditional artistic visual narratives in original picture books based on multimodal discourse analysis. *Publishing Journal*, *30*(3), 43–50. https://doi.org/10.13363/j.publishingjournal.20220517.012.

Qin, Y. (2022). Virtual body and semiotic consumption: A study of interactive mechanisms in virtual streamer livestreams. *Application Research*, *8*(2), 39–44. https://doi.org/10.16604/j.cnki.issn2096-0360.2022.02.006.

Shi, Y. (2024). Research on live broadcast discourse of network virtual uploader from a multimodal perspective. [Unpublished Master's thesis]. Shandong University.

Sun, N. (2024). A multimodal analysis on the interactive meaning of e-commerce live-streaming discourse. [Master's thesis, Harbin Engineering University]. https://doi.org/10.27060/d.cnki.ghbcu.2023.000833.

Tan, Y. H., & Greene, B. R. (2025). Can a 2D shark girl be an influencer? Uncovering prevailing archetypes in the virtual entertainer industry. *Journal of Business Research*, *186*, 114951. https://doi.org/10.1016/j.jbusres.2024.114951.

Wang, Y., & Pan, D. (2022). Research on multimodal interaction in Taobao live streaming. *Chinese Journal of Language Policy and Planning*, *7*(3), 34–46. https://doi.org/10.19689/j.cnki.cn10-1361/h.20220303.

Yang, H. (2011). Hyperreality, simulations and implosion — three key concepts in the late Jean Baudrillard's thought. *Jiangsu Social Sciences*, (4), 14–21. https://doi.org/10.13858/j.cnki.cn32-1312/c.2011.04.019.

Zhao, X. (2022). Ecological discourse analysis based on multimodal metaphor scenarios: a case of bioenergy political cartoons. *Foreign Languages in China*, *19*(6), 60–69. https://doi.org/10.13564/j.cnki.issn.1672-9382.2022.06.003.

Zhu, Y. (2007). Theory and methodology of multimodal discourse analysis. *Foreign Language Research*, *5*, 82–86. https://doi.org/10.16263/j.cnki.23-1071/h.2007.05.034.

Journal of Linguistics and Communication Studies
ISSN 2958-0412
www.paradigmpress.org/JLCS
Volume 5 Number 1 March 2026

Dimensional Differences in English Speaking Anxiety Across Physical and Online Contexts: A Study of Chinese EFL Undergraduates

Yang Xiaoying[1,2], Souba Rethinasamy[1] & Joseph Ramanair[1]

[1] Faculty of Education, Language and Communication, Universiti Malaysia Sarawak, Sarawak, Malaysia
[2] Weifang Institute of Technology, Shandong, China
Correspondence: Yang Xiaoying, Faculty of Education, Language and Communication, Universiti Malaysia Sarawak, Malaysia; Weifang Institute of Technology, Shandong, China.

doi:10.63593/JLCS.2026.03.05

Abstract

While previous research has established that online learning reduces language anxiety, less is known about how different dimensions of speaking anxiety respond to the shift from physical to digital classrooms. Grounded in foreign language classroom anxiety theory and situated within a communication studies framework, ESA is deconstructed into three dimensions: Communication Apprehension, Fear of Negative Evaluation, and Test Anxiety. This study employed survey design with 507 undergraduates in Shandong Province using parallel questionnaires for both contexts and paired samples t-tests. The findings show a significant contextual difference in overall ESA levels. Online classrooms (M=2.291) exhibit markedly lower anxiety than physical classrooms (M=2.791), with a mean difference of 0.500; Dimensional analysis reveals varying contextual sensitivities. Fear of Negative Evaluation shows the greatest contextual disparity (mean difference=0.579), descending from moderate to low levels across contexts. Test Anxiety follows closely with similar reduction patterns. Communication Apprehension exhibits the smallest contextual difference (mean difference=0.360), maintaining consistent moderate levels across contexts. This indicates its stability as a cognition-based construct; The digitally mediated environment proves particularly effective in alleviating social-evaluative anxiety linked to audience presence and immediate judgment. Yet it does little to reduce ability-focused cognitive anxiety. This study reveals the complex relationship between communication medium, learner psychology, and language performance. It provides empirical evidence for implementing "context-aware, dimension-sensitive" pedagogical frameworks in technology-enhanced language education. The study contributes to both linguistic and communication scholarship by demonstrating how media characteristics reconfigure specific affective components of language learning.

Keywords: English speaking anxiety, communication context, Chinese EFL learners, affective factors, blended learning

1. Introduction: The Intersection of Language Anxiety and Communication Media

Many Chinese EFL learners remain reluctant to speak English despite years of formal instruction—a phenomenon colloquially termed "Dumb English" (哑巴英语, literally "mute English"). This is more than a pedagogical challenge; it is a complex issue where language skill, emotion, and learning context meet (He, 2018; Li & Liang, 2022). This silence, students' unwillingness to speak despite years of English classes, stems largely from English Speaking Anxiety (ESA) (Horwitz et al., 1986; Liu & Jackson, 2008; Zhao et al. 2025). Many studies have viewed ESA through psychological factors like motivation or personality. Yet the rapid shift to online learning, accelerated by the pandemic, calls for rethinking anxiety through a different lens: communication context and media environment.

Physical and online classrooms are fundamentally different learning environments. The traditional physical classroom is a high-context (Hall, 1976), co-present (Goffman, 1963) environment rich in nonverbal cues, immediate feedback loops, and heightened social presence (Basille et al., 2025; Short et al., 1976; Zhang & Mohamed, 2024). This context makes the speaker more visible and the audience more reactive. It creates what communication scholars term a "high-social-presence" environment (Short et al., 1976). Conversely, the synchronous video-conferencing classroom offers potential visual anonymity, changes how people take turns, separates participants physically, and reduces social cues. These divergent environments create distinct "affective architectures", contextual setups that shape how learners feel and act (Chen & Zhang, 2022).

Studies show that online learning environments often reduce language anxiety (Alqarni, 2021; Lee & Chen Hsieh, 2019). But recent research adds nuances. Bárkányi and Brash (2025) found that while online settings can lower certain anxieties, features like breakout rooms may heighten emotional responses. Similarly, Wei (2025) reported that Chinese EFL learners feel less anxious online than in person, though teacher-related factors mattered less than expected.

However, few studies have examined how different types of speaking anxiety vary between physical and online classrooms (Alqarni, 2021; Lee & Chen Hsieh, 2019). The theoretical implications of such contextual differences also remain largely unexplored. Identifying which anxiety dimensions are most context-sensitive can inform the design of blended learning environments (Chen & Zhang, 2022).

Based on the above framework, this research addresses the following questions:

RQ1. How do overall English Speaking Anxiety (ESA) levels differ between physical and online classroom contexts among Chinese EFL undergraduates?

RQ2. How do Communication Apprehension, Fear of Negative Evaluation, and Test Anxiety respond to the shift from physical to online classrooms?

RQ3. What do these patterns reveal about Communication Apprehension, Fear of Negative Evaluation, and Test Anxiety?

By addressing these questions, this study aims to bridge communication media theory with language learning psychology and to offer practical guidance for affectively aware teaching in digital and blended spaces.

2. Theoretical Framework and Literature Review

2.1 Conceptualizing Classrooms as Distinct Communication Contexts

Physical and online classrooms are fundamentally different communication contexts, not just alternative delivery modes (Chen & Zhang, 2022). Each has its own media logic and interactional affordances. This view draws on communication and media studies to understand how learning environments shape learner emotion (Short et al., 1976; Daft & Lengel, 1986; Walther, 1992).

The traditional physical classroom is a high-social-presence, high-context environment (Hall, 1976; Garrison et al., 1999). Communication here is embodied and happens in shared physical space (Goffman, 1963). It offers a rich range of cues: paralinguistic signals (tone, volume, pace), nonverbal behaviors (facial expressions, gestures, posture, eye gaze), and immediate audience feedback (nods, smiles, confused looks, whispers). The speaker is physically central and highly visible to everyone present. This setup maximizes what Short et al. (1976) called "social presence": the sense of another person being "there" in the interaction (Richardson et al., 2017). High social presence typically heightens awareness of and sensitivity

to others' evaluations.

In contrast, the synchronous video-conferencing classroom (e.g., Zoom, Tencent Meeting, or DingTalk) is a digitally-mediated, leaner-cue environment (Kiesler et al., 1984; Sproull & Kiesler, 1986). Often called "virtual face-to-face," its media features create different interactional conditions (Walther, 1996). The interface reshapes communication: participants appear as tiles in a grid, spatial co-presence fades, and eye contact becomes misaligned (looking at the camera, not at others' eyes). Nonverbal cues are often limited to the upper body and face, framed by a screen (Hrastinski, 2008). Key features like "speaker view" (which highlights the current speaker) and the option to turn off one's camera fundamentally change communication dynamics (Chen & Zhang, 2022; Suler, 2004). These features can greatly reduce the audience's salience as an immediate, judging presence, creating the 'online disinhibition effect' (Bargh & McKenna, 2004; Tu & McIsaac, 2002).

This reframing lets established communication principles: social presence, media richness, and social information processing, generate specific predictions about affect in language learning (Daft & Lengel, 1986; Short et al., 1976; Walther, 1992; Timilsina, 2025). The idea is that the digital context's reduced cues and lower social presence will not uniformly dampen all anxieties. Instead, it will selectively ease those anxieties most tied to social-evaluative concerns (Walther, 1996; Suler, 2004; Peschka et al., 2025).

2.2 The Multidimensional Nature of Speaking Anxiety and Its Hypothesized Contextual Sensitivity

This study adopts and extends Horwitz et al.'s (1986) multidimensional model of Foreign Language Classroom Anxiety. It focuses specifically on the manifestation of anxiety in speaking (ESA). The three dimensions are treated as theoretically distinct constructs with different primary antecedents. They should therefore demonstrate different sensitivities to changes in communication context.

2.2.1 Fear of Negative Evaluation: High Contextual Sensitivity

This dimension is conceptualized as apprehension about others' assessments (Watson & Friend, 1969). It is associated with avoidance of evaluative situations and anticipation of negative judgments. Its core antecedent is the perceived presence and salience of an evaluating audience (He, 2018; Liu, 2006). This dimension is hypothesized to be highly sensitive to communication context because the contextual variables differ dramatically in the richness of social-evaluative cues they provide (Horwitz et al., 1986; Mak, 2011).

The physical classroom, as a high-social-presence context, provides a continuous stream of such cues. The speaker sees peers' faces, catches their glances, hears their reactions, and feels their physical presence. This makes the evaluating audience highly salient and the possibility of negative judgment feel immediate and tangible (Liu & Jackson, 2008; Tsui, 1996). Conversely, the digitally mediated classroom significantly filters these cues. When cameras are off, the audience becomes invisible. Even with cameras on, participants appear in small tiles, often with limited video quality. The spotlight of speaker view can paradoxically make the audience less visually prominent (Chen & Zhang, 2022). This leaner cue environment likely creates a psychological buffer or protective cloak against immediate social judgment. It reduces the felt salience of evaluation (Alqarni, 2021; Yaniafari & Rihardini, 2022). Recent research by Cheng and Sun (2025) provides direct empirical support for this claim, demonstrating that EFL college students experience significantly less cognitive and somatic speaking anxiety in synchronous online learning compared to traditional classroom settings. Therefore, the largest reduction in this dimension is predicted when moving from physical to online contexts.

2.2.2 Communication Apprehension: Moderate Contextual Sensitivity

This dimension encompasses the fear or anxiety associated with real or anticipated communication with others (McCroskey, 1977, 1984). In an L2 context, it includes both social-performance anxiety (the fear of being "on stage," closely related to Fear of Negative Evaluation) and cognition-based anxiety. This cognition-based anxiety arises from the real-time linguistic and cognitive demands of L2 speech production (MacIntyre, 1995; Woodrow, 2006).

This dual nature leads to the hypothesis of moderate contextual sensitivity. The online context may alleviate the social-performance aspect by reducing the sense of being physically "on stage" (Horwitz et al., 1986). However, the core cognitive challenge persists regardless of the communication medium. This challenge includes the pressured retrieval of vocabulary, the online

construction of syntax, and the monitoring of pronunciation and fluency under real-time constraints (MacIntyre & Gardner, 1994). The anxiety stemming from perceived linguistic inadequacy or processing difficulty is largely internally generated. It is less dependent on external audience cues (Gregersen & Horwitz, 2002). Furthermore, digital mediation introduces new anxieties: concerns about audio clarity, "talk-over" confusion due to network latency, or the cognitive load of managing technology alongside the language task (Resnik et al., 2022; Wang et al., 2021). While cognitive and somatic aspects of speaking anxiety decrease online, the behavioral dimension, reflecting actual avoidance of speaking, shows no significant difference between physical and online contexts (Cheng & Sun, 2025). This finding underscores the persistence of the core behavioral component of Communication Apprehension across environments. Thus, while some components of Communication Apprehension may decrease online, others may persist or even be exacerbated. This results in a predicted moderate overall reduction (Li et al., 2020).

2.2.3 Test Anxiety: Relative Contextual Stability

This dimension is defined as a form of performance anxiety stemming from the fear of failure in evaluative situations (Spielberger, 1983; Horwitz et al., 1986). Its primary trigger is the perceived stakes and consequences of the evaluation. It is not the specific medium through which the evaluation is delivered (Aida, 1994; He, 2018).

Relative contextual stability is hypothesized for this dimension based on traditional, Western-centric models (Li et al., 2020; Zeng & Liu, 2012). From this perspective, Test Anxiety is tied to the perceived stakes of assessment—a poor grade, negative feedback, or academic consequence—rather than its communication mode (Çağatay, 2015; Huang & Hwang, 2013). However, this hypothesis requires critical examination in the digitally-mediated and culturally specific context of Chinese EFL learning. First, the online environment introduces novel "technological reliability anxiety" (e.g., platform failure, poor connection), which may offset any affective benefits from reduced social presence (Wang et al., 2021; Resnik et al., 2022). Second, in China's collectivist culture, even formal tests carry a significant "face" (social standing, 面子) concern, potentially making Test Anxiety more context-sensitive than Western frameworks predict. Therefore, the contextual sensitivity of Test Anxiety remains an open empirical question.

3. Methodology

3.1 Participants and Communication Context Experience

Participants were 507 non-English major undergraduates recruited from 43 universities across Shandong Province, China. A key methodological feature was the selection criterion: all participants must have had substantive, recent, and comparable experience with both target communication contexts within the same academic year. Specifically, participants were required to have completed at least one full semester course in traditional, face to face university English classes. They were also required to have completed one full semester course in synchronous online English classes conducted via mainstream platforms (e.g., Tencent Meeting, Zoom, or DingTalk). This ensured valid within subjects comparisons. Each participant served as their own control across contexts, eliminating inter-individual differences as an alternative explanation for contextual effects.

The final sample comprised 208 males (41.0%) and 299 females (59.0%), with a mean age of 19.5 years (SD = 0.89). Participants represented diverse academic disciplines, including Engineering (24.9%), Arts (34.7%), Management (12.4%), Sciences (7.3%), Medicine (6.5%), and others. This diversity ensured the findings were not limited to students from specific academic backgrounds.

3.2 Instrument: A Context-Parallel Assessment Tool

A modified English Speaking Anxiety Questionnaire was developed based on the Foreign Language Classroom Anxiety Scale (FLCAS) by Horwitz et al. (1986). The principal innovation was its parallel design for both contexts. Two versions of the questionnaire were created: one referencing the physical classroom context and an identical one referencing the online classroom context. Instructions and every item were identically worded except for the specifying phrase "in physical class" or "in online class."

The questionnaire contained 46 items total per context, measuring the three theoretical dimensions:

Communication Apprehension (16 items): e.g., "In physical/online class, I get nervous when I

speak English"; "In physical/online class, I start to panic when I have to speak without preparation."

Fear of Negative Evaluation (16 items): e.g., "In physical/online class, I am afraid that other students will laugh at me while I am speaking English"; "In physical/online class, I feel self conscious about speaking English in front of other students."

Test Anxiety (14 items): e.g., "In physical/online class, I worry about the consequences of failing my English speaking class"; "In physical/online class, I am not at ease during English speaking tests."

All items employed a four-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree). The forced-choice format without a neutral midpoint was chosen to reduce central tendency bias and promote more decisive responses. This increased the discriminative capacity of the data (Chyung et al., 2017). For interpretation, mean scores were categorized into three anxiety levels following the framework established in comparable Chinese EFL studies (Liu & Jackson, 2008): scores below 2.4 indicate Low Anxiety, scores between 2.4 and 3.2 indicate Moderate Anxiety, and scores above 3.2 indicate High Anxiety. Given the four-point scale with no neutral midpoint, the theoretical neutral value is 2.5. Thus, the threshold of 2.4 represents the lower boundary of the moderate range. This classification system was applied consistently across both physical and online contexts to enable direct comparison.

The instrument was developed and validated following the model proposed by Meerah et al. (2012). This process included expert review for content validity and a pilot study (N=43). Psychometric analysis from the main study data (N=507) demonstrated excellent reliability: Cronbach' s α was 0.96 for the full scale. The sub-dimension α coefficients were 0.94 for Communication Apprehension, 0.93 for Fear of Negative Evaluation, and 0.92 for Test Anxiety. These values indicate "very good" to "excellent" internal consistency according to DeVellis & Thorpe's (2021) criteria.

3.3 Data Collection and Analytical Procedures

Data collection was administered online over a two-week period using the Wenjuanxing platform, a survey tool validated for academic research in China. Participants completed both context versions of the questionnaire in a single session. The context order was randomized to control for potential order effects.

Data were analyzed using SPSS 23.0 with the following sequential procedure:

Descriptive Statistics: Means and standard deviations were calculated for the overall ESA score and for each of the three dimension scores in both the physical and online contexts.

Anxiety Level Categorization: Following the interpretive framework used in comparable Chinese EFL studies (Liu & Jackson, 2008), mean scores were categorized into three levels: < 2.4 = Low Anxiety; 2.4-3.2 = Moderate Anxiety; > 3.2 = High Anxiety.

Paired Samples t-tests: To test for significant mean differences between the physical and online contexts, paired samples t-tests were conducted for (a) the overall ESA score, and (b) each of the three dimension scores. The significance level was set at $\alpha = 0.05$.

Analysis of Differential Sensitivity: The primary indicator for testing the core hypothesis of differential contextual sensitivity was the magnitude of the mean difference (M_diff) for each dimension. By comparing these M_diff values across dimensions, the analysis directly assessed which dimensions showed greater or lesser change across communication contexts.

This analytical approach allowed the study to first confirm the basic contextual effect (Are anxiety levels different online vs. offline?). It then addressed the central research question regarding differential dimensional sensitivity (Are some types of anxiety more affected by the context change than others?).

4. Results

4.1 The Foundational Contextual Effect: Overall Anxiety Reduction Online

The initial analysis confirmed a substantial, statistically significant contextual effect on overall ESA. As presented in Table 1, students' mean ESA score in the physical classroom context was 2.791 (SD = 0.554). This falls into the "Moderate Anxiety" category. In stark contrast, the mean ESA score in the online classroom context was 2.291 (SD = 0.588), categorizing as "Low Anxiety." The mean difference between contexts was 0.500. This represents a substantial shift in affective experience.

Table 1. Overall English Speaking Anxiety by Communication Context (N=507)

Communication Context	Mean (M)	Standard Deviation (SD)	Anxiety Level	Contextual Mean Difference
Physical Classroom	2.791	0.554	Moderate	0.500
Online Classroom	2.291	0.588	Low	—

A paired samples t-test rendered this difference unequivocally significant: $t(506) = 17.287$, $p < .001$. This confirms that the communication context, the medium through which instructional interaction occurs, exerts a powerful influence on learners' aggregate affective state. The digitally mediated environment is associated with significantly lower overall speaking anxiety among this population of Chinese EFL undergraduates.

4.2 Revealing Differential Sensitivity: A Dimensional Analysis

While the overall reduction is noteworthy, the dimensional analysis reveals a more nuanced and theoretically informative pattern. Table 2 presents the means, levels, and statistical comparisons for each of the three ESA dimensions across the two contexts.

Table 2. Dimensional Analysis of ESA Across Physical and Online Communication Contexts

Anxiety Dimension	Physical Context (Mean Level)	Online Context (Mean Level)	Contextual Mean Difference	t-value	p-value	Contextual Sensitivity Rank
Fear of Negative Evaluation	2.793 (Mod.)	2.214 (Low)	0.579	14.92	<.001***	1 (Highest)
Test Anxiety	2.827 (Mod.)	2.254 (Low)	0.573	14.743	<.001***	2
Communication Apprehension	2.761 (Mod.)	2.401 (Mod.)	0.360	11.018	<.001***	3 (Lowest)

The results demonstrate several critical findings:

First, all three dimensions showed statistically significant reductions in the online context (all p-values < .001). However, the magnitude of change, indexed by the mean difference (M_diff), varied dramatically across dimensions.

Second, Fear of Negative Evaluation and Test Anxiety both transitioned from the moderate anxiety range in the physical context to the low anxiety range in the online context. This represents a categorical shift in affective experience for these dimensions. In contrast, Communication Apprehension remained firmly within the moderate anxiety range in both contexts. Despite a statistically significant reduction, its level did not cross the threshold from moderate to low. This indicates a more persistent, context-resistant form of anxiety.

Third, the data robustly support the hypothesized gradient of contextual sensitivity. Fear of Negative Evaluation was the most context-sensitive dimension (M_diff = 0.579). This pattern is evident in the item-level data: "I am embarrassed to volunteer answers in front of other students" showed one of the largest contextual drops (physical M=2.842 vs. online M=2.185). Concerns about teacher correction showed more moderate reductions.

Fourth, Test Anxiety also showed high sensitivity (M_diff = 0.573). Its mean score dropped from 2.827 (moderate) in the physical context to 2.254 (low) online. This reduction is nearly identical in magnitude to that of Fear of Negative Evaluation.

4.3 Illustrative Item-Level Insights

Examining responses to individual items provides richer insight into the mechanisms behind the dimensional differences. Notable patterns include:

Social Spotlight vs. Digital Buffer: The Fear of Negative Evaluation item "I am embarrassed to volunteer answers in front of other students" showed one of the largest contextual drops. In the physical context, raising a hand and speaking into a silent room makes the student acutely

visible. In the online context, using a raise hand button or simply unmuting feels less performative and conspicuous. This demonstrates how the digital interface buffers against the spotlight effect (Gilovich et al., 2000).

The Persistent Cognitive Core: Despite the overall reduction, the single highest anxiety item in the online context was "I start to panic when I have to speak without preparation" (Communication Apprehension, M=2.550). This underscores that the core cognitive pressure of spontaneous speech remains a potent source of anxiety. This pressure—the rapid mental formulation of ideas in a linguistically imperfect system, persists regardless of whether the audience is physically present or digitally mediated.

Attenuated Social Comparison: Items directly related to peer comparison, such as "I'm worried that other students in class speak better than I do" (Fear of Negative Evaluation), showed pronounced decreases online. This supports the notion that digital mediation reduces the immediacy and salience of upward social comparison. By placing peers in separate visual frames and often depersonalizing them into icons or names, digital platforms reduce a key trigger for evaluative anxiety.

Shifting Anxiety Foci: In the physical context, the highest anxiety item overall was related to misunderstanding the teacher ("It frightens me when I don't understand what the teacher is saying in English", FNE, M=2.953). In the online context, while this item's anxiety decreased, concerns about being put on the spot without preparation (CA) became relatively more prominent. This suggests not just a reduction in anxiety online, but a potential reconfiguration of the anxiety profile. Social-evaluative concerns recede, while cognitive-linguistic challenges come to the fore.

5. Discussion

5.1 Communication Context as an Affective Architect: Validating the Differential Sensitivity Framework

The results provide robust empirical validation for the proposed theoretical framework. They establish the communication context, defined by its specific media characteristics, as a powerful architect of learner affect. The significant overall reduction in ESA online aligns with a growing body of literature. Online environments can create a less threatening space for performative tasks like language production (Chen & Zhang, 2022). These environments feature leaner cue systems and lower social presence. However, the true theoretical contribution of this study lies in the differential sensitivity pattern revealed by the dimensional analysis. This pattern moves beyond a generic "online reduces anxiety" conclusion to offer more nuanced insights into the nature of language learning anxieties themselves.

The finding that Fear of Negative Evaluation (FNE) is the most context-sensitive dimension offers strong confirmatory evidence for its conceptualization as a primarily social-evaluative construct. It reduces dramatically online because the medium makes the evaluating audience less salient, less immediate, and less tangible. This aligns directly with the "online disinhibition effect" (Suler, 2004), which posits that the reduced social cues and physical distance in digitally-mediated environments lower psychological barriers to self-expression by diminishing the perceived presence of an evaluating audience. This effect is further supported by core tenets of communication theory regarding social presence and cue multiplicity (Short et al., 1976; Daft & Lengel, 1986). The physical classroom makes the audience vividly present; the digital classroom, particularly when cameras are optional, renders that audience more abstract and distant. This interpretation aligns with Bárkányi and Brash's (2025) recent finding that online learners' fear of negative evaluation manifests differently than in face-to-face contexts. Technology serves as both an anxiety-inducing and anxiety-buffering mechanism. Their study revealed that students employ more nuanced avoidance strategies online, ranging from complete withdrawal to full engagement via text chat, a finding that complements the observation of Communication Apprehension's relative stability across contexts.

The near-equivalent reduction of Test Anxiety (TA) presents an intriguing finding that necessitates a revision of the original hypothesis. The initial prediction of TA's contextual stability was derived predominantly from Western theoretical frameworks that conceptualize assessment anxiety as consequence-focused—tied to fears of poor grades or academic failure (Aida, 1994; Çağatay, 2015; Horwitz et al., 1986). However, the empirical pattern compels recognition that in the Chinese educational context, TA is not purely consequence-focused; it is substantially infused with social-evaluative concerns rooted in collectivist culture and 'face'

dynamics (Wen & Clément, 2003; Liu, 2006). An oral test in a physical classroom is not merely an academic evaluation—it is a public performance where errors risk social embarrassment and loss of face before peers and the instructor. When the online context attenuates this social-evaluative component (through reduced audience salience, optional camera use, and physical distance), TA decreases alongside Fear of Negative Evaluation, yielding the near-equivalent reduction observed. Thus, rather than being context-stable, TA in the Chinese EFL context exhibits high contextual sensitivity—a finding that revises rather than merely complicates the original hypothesis.

This interpretation is supported by item-level response patterns: consequence-focused concerns dominated in the physical context, while preparation-related pressure became relatively more prominent online. Nevertheless, TA remained the highest among the three dimensions in the online context, indicating that the fundamental fear of academic consequences persists, potentially compounded by technology-related reliability concerns during digital assessments (Wang et al., 2021).

This heightened sensitivity can be further understood through the cultural lens of "face" (面子) in the Chinese educational context. In collectivist Confucian-heritage cultures, public performance is closely tied to social standing and the avoidance of public embarrassment (Goffman, 1955; Yum, 1988; Hwang, 2012). Research has shown that individuals from interdependent self-construal cultures—where the self is defined in relation to others—exhibit higher susceptibility to embarrassment than those from independent self-construal cultures (Singelis & Sharkey, 1995). The physical classroom, with its high social presence and immediate audience feedback, amplifies face-threatening possibilities. An incorrect answer or accented pronunciation risks not merely a low grade but public loss of face (Wen & Clément, 2003; Liu, 2006). The online environment attenuates these face threats through features such as optional camera use and reduced non-verbal cue visibility. This digital face buffer may be especially potent and liberating for Chinese EFL learners (Chen & Chew, 2021; Cheng & Sun, 2025). Recent qualitative work on Chinese learners' online experiences further supports this interpretation (Li et al., 2023).

Conversely, Communication Apprehension (CA) shows a smaller yet statistically significant reduction across contexts, a pattern that is equally theoretically significant. Its smaller reduction indicates that a substantial, core component of this anxiety is rooted in intra-individual cognitive and linguistic processes rather than in the external social context (MacIntyre, 1995; MacIntyre & Gardner, 1994). The anxiety linked to accessing lexical items, assembling grammatical structures, monitoring pronunciation, and managing fluency under real-time pressure appears to be a more stable trait (Woodrow, 2006; Gregersen & Horwitz, 2002). It is less malleable by simply changing the communication channel. This finding corroborates the view that ability-based or cognition-generated anxiety represents a more enduring challenge in language acquisition (MacIntyre, 1995; Trebits, 2025; Alkamel, 2025). It also highlights an important caveat for online language teaching: while the digital space may lower social barriers (Chen & Zhang, 2022; Resnik et al., 2022), it does not automatically lower the cognitive barriers to speaking.

5.2 Strategic Implications for Communication-Optimized Language Pedagogy

These findings carry implications for the design and implementation of blended language courses. They argue for moving beyond the logistical question of "what to put online vs. offline" toward a more principled, dimension-aware, context-strategic pedagogical approach. Instructional decisions should be informed by an understanding of which anxieties a given activity primarily engages and which communication context is best suited to manage that specific affective load.

Leveraging the Digital Affective Buffer for Social-Evaluative Goals: Instructional activities whose primary aim is to build initial confidence, encourage participation from reticent learners, or develop fluency through low-stakes practice should be strategically placed in the online context. This placement deliberately leverages the online environment's inherent capacity to dampen Fear of Negative Evaluation. Techniques such as asynchronous video responses, text-based or voice-message discussions in small breakout rooms, or the use of anonymous polling and response tools (Kohnke & Moorhouse, 2022) can maximize this affective benefit. These methods create a communicative space that allows learners to focus more on message formation and less on audience judgment. This is particularly important in the early stages of a

course or when introducing new, challenging topics.

Preserving the Physical Context for Complex Communication Competence: The development of higher-order communicative skills requires the rich, high-context environment of the physical classroom. Skills such as negotiating meaning in real-time group work, interpreting nonverbal cues, and managing the cognitive-affective load of spontaneous interaction are best cultivated face-to-face. The data suggest that the anxiety associated with these complex tasks (largely falling under Communication Apprehension) is less alleviated online anyway. Therefore, the physical classroom's unique affordances for embodied, multimodal practice should be preserved and prioritized for these advanced objectives. This argues for a flipped use of contexts: using online spaces for preparation, practice, and confidence-building, and reserving face-to-face time for the most interactionally complex and socially rich communicative activities.

Decoupling Assessment Anxiety from Delivery Mode through Design: To genuinely address Test Anxiety, the focus must shift from the delivery mode to the fundamental design of assessment itself. Strategies proven to reduce the threat value of evaluation should be employed in both contexts. These include implementing more frequent, low-stakes formative assessments (Black & Wiliam, 1998); using portfolio assessments that emphasize growth; providing transparent, criteria-based rubrics well in advance (Panadero & Jonsson, 2013); and conducting practice or mock assessments in both formats to build familiarity. For high-stakes summative assessments, ensuring clarity of task requirements, providing choice where possible, and guaranteeing technological robustness in the online context are essential (Yang, 2024).

5.3 Theoretical Contributions, Limitations, and Future Directions

This study makes a distinct contribution by bridging applied linguistics with communication and media studies. It demonstrates that established theories of media richness and social presence can effectively predict and explain nuanced patterns of language learner emotion. It validates a model of affective constructs as having differential sensitivity to environmental parameters. This challenges treatments of anxiety as a monolithic entity. The findings suggest that the common finding of lower anxiety online is largely driven by the reduction of one specific component: Fear of Negative Evaluation (FNE).

Several methodological limitations must be acknowledged. These reflect practical constraints and design choices common in survey-based EFL research:

First, the cross-sectional design captures a snapshot. It cannot trace how individual learners' dimensional anxiety profiles might adapt over time with sustained exposure to a blended environment. Does the online buffer effect for FNE persist, or do learners adapt?

Second, the study relied on self-report measures. While reliable, these cannot capture the real-time, moment-to-moment fluctuations of anxiety during actual speaking tasks.

Third, the sample, while sizable, was drawn from one province in China. The findings may be influenced by the specific cultural and educational context of Chinese universities.

These limitations do not undermine the core findings but instead point to valuable directions for future research.

Longitudinal & Dynamic Studies: Employing longitudinal designs or experience-sampling methods to track the co-adaptation of anxiety dimensions and context over a semester.

Mixed-Methods Integration: Combining quantitative surveys with qualitative methods (e.g., stimulated recall interviews, classroom observation) to unpack the lived experience of dimensional anxiety in different contexts.

Expanding the Nomological Network: Investigating how different anxiety dimensions mediate or interact with other key variables like Willingness to Communicate (WTC), self-efficacy, and self-regulation strategies. Such investigation could reveal their impact on ultimate communicative performance.

Platform-Specific Research: From a communication studies perspective, research on how specific platform features (e.g., virtual backgrounds, reaction emojis, immersive view vs. gallery view, the use of avatars) differentially modulate the three anxiety dimensions could provide actionable insights for educational technology design.

6. Conclusion

This investigation reveals that the communication context—the physical, co-

present classroom versus the digitally-mediated, online classroom—does not simply change the volume of anxiety; it fundamentally alters its composition. The online environment acts as a selective filter. It is particularly effective in muting social-evaluative anxiety (Fear of Negative Evaluation). However, it offers far less attenuation for cognition-based anxiety (Communication Apprehension) rooted in the linguistic act itself. Test Anxiety, tied to the perceived stakes of evaluation, shows significant reduction, though its high contextual sensitivity in the Chinese context revises initial expectations.

These findings deliver a clear message to educators and instructional designers. In a world moving toward blended learning, strategic context-design is essential. This involves moving beyond the binary logistics of modality choice. It requires making principled decisions based on the affective profile of learning objectives. By consciously matching activities that trigger social-evaluative fear with the buffering digital context, and reserving the rich, high-context physical environment for practicing the complex cognitive and interactive demands of real communication, educators can create more supportive and effective language learning journeys. Such an approach leverages understanding of both language and communication to better scaffold learners' path from anxious silence to confident, competent, and context-adaptive expression.

References

Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's Construct of Foreign Language Anxiety: The Case of Students of Japanese. *The Modern Language Journal*, *78*(2). http://www.jstor.org/stable/329005?seq=1&cid=pdf-reference#references_tab_contents

Alkamel, D. M. (2025). Beyond General Anxiety: A Systematic Review of the Distinct Roles of English Four-Skills Anxiety in Language Acquisition. Available at SSRN 5907944.

Alqarni, N. (2021). Language learners' willingness to communicate and speaking anxiety in online versus face-to-face learning contexts. *International Journal of Learning, Teaching and Educational Research*, *20*(11), 57–77. https://doi.org/10.26803/IJLTER.20.11.4

Bargh, J. A., & McKenna, K. Y. A. (2004). The Internet and social life. *Annu. Rev. Psychol.*, *55*(1), 573–590.

Bárkányi, Z., & Brash, B. (2025). Foreign language speaking anxiety online: Mitigating strategies and speaking practices. *ReCALL*, *37*(3), 421–440. https://doi.org/10.1017/S0958344025000060

Basille, A., Lavoué, É., & Serna, A. (2025). Impact of communication modalities on social presence and regulation processes in a collaborative game. *Journal on Multimodal User Interfaces*, 1–18. https://doi.org/10.1007/s12193-024-00450-z

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Çağatay, S. (2015). Examining EFL Students' Foreign Language Speaking Anxiety: The Case at a Turkish State University. *Procedia - Social and Behavioral Sciences*, *199*, 648–656. https://doi.org/https://doi.org/10.1016/j.sbspro.2015.07.594

Chen, Y., & Chew, S. Y. (2021). Speaking Performance and Anxiety Levels of Chinese EFL Learners in Face-to-Face and Synchronous Voice-based Chat. *Journal of Language and Education*, *7*(3), 43–57. https://doi.org/10.17323/jle.2021.11878

Chen, Y., & Zhang, Z. (2022). A Study of Willingness to Communicate Model of Chinese College Students in Extramural and Extracurricular Online English Learning Contexts. *Foreign Languages and Literature*, *38*(6), 151–160.

Cheng, Z., & Sun, P. P. (2025). Does L2 Speaking Anxiety Differ in Classroom and Synchronous Online Learning Environments? Evidence From EFL College Students. *Journal of Computer Assisted Learning*, *41*(5). https://doi.org/10.1111/jcal.70115

Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement*, *56*(10), 15–23.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 554–571.

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage

publications.

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2–3), 87–105.

Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: an egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology*, *78*(2), 211.

Goffman, E. (1955). On Face-Work: An analysis of ritual elements in social interaction. *Psychiatry*, *18*(3), 213–231. https://doi.org/10.1080/00332747.1955.11023008

Goffman, E. (1963). *Behavior in Public Places: Notes on the Social Organization of Gatherings*. Free Press of Glencoe.

Gregersen, T., & Horwitz, E. K. (2002). Language Learning and Perfectionism: Anxious and Non-Anxious Language Learners' Reactions to Their Own Oral Performance. *The Modern Language Journal*, *86*(4), 562–570. https://doi.org/10.1111/1540-4781.00161

Hall, E. T. (1976). *Beyond Culture*. Anchor Press.

He, D. (2018). Foreign Language Learning Anxiety in China. In *Foreign Language Learning Anxiety in China: Theories and Applications in English Language Teaching*. Springer Singapore. https://doi.org/10.1007/978-981-10-7662-6

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign Language Classroom Anxiety. *The Modern Language Journal*, *70*(2). https://www.jstor.org/stable/327317

Hrastinski, S. (2008). Asynchronous and synchronous e-learning. *Educause Quarterly*, *31*(4), 51–55.

Huang, P., & Hwang, Y. (2013). An Exploration of EFL Learners' Anxiety and E-learning Environments. *Journal of Language Teaching and Research*, *4*(1). https://doi.org/10.4304/jltr.4.1.27-35

Hwang, K.-K. (2012). *Foundations of Chinese Psychology: Confucian Social Relations* (Vol. 1). Springer New York. https://doi.org/10.1007/978-1-4614-1439-1

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, *39*(10), 1123.

Kohnke, L., & Moorhouse, B. L. (2022). Facilitating Synchronous Online Language Learning through Zoom. *RELC Journal*, *53*(1), 296–301. https://doi.org/10.1177/0033688220937235

Lee, J. S., & Chen Hsieh, J. (2019). Affective variables and willingness to communicate of EFL learners in in-class, out-of-class, and digital contexts. *System*, *82*, 63–73. https://doi.org/10.1016/j.system.2019.03.002

Li, C., & Liang, X. (2022). An Empirical Study on the Impact of Blended Learning Environment on College Students' Willingness for Second Language Communication. *Journal of Xi' an International Studies University*, *30*(4), 64–68. https://doi.org/https://doi.org/10.16362/j.cnki.cn61-1457/h.2022.04.016

Li, X., Liu, M., & Zhang, C. (2020). Technological impact on language anxiety dynamic. *Computers & Education*, *150*, 103839. https://doi.org/10.1016/j.compedu.2020.103839

Li, Y., Huang, Y., & He, Q. (2023). A Study of College English Blended Teaching Based on Attention Mechanism. *Journal of Xi'an International Studies University*, *31*(4), 80–85.

Liu, M. (2006). Anxiety in Chinese EFL students at different proficiency levels. *System*, *34*(3), 301–316. https://doi.org/10.1016/j.system.2006.04.004

Liu, M., & Jackson, J. (2008). An Exploration of Chinese EFL Learners' Unwillingness to Communicate and Foreign Language Anxiety. *The Modern Language Journal*, *92*(1), 71–86. https://doi.org/10.1111/j.1540-4781.2008.00687.x

MacIntyre, P. D. (1995). How does anxiety affect second language learning? A reply to Sparks and Ganschow. *The Modern Language Journal*, *79*(1), 90–99.

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, *44*(2), 283–305.

Mak, B. (2011). An exploration of speaking-in-class anxiety with Chinese ESL learners. *System*, *39*(2), 202–214. https://doi.org/10.1016/j.system.2011.04.002

McCroskey, J. C. (1977). Oral communication

apprehension: A summary of recent theory and research. *Human Communication Research*, *4*(1), 78–96.

McCroskey, J. C. (1984). *The communication apprehension perspective*. Avoiding Communication/Sage Publications.

Meerah, T. S. M., Osman, K., Zakaria, E., Ikhsan, Z. H., Krish, P., Lian, D. K. C., & Mahmod, D. (2012). Developing an Instrument to Measure Research Skills. *Procedia - Social and Behavioral Sciences*, *60*, 630–636. https://doi.org/10.1016/j.sbspro.2012.09.434

Moneypenny, D. B., & Aldrich, R. S. (2025). Foreign language anxiety in online college Spanish: Prevalence and effects on oral proficiency. *Language Teaching Research*, *29*(5), 2245–2262. https://doi.org/10.1177/13621688221112378

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, *9*, 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Peschka, L., Hock, M., Carbon, C.-C., Hajak, G., & Bergner-Köther, R. (2025). German adaptation and validation of the Factors of Online Disinhibition Scale. *Computers in Human Behavior Reports*, *18*, 100624. https://doi.org/10.1016/j.chbr.2025.100624

Resnik, P., Dewaele, J. M., & Knechtelsdorfer, E. (2022). Differences in the Intensity and the Nature of Foreign Language Anxiety in In-person and Online EFL Classes during the Pandemic: A Mixed-Methods Study. *TESOL Quarterly*. https://doi.org/10.1002/tesq.3177

Richardson, J. C., Maeda, Y., Lv, J., & Caskurlu, S. (2017). Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis. *Computers in Human Behavior*, *71*, 402–417. https://doi.org/10.1016/j.chb.2017.02.001

Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. (No Title).

Singelis, T. M., & Sharkey, W. F. (1995). Culture, Self-Construal, and Embarrassability. *Journal of Cross-Cultural Psychology*, *26*(6), 622–644. https://doi.org/10.1177/0022022195026006007

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)* (P. Alto, Ed.). Consulting Psychologists Press.

Sproull, L., & Kiesler, S. (1986). Reducing Social Context Cues: Electronic Mail in Organizational Communication. *Management Science*, *32*(11), 1492–1512. https://doi.org/10.1287/mnsc.32.11.1492

Suler, J. (2004). The Online Disinhibition Effect. *CyberPsychology & Behavior*, *7*(3), 321–326. https://doi.org/10.1089/1094931041291295

Timilsina, B. (2025). Beyond the Screen: Re-evaluating Non-Verbal Cues in Digital-Age Communication. *AWADHARANA*, *9*(1), 104–114. https://doi.org/10.3126/awadharana.v9i1.86198

Trebits, A. (2025). Foreign language anxiety and enjoyment concurrently shape pragmatic and grammatical awareness in learners of English as a foreign language—how communicative context modulates their relationship. *Language Awareness*, 1–22. https://doi.org/10.1080/09658416.2025.2504513

Tsui, A. B. M. (1996). Reticence and Anxiety in Second Language Learning. *Voices From the Language Classroom.*

Tu, C.-H., & McIsaac, M. (2002). The relationship of social presence and interaction in online classes. *The American Journal of Distance Education*, *16*(3), 131–150.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, *19*(1), 52–90.

Walther, J. B. (1996). Computer-Mediated Communication. *Communication Research*, *23*(1), 3–43. https://doi.org/10.1177/009365096023001001

Wang, H., Peng, A., & Patterson, M. M. (2021). The roles of class social climate, language mindset, and emotions in predicting willingness to communicate in a foreign language. *System*, *99*. https://doi.org/10.1016/j.system.2021.102529

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, *33*(4), 448.

Wei, X. (2025). Foreign Language Learning Enjoyment and Anxiety: The Effects of Teacher Variables and Online Class Environment. *Journal of English and Applied Linguistics*, *4*(1).

https://doi.org/10.59588/2961-3094.1141

Wen, W. P., & Clément, R. (2003). A Chinese conceptualisation of willingness to communicate in ESL. *Language, Culture and Curriculum*, *16*(1), 18–38. https://doi.org/10.1080/07908310308666654

Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC Journal*, *37*(3), 308–328. https://doi.org/10.1177/0033688206071315

Yang, L. (2024). Enhancing emotional health and engagement in Chinese English language learners: an approach from teachers' autonomy- supportive behavior, teachers' harmony, and peer support in a two-sample study. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1356213

Yaniafari, R. P., & Rihardini, A. A. (2022). Face-To-Face Or Online Speaking Practice: A Comparison of Students' Foreign Language Classroom Anxiety Level. *JEELS (Journal of English Education and Linguistics Studies)*, *8*(1), 49–67. https://doi.org/10.30762/jeels.v8i1.3058

Yum, J. O. (1988). The impact of Confucianism on interpersonal relationships and communication patterns in east Asia. *Communication Monographs*, *55*(4), 374–388. https://doi.org/10.1080/03637758809376178

Zeng, X., & Liu, Q. (2012). A study of English learning anxiety of science and engineering college students under multi-media environment: Based on the comparison of multi-media and traditional classroom teaching. *Computer-Assisted Foreign Language Education*, *9*, 50–55.

Zhang, Y., & Mohamed, H. B. (2024). The conceptions and dimensions of social presence: a systematic literature review. *Universidad y Sociedad*, *16*(5), 277–287.

Zhao, Y., Huang, X., & Hui, L. (2025). Academic buoyancy and academic engagement in English speaking learning among Chinese college students: the mediation of enjoyment and the moderation of anxiety. *Frontiers in Psychology*, *16*. https://doi.org/10.3389/fpsyg.2025.1680032

Zhou, Y. (2025). Aptitude, Anxiety, and Success in L2 Speech Development: A Longitudinal Study of Chinese EFL College-Level Learners. *Language and Speech*, *68*(2), 437–459. https://doi.org/10.1177/00238309241281741

Journal of Linguistics and Communication Studies
ISSN 2958-0412
www.paradigmpress.org/JLCS
Volume 5 Number 1 March 2026

Projecting Trajectories and Regulating Relations: Address Terms in Mandarin Initiating Actions

Ruiyang Ma[1]

[1] Ocean University of China, Shandong, China
Correspondence: Ruiyang Ma, Ocean University of China, Shandong, China.

doi:10.63593/JLCS.2026.03.06

Abstract

This study employs Conversation Analysis (CA) to explore the usage patterns and interactional functions of address terms in initiating actions in Mandarin Chinese telephone conversations. Based on naturally occurring audio recordings and transcriptions, the research focuses on the deployment of address terms at turn-initial, mid-turn, and turn-final positions, analyzing their roles in managing conversation progression, expressing stance and emotion, and regulating interpersonal relationships. The analysis reveals that at the turn-initial position, address terms are often used to project a shift in the sequence trajectory, indicating disalignment with the prior action. At the mid-turn position, address terms primarily function to intensify the speaker's affective stance and regulate the emotional tone. At the turn-final position, they are frequently used to construct specific identities and consolidate interpersonal solidarity. These findings reveal that address terms serve as complex interactional resources in Mandarin conversation. This study contributes to the interactional study of Mandarin address terms based on naturally occurring data, enriching the literature in CA and providing practical implications for cross-cultural communication and language teaching.

Keywords: address terms, conversation analysis, initiating actions, Mandarin Chinese conversation

1. Introduction

In everyday interaction, the use of address terms is a pervasive linguistic phenomenon that not only reflects social relationships and cultural norms but also plays a crucial role in the organization and management of interaction. Traditionally, address terms frequently appear in the opening and closing sequences of telephone conversations, functioning as a means of identity recognition and politeness markers. However, the sudden emergence of address terms in the middle of a telephone conversation is associated with specific sequential organizational features and interactional functions. Based on this observation, this study aims to explore the following questions: How are address terms used in the initiating actions of Mandarin Chinese conversations, and how does such usage impact the progression of the conversation and the relationship between participants?

Conversation Analysis (CA), an empirical sociological research methodology that respects linguistic facts (Yu & Li, 2009), provides an effective tool for investigating this issue. By conducting a detailed analysis of naturally occurring data, this study focuses on initiating actions in Mandarin telephone conversations, investigating the usage patterns and interactional

functions of address terms at turn-initial, mid-turn, and turn-final positions. This research not only helps us better understand the multiple roles address terms play in Mandarin conversation but also reveals how Chinese interlocutors skillfully utilize linguistic resources to achieve their interactional goals. Furthermore, this study offers valuable insights for applied fields such as cross-cultural communication and language teaching, facilitating a better understanding and application of this important linguistic resource.

2. Literature Review

As an important linguistic form in interpersonal communication, address terms have long garnered attention from multiple disciplines, including linguistics and sociology.

The study of address terms dates back to the 1950s. Brown and Gilman's (1960) pioneering research proposed the famous T/V model, arguing that the choice of address terms reflects the dimensions of power and solidarity between speakers and hearers. Building on this, scholars have conducted extensive sociolinguistic research on address terms across different languages and cultures. For instance, Ervin-Tripp (1972) examined the rules for using address terms in English and proposed a flowchart model of address term selection. Braun (2012) emphasized the multi-dimensionality of address term systems, noting that beyond power and solidarity, the selection of address terms is also influenced by multiple factors such as gender, age, and context. In the Chinese context, the systematic study of address terms was pioneered by Chao (1956), who provided a foundational description of modern Chinese appellations. Subsequently, researchers have explored the dynamic usage of address terms across various settings. For example, Li and Li (2013) examined the selection of address terms in e-commerce transactions, demonstrating how communicators dynamically construct relational identities.

From a pragmatic perspective, the use of address terms is closely related to face theory. Brown and Levinson (1987) suggested that address terms can serve as a positive politeness strategy used to mitigate the impact of face-threatening acts. Applying this to Mandarin Chinese, Gu (1990) systematically analyzed how politeness phenomena are realized through address practices. Furthermore, scholars have identified diverse pragmatic functions of address term conversions. Liang (2001) and Diao (2005) highlighted that shifting address terms serves as a pragmatic strategy to mark emotional changes and index interpersonal stances. Similarly, Jiang and Liu (2014) observed that in institutional mediation, adjudicators strategically alternate address forms to minimize social distance and foster a cooperative dialogue atmosphere.

With the deepening of research, scholars realized that address terms not only reflect social relations but also construct them. From a social interaction perspective, Sacks (1992) noted that address terms are a "Membership Categorization Device" (MCD). By selecting a specific address term, a speaker can categorize the recipient into a certain social member class. Recently, within the field of CA, the interactional perspective on address terms has diversified. For example, Lerner (2003) found that turn-initial address terms can be used to establish the recipient's attention. Rendle-Short (2007, 2010) revealed the role of address terms in turn organization and topic management in political interviews. Clayman (2010) found that in news interviews, address terms often appear in disaligning actions or expressive actions to emphasize the speaker's stance. Clayman (2013) also noted that address terms in responses express speaker agency. Butler et al. (2011) explored the role of turn-initial address terms in managing disalignment and disaffiliation in telephone counseling.

In summary, research on address terms has yielded fruitful results. However, limitations remain. First, most interactional studies focus on Western languages like English; systematic studies on Mandarin Chinese address terms are relatively insufficient. Second, while research has focused on address terms in responsive actions and turn-initial positions, there is a lack of systematic study on their use in mid-turn and turn-final positions within initiating actions. Through CA, this study investigates the patterns of address terms in different turn positions within initiating actions in Mandarin conversations, revealing the underlying connections between their positions and functions.

3. Research Methodology and Data Collection

CA is a sociological approach that investigates the orderly nature of social institution operations through meticulous observation of social members' talk-in-interaction (Wu & Yu, 2022). The scientific rigor of CA is reflected in several

aspects: First, CA uses naturally occurring authentic data. Second, it follows a strict analytical procedure: identifying a phenomenon, collecting instances, and inductively uncovering sequential patterns (Drew, 2008). Third, CA emphasizes understanding interactions from an *emic perspective* (Seedhouse, 2005). Finally, CA's findings are reproducible (Jacobs, 2012).

The data used in this study is drawn from the DMC (DIG Mandarin Conversations) corpus (Yu et al., 2024). Informed consent was obtained from all participants. All transcripts follow the Jefferson (2004) transcription conventions, adapted for the specific needs of Mandarin Chinese, documenting intonation, tone, pauses, and overlaps in detail to ensure the authenticity and objectivity of the research. Adopting this methodology, the present study examines the deployment of address terms across various turn positions within initiating actions in Mandarin telephone conversations, exploring the social actions they perform and their interactional functions in authentic encounters.

4. Address Terms in Mandarin Initiating Actions

4.1 Turn-Initial Position: Projecting Sequence Trajectory Shift and Disalignment

In Mandarin conversation, the use of address terms is not merely a form of politeness, but a complex interactional resource that plays a key role in conversational organization. In initiating actions, turn-initial address terms project a shift in sequence trajectory and structural disalignment with the prior action. The following two extracts demonstrate how address terms function to reset topic directions and negotiate interactional stances.

(1) [OUC-DMC-QYN_有一个急事_0000-0127]

01 小浅：你你>你知道<，

02 　　你知道哪个班他没有课吗？明天:.

03 琪琪：明天::.

04 琪琪：嗯:，学姐，要不就::

05 　　>因为他们考研[嘛<，

06 小浅：　　[°嗯°.

07 琪琪：有保研的人有小金:，有[小孙:，

08 小浅：　　[°嗯°

09 琪琪：有[小李:，然后小吴:.

10 小浅：　[°嗯°

11 琪琪：这些.

In Extract (1), Xiaoqian has an urgent matter and needs Qiqi's help, but Qiqi indicates she does not have time. Consequently, in lines 1-2, Xiaoqian initiates a new information-seeking sequence to find an alternative solution. In line 3, Qiqi first repeats the increment "tomorrow" ("明天") that Xiaoqian added after a structurally complete turn-constructional unit (Yu, 2021). The prosodic feature of the sound stretch indicates that this repetition displays Qiqi's thinking process rather than initiating a repair on line 2. Subsequently, in line 4, Qiqi's stretched "en:" ("嗯:") serves as a delay component, projecting the dispreferred nature of her upcoming response.

As expected, the word "how about" ("要不"), functioning as a conversational practice for making a proposal (Yu & Hao, 2020), demonstrates that Qiqi is about to abandon providing the information sought in line 2. She breaks out of the Question-Answer adjacency pair and provides a type-nonconforming response, initiating a proposal sequence herself to solve the other party's difficulty. Notably, Qiqi suddenly adds the address term "senior" ("学姐") at the beginning of this initiating action, projecting a shift in the sequence trajectory and forecasting that her response will disalign with the prior action. In line 5, Qiqi immediately provides an account for her relevant absence (i.e., most classmates are preparing for the graduate entrance exam), logically implying that even if they have no classes, they have no time to help. Thus, in lines 7-11, she offers an alternative solution by listing students who are guaranteed admission. The continuers "en" ("嗯") from Xiaoqian in lines 6, 8, and 10 demonstrate that she does not pursue the question from line 2 but aligns with Qiqi's newly set trajectory, treating the relevant absence as unproblematic and achieving mutual understanding.

(2) [OUC-DMC-DBY_系鞋带_0000-0251]

01 妈妈: 嗯,那个你不是要上姥姥家去嘛,

02 　　(0.6)

03 妈妈: 怎么不去呢?

04 壮壮: (去)((0.2 杂音))

05 　　(0.4)

06 妈妈: 嗯,

07 壮壮: 我爸说要-我爸说要教我系鞋带儿的.

08 　　(0.2)

09 妈妈: .tch 对:你不去姥姥家吗,

10 　　(2.0)

11 壮壮: 我去姥姥家,( )我爸要教我系鞋带儿.
12 (0.4)
13 妈妈: 啊:那就不系了呗让姥姥姥爷教你系.
14 (.)要不我现在打电话让姥姥去接你,
15 (2.0)
16 壮壮: °好°
17 (0.3)
18 妈妈: 啊:?
19 (0.7)
20 壮壮: 好:.
21 (0.2)
22 妈妈: >那为什么你们不上姥姥家去啊,
23 都九点了,上午.<
24 (1.0)
25 壮壮: 妈妈,
26 妈妈: 嗯:.
27 (0.6)
28 壮壮: 你现在在哪儿啊?
29 妈妈: 我还在妇女儿童医院呢.

In Extract (2), Mom uses a rhetorical question in line 1 to inquire about Zhuangzhuang's whereabouts. The "weren't you ... right" ("不是 …… 嘛") structure displays Mom's K+ (knowledgeable) epistemic status, as she already knows his original plan was to go to grandma's house. The 0.6-second silence projects the dispreferred nature of Zhuangzhuang's upcoming response. Therefore, Mom takes the floor again in line 3. Her turn design "why didn't you go" ("怎么不去呢") presupposes that he didn't go, carrying a blame-implicative tone. In line 4, Zhuangzhuang denies the premise of Mom's question with a positive response "go" ("去"), and subsequently provides an account in line 7 (his dad wants to teach him to tie his shoelaces). After a 0.2-second silence, possibly because Zhuangzhuang's voice in line 4 was obscured by noise, Mom seeks a second confirmation in line 9. Consequently, after a 2.0-second silence, Zhuangzhuang provides a full response in line 11 with a marked repetition, completely reiterating his account from line 7. Following a 0.4-second silence, Mom's "ah:" ("啊:") in line 13 marks a change of state, after which she immediately proposes having grandma pick him up.

After Zhuangzhuang agrees, Mom questions the reason for not going to grandma's house for the third time in line 22. The turn-final particle "ah" ("啊") displays Mom's stance that Zhuangzhuang's response is inadequate or problematic, suggesting that his verbal expression contradicts his actual behavior and causes comprehension difficulties. Furthermore, she adds the increment "it's already nine o'clock, in the morning" ("都九点了，上午") at the end of the turn, emphasizing the late hour and projecting blame. After a 1.0-second silence, instead of responding to the blame-implicative question, Zhuangzhuang suddenly uses the address term "Mom" ("妈妈") in line 25. This action re-establishes Mom's attention and prepares for a topic and sequence trajectory shift. Following Mom's continuer "en:" ("嗯:") in line 26, Zhuangzhuang initiates a disaligned information-seeking action in line 28 (asking where she is). Consequently, in line 29, Mom abandons her blame sequence and aligns with Zhuangzhuang's newly set trajectory, answering his question.

In summary, the deployment of turn-initial address terms in these initiating actions serves as a crucial interactional resource to project an imminent shift in sequence trajectory or disalignment. By co-occurring with other turn-design features like silences and delays, these address terms function as an attention-focusing device, explicitly marking a turning point in the interaction. Ultimately, they re-establish the recipient's attention and pave the way for a newly initiated sequence that deviates from the previously expected trajectory. Through such turn beginnings, interlocutors use linguistic resources to manage the ongoing interaction, demonstrating the complexity of sequence organization in Mandarin conversation.

4.2 Mid-Turn Position: Intensifying Affective Stance and Regulating Emotional Tone

Address terms also frequently appear in the mid-turn position during initiating actions, exhibiting a distinct interactional function compared to their turn-initial counterparts. In this position, rather than projecting trajectory shifts, address terms are primarily used to intensify the speaker's affective stance. This mid-turn deployment serves to highlight the speaker's emotional involvement, shaping the relational dynamics between interlocutors and the developmental path of the interaction.

(3) [OUC-DMC-LXJ_乱吃保健品_0000-0430]
01 蕾: .h 外外都不看病,就是吃保健品了,

02 吃的一个比一个坏,不是得了这癌,
03 就是得了那癌,.h 然后她就(.).h
04 觉得她自己得了这怪病,那怪病,然后.hh
05 °呃°都给她检查啦,浑身没一点儿毛病.
06 华: ￥噢::￥=
07 蕾: =°哎°就是一个<息肉>,她就是<那诶呀>.
08 华: 息肉°[小毛病,问题不大°.
09 蕾: [°噢:我妈°真是服啦°哎呀 h.°
10 华: °hehehe°[噢,
11 蕾: [你:们就不用,°呃°操那心,
12 >°三叔°<,把你们闹好就行啦.
13 华: ￥放心放心￥.

In Extract (3), Hua calls to inquire about Lei's mother's illness. In lines 1-5, Lei initiates a complaint sequence about her mother taking unnecessary health supplements, using detailing to legitimize her complaining action (Yu & Li, 2022). For example, lines 1-2 contrast "never sees a doctor" ("都不看病") with "just takes health supplements" ("就是吃保健品"), leading to worsening consequences. In lines 2-4, "this cancer" ("这癌"), "that cancer" ("那癌"), "this weird disease" ("这怪病"), and "that weird disease" ("那怪病") utilize extreme case formulations, emphasizing the severity of the mother's hypochondria through repetition. Conversely, in line 5, the mother's actual condition, "absolutely nothing wrong" ("浑身没一点儿毛病"), sharply contradicts the hypothetical scenarios. The lexical choice "got tested" ("检查") demonstrates that Lei's conclusion is fact-based. Since Hua is a relative, he does not join the complaint in line 6 but merely gives a smiling minimal response "oh" ("噢") to the stark contrast between imagination and reality.

In line 7, Lei clarifies the real diagnosis as a "polyp" ("息肉"), employing words like "just" ("就是"), "one" ("一个"), and incoherent expressions to downgrade the illness's severity, projecting her dissatisfaction towards her mother. When the talk touches upon the reason for the call (the mother's illness), Hua actively takes the floor in line 8, providing an affiliative response agreeing that the condition is "a minor issue" ("小毛病"). Meanwhile, Lei's overlapping talk in line 9 continues her complaint. In line 10, Hua again issues a smiling minimal response "oh" ("噢"), marking the end of the complaint sequence.

In line 11, Lei initiates a new comforting action, telling Hua's family not to worry. Here, Lei suddenly inserts the address term "Uncle" ("三叔") in the middle of her turn. Grammatically, Lei had already used "you guys" ("你们") as the subject, making the address term syntactically unnecessary. However, the insertion of "Uncle" ("三叔") significantly intensifies Lei's affective stance, reassuring Hua that her mother is fine and there is no need to worry. The kinship term closes the social distance between the two families and positively steers the sequence towards caring for Hua's family, advising them to "just take care of yourselves" ("把你们闹好就行啦"). In line 13, Hua immediately provides a preferred, repeated response "Don't worry, don't worry" ("放心放心").

(4) [OUC-DMC-LXJ_关心疱疹_0000-0152]

01 华: 我说一会儿°wo°,想过-,
02 >我说看看你害怕你<输液°了°是啥的,
03 英: 啊:,不用>不用不用<,一会儿我就输液走呀,
04 (0.2)
05 华: 输液走呀,
06 英: 噢.
07 英: >不用过来不用过来<,↑没事儿,°.hh°
08 华: ↑没事儿啊(Z 嗯:)晋英我可告你,
09 别>那么那<咋的,
10 (.)
11 华: 哈,
12 (.)
13 华: ↑我前两天正好我们同学在这儿玩儿了,
14 <也是>,(0.4)
15 他都十几天了还疼:了,>不过<好多了.

In Extract (4), Hua states the reason for his call in lines 1-2: wanting to visit the sick Ying without disturbing her, which corresponds to Hua's earlier use of a pre-sequence to confirm Ying's location. In line 3, Ying's turn-initial "ah" ("啊") marks a change in epistemic state. She then immediately rejects Hua's request to visit using a cross-cutting preference structure (Schegloff, 2007) – delivering a rapid, unmitigated, and thrice-repeated preferred organization, "no need, no need, no need" ("不用不用不用"), to perform the dispreferred action of rejection. Ying then rationalizes her rejection with an objective account, "I'm going to get an IV drip soon" ("一会儿我就输液走呀"), making it difficult for Hua to pursue his request.

After a 0.2-second silence, Hua completely repeats Ying's account in line 5. Ying responds affirmatively with "oh" ("噢") in line 6, treating Hua's repetition as an information confirmation. Because Hua does not explicitly accept the rejection, Ying redoes her rejecting action in line 7 with an accelerated, repeated emphasis, "no need to come" ("不用过来"), and provides another account, "I'm fine" ("没事儿"), thoroughly eliminating the necessity of the visit. In line 8, still not explicitly accepting or declining the rejection, Hua is triggered by Ying's "I'm fine" ("没事儿") and initiates a new advising/comforting sequence. Hua first repeats "fine" ("没事儿") in a high pitch. Then, in the middle of the turn, he suddenly adds Ying's name, "Jinying" ("晋英"), seamlessly connecting it to the following turn-constructional unit (TCU) without a pause, leading to his advice against being pessimistic. The use of the address term, combined with the strong phrase "let me tell you" ("我可告你"), effectively grabs Ying's attention, highlights the importance of the upcoming talk, and expresses deep concern. Subsequently, in lines 13-15, Hua uses a classmate's similar experience as evidence to further comfort Ying.

In summary, the insertion of address terms at the mid-turn position operates as a specialized turn-design practice to upgrade the speaker's affective stance. Rather than serving a referential or grammatical necessity, these mid-turn insertions regulate the emotional tone of the interaction, closing the social distance between interlocutors and intensifying the emotional resonance of the ongoing action. This deployment demonstrates how interlocutors use turn-internal positioning to accomplish nuanced relational work in real time.

4.3 Turn-Final Position: Constructing Identity and Consolidating Interpersonal Relationships

Besides initial and mid-turn positions, address terms also frequently appear at the end of a turn during initiating actions. In this position, they are frequently used to construct particular identities, consolidate interpersonal solidarity, and facilitate membership categorization.

(5) [OUC-DMC-LM_帮做 PPT_0000-0109]

01 东: 田 j-田径队儿:的能能,

02 °si°就推一次:这个行不行:?

03 (0.5)

04 我: h [可能想]找你帮个忙儿(.)

05 刘: [欧::]

06 东: 这边[儿:].

07 刘: [>哦]就咱班那个项目是吧.<

08 (0.8)

09 东：对,我想看你如果就(.)没别的事儿就-过来，

10 就是现在 PPT:°我°就想你不是((鼻子吸气))

11 参加过几次嘛想让你帮忙:打磨一下.

12 现成的东西都有,就是添添.h 加加:些东西.

13 刘: 哦:哦:.

14 (1.0)

15 东: 可以不老弟?=

16 刘：=嗯::行，行.

In Extract (5), Dong initiates a request in lines 1-2, hoping Liu will cancel his track team activity. The "okay or not" ("行不行") structure displays a structural preference for the response "okay" ("行") (Yu & Liang, 2018). A 0.5-second silence projects the potential dispreferred nature of Liu's response. Consequently, in line 4, Dong takes the floor to provide an account for his request. The sound stretch of "I" ("我") and the in-breath delay exhibit the dispreferred organizational features of the request action, while the turn design "might want to" ("可能想") lowers his entitlement to make the request. In line 7, Liu initiates a post-first insert expansion by confirming the specific task. After a 0.8-second silence, Dong provides an affirmative response in line 9 and further details the specific request of polishing a PPT in lines 9-12. Phrases like "I want to see if you" ("我想看你"), "if you have nothing else" ("如果没别的事儿"), and "I just thought" ("我就想") display his low entitlement and high contingency for the recipient's acceptance. Furthermore, he minimizes the cost of the request by stating "we have everything ready" ("现成的东西都有") and "just adding a few things" ("就是添添加加些东西"), increasing the likelihood of acceptance.

In line 13, Liu only provides a minimal response "oh" ("哦"), marking a change of state without explicitly accepting or rejecting the request. After a 1.0-second silence without a response from Liu, Dong actively takes the floor in line 15 to pursue a definitive answer. Notably, Dong suddenly adds the address term "bro" ("老弟") at the turn-final position. This term constructs his identity as Liu's "brother", displaying their close relationship. From a sociological standpoint, this

increases the likelihood of the request being accepted. Consequently, in line 16, Liu quickly latches onto the turn and gives a preferred, repeated response "okay, okay" ("行，行"), finally fulfilling Dong's communicative goal.

(6) [OUC-DMC-MXX_过年问候_0000-0233]

01 勇: 哎呀.我这不行啊.今天没挣下钱.

02 刘: 哎.没赔吧.

03 勇: 嗯.没-没回本呢.现在.

04 刘: 还没回本.

05 勇: 哎东西没弄完.

06 刘: 你看.要不发发小红书啥的.

07 借着元宵节中-正月十五这一波的.

08 再往外弄一点吧.

09 勇: 我-没事.我这还有点那个啥.我试试呗.

10 刘: 对发着发-想办法.

11 正月十五之前应该.出去就行对吧.

12 勇: 对.

13 刘: 行行行.那你玩吧.不影响你挣钱了兄弟.

14 勇: 行行行没事哥.[哎

15 刘: [好来好来.

16 勇: 我等.我给你发一个那个啥-公众号.

17 不是公众号.视频号.那个视频号

18 就是做那个酒-手工的那个酒精灯.

19 刘: 哦::.没问题.行行行.你发给我吧.好来.拜拜.

In Extract (6), Yong initiates a troubles-telling sequence in line 1 by stating "didn't make any money today" ("今天没挣下钱"). The turn-initial "aiya" ("哎呀") projects Yong's negative stance (Yu et al., 2020). In line 2, Liu follows up with a caring inquiry, designing the turn as "didn't lose money, right" ("没赔吧") to show a structural preference for "didn't lose" ("没赔"). In line 3, Yong's turn-initial delay "en" ("嗯") and cut-off project the dispreferred nature of his response; "haven't made the capital back" ("没回本呢") indicates he did lose money. However, the turn-final temporal increment "right now" ("现在") suggests the loss is temporary, projecting his expectation to recover the capital. In response, Liu repeats "haven't made the capital back" ("没回本") in line 4, adding the word "yet" ("还") at the beginning, similarly displaying that this state will not last.

After Yong provides an account in line 5, Liu initiates a proposal in lines 6-8 using "how about" ("要不"), suggesting social media promotion. In line 9, Yong's response "it's fine" ("没事") treats Liu's proposal as a comforting action, provides an account for being fine, and states his planned action, "I'll give it a try" ("我试试呗"). Liu provides an affiliative response "right" ("对") in line 10 and continues advising via a question format, where "right?" ("对吧") displays a structural preference for an affirmative response. As expected, Yong delivers a fast, unmitigated, and unelaborated preferred response "right" ("对") in line 12, marking the end of the advice sequence.

In line 13, Liu initiates a pre-closing sequence with "okay okay okay, then you go have fun" ("行行行，那你玩吧"). The phrase "won't affect you making money" ("不影响你挣钱了") echoes Yong's earlier troubles-telling. At the end of this turn, Liu suddenly adds the address term "brother" ("兄弟"). This not only projects the closing of the sequence but also constructs a profound brotherhood. In line 14, Yong provides an affiliative response to the closing sequence with "okay okay okay" ("行行行"), treats Liu's "affect you making money" ("影响你挣钱") as an apology using "it's fine" ("没事"), and finally adds the address term "bro" ("哥") at the turn-final position. This reciprocal use of address terms enables both parties to co-construct a harmonious social relationship. Subsequently, the conversation moves swiftly toward closing.

In summary, the use of turn-final address terms operates as a subtle interactional practice for constructing specific social identities and consolidating interpersonal relationships. Whether utilized to increase the likelihood of a request being accepted (as in Extract 5) or to mutually recognize and solidify a relationship during sequence closings (as in Extract 6), these turn-final terms are far more than simple nominal references. Within specific sequential environments, they help manage the social distance between interlocutors and steer the conversation toward the speaker's anticipated direction. This use of address terms highlights how Mandarin speakers use address terms to negotiate identities, manage relational dynamics, and ultimately achieve complex interactional goals.

5. Conclusion

Through CA, this study systematically examines the usage patterns and interactional functions of address terms in Mandarin Chinese telephone conversations, with a specific focus on their use in initiating actions. Address terms are not

merely forms of nominal reference; they are important interactional resources for managing conversational progression, expressing affective stances, and regulating interpersonal relationships.

At the turn-initial position, address terms project a shift in the sequence trajectory, forecasting disalignment with the prior action. At the mid-turn position, they are primarily utilized to intensify the speaker's affective stance, effectively regulating the emotional tone and closing the distance between interlocutors. At the turn-final position, they are frequently associated with identity work and the consolidation of interpersonal relationships, guiding the conversation toward the speaker's expected interactional goals.

These findings contribute to the broader understanding of turn design and sequence organization in CA, highlighting the relationship between micro-linguistic choices and macro-social actions. Uncovering these nuanced interactional practices not only enriches the empirical research on Mandarin conversation but also offers valuable insights for cross-cultural communication and language pedagogy.

While this study primarily focused on telephone interactions and initiating actions, future research could explore the deployment of address terms in face-to-face encounters, where multimodal resources (e.g., gaze, body posture, gestures) operate in tandem with address terms to accomplish social actions and manage the unfolding interaction. Additionally, examining address terms within responsive actions – such as answering questions, granting or declining requests – and across various institutional settings will further unveil their complex roles in diverse social interactions.

References

Braun, F. (2012). *Terms of address: Problems of patterns and usage in various languages and cultures* (Vol. 50). Walter de Gruyter.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Brown, R., & Gilman, A. (1960). The pronouns of power and solidarity. In T. A. Sebeok (Ed.), *Style in language* (pp. 253-276). MIT Press.

Butler, C. W., Danby, S., & Emmison, M. (2011). Address terms in turn beginnings: Managing disalignment and disaffiliation in telephone counseling. *Research on Language and Social Interaction, 44*(4), 338-358.

Chao, Y. R. (1956). Chinese terms of address. *Language*, *32*(1), 217-241.

Clayman, S. E. (2010). Address terms in the service of other actions: The case of news interview talk. *Discourse & Communication, 4*(2), 161-183.

Clayman, S. E. (2013). Agency in response: The role of prefatory address terms. *Journal of Pragmatics, 57*, 290-302.

Diao, S. (2005). Address form conversion and their pragmatic functions. *Journal of Hubei Normal University (Philosophy and Social Science Edition)*, *25*(3), 56-59.

Drew, P. (2008). Conversation analysis. In J. A. Smith (Ed.), *Qualitative psychology: A practical guide to research methods* (2nd ed., pp. 133-159). SAGE Publications.

Ervin-Tripp, S. (1972). On sociolinguistic rules: Alternation and co-occurrence. In J. J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 213-250). Holt, Rinehart and Winston.

Gu, Y. (1990). Politeness phenomena in modern Chinese. *Journal of Pragmatics*, *14*(2), 237-257.

Jacobs, S. (2012). Evidence and inference in conversation analysis. In M. E. McLaughlin (Ed.), *Communication yearbook 11* (pp. 433-443). Routledge.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13-31). John Benjamins.

Jiang, T., & Liu, J. (2014). Arbitrators' identity construction in mediation discourse: From the perspective of code-switching of terms of address. *Social Science Research*, (3), 91-96.

Lerner, G. H. (2003). Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society, 32*(2), 177-201.

Li, H., & Li, J. (2013). A study on address terms in online shopping conversations. *Journal of PLA University of Foreign Languages*, *36*(5), 11-15.

Liang, C. (2001). Pragmatic analysis of address form transformation. *Journal of Luoyang*

Normal University, *20*(6), 70-71.

Rendle-Short, J. (2007). "Catherine, you're wasting your time": Address terms in broadcast political interviews. *Journal of Pragmatics, 39*(9), 1503-1525.

Rendle-Short, J. (2010). 'Mate' as a term of address in ordinary interaction. *Journal of Pragmatics, 42*(5), 1201-1218.

Sacks, H. (1992). *Lectures on conversation* (G. Jefferson, Ed.). Blackwell.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis I*. Cambridge University Press.

Seedhouse, P. (2005). Conversation analysis as research methodology. In K. Richards & P. Seedhouse (Eds.), *Applying conversation analysis* (pp. 251-266). Palgrave Macmillan.

Wu, Y., & Yu, G. (2022). The essence and characteristics of conversation analysis: A sociological perspective. *Studies in Philosophy of Science and Technology, 39*(5), 102-107.

Yu, G. (2021). *What is conversation analysis*. Shanghai Foreign Language Education Press.

Yu, G., & Hao, Q. (2020). The proposal action performed by the turn-constructional component "yaobu" in everyday Mandarin conversations. *Linguistic Research*, (1), 6-18.

Yu, G., & Li, F. (2009). Conversation analysis: A sociological research method respecting linguistic facts. *Science Technology and Dialectics, 26*(2), 14-17.

Yu, G., & Li, Z. (2022). The social interactional significance of detailing in third-party complaint sequences. *Foreign Language and Literature Research*, (2), 86-100.

Yu, G., & Liang, H. (2018). Structural preference for X in the alternative question "X haishi (Y)." *Journal of Foreign Languages, 41*(1), 49-59.

Yu, G., Guo, H., & Wu, Y. (2020). "(You) mean + X" in the third turn of question-answer sequences. *Journal of Foreign Languages, 43*(2), 30-38.

Yu, G., Wu, Y., Drew, P., & Raymond, C. W. (2024). The DIG Mandarin Conversations (DMC) Corpus: Mundane phone calls in Mandarin Chinese as resources for research and teaching. *Chinese Language and Discourse*, *15*(1), 105-141.