

Beyond Pronunciation Correction: The Multidimensional Dialect–Vocal Integration Framework for Culturally Responsive Vocal Music Education

Li Qingyan^{*}, Shafizan Sabri

Faculty of Music and Performing Arts, Universiti Pendidikan Sultan Idris, 35900 Tanjung Malim, Malaysia

DOI Link: <http://dx.doi.org/10.6007/IJARPED/v15-i4/29108>

Published Online: 21 October 2026

Abstract

Dialectal interference in singing is commonly approached as a pronunciation problem requiring correction of non-standard speech patterns. Such a view, however, provides an incomplete explanation of the difficulties experienced by singers who move between regional dialects and Standard Mandarin in formal vocal performance. Singing requires the simultaneous coordination of linguistic articulation, auditory perception, musical pitch, rhythm, vocal control, memory, and expressive communication. This conceptual paper proposes the Multidimensional Dialect–Vocal Integration Framework (MD-VIF), a theoretically derived model for understanding dialectal interference within culturally responsive vocal music education. The framework integrates five dimensions: linguistic transfer, music–language processing, pedagogical mediation, vocal motor learning, and cultural-linguistic identity. Grounded in Language Transfer Theory (Odlin, 2022), Social Constructivist Theory (Vygotsky, 1978), and Experiential Learning Theory (Kolb, 1984), the framework conceptualizes dialectal interference as an interaction between established linguistic patterns and the simultaneous linguistic, musical, and vocal demands of singing. Experiential learning operates as the framework's learning mechanism through awareness, practice, feedback, reflection, application, and refinement. The framework distinguishes professional performance competence from cultural-linguistic identity and proposes culturally responsive vocal competence as the higher-order educational outcome. The Wu Dialect Vocal Enhancement Program (WDVEP) is positioned as a pedagogical application of MD-VIF rather than as the framework itself. The paper contributes a theoretically integrated perspective connecting dialect research, vocal pedagogy, music–language processing, educational psychology, and culturally responsive music education, while generating propositions for future empirical testing.

Keywords: Dialectal Interference, Wu Dialect, Language Transfer, Vocal Music Education, Culturally Responsive Education

Introduction

Vocal music education involves the development of linguistic, musical, technical, cognitive, and expressive competencies. Unlike instrumental performance, singing requires musical sound to be produced through the human voice while simultaneously communicating linguistic meaning. Clear articulation, appropriate vowel formation, consonant precision, tonal control, and intelligible diction therefore contribute to both textual communication and musical interpretation (Bu, 2014; Lu, 2005; Zhu, 2005).

The relationship between language and singing becomes particularly complex when learners' everyday linguistic systems differ from the pronunciation system required for formal vocal performance. In China, regional dialects represent established linguistic systems with distinctive phonological and articulatory characteristics. Studies of dialect influences on singing have identified pronunciation, articulation, resonance, tonal production, and textual clarity as relevant concerns (Guan, 2004; Liu, 2002; Shen, 2017; Xie, 2014; Yang, 2006). Although these studies establish the practical relevance of dialect-related difficulties, their emphasis on identifiable performance problems and corrective responses does not, on its own, explain how prior linguistic patterns interact with simultaneous musical demands or how instruction may support sustained change.

Within the Wu-speaking context, this issue is especially relevant because established dialectal habits may influence the production of Standard Mandarin in singing. The underlying PhD study identifies articulation inaccuracies, nasalization, retroflex consonant confusion, tonal deviations, and related vocal difficulties as relevant concerns. Guan (2004) examined Wu dialect pronunciation habits in vocal teaching, while Yang (2006) addressed negative transfer of Wu dialect on singing. Language transfer theory clarifies the influence of established linguistic knowledge (Odlin, 2022), whereas research on music–language processing explains potential interactions between musical and linguistic abilities (Patel, 2011; Choi & Chan, 2025; Choi et al., 2025; Thompson et al., 2025). However, evidence of music–language interaction does not itself specify how dialect-sensitive pronunciation is learned and maintained during singing. Similarly, culturally responsive approaches emphasize meaningful engagement with learners' linguistic and cultural backgrounds (Barton & Riddle, 2022; Weikle & Coles, 2025), but do not in themselves map the developmental relationship between professional Mandarin singing competence and dialect identity. Taken together, these perspectives address complementary aspects of the problem without specifying their relationships within one explanatory framework for vocal education.

Conceptualizing the issue merely as pronunciation correction is therefore limiting. Singing requires linguistic articulation to occur simultaneously with pitch, rhythm, phrasing, breathing, resonance, memory, and expression. A learner may reproduce a target Mandarin sound in isolation but struggle to maintain it within a demanding musical phrase. The WDVEP responds to this practical problem through four integrated components, Pronunciation Training, Layered Correction, Segmental Practice, and Multilingual Integration. The present paper moves one conceptual level higher by asking why these components need to operate together and through what mechanisms they may contribute to learning.

The central argument is that dialectal interference in singing is not solely a pronunciation problem. MD-VIF advances existing perspectives by specifying conceptual relationships

among five dimensions rather than merely combining them: linguistic transfer accounts for the emergence of interference; music–language processing explains the additional coordination demands of singing; pedagogical mediation identifies how instructional support responds to these demands; vocal motor learning accounts for increasingly stable performance; and cultural-linguistic identity situates professional development within a culturally responsive educational goal. Experiential learning connects pedagogical support with motor development through a recursive cycle of awareness, practice, feedback, reflection, application, and refinement. The conceptual novelty lies in this explicit linkage of problem origin, task demands, instructional support, developmental change, and identity-sensitive outcomes within one framework. On this basis, the paper proposes MD-VIF as a higher-order conceptual framework and positions WDVEP as its pedagogical application.

Conceptualizing Dialectal Interference

Dialectal interference can be understood initially through language transfer. Odlin (2022) explains cross-linguistic influence in terms of the effects that previously acquired linguistic knowledge can have on subsequent language learning and use. Learners therefore do not approach Standard Mandarin pronunciation without prior linguistic knowledge; they bring established phonological categories, articulatory habits, auditory expectations, and patterns of speech production into the learning environment.

Research on dialect and singing supports the relevance of this perspective. Guan (2004), Yang (2006), and Xie (2014) indicate that regional linguistic experience can influence the production of sung language. The educational implication is that dialectal pronunciation should not automatically be interpreted as an individual deficiency. It can instead be understood as an established linguistic system operating in a different communicative and performance context. The pedagogical problem is consequently one of linguistic adaptation and contextual control.

This distinction is central to MD-VIF. The framework does not argue against pronunciation standards in professional vocal performance. Rather, it conceptualizes the acquisition of Standard Mandarin singing pronunciation as an expansion of the learner's linguistic and vocal repertoire. Learners can develop professional pronunciation while retaining awareness and appreciation of their regional linguistic identity.

Existing studies further suggest that dialectal interference can extend beyond isolated pronunciation to articulation, resonance, tonal production, and other aspects of vocal performance (Guan, 2004; Liu, 2002; Xie, 2014; Shen, 2017; Yang, 2006; Luo, 2023; Zhang, 2023). A pronunciation difficulty may originate in linguistic transfer but become manifested through vocal motor behaviour and become more difficult when the learner must coordinate language with melody, rhythm, breathing, resonance, and expression. Effective intervention must therefore integrate linguistic, musical, pedagogical, embodied, and cultural dimensions.

Rationale for the Multidimensional Dialect–Vocal Integration Framework

MD-VIF is founded on the premise that dialectal interference in singing cannot be adequately explained by a single theoretical perspective. Five dimensions are therefore integrated: linguistic transfer, music–language processing, pedagogical mediation, vocal motor learning, and cultural-linguistic identity. Experiential learning is treated as the learning mechanism that

operates within pedagogical mediation and supports progression toward vocal motor learning.

Linguistic Transfer

Students from the Suzhou–Wuxi–Changzhou region enter vocal education with established Wu dialect speech patterns. The underlying PhD framework identifies articulation inaccuracies, inappropriate nasalization, retroflex consonant confusion, and tonal deviations as potential manifestations of established patterns during Mandarin singing. Odlin's (2022) Language Transfer Theory explains how previously acquired linguistic knowledge can influence subsequent production, particularly where the target system requires distinctions that differ from established patterns.

MD-VIF therefore conceptualizes the pathway as: linguistic background → established phonological patterns → cross-linguistic influence → dialectal interference. The observable pronunciation problem is treated as the behavioural manifestation of a deeper interaction between prior linguistic knowledge and the target performance context. Intervention is consequently conceptualized as reconstruction and adaptation rather than correction alone.

Music–Language Processing

Singing requires language to be produced within a musical structure. The singer must coordinate articulation with pitch, rhythm, timing, memory, vocal control, and expressive communication. Patel's (2011) OPERA hypothesis provides a theoretical basis for considering relationships between musical and speech processing. Recent studies further indicate interactions among musical ability, language learning, tone-language experience, pitch perception, and music-to-language transfer (Choi & Chan, 2025; Choi et al., 2025; Thompson et al., 2025). These studies do not demonstrate automatic transfer from musical training to pronunciation; rather, they support the proposition that music and language processing can interact.

MD-VIF therefore identifies three interacting systems: language processing, music processing, and vocal control. Their convergence is termed music–language integration. Pronunciation instruction should consequently progress from foundational isolated practice toward increasingly meaningful musical contexts.

Pedagogical Mediation

Pedagogical mediation is grounded in Vygotsky's (1978) Social Constructivist Theory. Vocal learning involves guided participation, scaffolding, modelling, interaction, and feedback. The teacher is therefore conceptualized not merely as an evaluator of pronunciation errors but as a mediator who helps learners move from existing performance toward increasingly independent performance.

Within MD-VIF, mediation includes target-pronunciation modelling, contrastive explanation, guided practice, corrective feedback, task segmentation, and gradual reduction of support. WDVEP's Layered Correction component illustrates this principle by progressing from basic articulation awareness toward more complex singing applications.

Experiential Learning as The Learning Mechanism

Kolb's (1984) Experiential Learning Theory provides the learning mechanism through which pedagogical support is transformed into performance development. Vocal learning involves experience, reflection, conceptual understanding, experimentation, and repeated application. MD-VIF translates this logic into the cycle: Awareness → Practice → Feedback → Reflection → Application → Refinement.

The cycle is recursive rather than strictly linear. A learner may return to practice after discovering a new difficulty during application, or return to awareness after feedback reveals a previously unnoticed problem. Experiential learning therefore connects pedagogical mediation with the development of increasingly stable vocal behaviour.

Vocal Motor Learning

Knowing how a Mandarin sound should be produced does not guarantee consistent production during singing. Pronunciation requires coordinated articulatory and vocal movement, and established dialectal speech habits may be highly familiar through repeated use. MD-VIF therefore distinguishes declarative awareness from embodied performance.

The framework proposes a progression from declarative awareness to deliberate practice, repeated coordination, contextual application, and increasingly stable vocal behaviour. Vocal motor learning consequently provides the bridge between linguistic knowledge and observable singing performance.

Cultural-Linguistic Identity

Regional dialects can function as markers of community, cultural history, personal experience, and identity. A pedagogy that frames dialect as inherently defective may create an unnecessary conflict between professional development and cultural identity. WDVEP's Multilingual Integration component addresses this issue by developing linguistic flexibility rather than requiring learners to abandon regional identity.

This orientation is consistent with culturally responsive music education, which emphasizes meaningful engagement with learners' cultural backgrounds and communication practices (Barton & Riddle, 2022; Weikle & Coles, 2025). MD-VIF therefore defines the desired outcome not as dialect elimination but contextual linguistic flexibility: the ability to use Standard Mandarin appropriately in professional singing while continuing to recognize and value Wu dialect as part of linguistic heritage.

Table 1

Theoretical and Construct Alignment of MD-VIF

MD-VIF developmental construct	Theoretical foundation	Core mechanism	Primary function
Linguistic transfer	Odlin (2022)	Transfer of established linguistic patterns	Explains the origin of dialectal interference
Music–language processing	Patel (2011); Choi & Chan (2025)	Concurrent linguistic and musical processing	Explains the complexity of pronunciation during singing
Pedagogical mediation	Vygotsky (1978)	Scaffolding, modelling, feedback, guided practice	Facilitates transformation of established patterns
Experiential learning mechanism	Kolb (1984)	Awareness, practice, feedback, reflection, application, refinement	Drives iterative learning
Vocal motor learning	Integrated mechanism	MD-VIF Repeated embodied coordination and contextual application	Stabilizes target vocal behaviour
Cultural-linguistic identity	Culturally responsive education	Contextual linguistic flexibility and identity affirmation	Connects professional competence with cultural recognition

The Central Argument of MD-VIF

The five dimensions of MD-VIF are not independent. They form an interconnected educational system through which learners move from an established linguistic background toward controlled, contextually appropriate, and culturally responsive vocal performance. The proposed conceptual sequence is presented in Figure 1.

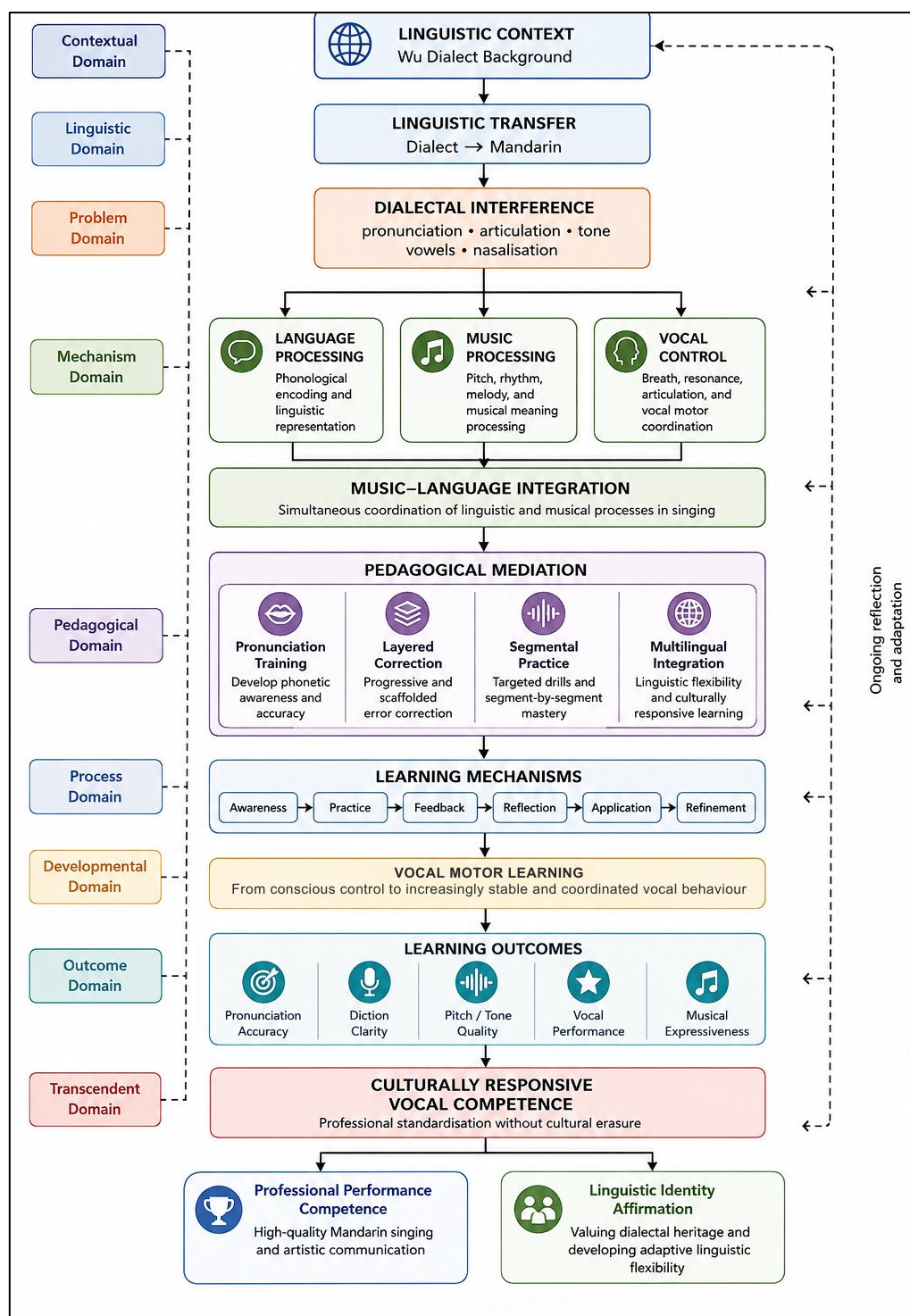


Figure 1. The Multidimensional Dialect–Vocal Integration Framework (MD-VIF) for Culturally Responsive Vocal Music Education.

The sequence is not a rigid linear pathway. Rather, it is recursive. Established linguistic patterns influence pronunciation; musical demands can intensify coordination requirements; pedagogical mediation provides support; experiential learning drives repeated adjustment; and vocal motor learning stabilizes new behaviours. Cultural-linguistic identity remains a contextual dimension throughout the process rather than an endpoint added only after technical learning.

The central proposition is therefore that effective dialect-sensitive vocal education requires integration across linguistic, musical, pedagogical, embodied, and cultural domains. Professional competence and linguistic identity affirmation are treated as complementary outcomes.

From WDVEP to MD-VIF

WDVEP and MD-VIF operate at different levels. WDVEP is an instructional programme comprising Pronunciation Training, Layered Correction, Segmental Practice, Multilingual Integration, instructional materials, learning activities, and assessment procedures. It answers the practical question: what can teachers do to support learners experiencing dialect-related difficulties in Mandarin singing?

MD-VIF is a theoretical-conceptual explanation of why an integrated programme is necessary and through what processes it may produce change. Linguistic transfer explains the initial interference; music–language integration explains the coordination demands; pedagogical mediation provides learning support; experiential learning drives repeated transformation; vocal motor learning stabilizes new performance patterns; and culturally responsive pedagogy ensures that professional development does not require linguistic identity erasure.

The distinction is important for originality. WDVEP is the pedagogical application, whereas MD-VIF is the higher-order conceptual model that explains, organizes, and potentially generalizes the processes activated by such an application.

Table 2

Conceptual Shifts Introduced by MD-VIF

Dimension of change	Conventional emphasis	MD-VIF proposition
Problem definition	The learner pronounces incorrectly.	Established linguistic patterns influence target-language production.
Intervention	Correct the pronunciation.	Integrate linguistic, musical, vocal, cognitive, and pedagogical processes.
Educational outcome	Replace dialectal pronunciation with Standard Mandarin.	Develop professional performance while expanding contextual linguistic repertoire and affirming linguistic identity.

MD-VIF Framework: Constructs and Relationships

MD-VIF is organized as a layered conceptual system comprising a contextual domain, linguistic domain, problem domain, mechanism domain, pedagogical domain, process domain, developmental domain, outcome domain, and transcendent domain. Within this architecture, five core theoretical dimensions organize the explanatory logic: linguistic transfer, music–language processing, pedagogical mediation, vocal motor learning, and cultural-linguistic identity. Experiential Learning Theory operates as the learning mechanism within the process domain rather than as a separate sixth dimension. This distinction allows the framework to remain theoretically coherent while showing how contextual linguistic

experience is progressively transformed into professional and culturally responsive vocal competence.

The domains in Figure 1 should not be interpreted as independent stages or as a deterministic one-way sequence. They represent different analytical levels within the same educational system. The contextual domain establishes the learner's Wu dialect background; the linguistic domain explains transfer; the problem domain identifies dialectal interference; the mechanism domain shows the interaction of language processing, music processing, and vocal control; the pedagogical domain identifies how teachers mediate learning; the process domain shows how learning is enacted; the developmental domain represents the stabilization of vocal behaviour; and the outcome and transcendent domains distinguish performance development from broader culturally responsive competence.

The first relationship is represented by the vertical movement from the contextual domain to the linguistic and problem domains. Learners enter formal vocal education with established Wu dialect patterns. Linguistic transfer provides the explanatory mechanism through which these prior patterns may influence Mandarin production, while dialectal interference represents the observable manifestation of that interaction in the target performance context. The figure therefore locates the pronunciation problem downstream from linguistic experience rather than treating it as an inherent learner deficit.

The second relationship is represented by the transition from the problem domain to the mechanism domain. Once dialectal interference occurs during singing, pronunciation is produced simultaneously with language processing, music processing, and vocal control. The learner must coordinate phonological and articulatory demands with pitch, rhythm, melody, musical meaning, breathing, resonance, and vocal motor coordination. The framework therefore treats pronunciation in singing as context-sensitive rather than as an isolated speech behaviour.

The third relationship links the mechanism domain with the pedagogical domain. Music–language integration creates the instructional condition in which pronunciation must be developed within increasingly musical contexts. Pedagogical mediation responds through the four WDVEP components shown in Figure 1: Pronunciation Training, Layered Correction, Segmental Practice, and Multilingual Integration. The framework thus explains why these components are most appropriately understood as coordinated forms of mediation rather than as unrelated techniques.

The fourth relationship is represented by the transition from pedagogical mediation to the process domain. The teacher's modelling, scaffolding, guided practice, and feedback activate a recursive learning cycle of awareness, practice, feedback, reflection, application, and refinement. This sequence operationalizes Kolb's (1984) experiential learning perspective within dialect-sensitive vocal education. It should be understood as recursive rather than strictly linear because learners may return to earlier processes when new difficulties emerge.

The fifth relationship connects the process domain with the developmental domain. Repeated cycles of awareness, practice, feedback, reflection, application, and refinement provide opportunities for learners to move from conscious control toward increasingly stable

and coordinated vocal behaviour. The framework deliberately avoids assuming complete automaticity. Vocal motor learning therefore represents the stage at which target pronunciation becomes increasingly integrated with the physical and musical requirements of singing.

The sixth relationship concerns the outcome and transcendent domains. Learning does not terminate with technical pronunciation accuracy. The framework distinguishes learning outcomes including pronunciation accuracy, diction clarity, pitch/tone quality, vocal performance, and musical expressiveness from the broader construct of culturally responsive vocal competence. This final construct integrates professional performance competence with linguistic identity affirmation, allowing learners to meet professional Standard Mandarin requirements without treating Wu dialect heritage as something that must be erased.

The layered structure also explains why MD-VIF is multidimensional rather than merely multidisciplinary. The domains are analytically distinct but educationally connected: linguistic theory explains transfer, music–language processing explains simultaneous performance demands, pedagogical theory explains mediation, experiential learning explains the learning process, vocal motor learning explains developmental stabilization, and culturally responsive education explains the identity-sensitive outcome. The framework integrates these perspectives around one educational problem.

The architecture also supports adaptive intervention. Different learners may require different forms or intensities of mediation depending on whether their primary difficulty occurs at the linguistic, musical, vocal, process, or performance level. A learner may require extensive pronunciation awareness but limited musical scaffolding, while another may demonstrate accurate isolated pronunciation but need additional support when target forms are embedded in demanding repertoire. MD-VIF therefore provides an organizing framework for pedagogical decision-making rather than a fixed instructional script.

The outcome domain distinguishes proximal and broader outcomes. Pronunciation accuracy, diction clarity, pitch/tone quality, vocal performance, and musical expressiveness represent observable learning outcomes. Professional performance competence and linguistic identity affirmation are then brought together within culturally responsive vocal competence as the higher-order educational outcome. This distinction allows future studies to examine immediate instructional effects without reducing the framework's ultimate purpose to pronunciation accuracy alone.

Figure 1 should therefore be read as a layered and recursive conceptual system rather than as a simple linear flowchart. The central downward pathway represents the principal developmental logic, while the ongoing reflection and adaptation pathway indicates that learners can revisit earlier domains as performance demands change. The figure also distinguishes the WDVEP pedagogical architecture from the broader MD-VIF explanation of how linguistic, musical, pedagogical, developmental, and cultural processes interact. To further clarify the relationships among these domains, Table 3 maps each core construct onto its central explanatory question, proposed mechanism, and potential empirical indicator. This alignment demonstrates how the conceptual framework can be translated into observable dimensions for future empirical investigation. Experiential learning is presented in Table 3 as

a learning mechanism rather than as a sixth core MD-VIF dimension, reflecting its role in connecting pedagogical mediation with vocal motor development.

Table 3

Construct–Mechanism–Outcome Alignment in the MD-VIF Framework

Construct/Mechanism	Key question	Mechanism in MD-VIF	Potential indicator
Linguistic transfer	Why does interference emerge?	Influence of established Wu dialect phonological and articulatory patterns on Mandarin production.	Dialect-sensitive pronunciation features; contrastive production accuracy
Music–language processing	Why can pronunciation become more difficult during singing?	Concurrent processing of linguistic, musical, and vocal demands.	Accuracy across isolated, phrase-level, and repertoire tasks
Pedagogical mediation	How is change supported?	Modelling, scaffolding, guided practice, feedback, and gradual release.	Quality/frequency of feedback; learner independence
Experiential learning (learning mechanism)	How does practice become learning?	Awareness → practice → feedback → reflection → application → refinement.	Reflective responses; improvement across repeated trials
Vocal motor learning	How does knowledge become stable behaviour?	Repeated embodied coordination and contextual application.	Consistency of target pronunciation during performance
Cultural-linguistic identity	How can professional learning coexist with identity?	Contextual linguistic flexibility and identity affirmation.	Attitudes toward dialect; perceived cultural inclusion; contextual repertoire

Table 3 reinforces the multidimensional logic of MD-VIF by showing that each construct addresses a different explanatory question while remaining connected to the overall developmental process. Linguistic transfer explains the emergence of interference, music–language processing explains the complexity of pronunciation during singing, and pedagogical mediation specifies how instructional support is provided. Experiential learning explains the process through which learners repeatedly adjust their performance, while vocal motor learning accounts for the development of increasingly stable behaviour. Finally, cultural-linguistic identity extends the framework beyond technical performance by positioning professional competence and linguistic identity affirmation as complementary outcomes. Thus, the table translates the conceptual architecture represented in Figure 1 into a set of relationships that can be examined empirically in future research.

Theoretical Contribution of MD-VIF

MD-VIF contributes to the literature by reframing dialectal interference from a narrow technical problem into a layered educational phenomenon. Its originality lies not in proposing that pronunciation, music, pedagogy, motor learning, or cultural identity are individually important, but in specifying how these domains can be conceptually related within dialect-sensitive vocal education. Figure 1 makes this contribution visible by organizing the framework from contextual conditions through mechanisms and pedagogical processes toward developmental and transcendent outcomes.

The first theoretical contribution is a transfer-based account of the problem domain. The contextual and linguistic domains establish that learners bring existing Wu dialect resources into formal vocal education, while the problem domain identifies dialectal interference as a possible manifestation of linguistic transfer. This positioning is consistent with Odlin's (2022) account of language transfer and shifts interpretation away from an unexplained learner deficit.

The second contribution is the explicit integration of language processing, music processing, and vocal control as interacting mechanisms. Rather than treating singing pronunciation as speech with a melody added to it, MD-VIF conceptualizes singing as a task in which linguistic and musical requirements are coordinated through vocal behaviour. The framework therefore offers a conceptual basis for examining why pronunciation control may vary according to musical and vocal demands.

The third contribution is the positioning of pedagogical mediation as the bridge between identified difficulties and learning processes. Vygotsky's (1978) social constructivism explains the role of modelling, scaffolding, guided practice, and feedback, while Kolb's (1984) experiential learning provides the process through which these supports can be converted into repeated learning cycles. In Figure 1, these ideas are represented through the pedagogical and process domains rather than as separate instructional theories.

The fourth contribution is the explicit connection between linguistic learning and vocal motor development. The developmental domain recognizes that pronunciation competence in singing requires more than declarative knowledge or isolated accuracy. Learners must increasingly coordinate target pronunciation with musical and vocal demands, making stable and coordinated vocal behaviour an important developmental outcome.

The fifth contribution is the placement of cultural-linguistic identity within the framework's transcendent outcome rather than treating culture as an optional addition to technical instruction. Culturally responsive vocal competence is defined as the capacity to meet professional performance expectations while affirming linguistic heritage and developing contextual linguistic flexibility. This formulation allows professional standardization and cultural recognition to be conceptualized as complementary educational goals.

Operationalizing the Framework Through Testable Relationships

The conceptual value of MD-VIF depends ultimately on whether its relationships can be tested. The framework is therefore intended to generate empirical questions rather than function only as a descriptive diagram. Its propositions can be investigated through

quantitative, qualitative, and mixed-methods designs, with different methods addressing different components of the model.

For the linguistic-transfer component, future studies could assess the relationship between the strength of learners' Wu dialect pronunciation patterns and their Mandarin singing performance. Measures may include contrastive pronunciation tasks, auditory discrimination tasks, or expert-rated singing diction. Such studies would help determine whether the proposed transfer pathway is observable and which linguistic features are most resistant to change.

For the music–language component, researchers could compare pronunciation performance across conditions with different levels of musical demand. The same target sounds could be assessed in isolated production, spoken phrases, rhythmically structured speech, melodic fragments, and complete repertoire. Differences across conditions would provide evidence about the interaction between linguistic accuracy and musical context.

For pedagogical mediation, experimental or quasi-experimental studies could compare different levels or forms of instructional support. One group might receive conventional pronunciation correction, while another receives layered mediation involving modelling, guided practice, feedback, reflection, and progressive transfer to repertoire. Such comparisons would directly test the framework's claim that integration is more educationally appropriate than isolated correction.

For vocal motor learning, longitudinal measurement is particularly important. Pronunciation improvement observed immediately after instruction may not represent stable learning. Repeated assessments across instructional phases could determine whether learners retain target production when musical complexity increases. This would allow researchers to distinguish short-term correction from more durable vocal motor learning.

Finally, the cultural-linguistic dimension requires measures that capture learner experience rather than performance alone. Interviews, reflective journals, or validated identity and culturally responsive learning measures could examine whether learners perceive Standard Mandarin acquisition as an additional repertoire or as a rejection of regional identity. Such evidence would help evaluate the framework's claim that professional competence and linguistic identity affirmation can coexist.

Theoretical Propositions of MD-VIF

The MD-VIF framework is intended not only to describe the relationship among linguistic, musical, pedagogical, developmental, and cultural dimensions, but also to generate propositions that can guide future empirical investigation. The propositions below translate the conceptual relationships represented in Figure 1 into theoretically testable statements. They follow the developmental logic of the framework, beginning with the influence of established linguistic experience and progressing toward professional and culturally responsive vocal competence.

Proposition 1: Linguistic Transfer and Dialectal Interference.

Learners' established Wu dialect linguistic patterns are expected to influence the accuracy and consistency of Standard Mandarin pronunciation in singing, particularly where the target language requires phonological or articulatory distinctions that differ from established dialectal patterns.

- Proposition 2: Music–Language Processing and Performance Demand.

Greater simultaneous linguistic, musical, and vocal demands are expected to increase the coordination requirements associated with target pronunciation. Consequently, pronunciation accuracy may differ across isolated pronunciation tasks, phrase-level singing tasks, and complete musical repertoire.

- Proposition 3: Pedagogical Mediation and Performance Development.

Pedagogical mediation through Pronunciation Training, Layered Correction, Segmental Practice, and Multilingual Integration is expected to facilitate learners' movement from awareness of pronunciation differences toward increasingly independent and contextually appropriate performance control.

- Proposition 4: Experiential Learning and Vocal Development.

The recursive learning process of awareness, practice, feedback, reflection, application, and refinement is expected to support the transformation of explicit pronunciation knowledge into increasingly stable and coordinated vocal behaviour.

- Proposition 5: Contextual Application and Vocal Motor Learning.

Progressive transfer from targeted segmental practice to increasingly authentic musical repertoire is expected to strengthen learners' ability to maintain target pronunciation under increasing linguistic, musical, and vocal demands.

- Proposition 6: Culturally Responsive Vocal Competence.

Culturally responsive pedagogy that recognises regional linguistic heritage while maintaining appropriate professional pronunciation expectations is expected to support both professional performance competence and linguistic identity affirmation.

Table 4

Theoretical Propositions and Empirical Implications of MD-VIF

Proposition	MD-VIF domain	Core relationship	theoretical	Potential empirical focus
P1: Linguistic Transfer and Dialectal Interference	Contextual Linguistic Problem	→ Established patterns Standard pronunciation singing.	Wu dialect influence Mandarin during	Relationship between dialect-sensitive phonological patterns and Mandarin singing pronunciation accuracy.
P2: Music–Language Processing and Performance Demand	Problem Mechanism	→ Simultaneous music, and vocal influence coordination.	language, demands pronunciation	Differences in pronunciation accuracy across isolated, phrase-level, and repertoire-based tasks.

P3: Pedagogical Mediation and Performance Development	Mechanism and Pedagogical	→	Structured mediation supports movement from pronunciation awareness toward independent performance control.	Effects of modelling, scaffolding, feedback, layered correction, and segmental practice on performance development.
P4: Experiential Learning and Vocal Development	Pedagogical Process	→	Recursive awareness, practice, feedback, reflection, application, and refinement facilitate learning.	Changes in pronunciation and vocal performance across repeated learning cycles.
P5: Contextual Application and Vocal Motor Learning	Process and Developmental	→	Progressive transfer into authentic musical contexts supports increasingly stable and coordinated vocal behaviour.	Maintenance of target pronunciation as musical and vocal complexity increases.
P6: Culturally Responsive Vocal Competence	Outcome Transcendent	→	Professional standardisation and linguistic identity affirmation can be developed as complementary outcomes.	Relationships among culturally responsive pedagogy, professional performance competence, and linguistic identity affirmation.

Table 4 clarifies the theoretical status of the propositions by linking each proposition to a specific domain relationship within MD-VIF and identifying potential directions for empirical examination. The propositions are not presented as confirmed causal relationships; rather, they represent theoretically derived expectations that require empirical testing. This distinction is important because MD-VIF is proposed in the present paper as a conceptual framework, while the empirical validation of its relationships remains a subject for future research.

Taken together, the six propositions establish the explanatory logic of MD-VIF. Propositions 1 and 2 explain the emergence and complexity of dialectal interference; Propositions 3 and 4 explain how pedagogical and experiential processes may facilitate change; Proposition 5 addresses the development and stabilisation of performance behaviour; and Proposition 6 extends the framework toward its broader culturally responsive outcome. The propositions therefore provide a bridge between the conceptual architecture presented in Figure 1 and future empirical studies designed to evaluate the framework.

Pedagogical Implications

MD-VIF suggests that teachers should interpret dialect-sensitive pronunciation difficulties as a linguistic and vocal learning issue rather than as isolated pronunciation errors. Instruction can begin by developing learners' awareness of contrasts between their established pronunciation patterns and the target Standard Mandarin forms. This initial awareness can then be followed by targeted practice, modelling, and feedback.

Second, pronunciation instruction should progressively move from isolated linguistic practice into increasingly authentic musical contexts. Segmental exercises remain useful as a foundation, but their purpose is to prepare learners for application in words, phrases, melodic

fragments, complete phrases, and ultimately musical repertoire. This progression enables learners to maintain pronunciation accuracy while simultaneously managing pitch, rhythm, vocal control, and expressive demands.

Third, feedback should be progressive and layered. Teachers may initially provide explicit modelling and corrective feedback before gradually encouraging learners to monitor their own pronunciation, reflect on performance, and make independent adjustments. This approach reflects the pedagogical mediation and experiential learning dimensions of MD-VIF and supports the transition from teacher-supported performance toward greater learner independence.

Fourth, culturally responsive practice should avoid framing Wu dialect as a linguistic deficit. Teachers can acknowledge learners' regional linguistic backgrounds while clearly distinguishing between dialect use and the professional contexts in which Standard Mandarin pronunciation is required. The pedagogical objective is therefore not dialect elimination but the development of contextual linguistic flexibility, enabling learners to meet professional performance standards while maintaining an understanding and appreciation of their linguistic heritage.

At the curriculum level, MD-VIF supports a progression from foundational pronunciation awareness to increasingly authentic vocal performance. Curriculum designers can identify pronunciation features that may be influenced by learners' dialect backgrounds, develop targeted learning activities, and progressively embed these features within musical tasks. Such sequencing can help manage the linguistic, cognitive, and motor demands placed on learners while maintaining a clear connection between pronunciation instruction and authentic performance.

Assessment should similarly reflect the multidimensional nature of MD-VIF. A single overall pronunciation score may conceal differences between isolated pronunciation accuracy, musical-context accuracy, vocal coordination, and expressive performance. Assessment can therefore distinguish among pronunciation accuracy, diction clarity, pitch/tone quality, vocal performance, and musical expressiveness, consistent with the learning outcomes identified in the framework. Such differentiation can help teachers determine whether difficulties originate primarily from linguistic transfer, musical integration, pedagogical support, or the transfer of learned pronunciation into performance.

Teacher education also has an important role. Teachers require sufficient linguistic knowledge to recognize dialect-sensitive pronunciation features, musical knowledge to understand how pronunciation interacts with repertoire, and pedagogical knowledge to provide appropriately sequenced mediation. Equally important is culturally responsive competence, enabling teachers to maintain professional pronunciation expectations while communicating correction in ways that do not position regional linguistic identity as inferior.

Implications for Future Empirical Research

MD-VIF provides a basis for empirical investigation because its conceptual relationships can be examined through quantitative, qualitative, and mixed-methods approaches. Future studies could examine whether the strength of dialectal influence is associated with specific

pronunciation and vocal outcomes, whether pronunciation accuracy changes as musical demands increase, and whether integrated music–language instruction produces different outcomes from isolated pronunciation training.

Experimental and quasi-experimental studies could compare conventional pronunciation-focused instruction with integrated interventions informed by MD-VIF. Such studies could examine whether the integration of Pronunciation Training, Layered Correction, Segmental Practice, and Multilingual Integration produces stronger transfer from pronunciation knowledge to singing performance. Longitudinal studies could further investigate whether repeated cycles of awareness, practice, feedback, reflection, application, and refinement contribute to increasingly stable and coordinated vocal behaviour.

The framework also lends itself to mixed-methods investigation. Quantitative measures could examine changes in pronunciation accuracy, diction clarity, pitch/tone quality, and vocal performance, while interviews, reflective journals, or learner narratives could explore how students experience the relationship between professional Standard Mandarin competence and their regional linguistic identity. This combination would be particularly important for examining the framework's proposition that professional performance competence and linguistic identity affirmation can develop as complementary outcomes.

Future research could also examine the applicability of MD-VIF beyond the Wu dialect context. Other regional Chinese dialects and multilingual or dialect-sensitive singing environments could be investigated to determine whether the proposed relationships remain consistent or require contextual modification. Comparative research across dialect groups could consequently help determine which elements of MD-VIF represent broader principles of dialect-sensitive vocal education and which require adaptation to specific linguistic and cultural contexts.

Ultimately, empirical research can move MD-VIF from a contextually grounded conceptual framework toward an empirically evaluated theoretical model. Such research should not assume that the propositions are already established; rather, the value of MD-VIF lies in providing a structured set of relationships and propositions that can be systematically tested, refined, and potentially extended.

Synthesis of the Framework

MD-VIF conceptualises dialect-sensitive vocal education as an interconnected and recursive educational system rather than a sequence of isolated instructional activities. As illustrated in Figure 1, learners enter the learning process with established linguistic resources shaped by their Wu dialect background. These resources may influence Mandarin pronunciation through linguistic transfer, producing dialectal interference when established phonological and articulatory patterns differ from target performance requirements. During singing, these linguistic demands interact with music processing and vocal control, creating the conditions for music–language integration.

Pedagogical mediation provides the structured support through which learners respond to these demands. The four WDVEP components, Pronunciation Training, Layered Correction, Segmental Practice, and Multilingual Integration provide practical mechanisms for guiding this process. Through the experiential learning cycle of awareness, practice, feedback, reflection, application, and refinement, learners progressively apply target pronunciation within increasingly complex musical contexts. Vocal motor learning then represents the developmental process through which target pronunciation becomes increasingly stable and coordinated with vocal and musical performance.

The framework ultimately distinguishes between learning outcomes and culturally responsive vocal competence. Pronunciation accuracy, diction clarity, pitch/tone quality, vocal performance, and musical expressiveness represent immediate or proximal outcomes, while professional performance competence and linguistic identity affirmation represent complementary dimensions of the broader educational outcome. In this sense, culturally responsive vocal competence does not require learners to abandon their regional linguistic identity in order to achieve professional Standard Mandarin performance.

This synthesis demonstrates that MD-VIF is more than a combination of established theories. Its conceptual contribution lies in specifying how linguistic transfer, music–language processing, pedagogical mediation, experiential learning, vocal motor development, and cultural-linguistic identity can be understood as interconnected dimensions of one educational phenomenon. The framework therefore provides a common conceptual language for examining dialect-sensitive singing across linguistics, vocal pedagogy, music education, educational psychology, and culturally responsive education.

Importantly, MD-VIF remains a theoretical-conceptual framework rather than an empirically validated causal model. The relationships proposed in the framework and formalised in the theoretical propositions require future empirical investigation. Its contribution at this stage is to provide a coherent explanatory structure from which such investigation can proceed.

Conclusion

Dialectal interference in singing should not be reduced to a problem of incorrect pronunciation. It emerges from the interaction between established linguistic patterns and the simultaneous linguistic, musical, vocal, pedagogical, and cultural demands of singing. The Multidimensional Dialect–Vocal Integration Framework responds to this complexity by integrating five dimensions: linguistic transfer, music–language processing, pedagogical mediation, vocal motor learning, and cultural-linguistic identity. Experiential learning provides the recursive mechanism through which awareness, practice, feedback, reflection, application, and refinement support movement from knowledge toward embodied performance.

The conceptual distinction between WDVEP and MD-VIF is central to the contribution of this paper. WDVEP represents a structured pedagogical application, whereas MD-VIF provides a higher-order explanation of why integrated intervention is necessary and how its components may interact. The framework consequently shifts the field from error to transfer, from correction to integration, and from standardization to culturally responsive competence.

MD-VIF does not reject the importance of professional pronunciation standards. Instead, it proposes that such standards can be developed as an additional linguistic and vocal repertoire while learners retain recognition of their regional linguistic identity. The framework therefore offers a foundation for future empirical research and for the development of culturally responsive approaches to dialect-sensitive vocal music education.

References

- Barton, G., & Riddle, S. (2022). Culturally responsive and meaningful music education: Multimodality, meaning-making, and communication in diverse learning contexts. *Research Studies in Music Education*, 44(2), 345–362. <https://doi.org/10.1177/1321103X211009323>
- Bu, Y. M. (2014). Interpretation of linguistic elements in vocal art. *Music Time and Space*, (6), 178–179.
- Choi, W., & Chan, L. Y. (2025). The comparative effects of Cantonese and Mandarin tone language backgrounds on musical pitch perception. *Scientific Reports*, 15, Article 31558. <https://doi.org/10.1038/s41598-025-17171-2>
- Choi, W., Lai, V. K. W., Kong, S. H., & Bautista, A. (2025). Examining the cognitive and perceptual perspectives of music-to-language transfer: A study of Cantonese–English bilingual children. *Journal of Experimental Child Psychology*, 249, Article 106069. <https://doi.org/10.1016/j.jecp.2024.106069>
- Guan, Y. (2004). Correction of Wu dialect pronunciation habits in vocal music teaching. *Symphony: Journal of Xi'an Conservatory of Music*, (2), 80–82.
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.
- Liu, C. (2002). A brief discussion on the influence of dialect on the art of singing. *Journal of Ningbo Education College*, (4), 57–59.
- Lu, Y. (2005). An important aspect of singing language training: *Reading aloud*. *Playwright*, (6), 128–129.
- Luo, S. (2023). *The influence of Jiujiang dialect phonation habits on vocal learning articulation and solutions [Master's thesis, Shanghai Conservatory of Music]*. China National Knowledge Infrastructure (CNKI). <https://cdmd.cnki.com.cn/Article/CDMD-10278-1023507082.htm>
- Odlin, T. (2022). *Explorations of language transfer*. Multilingual Matters. <https://doi.org/10.21832/ODLIN9547>
- Patel, A. D. (2011). *Music and language: From behavior to neuroscience*. Oxford University Press.
- Shen, K. (2017). *The influence of dialect on vocal performance [Master's thesis, Central China Normal University]*. China National Knowledge Infrastructure (CNKI). <https://cdmd.cnki.com.cn/Article/CDMD-10511-1017264688.htm>
- Thompson, R. M., Salig, L. K., & Slevc, L. R. (2025). Is musical ability related to second-language acquisition? A meta-analysis. *Royal Society Open Science*, 12(1), Article 241193. <https://doi.org/10.1098/rsos.241193>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Weikle, E., & Coles, D. (2025). Developing music-specific culturally responsive teaching professional development: An autoethnographic action research study. *Journal of Teacher Development and Education*, 3(1), 1–13. <https://doi.org/10.29329/journaltded.40>

- Xie, Y. L. (2014). *The influence of local dialects on the performance of Chinese songs and solutions* [Master's thesis, South China University of Technology]. China National Knowledge Infrastructure (CNKI). <https://cdmd.cnki.com.cn/Article/CDMD-10561-1014065936.htm>
- Yang, C. H. (2006). *The causes and countermeasures of negative transfer of Wu dialect on singing* [Master's thesis, Fujian Normal University].
- Zhang, C. Y. (2023). Analysis of common dialect habit issues for vocal learners in the Longdong area. *Drama House*, (12), 106–108.
- Zhu, W. H. (2005). The complete process of vocal language phonation: Articulation and enunciation. *Chinese Character Culture*, (1), 44–46.