STUDIA UBB PSYCHOL.-PAED., LXXI, 1, 2026, p. 303-320
(Recommended Citation)
DOI:10.24193/subbpsyped.2026.1.13

The Vowel as an Acoustic, Emotional, and Biological Stimulus in Music Pedagogy

Milena PETROVIĆ[1]

ABSTRACT. This research explores the multidimensional role of vowels as acoustic, emotional, and biological stimuli within the context of music pedagogy. The study argues that vowels are not neutral phonological units but active triggers of affective and physiological processes. Using the formant architecture of the standard Serbian language as a baseline, the analysis demonstrates how specific vowel configurations are instinctively associated with spatial and emotional archetypes. Through the lens of the facial feedback hypothesis, the research shows that the mechanical activation of orofacial muscles during vowel production functions as a somatic stimulus, modulating the performer's internal affective state. Neurophysiological evidence further confirms that vowels elicit stable cortical responses, while empirical studies highlight their capacity to regulate biomarkers of stress, immunity, and social bonding through collective singing. By integrating insights from acoustic phonetics, biomechanics, and psychophysiology, this work repositions the vowel as a fundamental pedagogical tool – one that simultaneously shapes technical vocalization, authentic emotional expression, and biological regulation. The findings advocate for a holistic approach to music education, particularly for children and adolescents, where vowel articulation fosters systemic balance, resilience, and artistic integrity.

Keywords: vowels, phonoaesthetics, formant architecture, music pedagogy, biological regulation, emotional resonance

INTRODUCTION

Human speech and singing have traditionally been conceptualized as systems for transmitting lexical and musical information. However, contemporary research in phonetics, psychology, and musicology demonstrates that vowels

[1] Faculty of Music, University of Arts in Belgrade, Serbia. Email: milena.petrovic@fmu.bg.ac.rs

possess an intrinsic expressive and biological power that transcends their conventional linguistic functions. Building upon the foundations established in *The Importance of Vowels in Music Education* (Petrović, 2017), this study deepens the inquiry by positioning vowels not only as acoustic symbols of affect but also as biological stimuli capable of modulating emotional and physiological states.

The central premise is that vowels are autonomous acoustic signals whose articulation engages the orofacial musculature, respiratory system, and neurophysiological pathways. Their spectral architecture, defined by the first (F1) and second (F2) formants, provides a stable framework for cross-modal associations with spatial, cognitive, and emotional archetypes, as evidenced by recent behavioral and motor-response studies (Vainio, Wikström & Vainio, 2023; Vainio, Myllylä & Vainio, 2024). At the same time, recent neurophysiological and neuroimaging evidence confirms that vowel production stabilizes cortical responses and directly modulates autonomic nervous system function (Morales-Luque et al., 2025; Saus, Seither-Preisler & Schneider, 2025), while psychophysiological studies show that collective singing and vowel-centered chanting regulate biomarkers of stress, immunity, and social bonding (Canessa-Pollard, Anikin & Reby, 2025; Good & Russo, 2021).

To address this complex, triadic nature of vowels, this study adopts an integrative, interdisciplinary methodology. Rather than limiting the analysis to a single analytical framework, this approach synthesizes empirical data and theoretical insights from acoustic phonetics, vocal biomechanics, affective neuroscience, and music pedagogy. This specific methodology was chosen because the phenomenon of vowel articulation in singing cannot be fully captured by any single discipline. An isolated acoustic analysis would overlook the somatic, emotional modulation experienced by the performer, while a purely physiological approach would fail to account for the phonoaesthetic and aesthetic implications in vocal artistry. Bridging these diverse fields offers a deeper, more unified perspective that justifies treating vowel production as a central, multi-layered practice in music education.

This research is structured into four sections, followed by a conclusion. Section 1 establishes the acoustic foundation of vowels and their phonoaesthetic associations with size and space. Section 2 examines the biomechanical constraints of the vocal apparatus, highlighting the physiological architecture that shapes vowel identity. Section 3 explores the somatization of emotion through the facial feedback hypothesis, demonstrating how vowel articulation modulates affective states. Section 4 investigates vowels as archetypes of affective meaning and their musical manifestation, presenting empirical evidence on their role in emotional and biological regulation. The concluding section synthesizes these findings, underscoring their implications for music pedagogy and the holistic role of vowels as acoustic, emotional, and biological stimuli.

By integrating insights from acoustic phonetics, biomechanics, neuroendocrinology, and pedagogy, this study repositions the vowel as a holistic instrument – simultaneously technical, emotional, and biological. The thesis advanced here is that in music education, vowels must be treated as multidimensional stimuli: they are technical elements of vocalization, carriers of emotional resonance, and biological regulators. Recognizing this triadic role enables pedagogy to cultivate authentic vocal expression, systemic balance, and resilience in the next generation of performers.

PHONESTHETICS AND THE ACOUSTIC STRUCTURE OF SOUND SYMBOLISM

Phonoesthetics explores the aesthetic qualities of sounds and their profound influence on human perception and semantic attribution. Within this framework, auditory signals are conceptualized not merely as neutral conduits of information, but as carriers of latent symbolic values. These values manifest through euphony – the inherent pleasantness of vocal sounds – where specific configurations of vowels and consonants generate harmonic resonances. Beyond mere acoustics, the qualitative dimensions of these sounds, including timbre, duration, and resonance, dictate the perceived structural harmony of a linguistic or musical utterance.

The scope of phonoesthetics transcends linguistic boundaries, intersecting with the broader fields of musical and spatial perception. As argued in Petrović (2017), the vowel functions as the primary vehicle for this intersection. This is supported by Milanković et al. (2003), who suggest that tonal perception possesses an inherent spatial dimension: high-pitched tones are spontaneously experienced as being „above" and „to the right", while lower frequencies are perceived as „below" and „to the left". More recent behavioral studies confirm sound-symbolic associations between pitch and spatial orientation, with high tones consistently linked to „up" and low tones to „down" (Vainio, Wikström & Vainio, 2023). Recent acoustic studies (Brandner, Frank & Sontacchi, 2022) further demonstrate that sung vowels exhibit distinct horizontal and vertical directivity patterns, with /i:/ and /u:/ projecting energy forward. These findings confirm that vowel shaping is central not only to resonance but also to the spatial projection of the singing voice. Such insights confirm that vowels, through their unique acoustic architecture, actively construct spatial associations that bypass conventional linguistic functions.

The acoustic identity of a vowel is fundamentally governed by its formants – the resonant frequencies generated within the vocal tract. As illustrated in Table 1, these frequencies serve as the physical substrate for phonoesthetic associations. In standard Serbian, following the foundational research of Kostić et al. (1964), vowels are categorized by two primary parameters: tongue height, which correlates with the first formant (F1 – degree of openness), and the place of articulation, which determines the second formant (F2 – front/back dimension). A defining characteristic of this architecture is the inverse relationship between formants: F1 reaches its maximum in the open vowel /a/ and its minimum in the closed vowels /i/ and /u/. Conversely, F2 peaks in front vowels (/i/, /e/) and diminishes in back rounded vowels (/u/, /o/). While these mean values fluctuate across genders and individual speakers, they provide a stable spectral framework for the symbolic and emotional qualities explored in the subsequent sections of this research.

Table 1. *Acoustic Characteristics of Standard Serbian Vowels (Kostić et al., 1964).*

Vowel	F1 (Hz) – Vowel Openess	F2 (Hz) – Place of Articulation	Position Description
/**i**/	250–350	2100–2500	High, front
/**e**/	450–550	1700–2000	Mid, front
/**a**/	700–850	1200–1400	Low, central
/**o**/	450–550	900–1100	Mid, back
/**u**/	250–350	700–900	High, back

Howell (2015) introduced the concept of Absolute Spectral Tone Color (ASTC), later expanded by Bozeman (2017), as a psychoacoustic framework explaining how listeners perceive frequencies as consistent „tone colors“ analogous to vowel qualities and brightness. This framework situates vowel perception not only within the physical domain of formants but also within the perceptual domain of tonal coloration, bridging acoustics and phonoesthetic experience.

Integral to this acoustic structure is sound symbolism (or phonosemantics), which investigates how meaning arises directly from phonetic properties. Cross-culturally, vowels have often been mapped onto dimensions of magnitude: Ultan (1978) discovered the front-small and back-large relation in 136 languages, while Sapir (1929) experimentally demonstrated that speakers link /i/ with diminutive objects and /a/ with larger ones. Newman (1933) expanded this into a systematic scale of phonetic magnitude, showing that back vowels are perceived as „larger" than front vowels. From a musicological and physiological perspective, Nettl (1954) observed that in American Indians songs high tones are sung on syllables with front vowels, whereas lower tones alins with back vowels.

Recent research has confirmed and extended these early findings. Ekström (2022) reviewed the field of size-sound symbolism, highlighting how vowel quality, pitch, and phonation type interact to produce consistent associations of magnitude across languages. Vainio, Myllylä, and Vainio (2024) demonstrated that phoneme-response interactions extend beyond perception into motor behavior: vowels such as /i/ facilitate precise grasping actions, while /ɑ/ supports stronger, forceful grips, linking sound symbolism to embodied cognition. Taken together, scholars such as Paget (1930) emphasized that sound symbolism is grounded in the physical volume of the oral cavity during articulation. This perspective suggests a complex, embodied relationship between physiological space and the perception of magnitude – a transition from abstract symbolism to phonetic reality shaped by the biomechanical constraints of the vocal apparatus.

Phonoesthetics posits that vowels serve as vehicles for emotional signaling, where their inherent sound symbolism – defined by shape, duration, and resonance – can trigger affective responses. During articulation, the activation of facial musculature is not neutral; these physiological movements are linked to internal emotional states. For instance, the production of open vowels such as /a/ requires a wider oral aperture, a configuration often associated with expressions of joy or surprise. Consequently, vowels are commonly described as carrying specific phonoesthetic qualities, like /a/ as being open and robust, /i/ bright and acute, and /o/ closed and resonant. These theoretical postulates of phonoesthetics find their biological substrate in the interplay between facial expression, respiratory modulation, and vocal articulation.

These acoustic associations, however, are not merely abstract constructs; they are deeply rooted in the physical body. The bridge between sound symbolism and phonetic reality is built upon the biomechanical constraints of the vocal apparatus. Therefore, understanding these sounds as stable symbolic markers requires a rigorous examination of the physiological architecture that governs their articulation.

BIOMECHANICAL CONSTRAINTS AND PHYSIOLOGICAL MODULATION

The transition from abstract acoustic symbolism to phonetic reality is mediated by the specific biomechanical architecture of the vocal apparatus. While vowels serve as primary vehicles for emotional signaling, their efficacy is strictly governed by the physical parameters of the vocal tract. The relationship between vowel identity and affective expression is not uniform; rather, it is shaped by articulatory constraints that either facilitate or inhibit emotional contrast. Within this hierarchy, the high-front vowel /i/ shows relative resistance to significant acoustic-emotional modulation, maintaining a stable profile compared to the more expansive and variable configurations of /a/ or /u/. This perspective extends Flemming's (2004) account of articulatory and prosodic constraints on vowel stability to affective expression, and is consistent with Lee et al. (2005), who found that /i/ exhibits comparatively less variation across affective states due to the physical limitations of lingual movement and oral geometry. In this sense, the configuration of /i/ offers reduced acoustic space for contrast, making it a comparatively consistent marker within the vocal spectrum. Such biomechanical stability can serve as a perceptual anchor for listeners, helping phonetic identity remain reliably recognizable even under emotional arousal or physiological constraints.

The manifestation of emotional affect through vocalization is fundamentally predicated upon the biomechanics of the articulatory system. Distinct vowels are shaped through the coordinated activity of the pharyngeal, oral, and nasal cavities, which collectively modulate the phonatory airstream. This modulation functions not merely as a mechanical filter but as a sophisticated process of acoustic shaping, wherein the geometry of the vocal tract determines the spectral characteristics of the sound. Helmholtz (1863) demonstrated that vowel timbre (Klangfarbe) arises from the resonant configuration of the vocal tract, grounding the metaphor of vowel color in the physical acoustics of the human body. From this foundation, two trajectories emerge: embodied accounts in phonetics and cognitive science (Sapir, 1929; Newman, 1933; Vainio et al., 2024; Ekström, 2022) emphasize the inseparability of vowel identity from physiological and affective dimensions of vocal production; Howell (2015) extends Helmholtz's acoustic legacy into vocal pedagogy through the psychoacoustic concept of Absolute Spectral Tone Color, framing vowel timbre as a perceptual analogue to color perception in light, and thereby reinforcing its embodied and pedagogical significance.

Variations in articulatory configurations produce diverse psychoacoustic effects: front vowels are associated with greater brightness than back vowels (Newman, 1933). Beyond individual vowel targets, the global positioning of the tongue and the tension within the orofacial complex can be understood as indicators of the articulatory state. Ladefoged (1964) was exploring the anatomical constraints of speech and collected data on tongue muscle positions to hypothesize how muscular forces shape vowels. Nastase et al. (2007) reported that different vowels produce distinct effects, likely due to articulatory constraints: open low vowels such as /a/ tend to show stronger effects than high vowels like /i/.

In this regard, formant frequencies generated by articulatory movements function not only as linguistic markers but also as acoustic indicators of internal states. Subtle shifts in formant values – induced by emotional arousal and muscular tension – reflect the affective state of the speaker or singer. While primary formant structures provide phonetic identity, their dynamic variation parallels musical phenomena. Vowel formant configurations can thus be understood as producing an „acoustic harmony" that encodes and transmits affective signals. Petrović (2017) develops this perspective further, framing vowels as carriers of embodied resonance whose spectral organization contributes to emotional communication.

For vocal pedagogy, this understanding is of critical importance. Recognizing that the emotional „color" of a vowel results from the negotiation between phonological targets and physiological arousal allows for more sophisticated development of vocal technique. When students grasp the affective resonance of each vowel, they can better navigate the biomechanical limits of their instrument. This awareness leads to improved vocal quality, fostering precise intonation and more authentic interpretation. By mastering the relationship between vowel articulation and its affective weight, students can reduce unnecessary physical tension, ensuring that vocal production is both technically sound and emotionally resonant. Articulation is therefore never a purely mechanical event. The tension and movement of the vocal apparatus do not merely shape sound; they serve as an interface for the manifestation of internal affective states. Consequently, the somatic reality of the voice finds its most profound expression through the orofacial complex, where the physical act of shaping a vowel becomes inseparable from the biological signaling of emotion. Thus, the pedagogy of the voice must be grounded not only in technical precision but also in the embodied reality of emotional expression.

The exploration of biomechanical constraints and physiological modulation has shown that vowel articulation is inseparable from the embodied processes of vocal production. Yet, the voice does not operate in isolation: its affective resonance is amplified and complemented by facial expression and respiratory dynamics. Emotional communication is therefore not confined to the acoustic domain but emerges from a multimodal system in which sound, breath, and facial movement converge. To fully understand how emotional meaning is transmitted, it is necessary to move beyond the internal mechanics of the vocal tract and examine the integrated pathways through which vocal and facial signals interact. Accordingly, the following section investigates this broader system of emotional communication, highlighting how acoustic and visual cues work together to encode and convey affective states.

VOCAL AND FACIAL PATHWAYS OF EMOTIONAL EXPRESSION

Vowels play a central role in sound production, early language development and emotional communication. Spencer (1855) and Darwin (1872) emphasized that singing possesses a unique capacity to externalize complex psychological states. Their importance is evident in universal affective vocalizations such as crying, laughter, and shouting, and in the pre-linguistic phase of infant development. Jakobson and Halle (1956) highlighted the primacy of vowels in early phonological development, arguing that they constitute the foundational elements of the child's emerging sound system. Empirical studies confirmed that infants' earliest vocalizations are predominantly vocalic (Oller, 1980; Stark, 1981), and that these sounds serve communicative functions before lexical precision is achieved. Analyses of infant crying revealed that variations in formant structure convey distinct emotional states such as hunger, distress, or contentment (Wasz-Höckert et al., 1985). More recent work has shown that infants' perception and production of vowels are closely linked to emotional and social development (Tsao et al., 2004). Philippot (2002) highlighted mechanisms such as the facial feedback hypothesis, where the physical shaping of sound can influence affective states.

Within this framework, a functional distinction emerges: consonants primarily provide lexical precision and structural boundaries, while vowels carry tone and affect. The evolutionary descent of the human larynx, discussed by Fitch (2000) and Lieberman (1988), increased the length and flexibility of the vocal tract, enabling a wider range of formant patterns. This anatomical change allowed humans to transmit subtle shifts in physiological arousal through vowel production. Historical observations support these biological findings. In his work on rhetoric, Mikhail Lomonosov (1755) categorized vowels according to their acoustic qualities, associating high-frequency vowels such as /i/ and /e/ with tenderness and affection, and darker vowels such as /u/ and /o/ with gravity and fear. Darwin (1872) noted that vowels such as /o/ and /a/ accompany laughter, while /e/ and /i/ are linked to amusement or grief. As noted by Jespersen (1922), back vowels are often found in words expressing disgust. Morton (1977), in his study of bird and mammal vocalizations, demonstrated that low, harsh sounds are associated with aggression, while high-frequency sounds signal affiliation and care – findings that have since been extended to human vocal and musical communication as evidence of universal acoustic coding. Cross-cultural studies by Scherer and Wallbott (1994) further showed that high-intensity emotions like anger are associated with expansive vocalic configurations, while sadness correlates with more constricted acoustic profiles.

In musical aesthetics, acoustic configurations are consistently mapped onto emotional archetypes. Murray and Arnott (1993) identified sadness with lower registers and subdued dynamics, while Davitz (1964) described joy and anger as expressed through elevated pitch and expanded tonal ranges. Kienast and Sendlmeier (2000) found that anger generates the highest degree of articulatory precision in both speech and song, producing acoustically dominant phonemes resistant to auditory masking. Yildirim et al. (2004) noted that both anger and sadness can employ extreme registers to convey intensity, suggesting that register extremity functions as a meta-signal of affective depth.

To synthesize this relationship between phonetic structure and psychological response, the following illustration maps these vowel configurations onto their emotional archetypes (Table 2).

Table 2. *The Phono-Symbolic Nexus: Mapping Vowel Articulation to Emotional Affect.*

Vowel	Acoustic/Articulatory Marker	Primary Mechanism	Emotional Archetype	References
/i/	High F2 (Front vowel)	Zygomaticus major (Smiling muscle)	Joy / Pleasure	(Rummer et al., 2014; Darwin, 1872)
/e/	Mid-High F2	Facial affiliation tension	Tenderness / Affiliation	(Lomonosov, 1755; Morton, 1977)
/a/	High F1 (Open aperture)	Mandibular aperture (Wide jaw)	Surprise / Anger	(Nordstrand et al., 2004; Sapir, 1929)
/o/	Low F2 (Back vowel)	Orbicularis oris (Lip rounding)	Sadness / Dysphoria	(Rummer et al., 2014; Lomonosov, 1755)
/u/	Lowest F2 (Back/Close)	Maximum oral constriction	Fear / Dread	(Morton, 1977; Darwin, 1872)

The biological substrate of emotional vocalization lies in the interplay between respiratory modulation, facial expression, and vocal articulation. Emotions manifest through a synchronized system of acoustic and visual signals, driven by specific respiratory patterns. Research indicates that mimicking these patterns can even induce the corresponding emotional state (Philippot, 2002). Specifically, fear is often associated with rapid, shallow respiration, while sadness is characterized by slower, deeper cycles (Philippot et al., 2002). These respiratory shifts are transmitted to the glottal source, rendering the vocal apparatus a reliable indicator of internal emotional states. Supporting this, Scherer (1986) argued that acoustic cues often provide a more transparent reflection of affective states than visual ones, as vocal parameters such as intonation, tempo, and intensity are highly sensitive to physiological arousal and less subject to conscious suppression than facial expressions.

The relationship between articulation and emotion is driven by orofacial mechanics. Speech motor control studies show increased jaw displacement during high-arousal states, such as anger, to enhance emotional saliency (Nordstrand et al., 2004). Furthermore, prosodic emphasis through hyper-articulation requires wider jaw opening, marking stressed syllables as critical, multisensory indicators of emotion (De Jong, 1995). Rummer et al. (2014) demonstrated that vowels function as potent emotional signals through a psychophysiological interplay. Their findings establish that facial musculature activity does not merely reflect internal states but actively modulates emotional experience via the facial feedback hypothesis. For instance, the articulation of /i/ requires the contraction of the zygomaticus major, the primary muscle involved in smiling, thereby enhancing positive mood states. Conversely, vowels requiring a more constricted orofacial posture, such as /o/, can induce a shift toward negative valence through the activation of the orbicularis oris.

Ringeval and colleagues (2008) demonstrated that prosodic features of vowels (duration, rhythm, intonation) contribute to emotion recognition. Blacking (1973) emphasized the cultural and musical dimensions of vocalization, showing that vowels and musical intervals can function as carriers of human feeling. Curtis and Bharucha (2010) demonstrated that the minor third, traditionally associated with melancholia, has a direct equivalent in prosodic contours of speech, confirming that music and language share a bio-acoustic lexicon. Following these findings, research indicates that both vocal cues (Scherer, 1986) and facial expressions (Ekman, 1982) constitute reliable pathways for decoding emotional states, with each modality offering distinct advantages depending on context. Collectively, these studies indicate that vowels and facial expressions convey patterns that connect linguistic structure with human affect.

This deep-seated connection between vowel archetypes and affective valence suggests that the human voice does more than just represent emotion; it translates internal states into structured acoustic signals. When these signals transcend speech and enter the domain of musical expression, they activate a shared set of universal codes. Accordingly, the transition from linguistic vocalization to singing marks a sophisticated escalation of the affective system, where the biomechanics of articulation and the aesthetics of music converge to produce a profound impact on both the listener’s perception and the singer’s biological reality. From the perspective of music pedagogy, articulation is not merely a technical requirement but a tool for emotional regulation, enabling performers to embody emotion authentically and transmit it through sound in ways that bridge linguistic structure, cultural codes, and biological reality. This convergence of biomechanics, aesthetics, and pedagogy sets the stage for the following section, where vowels are examined as archetypes of affective meaning and as stimuli with profound biological implications.

VOWELS AS ARCHETYPES OF AFFECTIVE AND BIOLOGICAL MEANING

The human voice functions as a biological transducer, converting physiological states into acoustic signals through coordinated laryngeal, respiratory, and resonant activity (Sundberg, 1987; Titze, 1994). Within this framework, vowels emerge not only as linguistic units but as archetypal carriers of affective and biological meaning. Neurocognitive studies demonstrate that vowels are processed simultaneously as perceptual and motor events, activating auditory, motor, and limbic networks (Obleser et al., 2003; Pulvermüller et al., 2006). This dual activation situates vowel articulation at the intersection of sensory input, motor output, and emotional regulation. Recent findings confirm that vowel production can stabilize cortical resonance (Saus et al., 2025), modulate autonomic function (Morales-Luque et al., 2025), and foster relaxation and social cohesion through collective chanting dominated by mid-central vowels (Canessa-Pollard et al., 2025). These results highlight vowels as biological stimuli with measurable impact on stress regulation, wellbeing, and expressive authenticity. Welch (2021) situates such findings within a broader framework, arguing that music optimizes wellbeing and human development by integrating biological regulation, emotional expression, and social cohesion.

This body of evidence – ranging from neurocognitive activation of auditory, motor, and limbic networks to findings on cortical resonance, autonomic modulation, and collective vowel chanting – suggests that the impact of vocal music cannot be reduced to semantic or lexical content alone. As Stevens (1998) argues, the expressive meaning of a vocal work is significantly expanded through its acoustic properties, where sound itself functions as a primary carrier of meaning. Consequently, vowels must no longer be treated as secondary technical elements or mere vehicles of lyrics; instead, they must be recognized as integral components of the emotional architecture of singing. The technical shaping of a vowel, characterized by its specific formant structure, acts as a physiological „trigger" that initiates a neuroendocrine response, directly influencing the singer's biological state.

Biological studies corroborate these effects: singing modulates immune markers (Kreutz et al., 2004), reduces stress hormones, and enhances oxytocin-mediated bonding (Good & Russo, 2021; Bowling et al., 2022). More recent evidence confirms that singing interventions directly influence endocrine and immune responses, lowering cortisol levels, increasing oxytocin release, and modulating cytokine activity (Fancourt & Perkins, 2018). Complementary findings demonstrate that choir singing modulates both oxytocin and cortisol, suggesting reduced stress and dynamic social-bonding effects. The sharper oxytocin drops observed in older children, alongside age-linked cortisol shifts, point to developmental differences in how collective singing shapes biological regulation (Petrović, 2025). Taken together, these mechanisms highlight the endocrine dimension of vocalization, positioning vowel articulation not merely as a technical act but as a biological regulator that fosters systemic balance, resilience, and emotional growth. As Reybrouck (2021) and Habibi & Damasio (2014) emphasize, music engagement operates as a primary biological and emotional regulator, assisting the organism in maintaining internal stability against environmental stressors.

From a pedagogical perspective, vowel training must be understood as a multidimensional practice: technical in its precision, emotional in its expressive depth, and physiological in its regulatory function. Within music pedagogy, this positions vowel training as a uniquely integrative practice, where formant shaping becomes a conduit for both artistic authenticity and embodied balance. This integrative view underscores that the shaping of vowels is not a peripheral technical exercise but a central pedagogical strategy that unites the scientific, emotional, and artistic dimensions of singing.

Contemporary pedagogy reflects this convergence. American institutions such as NATS and AATS have institutionalized science-informed frameworks, emphasizing resonance, vowel shaping, and standardized terminology, while scholars highlight inclusivity and psychophysiological safety (Bozeman, 2017; Benson, 2020; Benson & Rosen, 2023). Applied science further confirms the pedagogical efficacy of biofeedback and acoustic visualization (Nair, Schellenberg & Gick, 2015), demonstrating that the integration of technology into vocal training enhances both precision and awareness. Globally, Igna & Ma (2025) define pedagogy as a scientific discipline, integrating acoustics, physiology, and psychology into teaching, while Li, Mazlan & Safian (2026) situate it within the framework of 21st-century skills, integrating creativity, collaboration, and digital literacy. These developments collectively affirm that vowel training is not only a matter of vocal technique but a pedagogical practice that embodies the principles of holistic education.

Taken together, these insights reposition vowels as archetypes of affective meaning. They are not passive carriers of text but active stimuli that shape emotional resonance, hormonal regulation, and social connection. In music pedagogy, this recognition transforms vocal instruction into a holistic discipline, where vowel shaping becomes a conduit for both artistic authenticity and systemic wellbeing, situating the singer as both an expressive agent and a biological actor whose voice fosters harmony, resilience, and human development.

CONCLUSION AND PEDAGOGICAL IMPLICATIONS

This study has traced the role of vowels across historical, neurophysiological, biological, and pedagogical dimensions, demonstrating their function as archetypes of affective and systemic meaning. Vowels possess an intrinsic phonoesthetic quality: their articulation requires specific configurations of the orofacial musculature that are biologically and neurologically tethered to emotional states. In this sense, they transcend their status as mere phonological units of language and emerge as autonomous acoustic signals of affect. This profound interconnectedness was recognized as early as the medieval period, when Guido d'Arezzo established a systematic liturgical practice in which vowels were strategically mapped onto specific pitch degrees within the hexachordal system, underscoring the enduring link between phoneme identity, tonal positioning, and affective resonance (Petrović, 2017).

This study has shown that vowels, through their phonoesthetic qualities and sound symbolism, construct associations of size, brightness, and spatial orientation. Their identity is inseparable from the biomechanical architecture of the vocal tract, where articulatory stability and variation shape affective resonance. Vowels dominate early language development and emotional communication, functioning as multimodal signals that integrate sound, breath, and facial movement. At the same time, neurocognitive and biological evidence confirms that vowel articulation activates auditory, motor, and limbic networks, regulates endocrine and immune responses, and serves as a multidimensional pedagogical stimulus – technical, emotional, and physiological.

The pedagogical implications of these findings are profound. Vowel training must be recognized as a holistic practice that unites acoustics, physiology, psychology, and artistry. In practical terms, these finding translate into specific pedagogical strategies within the vocal pedagogy. For instance, vocal teachers can deliberetely utilize the high-front vowel /i/ not only as a technical tool for resonance and forward vocal placement, but as a psychological catalyst to stimulate positive affect and reduce performance anxiety via the activation of the zygomaticus major muscle. Conversely, open vowels like /a/ can be strategically introduced in repertoire coaching to facilitate emotional release and high-arousal expressions such as joy or dramatic intensity, optimizing jaw displacement to enhance acoustic projection without causing vocal strain. By applying these vowel-specific interventions, educators can design targeted vocal exercises and warm-ups that directly bridge somatic awareness with scientific precision, simultaneously supporting the students' vocal health and emotional resilence.

Contemporary frameworks already reflect this convergence, integrating science-informed methods, positioning the singer as an embodied communicator whose voice regulates biological states while transmitting aesthetic meaning. Globally, music and vocal pedagogy is increasingly defined as a discipline that bridges acoustics, physiology, and creativity, situating vowel shaping at the heart of 21st-century skills.

Taken together, this research repositions vowels as holistic instruments – simultaneously acoustic, emotional, and biological. They are not passive carriers of text but active stimuli that shape resonance, systemic wellbeing, and social connection. Ultimately, understanding the symbiotic relationship between pitch, vowel articulation, and emotional response enhances the pedagogical value of music education. It empowers the student to perceive sound as a medium that transcends both linguistic and purely musical structures, fostering a deeper, more embodied engagement with the art of singing. By repositioning the vowel as a psychophysiological tool, music pedagogy can cultivate a new generation of performers capable of authentic, scientifically-informed, and emotionally resonant vocal expression, fostering systemic biological balance and artistic integrity through the mastered art of phonation.

REFERENCES

Benson, E. A. (2020). *Training Contemporary Commercial Singers*. Compton Publishing.

Benson, E. A. & Rosen, K. (2023). Anti-Fat Bias in the Singing Voice Studio, Part Two: How to Make a Size Inclusive Voice Studio. *Journal of Singing*, 79(5), 653–660.

Berlin, B. (2006). Evidence for pervasive sound symbolism in ethnozoological nomenclature. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound Symbolism* (pp. 76–93). Cambridge University Press.

Blacking, J. (1973). *How Musical is Man?* University of Washington Press.

Bowling, D.L., Gahr, J., Ancochea, G., Hoeschele, M., Canoine, V., Fusani, L. & Fitch, W.T. (2022). Endogenous oxytocin, cortisol, and testosterone in response to group singing. *Hormons and Behaviour*, doi: 10.1016/j.yhbeh.2021.105105.

Bozeman, K. (2017). The pedagogic use of absolute spectral tone color theory. *Journal of Singing*, 74(2), 179–183.

Brandner, M., Frank, M. & Sontacchi, A. (2022). Horizontal and Vertical Voice Directivity Characteristics of Sung Vowels in Classical Singing. *Acoustics*, 4(4), 849–866. https://doi.org/10.3390/acoustics4040051

Canessa-Pollard, V., Anikin, A. & Reby, D. (2025). Chants across seven traditions share acoustic traits that enhance subjective relaxation. *Proceedings of the National Academy of Sciences of the United States of America*, 122(35): e2506480122. doi: 10.1073/pnas.2506480122.

Curtis, M. E. & Bharucha, J. J. (2010). The Minor Third Communicates Sadness in Speech, Mirroring Its Use in Music. *Emotion*, 10(3), 335–348.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray.

Davitz, J. R. (1964). *The communication of emotional meaning*. New York: McGraw-Hill.

De Jong, K. J. (1995). The Supraglottal Articulation of Prominence in English: Linguistic Stress as Localized Hyperarticulation. *Journal of the Acoustical Society of America*, 97: 491–504.

Ekman, P., Friesen, W. V. & Ellsworth, P. (1982). What are the relative contributions of facial behavior and contextual information to the judgment of emotion? In P. Ekman (Ed.), *Emotion in the human face* (pp. 111–127). Cambridge: Cambridge University Press.

Ekström, A. G. (2022). What's next for size-sound symbolism? *Frontiers in Language Sciences*, 1, Article 1046637. https://doi.org/10.3389/flang.2022.1046637

Fancourt, D., & Perkins, R. (2018). The effects of singing interventions on cortisol, oxytocin, and cytokine levels. *Evidence-Based Complementary and Alternative Medicine*, 1–12.

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, 4(7), 258–267.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner & D. Steriade (Eds.), *Phonetically based phonology* (pp. 232–276). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486401.008

Good, A. & Russo, F. A. (2021). Changes in mood, oxytocin, and cortisol following group and individual singing: A pilot study. *Psychology of Music*, 50(4). https://doi.org/10.1177/03057356211042668

Habibi, A. & Damasio, A. (2014). Music, feelings, and the human brain. *Psychomusicology: Music, Mind, and Brain*, 24(1), 92–102.

Helmholtz, H. von (1863). *Sensations of Tone as a Psychological Basis for the Theory of Music*. London: Longmans, Green and Co.

Howell, I. (2015). *Absolute Spectral Tone Color: A Psychoacoustic Concept Applied to Voice Pedagogy*. New England Conservatory.

Igna, O. N. & Ma, H. (2025). Vocal pedagogy as a contemporary scientific direction. *Education and Pedagogy Journal*, 2, 41–58. doi: 10.23951/2782-2575-2025-2-41-58

Jakobson, R. & Halle, M. (1956). *Fundamentals of Language*. Mouton.

Jespersen, O. (1922). *Language; It's Nature, Development and Origin*. London: George Allen & Unwin Ltd.

Kostić, D., Nestorović, M. & Kalić, D. (1964). *Akustička fonetika srpskohrvatskog jezika*. Beograd: Institut za eksperimentalnu fonetiku i patologiju govora.

Kienast, M. & Sendlmeier, W. F. (2000). *Acoustical analysis of spectral and temporal changes in emotional speech*. Institute of Communication science, Technical University of Berlin, Germany.

Kreutz, G., Bongard, S., Rohrmann, S., Hodapp, V. & Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin A, cortisol, and emotional state. *Journal of Behavioral Medicine*, 27(6), 623–635.

Ladefoged, P. (1964). A Phonetic Study of West African Languages – An Auditory-Instrumental Survey. Cambridge: Cambridge University Press.

Lee, S., Yildirim, S., Kazemzadeh, A. & Narayanan, S. S. (2005). An articulatory study of emotional speech production. In: *Proceedings of Interspeech*, pp. 497–500.

Li, W., Mazlan, C.A.N. & Safian, A.R (2026). 21st-century skills in vocal pedagogy: a scoping review. *Frontiers of Education*, 11. https://doi.org/10.3389/feduc.2026.1744952

Lieberman, P. (1988). *The Biology and Evolution of Language*. Harvard University Press.

Lomonosov, M. (1755). *Rossiyskaya grammatika* [Russian Grammar]. Imperial Academy of Sciences.

Milankovic, V., Petrovic, M. and Petrovic, J. (2003). Tone constellation: a personal spatial scale presentation. In R. Kopiez, A. C. Lehmann, I. Wolther and C. Wolf (Eds.), *Proceedings of the 5th Triennial ESCOM Conference* (pp. 413–415), Hanover University of Music and Drama, Germany.

Morales-Luque, C., Carrillo-Franco, L., López-González, M.V., González-García, M., Dawid-Milner, M.S. (2025). Mapping the Neurophysiological Link Between Voice and Autonomic Function: A Scoping Review. *Biology (Basel)*, 14(10):1382. doi: 10.3390/biology14101382.

Morton, E. S. (1977). On the occurrence and significance motivation-structural rules in some bird and mammal sounds. *American Naturalist*, 111, 855–869.

Murray, I. R. and Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93, 1097–1108.

Nair, A., Schellenberg, M. & Gick, B. (2015). A Case Study on the Efficacy of Ultrasound Biofeedback in Voice Pedagogy. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, J. M. Scobbie (Eds.), *18th Proceedings of the International Congress of Phonetic Sciences (10-14 August 2015)*. Glasgow, Scotland UK: University of Glasgow.

Nastase, V., Sokolova, M. & Sayyad Shirabad J. (2007). Do happy words sound happy? A study of the relation between form and meaning for English words expressing emotions. In: *Proceedings of the Recent Advances in Natural Language Processing* (pp. 406–410).

Nettl, B. (1954). Text-music relationships in Arapaho songs. *Southwestern Journal of Anthropology*, 10m, 192–199.

Newman, S. S. (1933). Further experiments in phonetic symbolism. *American Journal of Psychology*, 45(1), 53-75.

Nordstrand, M., Svanfeldt, G., Granstrom, B. & House, D. (2004). Measurements of articulatory variation in expressive speech for a set of Swedish vowels. *Speech Communication*, 44, 187–196.

Obleser, J., Elbert, T., Lahiri, A., & Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Brain Research. Cognitive Brain Research*, 15(3): 207-13. doi: 10.1016/s0926-6410(02)00193-3.

Oller, D. K. (1980). The Emergency of the Sounds of Speech in Infancy – Chapter 6. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), *Child Phonology* (pp. 93–112).
https://doi.org/10.1016/B978-0-12-770601-6.50011-5.

Paget R. (1930). *Human Speech – Some Observations, Experiments and Conclusions as to the Nature, Origin, Purpose and Possible Improvement of Human Speech*. London: Kegan Paul, Trench, Trubner & Co.

Petrović, M. (2017). *The Importance of Vowels in Music Education*. London: iMerc.

Petrović, M., Pekić, S., Soldatović, I. & Savović, S. (2025). Oxytocin and Cortisol in Children in Response to Group Singing. In A. Branković & S. Dabić (Eds.), *Program and Book of Abstracts of the 28th International Pedagogical Forum of Performing Arts „In Search of a Model of Sustainable Music Education"* (pp. 63–66). Belgrade: Faculty of Music.

Philippot, P., Chapelle, G. & Blairy, S. (2002). Respiratory feedback in the generation of emotion. *Cognition and Emotion*, 16(5), 605–627.

Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin F. M., Hauk, O. & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, 103(20), 7865-70. doi: 10.1073/pnas.0509989103.

Reybrouck, M. (2021). *Musical Sense-Making. Enaction, Experience, and Computation.* Routledge: Abingdon, UK; New York, NY, USA.

Ringeval, F., Chetouani, M. et al. (2008). Automatic Prosodic Disorders Analysis for Impared Communication Children. *Paper presented at the workshop on Child, Computer and Interaction*. Chania, Greece.

Rummer, R., Schweppe, J., Schlegelmilch, R. & Grice, M. (2014). Mood is linked to vowel type: The role of articulatory movements. *Emotion*, 14(2), 246–250.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12(3), 225–239.

Saus, W., Seither-Preisler, A. & Schneider, P. (2025). Harmonic vowels and neural dynamics: MEG evidence for auditory resonance integration in singing. Frontiers of Neuroscience, 19. https://doi.org/10.3389/fnins.2025.1625403

Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, 99, 143–165.

Scherer, K. R. & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66(2), 310–328. https://doi.org/10.1037/0022-3514.66.2.310

Spencer, H. (1855). *Principles of Psychology*, Third edition, Vol 2. New York: Appleton.

Stark, R. E. (1981). Infant vocalization: A comprehensive view. *Infant Mental Health*, 2(2), 118–128. https://doi.org/10.1002/1097-0355(198122)2:2<118::AID-IMHJ2280020208>3.0.CO;2-5

Sundberg, J. (1987). *The Science of the Singing Voice*. Dekalb, Illinois: Northern Illinois University Press.

Stevens, K. (1998). *Acoustic Phonetics*. The MIT Press.

Titze, I. R. (1994). *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice Hall.

Tsao, F-M., Huei-Mei Liu, H-M. & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Development*, 75(4):1067-84. doi: 10.1111/j.1467-8624.2004.00726.x.

Ultan, R. (1978). Size-sound symbolism. *Universals of Human Language*, 2, 525–568.

Vainio, L., Myllylä, I. L. & Vainio, M. (2024). Sound symbolism in manual and vocal responses: phoneme-response interactions associated with grasping as well as vertical and size dimensions of keypresses. *Cognitive Processing*, 25, 363–378. https://doi.org/10.1007/s10339-024-01188-y

Vainio, L., Wikström, A., Repetto, C. & Vainio, M. (2023). Sound-symbolic association between speech sound and spatial meaning in relation to the concepts of up/down and above/below. *Language and Cognition: An Interdisciplinary Journal of Language and Cognitive Science*, 15(4), 884–903. https://doi.org/10.1017/langcog.2023.31

Wasz-Höckert, O., Michelsson, K. & Lind, J. (1985). Twenty-five years of Scandinavian cry research. In Lester, B. M. & Boukydis, C. F. Z. (Eds.), *Infant Crying: Theoretical and Research Perspectives* (pp. 83–104). New York, NY: Plenum Press.

Welch G. F., Biasutti M., MacRitchie J., McPherson G. E. & Himonides E. (2020). Editorial: The impact of music on human development and well-being. *Frontiers of Psychology,* 11:1246. 10.3389/fpsyg.2020.01246

Yildirim, M., Bulut, C., Kazemzadeh, L., Busso, C., Deng, Z., Lee, S. & Narayanan, S. (2004). An acoustic study of emotions expressed. In: *Proceedings of the 8th International Conference on Spoken Language Processing*, Jeju Island, Korea.