The Evolution of Human Language from Musical Protolanguage:

A High-Dimensional Communication Theory

A Formal Interdisciplinary Monograph

Abstract

This monograph develops the Sung–Speech Protolanguage (SSP) hypothesis: the proposal that early human communication may have relied on a **multidimensional melodic system** integrating pitch, rhythm, timbre, gesture, emotional contour, and speaker identity. Rather than assuming that spoken language emerged first through discrete symbolic units, the SSP hypothesis explores the possibility that musical communication served as a **high-bandwidth expressive system** well-suited to small, cohesive communities that relied on synchrony, emotional alignment, and coalition signaling.

Drawing on research from linguistics, music cognition, anthropology, developmental psychology, and evolutionary neuroscience, this work outlines how such a system could have functioned and how it might later have undergone **cultural compression**. As human societies expanded, the pressures of secrecy, standardization, cognitive efficiency, and restricted knowledge circulation may have favored the emergence of a **reduced-dimensional, symbolic spoken mode**, redirecting components of the ancestral melodic system into specialized cultural domains such as ritual and music.

The monograph provides a cautious, interdisciplinary argument for this evolutionary trajectory, a timeline informed by archaeological and mythological patterns, and a computational framework for modeling multidimensional token structures. While the SSP hypothesis does not attempt to reconstruct a definitive ancestral language, it offers a coherent perspective on how musical and linguistic systems may have diverged and how traces of early melodic communication persist in prosody, emotional expression, and human musicality.

---

1. Introduction

The question of how human language originated remains an enduring challenge spanning linguistics, anthropology, evolutionary biology, psychology, and cognitive science. Traditional accounts often begin with the emergence of segmented phonemes or proto-syntactic structures, treating symbolic speech as the foundational substrate from which other forms of communication evolved. Yet a

parallel line of inquiry—originating with Darwin and developed in modern studies of music and vocal learning—suggests an alternative possibility: that early human communication may have been **fundamentally musical**.

This monograph explores that possibility through the **Sung–Speech Protolanguage (SSP) hypothesis**, which posits that ancestral communication could have operated as a **multidimensional melodic system**. Such a system might have drawn simultaneously on pitch, rhythm, timbre, emotional contour, and gesture, functioning as a multimodal expressive channel rather than a purely sequential symbolic code. Modern humans continue to exhibit many of these capabilities: we coordinate speech with gesture, modulate intention through melodic prosody, engage in collective musical synchrony, and use vocal contours to signal affect and identity.

The focus of this hypothesis is not to assert a definitive historical reconstruction, but to examine how such a system fits with known human capacities and to consider why it might later have been **compressed into a reduced-dimensional spoken mode**. As communities grew in size and complexity, cultural pressures—ranging from secrecy and social stratification to cognitive efficiency and the need for standardized communication—could have favored symbolic speech over multidimensional melodic expression.

The chapters that follow develop this idea in four stages:

1. outlining the structure and plausibility of a multidimensional protolanguage;
2. proposing reasons for its reduction into spoken language;
3. situating these arguments within archaeological and mythological contexts; and
4. presenting a computational framework for modeling melodic token systems.

Together, these components offer a theoretical lens through which to reinterpret the relationship between music and language, treating spoken language not as the origin of human communication but as a **scaled, specialized descendant** of a richer expressive heritage.

---

2. Background and Literature Review

Research into language evolution has long grappled with the complexity of human vocal and cognitive abilities. Several strands of scholarship provide a foundation for the SSP hypothesis while also illuminating its limitations and alternative interpretations.

2.1 Darwinian and Early Musicality Hypotheses

Darwin (1871) suggested that musical expression may have preceded linguistic articulation, particularly in the context of courtship and emotional signaling. While speculative, this view opened the possibility that melodic communication served as an early scaffold for more structured linguistic systems.

2.2 Brown's Musilanguage and Related Models

Brown's (2000) musilanguage model introduced the idea of a shared ancestor between music and language—a communicative system possessing both tonal and proto-semantic properties. Though Brown did not propose a fully musical linguistic system, his work supports the idea that prosodic and melodic contours could have played a central role in early communication.

2.3 Mithen's "Hmmmmm" System

Mithen (2005) described early communication as **Holistic, Manipulative, Multi-modal, Musical, and Mimetic**. This framework aligns with SSP's emphasis on multimodality and high-dimensional expression while also noting that melodic communication may have coexisted with gesture and bodily rhythm.

2.4 Speech-First Accounts

Other theories (Pinker, 1994; Jackendoff, 2002; Bickerton, 2009) posit that early language was based on discrete symbols or proto-syntax. While these models remain influential, they often assume that musical communication is inherently inefficient—a premise the SSP hypothesis examines critically.

2.5 Current Gaps and Motivations

Across these models, several issues remain unresolved:

- How did multimodal communication scale as societies grew larger and more complex?
- What pressures might lead a multidimensional system to become reduced-dimensional?
- How do cognitive load, transmission fidelity, and standardization influence language evolution?
- Why are musical and spoken communication now functionally separated in most cultures?

These questions motivate the SSP hypothesis and shape the analyses in later sections.

---

3. The Multidimensional Structure of Musical Protolanguage

The SSP hypothesis proposes that early human communication may have relied on a **multidimensional acoustic and multimodal feature space**, organized around musical principles. This section outlines the structure of such a system while maintaining appropriate scientific caution.

3.1 Acoustic and Multimodal Feature Space

A melodic protolanguage could theoretically integrate multiple simultaneous channels of meaning, including:

1. pitch and pitch contour
2. intervallic transitions
3. rhythmic structure and duration
4. temporal grouping and phrasing
5. octave register
6. timbre or vocal color
7. dynamic intensity
8. articulation patterns
9. prosodic and affective contour
10. speaker identity cues
11. grammatical role tendencies
12. gestural accompaniment (inferred but not reconstructable)

These dimensions may have functioned in parallel, forming "chords" of semantic and emotional information rather than sequences of discrete units.

3.2 Information-Theoretic Considerations

A multidimensional token system could encode far more discriminable states than a typical spoken syllable (Patel, 2008). However, this richness also entails greater cognitive and perceptual demands, a point developed further in Section 4. The present model therefore represents a **functional possibility**, not a claim of universally adopted structure.

3.3 Cognitive and Developmental Plausibility

Infants produce pitch-modulated protophones before consonant–vowel segmentation (Oller, 2000), suggesting that melodic control is early-emerging and neurologically foundational. This developmental pattern is consistent with the plausibility of an early melodic communicative system.

3.4 Multimodality and Transmission Fidelity

Gesture and bodily coordination were almost certainly involved in early communication. However, melodic contours show greater **intergenerational stability**, offering a partial rationale for why the vocal melodic layer is the most viable target for conceptual reconstruction.

---

4. Why Abandon Musical Language? A Multifactorial Account of Linguistic Reduction

The Sung-Speech Protolanguage (SSP) hypothesis proposes that early human communication was fundamentally melodic, expressive, and multidimensional. If this system offered such rich semantic bandwidth and such clear advantages for emotional signaling and group cohesion, a natural question arises: **Why would humans abandon a powerful multidimensional code in favor of a lower-bandwidth symbolic system?**

This section outlines a **multifactorial explanation**, integrating cultural, cognitive, social, and structural pressures. These factors should not be interpreted as definitive causal claims but rather as **plausible interacting influences** consistent with comparative, ethnographic, and archaeological evidence.

4.1 Cultural Pressures Toward Opacity and Restricted Knowledge

As human societies transitioned from small foraging bands to increasingly stratified agricultural communities (c. 10,000–5,000 BCE), communication occurred within new social environments characterized by:

- resource accumulation
- hereditary status
- priestly knowledge and ritual authority
- territorial conflict and inter-group competition

In such contexts, groups benefited from the ability to **restrict knowledge** and maintain in-group exclusivity. A multidimensional melodic system—where affective intent, relational stance, and emotional nuance are readily interpretable even by outsiders—offered little protection against eavesdropping or cultural diffusion.

By contrast, spoken language, with its **arbitrary symbolic conventions**, provides:

- opaque mappings of form to meaning

- steep learning requirements for outsiders
- rapid potential for dialectal divergence
- tools for cryptic or specialized jargon

Thus, spoken language can be understood as a **culturally selected compression** of an earlier melodic system—one optimized less for expressivity and more for **controlled, restricted, and group-specific information flow**.

This explanation remains hypothetical but aligns with broader theories of linguistic divergence in complex societies.

4.2 Gesture, Multimodality, and the Problem of Transmission Fidelity

Musical communication in early Homo sapiens was unlikely to be purely acoustic. Modern humans routinely integrate **gesture with vocalization**, and comparative primate communication suggests that multimodality is ancestral rather than recent.

Such multimodal expression may have included:

- coordinated hand and arm movements
- beat and iconic gestures
- posture and bodily rhythm
- facial expression
- coordinated breathing patterns

However, gesture presents a major challenge for **intergenerational fidelity**:

- There is no strong cultural expectation to replicate gestures with precision.
- Gestural variation proliferates more quickly than melodic variation.
- Archaeological and ethnographic evidence preserves songs far more reliably than movement patterns.
- Melodic contours can be memorized and transmitted; subtle gestures cannot.

Thus, while gesture may have been **integral** to Protolanguage, it is largely **lost to history**, leaving the vocal channel as the only reconstructable substrate. As communities grew and standardized communication norms emerged, features that could not be reliably transmitted—especially complex gestural complements—may have been gradually deemphasized in favor of **more easily stabilized symbolic systems**.

4.3 Social Bonding, Synchrony, and Coalition Signaling

Musical communication originally offered profound social advantages. Research on the evolution of music highlights several interrelated functions that would have strengthened early human groups:

4.3.1 Synchrony

Collective rhythmic activity supports:

- entrainment of physiological states
- shared timing during group tasks
- heightened prosociality

Synchronous melodic communication likely enhanced coordination in hunting, foraging, ritual, and conflict.

4.3.2 Coalition Signaling

Group singing or chanting creates:

- high-energy collective displays
- markers of group identity
- signals of unity and mutual obligation

Such displays are difficult to fake and function as **reliable coalition signals**, potentially deterring rivals.

4.3.3 Group Cohesion and Emotional Alignment

Melodic systems facilitate:

- emotional alignment among group members
- shared intentionality
- encoding of collective memory in recurring motifs

These benefits reinforce the plausibility of early melodic protolanguage.

Yet these same properties also highlight the **tension** that emerged as societies expanded. The very features that enhance cohesion within small groups may reduce **communicative control** in larger, more hierarchical populations. Ritual and music may have retained their bonding functions even as *everyday communication* shifted toward more efficient, standardized, and opaque modes.

4.4 Cognitive Load, Efficiency Trade-Offs, and the Pressure to Reduce Dimensionality

Cultural pressures alone cannot explain the transition. A fully multidimensional communication system—even if effective—incurs substantial **cognitive**

and computational costs.

4.4.1 Cognitive and Perceptual Demands

A system encoding pitch, rhythm, timbre, octave, grammar, emotion, and speaker identity simultaneously requires:

- high working-memory capacity
- fine-grained auditory discrimination
- complex motor coordination
- rapid parallel processing

Although early humans possessed such abilities, simpler systems reduce **processing load**, especially in fast-paced or high-density social communication.

4.4.2 Efficiency and Pragmatic Constraints

Musical tokens can encode rich meaning but:

- take longer to articulate
- demand greater precision
- are less tolerant of noise or interruption
- require more energy and vocal control

In daily exchanges—negotiations, instructions, coordination—short, discrete, low-effort symbolic utterances are **faster and more robust**.

4.4.3 Standardization in Growing Populations

As populations grew beyond kin networks, communication required:

- stable conventions
- predictable learning pathways
- high fidelity across thousands rather than dozens of individuals

Musical systems are prone to variation due to ornamentation, personal style, and emotional contour. Spoken phoneme inventories, despite lower bandwidth, offer **better standardization**:

- less expressive drift
- more consistent segmentation
- easier acquisition by children
- clearer boundaries between linguistic and paralinguistic cues

These factors may have contributed to the **gradual cultural selection of reduced-dimensional speech** as the dominant communicative mode.

4.5 Synthesis: A Multifactorial Evolutionary Shift

The move from musical to spoken language is best understood not as the triumph of one system over another, but as a **trade-off** shaped by interacting pressures:

1. **Cultural opacity and secrecy** favored arbitrary symbolic codes.
2. **Multimodal components** (especially gesture) lacked stable transmission across generations.
3. **Coalition signaling and synchrony** remained valuable but shifted into ritual contexts.
4. **Cognitive simplification** reduced per-interaction demands on speakers and listeners.
5. **Efficiency pressures** favored shorter, more adaptable utterances.
6. **Standardization requirements** in large groups led to compression of expressive dimensions.

Under this view, spoken language did not replace musical protolanguage because it was inherently superior; rather, it emerged as the most sustainable **solution to the communicative demands of large, hierarchical, culturally diverse societies**.

Melodic communication likely persisted—but increasingly in specialized domains such as ritual, song, performance, and emotional expression—while everyday linguistic exchange shifted toward a more compact, symbolic mode.

---

5. Archaeological and Mythological Correlations

The Sung–Speech Protolanguage (SSP) hypothesis proposes a transition from a multidimensional melodic system to a reduced-dimensional spoken system. While direct evidence of early vocal practices is inherently limited, archaeological, ethnographic, and mythological materials provide **indirect but suggestive correlations** that help situate musical protolanguage within a plausible evolutionary and cultural timeline. These correlates should be interpreted cautiously—as converging lines of inference rather than definitive proof.

5.1 A Plausible Prehistoric Timeline for Linguistic Evolution

The archaeological record supports a sequence of developments that is broadly compatible with the SSP trajectory, particularly when viewed through the lens of **social complexity, transmission fidelity, and communicative efficiency**.

>300,000 years ago — Anatomical Foundations for Melodic Communication

Fossil evidence suggests that *Homo sapiens* possessed a modern or near-modern vocal tract and thoracic breath control long before the appearance of symbolic writing or large-scale social hierarchies. These capacities, combined with rhythmic motor coordination, would have supported a **multimodal melodic protocommunication system** involving voice, gesture, and bodily synchrony.

100,000–10,000 BCE — Symbolic Culture and High-Fidelity Music Transmission

Archaeological and ethnographic analogues show complex musical traditions, coordinated communal performance, and early ritual structures. Musical forms often display **high intergenerational stability**, consistent with Section 4's argument that vocal melodic features transmit more reliably across generations than fine-grained gesture or complex motor patterns.

During this period, melodic communication may have served essential functions in **social bonding, coalition signaling, ritual coordination, and emotional alignment**, aligning with the high-dimensional expressive framework proposed by SSP.

10,000–3,000 BCE — Agricultural Revolution and Cultural Pressures Toward Opacity

The emergence of agriculture introduced fundamentally new social structures: villages, property, specialized labor, ritual authority, and conflict between neighboring polities. These contexts likely generated:

- incentives for **restricted knowledge systems**
- rapid dialectal divergence
- codified ritual languages
- and the rise of symbolic conventions that favored **standardization and communicative control**

Such pressures align with Section 4's argument that **opaque, learnable, low-dimensional spoken systems** became advantageous as populations expanded and social stratification intensified.

3,000–1,000 BCE — State Formation and Large-Scale Linguistic Fragmentation

The development of writing, centralized governance, and formal education systems introduced further forces toward **linguistic compression and standardization**. Spoken language became embedded in administrative, legal, and ritual institutions, reinforcing constraints on style, ornamentation, and prosodic variation.

This period also marks the increasing separation of **music** (retained primarily in ritual, performance, and identity signaling) from **speech** (dominant in transactional, administrative, and daily communication). The divergence of these domains mirrors the SSP proposal that the multidimensional system was gradually compartmentalized rather than entirely lost.

5.2 Myth as Cultural Memory: Interpreting the Tower of Babel

The Genesis 11 "Tower of Babel" narrative provides a compelling case study of how ancient societies conceptualized linguistic change. Though mythological rather than historical, the story encodes several themes relevant to the SSP framework:

- a recollection of a **previously unified mode of communication**
- a monumental collective project requiring shared intent
- a sudden **loss of mutual intelligibility**
- the dispersal of populations and the diversification of languages

While the text should not be taken as literal evidence of a melodic protolanguage, it may reflect **Bronze Age cultural awareness** of:

- rapid linguistic divergence
- the relationship between social cohesion and shared communication
- the political or ritual implications of language separation
- the breakdown of standardized communicative systems under conditions of expansion

This interpretation aligns with Section 4's position that linguistic change is deeply entangled with **social scale, cultural identity, and the management of knowledge**.

5.3 Archaeology, Myth, and the SSP Hypothesis

Taken together, the archaeological timeline and mythological motifs provide **contextual coherence** for the SSP hypothesis:

1. **Early humans possessed the anatomical and cognitive capacities** for multidimensional melodic communication.
2. **Musical practices exhibit long-term stability**, consistent with the reconstructability of vocal melodic features.
3. **Gesture and multimodal elements likely existed but were poorly preserved**, aligning with the transmission limitations discussed in Section 4.

4. **Increasing social complexity introduced pressures favoring reduction, opacity, and standardization**, consistent with the emergence of spoken language.
5. **Mythological narratives reflect lived experiences of linguistic fragmentation**, reinforcing the connection between communication systems and sociopolitical structure.

These correlations do not prove the SSP model but support its plausibility within an interdisciplinary understanding of human prehistory. They suggest that the shift from melodic to spoken communication was shaped not only by cultural dynamics but also by **scaling constraints**, **transmission pressures**, and **changing social ecologies**.

---

6. Model Reconstruction of ProtoLanguage

The preceding sections have outlined a plausible evolutionary trajectory in which early human communication operated within a **multidimensional melodic space**, later constrained by cultural, cognitive, and societal pressures into a reduced-dimensional symbolic system. While the original multimodal performance—including gesture, posture, movement, and timbre—cannot be directly recovered, certain components of the **vocal melodic channel** exhibit enough structural persistence to support partial, theoretically grounded reconstruction.

The goal of this section is therefore not to claim a literal restoration of a lost ancestral language, but to present a **conceptual and computational framework** for modeling how a melodic protolanguage *might* have encoded semantic relations, speaker identity, emotional stance, and grammatical structure. This reconstruction should be understood as a **hypothesis-driven simulation**, informed by modern linguistic, musical, and cognitive evidence rather than direct historical data.

6.1 Multidimensional Token Architecture

Under the SSP hypothesis, ProtoLanguage "tokens" can be viewed as **composite musical units**, each combining several feature dimensions simultaneously. These features do not map one-to-one onto modern linguistic categories but may have supported distinctions such as:

- **Actor roles** (e.g., agent, patient, experiencer)
- **Basic relational structures** (e.g., movement, possession, presence, change)
- **Emotional or affective coloration**
- **Speaker identity and social stance**

- **Sentence-final or phrase-level inflections**

In a multidimensional communicative system, these components would likely be **layered**, with different dimensions conveying information in parallel:

- pitch contour for core semantics
- rhythmic pattern for relational or temporal structure
- timbre or octave for speaker identity
- dynamic intensity for emotional force
- micro-prosodic shifts marking phrase boundaries

The resulting "token" resembles a **musical chord with temporal structure**, rather than a discrete phonemic unit. This architecture aligns with the acoustic and cognitive capacities outlined in Section 3 and with the social functions discussed in Section 4.

6.2 Limits of Reconstruction and the Role of Abstraction

As argued in Section 5, only certain aspects of melodic communication are likely to have preserved enough structural regularity across large timescales to be meaningfully modeled. Gesture, motor patterns, and embodied synchrony were almost certainly essential contributions to early communication, but these elements are:

- highly variable
- poorly preserved across generations
- culturally flexible
- and archaeologically silent

Thus, **reconstruction focuses on the most stable and inferentially accessible components**—the acoustic dimensions that modern humans continue to use for prosody, emotion, identity marking, and expressive nuance even within spoken language.

The model presented here is therefore **abstract**: it does not attempt to respecify exact melodies used by early humans, but rather to illustrate how a multidimensional system *could* organize meaning using the dimensions available to the human vocal apparatus.

6.3 Diagram 1 (Textual Description): High-Dimensional Acoustic Feature Space

To visualize this architecture, we adopt a simplified three-axis representation of the melodic feature space:

- **X-axis: pitch and interval structure**

- **Y-axis: duration, rhythm, and temporal grouping**
- **Z-axis: timbre and dynamic intensity**

Additional overlays represent:

- **emotional valence** (e.g., tension, calmness, urgency)
- **speaker identity features** (e.g., habitual octave, vocal color)
- **grammatical role tendencies** (e.g., rising contours for agents, falling contours for objects—only hypothetical categories)

These overlays should be understood as **conceptual guides**, not empirical mappings. Their purpose is to illustrate how multiple channels might have encoded information concurrently, consistent with the theoretical framework established in Sections 1–4.

6.4 Diagram 2 (Textual Description): Cultural Compression from Song to Speech

To align with the transition described in Section 4, Diagram 2 illustrates how a richly multidimensional token system might undergo **cultural compression**:

- parallel feature streams → reduced to
- largely sequential phonemic strings

This conceptual diagram highlights:

- **loss of multidimensional simultaneity**
- **greater reliance on discrete symbolic units**
- **reduced expressive bandwidth**
- **increased learnability and standardization**
- **growing separation between speech and music**

This reduction parallels the social and cognitive pressures described earlier and provides a visual metaphor for how spoken language may have emerged from melodic prototypes.

6.5 Computational Framework for ProtoLanguage Simulation

The JSON formalism included in **Appendix A** operationalizes the conceptual model by specifying:

- instruments (i.e., timbral profiles)
- speakers (identity-layer defaults)
- melodic dictionary entries (pitch/duration sequences)
- grammar-role metadata

- emotional presets that bias performance parameters

This framework does **not** claim to reproduce an ancestral language. Rather, it offers a **testbed** for exploring:

- how multidimensional tokens combine
- how emotional inflection shapes melodic meaning
- how identity markers affect interpretation
- how phrase-final pitch movements encode pragmatic force
- how compression into symbolic sequences changes expressive capacity

It serves as a **hypothesis-generation tool**, enabling computational comparison between multidimensional and reduced-dimensional communication systems.

6.6 Transition to Section 7

Section 7 extends this modeling framework by considering its implications for linguistics, anthropology, and neuroscience. Whereas Section 6 focuses on **how** a melodic protolanguage might be structured, Section 7 addresses **what it would mean** for theories of grammar, cultural evolution, and cognitive architecture if such a system once existed.

Together, Sections 3–6 now form a coherent narrative:

- **Section 3:** explores the structure and plausibility of a multidimensional system
- **Section 4:** explains why such a system may have shifted toward reduced-dimensional speech
- **Section 5:** situates this hypothesis within archaeological and mythological correlates
- **Section 6:** proposes a cautious, computationally grounded framework for modeling the melodic substrate

This trajectory sets the stage for Section 7's broader theoretical implications.

---

7. Implications

The Sung–Speech Protolanguage (SSP) hypothesis provides a theoretical framework for considering how early human communication may have integrated musical, emotional, and identity-rich features in ways that modern spoken language only partially preserves. While the model presented here remains speculative, it invites several lines of inquiry across linguistics, anthropology, neuroscience, and cognitive science. These implications should be seen as **potential avenues for investigation**, not as definitive consequences of the hypothesis.

7.1 Linguistics: Rethinking the Origins of Structural Complexity

If early communication relied on multidimensional melodic tokens, then certain features of modern linguistic structure—syntax, prosody, and phonological segmentation—may reflect **secondary developments** emerging under pressures for standardization, efficiency, and communicative opacity (Section 4).

Possible implications include:

- **Syntax as a derived organizational principle** Syntax may have evolved as a tool for *sequencing* meanings that were once expressed concurrently through melodic overlay.
- **Prosody as a residual channel** Modern intonation patterns (falling declaratives, rising interrogatives) may represent attenuated forms of earlier melodic distinctions.
- **Phoneme inventories as compression artifacts** Discrete phonemic categories could be cultural solutions for stabilizing communication in expanding populations.

These possibilities suggest that linguistic structure may reflect a **historical narrowing** of expressive bandwidth, not an initial expansion from minimal symbolic units.

7.2 Anthropology: Communication as a Scalable Cultural Technology

From an anthropological perspective, the SSP hypothesis reinforces the idea that communication systems evolve in tandem with social complexity. Multidimensional melodic communication may function well in:

- small, cohesive groups
- kin-based bands
- ritual or ceremonial contexts
- emotionally synchronized communities

But as societies grow, pressures toward:

- rapid transmission
- restricted knowledge
- administrative uniformity
- supra-household cooperation

may favor lower-dimensional symbolic systems.

This framing situates linguistic evolution as a **culturally mediated response** to challenges of scale, identity, and social coordination rather than purely biological change.

7.3 Neuroscience: Music and Language as Divergent Outcomes of Shared Substrates

Neuroscientific research often identifies overlapping but distinct neural circuits for music and language. The SSP hypothesis offers one possible explanation: these domains may represent **divergent specializations** emerging from a once more unified communicative system.

Implications include:

- **Shared ancestral circuitry** might have supported general-purpose melodic communication.
- Modern language could reflect functional **streamlining** and **specialization** of this circuitry.
- Music may retain features of the ancestral communicative mode—especially emotional contagion, synchrony, and identity signaling.
- Speech might have co-opted certain melodic scaffolds (e.g., pitch accents, prosody) while reducing their expressive range.

This view remains conjectural but aligns with evidence that musical and linguistic processing share deep developmental and neurocognitive connections.

7.4 Interdisciplinary Research Directions

The SSP framework suggests several empirical pathways:

- **Computational modeling** of multidimensional vs. reduced-dimensional communicative efficiency
- **Cross-cultural studies** of prosody, chant, and ritual vocalization as potential analogues
- **Developmental research** exploring how infants integrate melody and gesture
- **Cognitive load experiments** comparing parallel vs. sequential encoding channels
- **Historical linguistics** examining when phonological compression becomes evident in expanding populations

These research directions do not test ancient protolanguage directly but can illuminate whether **SSP-like systems are plausible, cognitively tractable, and socially functional**.

---

8. Conclusion

This monograph has explored the hypothesis that early human communication may have operated within a **multidimensional melodic system**, later compressed into the reduced-dimensional symbolic mode characteristic of modern spoken language. The Sung–Speech Protolanguage (SSP) hypothesis does not claim to reconstruct a definitive ancestral language, nor does it propose a singular or deterministic evolutionary path. Rather, it offers a coherent framework for understanding how:

- biological capacities for melody, rhythm, gesture, and emotional expression
- multimodal group coordination and synchrony
- pressures of cultural opacity, secrecy, and identity maintenance
- cognitive and perceptual demands
- and the need for standardization in growing societies

may have interacted to shape the communicative systems observed today.

The archaeological and mythological correlates discussed in Section 5 provide **contextual plausibility** rather than empirical proof, while the computational model in Section 6 demonstrates how multidimensional communication can be formalized for exploratory analysis. Taken together, these components illustrate how a melodic protolanguage could emerge, flourish, adapt, and ultimately yield to a more constrained but scalable symbolic system.

The SSP hypothesis thus serves as a **conceptual bridge** connecting insights from music cognition, language evolution, anthropology, and the cognitive sciences. It encourages an expanded view of human communication—one in which speech is not the starting point, but a **specialized derivative** of a richer expressive heritage.

Future research may clarify which elements of this model correspond to historical processes and which represent theoretical approximations. Regardless of the outcome, studying the interplay between music, language, social structure, and cognition promises to deepen our understanding of the forces that shaped the human communicative landscape.

Appendix A — ProtoLanguage JSON Placeholder

A fully revised JSON-based computational specification for ProtoLanguage tokens will be inserted here after manuscript completion.

References

- Bickerton, D. (2009). *Adam's Tongue*. Hill & Wang.
- Brown, S. (2000). "The 'musilanguage' model of music evolution." In *The Origins of Music*.
- Darwin, C. (1871). *The Descent of Man*. John Murray.
- Jackendoff, R. (2002). *Foundations of Language*. Oxford University Press.
- Mithen, S. (2005). *The Singing Neanderthals*. Harvard University Press.
- Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Lawrence Erlbaum.
- Patel, A. (2008). *Music, Language, and the Brain*. Oxford University Press.
- Pinker, S. (1994). *The Language Instinct*. William Morrow.
- Speiser, E. (1964). *Genesis: Introduction, Translation, and Notes*. Anchor Yale Bible.
