The Voice That Leaves No Signature: Prosody and Hidden Coherences Between Intention, Emotion, and Power

(ORCID: 0009-0005-4079-7418)

Andrés Calvo Espinosa

Email:
ace@optimeflow.com

Note:
This essay forms part of the Universal Coherence (UC) Project, a structural framework for reading and reorganizing reality through observable patterns across layers (physical, emotional, symbolic). It serves as a practical bridge between linguistics and institutional ethics, applying UC to detect hidden incoherences in prosody—the audible "form" through which intention, emotion, and power are filtered—using measurable offline tools such as a prototype PWA, with the aim of fostering relational congruence and the "hospitality of communicative forms" without turning it into persecution.

Abstract:
This preprint explores prosody—defined as the suprasegmental elements of speech (intonation, rhythm, pauses, stress, volume, speed, and timbre)—as a revealing bridge between underlying intention, emotional dynamics, and structures of power in human communication. In contexts where verbal content can be manipulated or denied, prosody acts as a "non-explicit signature" through which hidden coherences become perceptible, such as sarcasm, irony, exaggerated neutrality, double binds, or repressed emotionality, which often mask hostility in formal institutional settings.

Bringing together perspectives from linguistics, social psychology, institutional ethics, and acoustic measurement, the essay proposes a conceptual framework for detecting these asymmetries through observable temporal patterns, while avoiding subjective interpretations or "emotional polygraphs." It links this approach to the concept of Universal Coherence (UC), interpreting prosody as a form that reveals incoherences across layers (what is said vs. how people are treated), and uses a "fruit" criterion to assess impacts (clarity vs. fear). It also outlines an offline PWA prototype pipeline for replicable measurement (acoustic features such as F0, RMS, duration, and voiced ratio; a taxonomy of pragmatic indicators; ethical thresholds), with applications in fields such as politics, justice, education, business, healthcare, and media. It addresses risks (false positives, cultural/gender/neurodiversity biases) and proposes a research agenda to validate observable patterns, promoting a "hospitality of communicative forms" that prioritizes protection over persecution. Its ethical approach emphasizes measuring patterns rather than intentions, in order to foster congruence in social interactions.

Keywords:
prosody, hidden coherences, intention-emotion-power, Universal Coherence (UC), acoustic measurement, systemic harassment, container error, PWA pipeline, institutional ethics, communicative hospitality.

Safe Creative ID:
2603094813648

DOI (Zenodo):
10.5281/zenodo.18913483

The Voice That Leaves No Signature: Prosody and Hidden Coherences Between Intention, Emotion, and Power

Table of contents

Table of contents

Table of contents

1. Prelude

There is a characteristically contemporary scene: a conversation whose literal transcript appears impeccable—polite, professional, even "careful"—and which nevertheless leaves an unmistakable residue in lived experience: tension, shame, a constriction of one's own voice, withdrawal, or a kind of anger that is difficult to locate.

The phenomenon is recognizable in contexts of hierarchy, bureaucracy, or public exposure: the content may be correct and yet the treatment may still be corrosive. That corrosion rarely travels in strict semantics; it travels in form.

1.1. Why write about prosody today: when "what is said" does not match "what is done"

Prosody(1)—the architecture of rhythm, pitch, pauses, emphasis, intensity, and breathing—is the plane on which language becomes embodied. On that plane, the utterance ceases to function as abstract text and becomes a concrete relation: attention is modulated, the implicit permission to intervene is distributed, cost or value is assigned to a word, and the breathing room available for alternative versions of reality is delimited.

In times of protocols, "aseptic" communication, impeccable emails, and polished institutional statements, prosody becomes especially relevant for a simple reason: when a system wishes to preserve its image, it learns to "clean up" the text. Yet cleaning up the text is not the same as cleaning up intention; often, it means disguising it. What is harder to neutralize is the temporal and affective dynamic that filters through actual speech: the cut-off, the selective chill, the silence that follows, the emphasis that ridicules without insulting, or the tiny laugh that undermines without explicitly declaring it.

This does not mean romanticizing "tone" as though it were revealed truth. It means recognizing an operational fact: in social life, a substantive part of meaning(2) circulates through channels that do not pass through the dictionary. And when a conflict appears between the literal and the performative(3), interpretation is not played out solely on the rational plane: the bodily and social response is activated on both levels at once.

For this reason, writing about prosody today is not a technical whim; it is a cultural necessity. Wherever symbolic violence(4) requires plausible deniability(5), prosody becomes one of the few places where the pattern can be observed without leaving a signature.

(1) Lehiste (1970) systematizes suprasegmental features (duration, pitch, intensity, pauses) and their linguistic function in the organization of speech. Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

(2) Gumperz (1982) introduces the notion of contextualization cues and shows how signals such as prosody and rhythm guide inferences about framing and intention beyond lexical content. Gumperz, J. J. (1982). *Discourse Strategies*. Cambridge: Cambridge University Press.

(3) Austin (1962) formulates speech act theory and distinguishes between what is said and what is done in saying it, providing the basis for treating the "performative" as a pragmatic dimension. Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

(4) Bourdieu (1991) analyzes the symbolic power of language and how social authority is inscribed in forms of speech capable of exercising domination without declaring it explicitly. Bourdieu, P. (1991). *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.

(5) Pinker, Nowak, and Lee (2008) explain the logic of indirect speech and the role of plausible deniability when the literal content remains defensible while another intention is being communicated. Pinker, S., Nowak, M. A., & Lee, J. J. (2008). *"The logic of indirect speech." Proceedings of the National Academy of Sciences of the United States of America, 105*(3), 833-838.

1.2. Thesis: prosody as a bridge between intention, emotion, and social structure

This essay advances a central thesis: prosody functions as an operative bridge[6] between three planes that are often separated for the sake of analytical convenience:

· Intention: what one seeks to do with what is said (draw closer, dominate, neutralize, care for, humiliate, test loyalty, close down a topic, open it up, etc.).

· Emotion[7]: not as an isolated "inner feeling," but as an embodied dynamic that organizes attention, energy, threat, and safety.

· Social structure: the order of turns[8], the permission to dissent[9], the differential weight of voices, and whatever an institution rewards or punishes.

Prosody matters because it turns a sentence into an act. Saying "okay" as a careful closing is not equivalent to saying "okay" as a verdict; "I understand" as recognition is not equivalent to "I understand" as a cage. The word may be the same; the act is not.

Minimal example (same words, different acts):

A: "*Shall we review it tomorrow?*"

B (opening): "*Okay*." (soft falling tone + micro-pause + "yes, let's look at it")

B (closing): "*Okay*." (clipped delivery + abrupt drop + silence afterward)

In literal transcription, both responses are "Okay." In interaction, the permission to continue speaking is not the same.

From this thesis comes a decisive notion for the rest of the essay: prosody operates as a signature of coherence across layers. At times it reveals benign coherence—when intention, emotion, and form align and are perceived as clarity—but it can also reveal hostile coherence: when the real intention aligns with prosody while the content presents itself as neutral. In that case, what remains hidden is not the coherence itself, but where that coherence actually resides—which layer is truly sustaining the act.

In practical terms: there are conversations in which harassment is the coherent structure, yet it is camouflaged beneath correct wording. Prosody, by its temporal and embodied nature, thus becomes an especially relevant clue for detecting that coherence "between the lines."

(6) Couper-Kuhlen and Selting (1996) bring together studies of prosody in conversation and show how intonation, rhythm, and pauses operate as resources for social action and interactional organization. Couper-Kuhlen, E., & Selting, M. (Eds.). (1996). *Prosody in Conversation: Interactional Studies*. Cambridge: Cambridge University Press.

(7) Scherer (2003) reviews research paradigms on vocal communication of emotion and relates acoustic parameters (F0, intensity, tempo, voice quality) to affective states and evaluations. Scherer, K. R. (2003). "Vocal communication of emotion: A review of research paradigms." *Speech Communication, 40*(1-2), 227-256.

(8) Sacks, Schegloff, and Jefferson (1974) describe the system of turn organization in conversation and its practical rules, which are key to conceptualizing "turn order" as social structure. Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). "A simplest systematics for the organization of turn-taking for conversation." *Language, 50*(4), 696-735.

(9) Edmondson (1999) defines psychological safety in teams and links it to learning behaviors and voice, including the willingness to speak up, disagree, and report problems. Edmondson, A. (1999). "Psychological safety and learning behavior in work teams." *Administrative Science Quarterly, 44*(2), 350-383.

1.3. Scope and limits: linguistics + social psychology + institutional ethics + measurement

This essay does not aim to be an exhaustive manual of phonetics, a clinical treatise, or a legal tool. It is deliberately situated at a crossroads of fields:

· **Linguistics:** to specify what prosody is and how it participates in the construction of meaning.

· **Pragmatics:** to analyze what a sentence does in context (act, function, effect).

· **Social and organizational psychology:** to situate prosody within hierarchies, groups, and climates.

· **Institutional ethics:** to assess the "fruit" of communicative forms: whether they generate clarity and repair or whether they generate fear and submission.

· **Measurement:** to translate intuitions into observables without turning them into verdicts.

The limits are also explicit:

· **Prosody does not demonstrate intention on its own:** it suggests probabilities[10] and patterns, not inner certainties.

· **Prosodic norms vary** according to culture[11], region, class, age, gender, neurodiversity[12], and context; **there is no single universal reading.**

· **Not every unusual tone amounts to harm:** it may reflect fatigue, pain, shyness, style, professional training, or differences in processing.

· **Prosody does not replace contextual, documentary, or behavioral evidence in real conflicts.**

The wager is a modest one: if we want to care for complex human spaces, we need to learn to observe form, not only content. And that observation must be trainable, debatable, and, when useful, measurable.

1.4. Methodological note: from qualitative analysis to indicators (without emotional "polygraphs")

When speaking of prosody, two symmetrical dangers appear:

1.- The mystical danger: assuming that "the voice never lies" and turning prosody into an oracle.

(10) Ladd (2008) develops the phonology of intonation and emphasizes the contextual and probabilistic character of its interpretation, avoiding direct readings of intention from isolated signals. Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press.

(11) Jun (2005) compiles prosodic typologies across multiple languages and shows systematic variation in patterns of intonation and phrasing, supporting the idea that there is no single universal reading. Jun, S.-A. (Ed.). (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

(12) McCann and Peppé (2003) review the evidence on prosody in autism spectrum disorders, discuss atypical prosodic profiles, and highlight interpretive challenges relevant to considering neurodiversity in prosodic analysis. McCann, J., & Peppé, S. (2003). "Prosody in autism spectrum disorders: a critical review." *International Journal of Language & Communication Disorders, 38*(4), 325-350.

2.- The cynical danger: reducing everything to "subjective interpretation" and denying the performative dimension of speech.

This essay proposes a middle path: moving from "it is felt" to "it is observed" without leaping to "it is condemned."

The methodology adopted is organized around four complementary movements:

A) Qualitative reading of episodes. Concrete scenes are described (meetings, interviews, classes, trials, debates), and what the exchange produces is identified: an expansion of clarity or a shutting down of voice.

B) Coding[13] observable signals. The analysis moves from moral judgment down to data: pauses, changes in speed, variation in pitch (F0), energy (intensity), distribution of turns, patterns of interruption, silences following a contribution, and so forth.

C) Triangulation[14] with context and effect. The same signal can mean different things depending on context. The guiding question is not "What did they mean?" but "What structure did this produce here?": Did psychological safety increase or diminish? Did agency expand, or was submission induced? Did it open the way to repair, or did it generate fog?

D) Construction of pattern indicators[15]. Instead of "detecting emotions," repeated forms are described: asymmetries, streaks, outliers, recurring motifs. What is measured is not "the person"; what is measured is the dynamic.

Along these lines, the possibility of incorporating technical instrumentation is justified as long as the guiding principle is preserved: measure patterns, not inner states. The aim is not to infer an "inner truth" of the participants; it is to make visible what a system usually denies, so that it can be discussed with less ambiguity and greater care.

(13) Miles, Huberman, and Saldaña (2014) systematize qualitative analysis and coding techniques, including pattern coding, in order to convert observations into comparable and traceable categories. Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). Thousand Oaks, CA: SAGE.

(14) Denzin (1978) formalizes triangulation (of data, methods, researchers, and theories) as a strategy for strengthening inferences in social research and reducing the bias of single-interpretation readings. Denzin, N. K. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods* (2nd ed.). New York, NY: McGraw-Hill.

(15) DeVellis (2017) describes the development of scales and indicators (generation and refinement of items; evaluation of reliability and validity), which is useful for operationalizing constructs without turning them into verdicts. DeVellis, R. F. (2017). *Scale Development: Theory and Applications* (4th ed.). Thousand Oaks, CA: SAGE.

2. General Introduction

If the prelude argued that prosody is the plane where language "becomes embodied" —where a sentence ceases to be mere text and begins to operate as a relationship— this introduction sets the conceptual ground: it defines what counts here as prosody, what does not, and why that layer becomes decisive when formality functions as a mask.

The aim is not to settle the subject, but to establish an operational definition and a guiding intuition for what follows: history, conceptual framework, hidden coherences, organizational power, and, finally, measurement.

2.1. Operational definition of prosody (intonation, rhythm, pauses, stress, volume, speed, timbre...)

In this essay, "prosody" is not used as a synonym for "pleasant intonation," nor as a stylistic ornament. It is used in an operational sense: as the set of suprasegmental[16] and paralinguistic[21] features that modulate speech and that, in interaction[18], alter what a sentence does.

From this point on, the term prosody will include, at a minimum:

1.- Intonation (melody / perceived pitch)

· Rises and falls.

· Contours[17] (rising, falling, level, broken).

· Range used (wide vs. narrow; expressiveness vs. restraint).

2.- Stress and prominence (emphasis)

· Which syllable or word becomes the focus.

· Effects of shifting emphasis (the same verbal sequence can change the act it performs).

3.- Rhythm and tempo

· Overall speed and local variations.

· Accelerations, slowdowns, opening or closing cadences.

4.- Pauses and silences

(16) Crystal (1969) systematizes "prosodic systems" and describes prosody as the suprasegmental organization of speech (beyond the segment), providing a classic basis for an operational definition such as the one used in this essay. Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge, UK: Cambridge University Press.

(17) Gussenhoven (2004) describes how tonal contours and F0 range are organized in intonation systems and how they relate to pragmatic interpretations (continuity, closure, attitudes), which is useful for supporting the description of contours and ranges. Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge, UK: Cambridge University Press.

(18) Couper-Kuhlen and Selting (1996) gather studies showing prosody as an interactional resource (delimiting units, projecting closures, managing turn-taking and alignment), supporting the idea that it is not merely a "property of the speaker." Couper-Kuhlen, E., & Selting, M. (Eds.). (1996). *Prosody in Conversation: Interactional Studies*. Cambridge, UK: Cambridge University Press.

(21) Poyatos (1993) describes paralanguage as a systematic component of oral communication (non-lexical sounds, vocalizations, micro-signals) that is integrated into the turn and alters the communicative act beyond literal content. Poyatos, F. (1993). *Paralanguage: A Linguistic and Interdisciplinary Approach to Interactive Speech and Sound*. Amsterdam/Philadelphia: John Benjamins.

· Intra-phrase pauses and inter-turn pauses.

· Silences of containment, punishment[19], or freezing the interactional field.

5.- Intensity (energy / volume)

· Rises that impose, falls that humiliate, "aseptic" stability that chills.

· Intensity is not, by itself, equivalent to aggressiveness; it acquires meaning when it is differential and systematic.

6.- Duration

· Lengthenings ("suuuure…") or clipped forms ("fine," "okay") that function as closure, verdict, or control.

7.- Timbre / voice quality

· Roughness[20], nasality, breathy voice, tense voice, "smiling" voice, "metallic" voice, etc.

· This is an area especially vulnerable to misinterpretation if context is not controlled for (fatigue, health, environment, professional style).

8.- Integrated paralinguistic signals (when they operate within the turn and affect the act)

· Minimal laughter, sighs, throat-clearing, micro-sounds of disapproval.

· They are not always "prosody" in the strict phonological sense, but they form part of the channel through which what is said becomes an act.

It is worth stressing that prosody is not only a property of the speaker; it is also a property of the interaction. Its significance intensifies when it is combined with:

· **Turn organization** (who speaks, for how long, who interrupts, who closes).

· **Institutional context** (what can be said and what cannot; who pays a cost for dissent).

· **Effect** (what the interaction leaves behind: clarity or threat).

2.2. The central intuition: prosody as a signature between form and substance

The intuition guiding this essay is simple, but far-reaching: prosody functions as a signature. It does not certify "truth" in any mystical sense (it is not an infallible detector), but it does register

(19) Ephratt (2008) analyzes silence as a pragmatic behavior with conversational functions (e.g., suspension, containment, sanction/withdrawal), which is useful for treating pauses and silences as part of the prosodic-pragmatic repertoire. Ephratt, M. (2008). "The functions of silence." *Journal of Pragmatics, 40*(11), 1909-1938.

(20) Kreiman and Sidtis (2011) review voice quality (timbre/voice quality) from the perspectives of production and perception, showing its multicausal nature (physical state, technique, context), which is key to interpreting "roughness" and other qualities without turning them into a diagnosis. Kreiman, J., & Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Chichester, UK: Wiley-Blackwell.

something more operational: the actual alignment[22] between layers.

In particular, between:

· What is said (literal content).

· What is done in saying it (the pragmatic act).

· What is sustained bodily (the organization of threat or safety).

In the prelude, this idea was formulated as a bridge between intention, emotion, and social structure. In methodological terms, what matters is that this alignment can be observed as a temporal and relational pattern, even when the content presents itself as neutral.

2.3. "Hidden coherences": sarcasm, irony, double bind, exaggerated neutrality, repressed emotionality

The concept of "hidden coherences" is used here to name a specific phenomenon: the convergence between intention and form that remains partly hidden at the literal level of content. In such cases, the text may be formally correct, yet the act performed by the speech (and its effect) is organized on another layer.

Among the most frequent mechanisms are:

· **Sarcasm:** the literal content affirms A while the prosodic contour[23] indicates "not-A" (discrediting under verbal cover).

· **Irony:** literalness distances itself from a shared code; it operates as belonging for some and exclusion for others.

· **Double bind:** messages of care formulated with a prosody of control or punishment, generating submission[24] without the need for an explicit command.

· **Exaggerated neutrality:** withdrawal of affect while retaining power; formal correctness that can function as dehumanization.

· **Repressed emotionality:** intense containment that does not eliminate the charge, but displaces it into micro-signals (tension, clipped delivery, silences, timbre).

These coherences are not proposed as psychological "proofs," but as pragmatic-prosodic patterns that, under certain conditions, help explain why an exchange is experienced as hostile even when the text appears innocuous.

(22) Wennerstrom (2001) develops a framework of prosody and discourse analysis in which temporal patterns (prominence, rhythm, intonation) guide listener inferences and connect sound form with interactional function, providing a basis for the idea of observable alignments beyond content. Wennerstrom, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford, UK: Oxford University Press.

(23) Bryant and Fox Tree (2005) investigate whether there is an "ironic tone of voice" and which prosodic signals help listeners recognize irony/sarcasm in speech, supporting the idea that the prosodic contour can contradict literal content. Bryant, G. A., & Fox Tree, J. E. (2005). "Is there an ironic tone of voice?" *Language and Speech, 48*(3), 257-277.

(24) Bateson, Jackson, Haley, and Weakland (1956) introduce the concept of the double bind as a pattern of incompatible messages that traps the receiver's response, which is useful for describing verbal care delivered with a prosody of control/punishment and producing submission without an explicit order. Bateson, G., Jackson, D. D., Haley, J., & Weakland, J. (1956). "Toward a theory of schizophrenia." *Behavioral Science, 1*(4), 251-264.

2.4. Practical objective: measurable environments where hostility hides behind formality

The practical objective of this essay is twofold:

1.- To describe precisely how covert hostility is organized in environments where the text is regulated (institutions[25], organizations, public settings).

2.- To open a path toward observables that make it possible to discuss these phenomena with less ambiguity: not in order to issue verdicts, but to improve diagnosis, prevention, and intervention.

In fields such as education, politics, justice, business, or public-facing services, symbolic violence rarely presents itself explicitly. Its effectiveness depends precisely on sustaining formal "correctness." For that reason, prosody —and, inseparably, the management of time and turn-taking— becomes an especially relevant field for detecting relational dysfunctions that the text alone does not capture.

2.5. Guiding hypothesis: neutral content + prosody/silences reveal asymmetries

The guiding hypothesis that runs through this essay can be formulated operationally:

· When institutional or hierarchical hostility exists, literal content tends to preserve neutrality or correctness, while prosody, silences, and the organization of turns display consistent asymmetries.

In other words: wherever the text protects the speaker (plausible deniability[26]), the temporal and affective pattern tends to reveal the real order of the exchange (who is allowed to speak, who pays the cost, who is reduced). This hypothesis does not seek to replace contextual evidence; it seeks to direct observation toward that which often remains outside minutes, emails, and protocols.

(25) Drew and Heritage (1992) compile studies of interaction in institutional settings, showing how roles, procedures, and goals organize speech (turn rights, formats, constraints), which provides a basis for analyzing covert hostility where the text is regulated. Drew, P., & Heritage, J. (Eds.). (1992). *Talk at Work: Interaction in Institutional Settings*. Cambridge, UK: Cambridge University Press.

(26) Eisenberg (1984) describes strategic ambiguity in organizational communication as a resource for maintaining flexibility and avoiding explicit commitments, which connects with the notion of plausible deniability when the content remains "correct" and the asymmetry shifts into form. Eisenberg, E. M. (1984). "Ambiguity as strategy in organizational communication." *Communication Monographs, 51*(3), 227-242.

3. From Body to Sound

This chapter establishes the minimum bridge between a social intuition ("the form of speech alters the act") and a descriptive-observational basis: which acoustic and temporal variables underpin what is perceived as prosody. The aim is not to turn the voice into a clinical object, but to specify what is being measured when one speaks of pitch, intensity, rhythm, pauses, or voice quality.

3.1. Vocal production: respiration, phonation, resonance, articulation

The voice signal emerges from a chain of bodily processes[27] that are continuously coupled:

· **Respiration (support):** control of the airflow regulates the duration of phrases, the placement of pauses, the capacity to sustain emphasis, and the stability of energy. Changes in respiratory support translate into variations in intensity and phrasing patterns.

· **Phonation (laryngeal source):** the vibration of the vocal folds produces a periodic signal whose fundamental frequency is related to the perception of pitch. Laryngeal tension, adduction, and glottal control affect pitch, pitch stability, and voice quality (tense, breathy, creaky voice).

· **Resonance (filter):** the vocal tract (oral and nasal cavities) filters the signal, shaping timbre. Resonance influences the perceived "presence" or "metallic quality" of the voice, and it can change with posture, fatigue, emotion, or communicative strategy.

· **Articulation (segmental configuration):** the tongue, lips, and jaw determine consonants and vowels, but they also affect the relative duration of syllables, articulatory clarity, and the perception of precision or looseness (factors relevant to the pragmatic interpretation of confidence, doubt, or control).

In practice, prosodic features are not "added" to words from the outside; they are organized through these mechanisms as a continuous regulation. For that reason, any prosodic measurement should be understood as an approximation to a dynamic system, not as an exhaustive capture of the phenomenon.

3.2. What "intonation" means in acoustic terms: F0/pitch and its limits

In acoustic terms, intonation is approximated through estimates of fundamental frequency (F0[28]), which is related to the perception of pitch. This relationship is useful, but it should not be mistaken for a perfect equivalence:

· **F0 is a physical measure; pitch is a perception.** The perception of pitch also depends on context, the speaker's habitual range, and local contrast.

(27) Titze (1994) develops a physical-physiological explanation of voice production (respiration, phonation, and laryngeal control) and its relation to parameters such as F0, intensity, and voice quality. Titze, I. R. (1994). *Principles of Voice Production*. Prentice Hall.

(28) Stevens (2000) presents an acoustic theory of speech production (including the source-filter model) and describes the physical basis of measures such as fundamental frequency (F0) and its perceptual correlate. Stevens, K. N. (2000). *Acoustic Phonetics*. The MIT Press.

· **F0 exists in a stable form only in voiced segments.** In voiceless consonants, whispers, or noisy segments, F0 estimation may be absent or erratic.

· **Estimation errors are common.** Automatic algorithms may "double" or "halve" the real value (octave errors)[29], especially with very low or very high voices, background noise, or low-quality microphones.

· **Comparisons across speakers require normalization.** Range, register, and pitch baseline vary among speakers; for that reason, relative changes (rises, falls, internal ranges) within the same episode or within the same speaker are usually more informative than absolute values.

For the purposes of this essay, the key point is operational: intonation is not interpreted through the F0 "number" itself, but through patterns of contour[30], range, prominence, and temporal position (for example, how closure is marked, how irony is sustained, or how repair is introduced or denied).

3.3. Energy/intensity: RMS/dB and how it relates to emphasis and control

Perceived intensity (volume) is commonly approximated through measures of signal energy (for example, RMS[31]) and its expression on logarithmic scales (dB). These measures make it possible to describe patterns, but they also present important limits:

· **Channel dependence:** distance from the microphone, automatic compression, echo, device type, and background noise can substantially alter energy values.

· **Intensity does not by itself equal dominance:** a voice may be quiet because of style, fatigue, or care; a voice may be loud because of acoustic context or non-hostile expressiveness.

· **The relevant information is usually relative and differential:** local rises (emphasis), aseptic stability, abrupt drops, or systematic asymmetries between interlocutors (who "rises" to impose and who "drops" to withdraw).

In pragmatic terms, intensity becomes significant when it participates in the organization of the turn: to impose, cut off, occupy the floor, intimidate, or, conversely, to soften, facilitate, and sustain listening. For that reason, energy measurement acquires value when it is combined with temporal analysis and with the distribution of turns.

3.4. Time: duration, rhythm, variability, pauses (and why time bears the signature of intention)

If pitch and energy are usually associated with "tone" and "emphasis," time is often the plane on which the architecture of the act becomes most visible. In human interaction, time organizes

(29) de Cheveigné and Kawahara (2002) propose the YIN F0 estimator and describe typical errors in automatic estimation, including "octave errors." de Cheveigné, A., & Kawahara, H. (2002). "YIN, a fundamental frequency estimator for speech and music." *The Journal of the Acoustical Society of America, 111*(4), 1917-1930.

(30) Gussenhoven (2004) explains how intonation is interpreted as a configuration of contours, range, and prominence, and how its temporal alignment contributes to pragmatic meaning. Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.

(31) Kent and Read (2002) synthesize methods for acoustic analysis of speech and discuss energy/intensity measures (e.g., RMS/dB), together with practical limitations related to channel, distance, and noise. Kent, R. D., & Read, C. (2002). *The Acoustic Analysis of Speech* (2nd ed.). Singular/Thomson Learning.

permission, priority, pressure, and the possibility of response[32].

From an observational perspective, time can be described through:

· **Segment duration:** length of words, syllables, or prosodic units (lengthenings, truncations).

· **Speech rate:** rate of production (for example, syllables per second) and local variations.

· **Rhythm and variability:** alternation of fast/slow stretches; stability vs. irregularity.

· **Pauses:** location and duration of intra-phrase and inter-turn silences.

· **Response latency:** reaction times after another person's intervention, which may function as acknowledgment, containment, rejection, or punishment.

· **Turn management:** overlaps, interruptions, abrupt closures, and reopenings.

The reason for its relevance is simple: time is difficult to falsify in real time when there is emotional charge or an intention to control. Not because it reveals an inner truth, but because time is the very medium through which conversational power is exercised: one accelerates to dominate, slows down to humiliate, interrupts to remove agency, leaves silence[33] to sanction, or creates space to enable repair.

Methodologically, this temporal dimension is especially suited to a cautious approach: it allows patterns to be detected (for example, consistent asymmetries) without claiming to read "mental states."

3.5. Voiced/unvoiced: "voiced ratio" and noise (fatigue, microphone, environment, neurodiversity)

In addition to pitch and intensity, it is useful to distinguish between stretches with periodic phonation and stretches without it. A common metric for approximating this is the voiced ratio[34] (the proportion of time with a clearly voiced signal as opposed to unvoiced or noisy time). Its usefulness is twofold:

· **It describes vocal texture:** a higher proportion of periodic voicing is usually associated with stable phonation; a lower proportion may reflect whispering, breathiness, creakiness, noise, or cut-outs.

· **It helps detect signal conditions:** background noise, compression, poor capture, or fatigue can reduce the reliability of other measures (F0, energy), and it is useful to register this.

However, it is an indicator especially sensitive to context and diversity:

(32) Sacks, Schegloff, and Jefferson (1974) formulate a model of turn-taking in conversation, showing how pauses, overlaps, and latencies structure rights and obligations of participation. Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). "A simplest systematics for the organization of turn-taking for conversation." *Language, 50*(4), 696-735.

(33) Jaworski (1993) analyzes silence as a pragmatic and social resource (containment, power, sanction), emphasizing that its function depends on the interactive frame. Jaworski, A. (1993). *The Power of Silence: Social and Pragmatic Perspectives*. Sage.

(34) Kreiman and Van Lancker Sidtis (2011) offer an interdisciplinary framework on the production and perception of voice quality, including the distinction between periodic/aperiodic components and their perceptual correlates. Kreiman, J., & Van Lancker Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Wiley-Blackwell.

· **Fatigue and vocal health** alter voicing stability.

· **Acoustic environment** (reverberation, noise) can confuse "unvoiced" stretches.

· **Communicative styles and neurodiversity** may involve more restrained or more tense voice patterns, or variations in voice quality that are not, in themselves, hostility.

For that reason, the responsible use of these variables requires one rule: every acoustic signal must be interpreted as data that require contextualization, not as a diagnosis. Its value emerges when it is triangulated with the temporal pattern, the organization of turn-taking, and the relational effect observed.

4. The Long History of Prosodic Power

If prosody is the temporal and embodied dimension of language—its rhythm, contour, and energy—then its history cannot be reduced to a history of "styles." From a very early stage, prosody has functioned as a social technology: it organizes shared attention, distributes authority, marks belonging, and regulates collective affective states. Before writing existed as a stable archive, the voice was already a living archive[35]: repeatable, trainable, and capable of producing public effects.

This section does not seek to reconstruct an exhaustive chronology (an impossible task in an essay of interdisciplinary scope), but rather to establish a historical backbone: to show that the prosodic dimension predates modern normative frameworks and that, precisely because of that, it retains a particular power in contexts where the text is cleaned up and intention becomes "deniable." In practical terms, the genealogy proposed here supports a simple thesis: as language becomes institutionalized, prosody becomes a privileged site for reading the implicit, because verbal content learns to cloak itself in formality while interaction continues to filter asymmetries.

4.1. Prehistory and proto-ritual: voice, rhythm, and trance as social technology

At the origins, there is no "discourse" separate from "body." The voice is shared breathing and coordination; rhythm is a way of synchronizing[36] multiple organisms; repetition is a way of stabilizing memory and belonging. From this perspective, prosody appears as an infrastructure of cohesion: before it is "individual expression," it is a tool of group attunement.

In proto-ritual[37] contexts (and in ritual practices that persist across many cultures), rhythmic repetition fulfills at least three socially relevant functions:

- **Fixation and transmission.** Repetition—cadences, formulas[38], stress patterns—allows content to travel without depending on an external medium. A group remembers because a group repeats. Prosody does not merely "beautify" an utterance: it makes it transportable, turning it into a memorable form.

- **Affective modulation.** Gradual intensification, shifts in tempo, or alternation between silence and rhythmic beat regulate states of alertness, calm, mourning, or euphoria. Prosody operates as a mechanism of emotional guidance: it does not describe a state; it produces or channels it.

- **Boundary production.** Shared song or vocal pattern separates "inside" from "outside": whoever knows the rhythm belongs; whoever does not is left out. Here, prosody is not merely communication: it is a sign of belonging, a marker of access.

(35) Ong (1982) analyzes the difference between orality and literacy and describes how, prior to written record, memory and vocal repetition function as a social "archive" in oral cultures. Ong, W. J. (1982). *Orality and Literacy: The Technologizing of the Word*. Methuen.

(36) McNeill (1995) studies how rhythmic synchrony (dance, marching, "drill") produces group cohesion and coordination, linking shared time with social organization. McNeill, W. H. (1995). *Keeping Together in Time: Dance and Drill in Human History*. Harvard University Press.

(37) Durkheim (1912/1995) explains ritual as a mechanism of social integration and the generation of collective effervescence, which is useful for grounding the cohesive function of shared and repeated patterns. Durkheim, E. (1995). *The Elementary Forms of Religious Life* (K. E. Fields, Trans.). *Free Press*. (Original work published 1912)

(38) Lord (1960) describes the transmission of oral poetry through formulas and repetition, showing how rhythmic and formulaic patterns make content transportable without external support. Lord, A. B. (1960). *The Singer of Tales*. Harvard University Press.

Within that early matrix, the voice also participates in the management of crisis: mourning, external threat, illness, scarcity. Laments, war cries, call-and-response, and voices of care (lullabies, soothing vocalizations) show a constant: the group organizes its experience through temporal patterns. Prosody thus appears as a technology of coordination and, at the same time, as a space where hierarchy is inscribed: who begins the chant, who sets the pulse, who "carries" the voice.

From this stage emerges a first historical thesis useful for the essay: prosody is not a late ornament of language; it is a primary tool of collective direction. And if it is primary, then its reading cannot be treated as a mere "detail": it is often there that the kind of relationship speech establishes is decided.

4.2. Antiquity: oratory, authority, and public truth

With the emergence of more complex political forms—assemblies, courts, public deliberation—the voice enters a new scene: that of public truth. In that space, it is not enough to "be right"; one must appear credible, sustain presence, manage the turn, capture attention, and maintain an energy that persuades. Prosody ceases to be only ritual coordination and becomes a technique of persuasion.

The classical rhetorical tradition (in all its internal diversity) is, at bottom, a formalization of something that was already operative: the relationship between manner of speaking and social effect. Within that framework, prosody becomes part of training[40]. It is not only a matter of "ideas"; it is a matter of how those ideas become effective before an audience.

Three dimensions are especially relevant for this essay:

· **Prosody as the construction of authority (*ethos*[39]).** Credibility is embodied: a steady rhythm, a "confident" cadence, a well-placed pause can produce an effect of control even when the content is fragile. Conversely, vocal tremor or excessive acceleration can destroy credibility even when the argument is sound. The voice is not a neutral channel: it is a component of the public perception of competence and legitimacy.

· **Prosody as affective activation (*pathos*).** Ancient rhetoric recognizes—explicitly or implicitly—that audiences are not moved by abstract reasons alone. The voice is an instrument for guiding attention, intensifying indignation, producing compassion, signaling threat, or conveying gravity. In contemporary terms, prosody functions as a marker of emotional framing.

· **Prosody as a form of order (*logos* on stage).** Even a "rational" argument needs temporal structure in order to be received. Prosody segments, hierarchizes, anticipates, and concludes. A pause can turn a phrase into a verdict; a final rise can turn it into an invitation or a challenge. In this way, prosody organizes logic not only as content, but as the listener's temporal experience.

The historical result is decisive: when public life becomes institutionalized, a structural tension appears between "what is said" and "what the voice does." In assemblies, forums, or courts, prosody begins to operate as a device of legitimation: solemnity, threat, mockery, or condescension can be

(39) Aristotle (2007) develops the triad ethos-pathos-logos and the role of credibility and emotion in public persuasion, providing the classical basis for linking manner of speaking and audience effect. Aristotle. (2007). *On Rhetoric: A Theory of Civic Discourse* (G. A. Kennedy, Trans., 2nd ed.). Oxford University Press.

(40) Quintilian (2001) details the role of actio (delivery) and the training of voice, pauses, and cadence in oratory, as a historical codification of prosodic control in rhetorical education. Quintilian. (2001). *The Orator's Education* (Institutio Oratoria) (D. A. Russell, Trans.). Harvard University Press. (Original work published c. 95 CE)

conveyed without the need for explicit insult. The wording may remain formal; the voice may carry the act.

This stage introduces a principle that will reappear later: symbolic violence[41] does not require explicit verbal violence. It can function through tone, rhythm, interruption, or control of the turn. And the more public language becomes formalized, the more refined becomes the ability to deny ("I did not say that") while preserving the effect ("but it was made felt").

4.3. Religious and poetic traditions: prosody as a container for the sacred

If public oratory links prosody to political authority, religious and poetic traditions reveal another axis: prosody as a container of the sacred and as a technology of subjective transformation. Here, the voice is not only persuasion; it is a mode of access: to a text, to a community, to an inner state, to a shared experience.

Across many traditions, recitation is not silent reading, but a regulated vocal act: prosody matters because it sustains a mode of relation both to the text and to the group. Patterns of intonation, duration, ritual silences, and the distribution of stresses all configure a form of presence. Content matters, but its efficacy depends on the form that carries it.

This dimension has several implications for the line of argument developed in this essay:

· **Saying as doing.** In religious contexts, certain formulas do not merely describe; they perform[42] (to bless, promise, consecrate, absolve, swear). Prosody becomes part of the perceived "validity" of the act: not because of superstition, but because the act is social and requires marks of shared recognition. An oath without prosodic solemnity may be perceived as empty; a blessing without a tone of care may be perceived as a mechanical gesture.

· **Rhythm-breathing-state.** Prosodic repetition and rhythm influence breathing[43], and breathing influences state. This does not imply mysticism; it implies basic physiology: the body responds to temporal patterns. For that reason, prosody in religious and poetic contexts functions as an engineering of attention: it sustains concentration, reduces dispersion, and can induce calm or intensification.

· **The voice as a link between symbol and body.** Poetry and recitation show that prosody is not an "ornament" added to meaning: very often, it is the mode through which meaning becomes perceptible. A verse is not only semantics; it is temporal tension. A psalm is not only doctrinal content; it is a form of shared breathing.

This tradition contributes a key idea to the essay: prosody functions as a container of meaning. It not only transmits; it organizes, protects, intensifies, or cools. And that capacity to "contain" will be

(41) Bourdieu (1991) conceptualizes the symbolic power of language and shows how "legitimate" styles and registers can exercise domination without explicit insult, supporting the idea of symbolic violence through form. Bourdieu, P. (1991). *Language and Symbolic Power* (J. B. Thompson, Ed.; G. Raymond & M. Adamson, Trans.). Harvard University Press.

(42) Austin (1962) formulates speech act theory and performatives, clarifying how certain institutional or ritual formulas "do" things (swear, absolve, consecrate) in addition to describing them. Austin, J. L. (1962). *How to Do Things with Words*. Oxford University Press.

(43) Bernardi et al. (2001) show that repetitive recitation (rosary/mantras) can couple with breathing rhythm and modulate autonomic cardiovascular rhythms, supporting the link between rhythm, breathing, and state. Bernardi, L., Sleight, P., Bandinelli, G., Cencetti, S., Fattorini, L., Wdowczyc-Szulc, J., & Lagi, A. (2001). "Effect of rosary prayer and yoga mantras on autonomic cardiovascular rhythms: Comparative study." *BMJ, 323*(7327), 1446-1449. doi:10.1136/bmj.323.7327.1446

crucial when, in institutional contexts, correct form is used to contain something else: not the sacred, but control.

4.4. Modernity: bureaucracy, institutions, and performative neutrality

Modernity introduces two developments that are especially relevant to the history of prosodic power. On the one hand, there is the expansion of writing(44), the archive, and formal documentation; on the other, the consolidation of institutions that must appear impartial(45): courts, public administrations, educational systems, and complex organizations. Within this framework, a specific aesthetic emerges: performative neutrality.

Performative neutrality is not only an ethical norm ("it should be impartial"), but also a style that is learned and reproduced: stable tone, contained emotion, technical vocabulary, controlled rhythm. At times, this style serves legitimate functions: it reduces arbitrariness, protects procedures, and prevents momentary emotion from capturing decisions. But it also has a reverse side: it can become a mask of power, especially when procedure replaces care and repair.

At this point, a precise distinction is necessary: the problem is not neutrality as an ideal; the problem is neutrality as an aesthetic device that produces effects of hierarchy without naming them. Some typical forms of this performative neutrality are:

· **A "technical" voice in the face of suffering.** When someone presents harm and the institutional response remains in a cold register, the content may be correct ("we take note," "a case file will be opened"), but the prosody may produce dehumanization: the person does not feel heard; they feel managed.

· **Formality as closure.** The use of final cadences, pauses that do not open space, and a tempo that blocks reply can turn a formal sentence into a verdict. There is no need to shout in order to close a conversation; a prosody that leaves no air is enough.

· **Condescending correction.** Modernity organizes knowledge into professional hierarchies. Prosody can translate that hierarchy into treatment: corrections that, by the way they are delivered, do not seek to clarify, but to mark inferiority. It is a power that is not declared, but enacted.

Moreover, modernity multiplies the settings in which the voice becomes political without naming itself as such: the classroom, the job interview(46), the corporate meeting, customer service, the institutional statement, the parliamentary session, crisis communication. In all these spaces, prosody participates in the production of social reality: it enables or inhibits agency, distributes turns, and defines which emotion is legitimate and which counts as "excess."

The paradox that matters for this essay is the following: the more institutional the context, the

(44) Goody (1986) analyzes how writing, record-keeping, and textualization reconfigure social organization and institutions, linking archive/documentation to new forms of control and administration. Goody, J. (1986). *The Logic of Writing and the Organization of Society*. Cambridge University Press.

(45) Weber (1922/1978) describes bureaucracy and rational-legal authority, emphasizing regulated procedures and impersonality as the basis of the ideal/appearance of institutional impartiality. Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology* (G. Roth & C. Wittich, Eds.). University of California Press. (Original work published 1922)

(46) Heritage and Clayman (2010) synthesize research—especially from conversation analysis—on institutional interaction, showing how asymmetries, turn design, and question formats structure contexts such as interviews and other regulated situations. Heritage, J., & Clayman, S. (2010). *Talk in Action: Interactions, Identities, and Institutions*. Wiley-Blackwell.

"cleaner" the text can become, and the more relevant prosody becomes as a clue to what is real. Institutions learn to correct words, to draft texts without adjectives, to avoid compromising statements. And yet interaction still takes place in real time, through bodies: someone interrupts, someone sustains silences, someone turns selectively cold, someone speeds up to dominate, someone slows down to humiliate. Content may be shielded; the temporal dynamic continues to filter structure.

That is why this genealogy prepares the ground for what follows: if prosody has historically been a tool of cohesion, authority, belonging, and control, then it is no surprise that, in contemporary contexts, it also functions as a vehicle of hidden coherences: stable patterns of treatment that are not signed in the text, but enacted in the form.

The next chapter proposes a conceptual framework for describing precisely that intersection between semantics, pragmatics, and prosody, and for delimiting the role of micro-signals and silences in the production of subtext.

5. Conceptual Framework

The preceding chapters have delimited the object (prosody), its material basis (body-sound), and its social and historical force (the voice as a technology of coordination, authority, and control). What remains now is to establish a conceptual framework that makes it possible to analyze, with sufficient precision, which part of meaning belongs to literal content, which part belongs to the social action performed in speaking, and how the temporal sound-form makes that action "real" in interaction.

This chapter serves an operative function: to organize levels and concepts so as to avoid two frequent reductions. The first is to semanticize everything (to believe that what matters lies only in "what was said"); the second is to psychologize everything (to believe that prosody directly reveals a person's interior). The goal is more sober: to describe how prosody, in context, participates in the production of meaning, in the distribution of conversational power, and in the generation of coherences (or incoherences) between discourse and treatment.

5.1. Distinguishing levels: semantics vs. pragmatics vs. prosody

A conversation is not merely a sequence of propositions: it is a sequence of situated acts[47]. It is therefore useful to distinguish three analytical levels, while keeping in mind that in practice they appear intertwined.

1) Semantics (what is said). Semantics refers to the propositional content[48] of the utterance: that which, approximately, can be transcribed without losing its logical structure. In everyday terms: "the words." It is the level at which one can discuss whether a sentence is informative, contradictory, ambiguous, or grammatically correct.

2) Pragmatics (what it is said for; what is done in saying it). Pragmatics refers to the use of language in context[49]. It includes, among other things:

· Speech acts: requesting, ordering, implying, warning, apologizing, threatening, ridiculing, legitimizing, invalidating.

· Implicatures:[50] what is communicated without being said explicitly (for example, through politeness, strategy, or plausible deniability).

· Face-work[51] and status management: the ways in which the other's dignity is protected or eroded, and how each person's place is negotiated (authority, competence, belonging).

· Situational frame: who is speaking, before whom, with what roles, under what formal rules, and under what risks.

(47) Austin (1962) develops speech act theory and shows that to utter is to perform actions within a situational framework, providing the basis for treating conversation as a sequence of situated acts. Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

(48) Lyons (1977) presents semantics as the study of linguistic meaning and propositional content, which is useful for delimiting the level of "what is said" as distinct from contextual use. Lyons, J. (1977). *Semantics* (Vols. 1-2). Cambridge: Cambridge University Press.

(49) Huang (2014) synthesizes pragmatics as the study of meaning in use: how context, roles, and inferences turn a sentence into a social act. Huang, Y. (2014). *Pragmatics* (2nd ed.). Oxford: Oxford University Press.

(50) Grice (1975) formulates the cooperative principle and conversational maxims in order to explain implicatures, that is, what is communicated without being literally stated. Grice, H. P. (1975). "*Logic and conversation.*" In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics* (Vol. 3: Speech Acts, pp. 41-58). New York, NY: Academic Press.

(51) Goffman (1967) introduces the notion of face and face-work in interaction, describing how dignity/status is maintained or eroded beyond literal content. Goffman, E. (1967). *Interaction Ritual: Essays on Face-to-Face Behavior*. Garden City, NY: Anchor Books.

Pragmatics makes it possible to understand why two semantically similar sentences can be socially opposite. “We should review it” can be help, evasion, threat, or closure; the difference does not lie in the dictionary, but in the context and in the signs that turn the sentence into an act.

3) Prosody (how what is said becomes real). In this essay, “prosody” designates the set of temporal and acoustic features[52] (intonation, rhythm, pauses, stress, intensity, speed, variability, timbre in a broad sense) that shape the realization of speech and alter its interactive effect. Prosody is especially relevant because it:

· Connects language with the body (breathing, tension, control, available energy).

· Participates in the organization of turns (the beginning, maintenance, and closure of interventions).

· Carries signals of relational stance (care, distance, impatience, disauthorization, mild mockery, concealed threat).

· Modulates the degree of “openness” or “closure” that a sentence generates in the other person.

This distinction allows one to formulate a central idea of the essay more precisely: the incoherence that causes the most harm in institutional contexts is usually not semantic, but pragmatic and prosodic. The text may be formally correct; the conversational action and the sound-form may operate as control, harassment, or dehumanization.

From this perspective, the analytical question changes. Instead of asking only “What did they say?”, it becomes necessary to ask:

· ***What did they do with what they said? (pragmatics)***

· ***How did they make it audibly and bodily effective? (prosody)***

· ***What did it produce in the structure of the interaction? (observable outcome)***

5.2. Prosody as the temporal structure of language

Prosody can be understood as the temporal dimension of language: that which organizes the listener’s experience in time. Unlike a text, which can be “grasped” visually all at once, speech unfolds in an irreversible sequence. That sequence creates expectations, confirms or breaks patterns, and regulates the possibility of intervention.

Time is not an accessory; it is the medium of relation.

In interaction, time organizes at least four decisive functions:

1. Segmentation and informational hierarchy. Prosody marks which part of what is said

(52) Cutler, Dahan, and van Donselaar (1997) review evidence showing that prosody (stress, rhythm, intonation, pauses) guides speech comprehension and alters interactive effect in real time. Cutler, A., Dahan, D., & van Donselaar, W. (1997). “Prosody in the comprehension of spoken language: A literature review.” *Language and Speech, 40*(2), 141-201.

is presented as focus[53], which part as background, what is emphasized, what is minimized, and where the boundaries of interpretable units lie. An emphasis can turn a neutral sentence into an accusation; a pause can turn an assertion into an insinuation; a falling contour can turn an invitation into a verdict.

2. Turn management[54] and control of conversational space. Speech is not only emission; it is negotiation of access. Prosody helps signal continuation ("I'm still going"), closure ("I'm done"), yielding ("your turn"), or blockage ("there is no room for your reply"). At this point, temporality is not merely acoustic: it is a form of micro-social governance.

3. Regulation of threat[55] and safety. Prosody participates in the affective calibration of the environment. Accelerated rhythms, abrupt cut-offs, silences following someone's intervention, or disproportionate intensities can function as signals of threat, even when the content is formally benign. Conversely, stable prosody and a cadence that offers response time can function as signals of safety.

4. The transport of intention under plausible deniability. There are reasons why time "signs" more than text: sustaining a semantic mask is relatively easy; sustaining, for minutes or hours, a temporal mask coherent with that semantics is more difficult. Many strategies of conversational domination are not expressed as explicit insult, but as dynamics: speeding up when the other person speaks, prolonging silences when they ask, cutting off closures, or colonizing the total time.

For that reason, in an analysis aimed at detecting hidden coherences, prosody is not treated as "emotional color," but as the temporal architecture of interaction. And an architecture is recognized through patterns: the distribution of pauses, the repetition of stresses, the stability or instability of rhythm, and the consistency of sonic treatment toward different interlocutors.

5.3. Micro-signals and meta-signals: pause, minimal laughter, emphasis, technical coldness

One of the risks in speaking about prosody is overreacting to an isolated signal. In complex contexts, a long pause may be due to cognitive load, legitimate doubt, environmental noise, fatigue, or communicative style. To avoid hasty interpretations, it is useful to distinguish micro-signals from meta-signals[56].

Micro-signals

These are localized events, short in temporal scale, that appear as "marks" in the flow of speech. Some frequent examples in the kinds of phenomena that concern this essay are:

(53) Lambrecht (1994) develops the framework of information structure (topic/focus) and its encoding, supporting the idea that focus/background and hierarchy are also marked through prosody. Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

(54) Duncan (1972) describes signals and rules for taking and yielding turns in conversation, including prosodic cues that project continuation, closure, or yielding, and that organize access to conversational space. Duncan, S. (1972). "Some signals and rules for taking speaking turns in conversations." *Journal of Personality and Social Psychology, 23*(2), 283-292.

(55) Juslin and Laukka (2003) review how the voice communicates emotion through patterns of tempo, pitch, intensity, and timbre, supporting prosody as a calibration of threat/safety and affective climate. Juslin, P. N., & Laukka, P. (2003). "Communication of emotions in vocal expression and music performance: Different channels, same code?" *Psychological Bulletin, 129*(5), 770-814.

(56) Goodwin (1981) analyzes conversational organization and how local marks acquire meaning within sequential patterns, which is useful for distinguishing punctual micro-signals from meta-signals as second-order regularities. Goodwin, C. (1981). *Conversational Organization: Interaction between Speakers and Hearers*. New York, NY: Academic Press.

· **Pauses:** before responding, after someone else's intervention, in positions where continuity would normally be expected.

· **Shifted emphasis:** strong stress on words that, semantically, would not require it, thereby generating insinuation or irony.

· **Lengthenings:** prolonging vowels or consonants as a gesture of mild mockery, disbelief, or pressure.

· **Minimal laughter or laughter without joy:** micro-laughs, exhalations, or brief "ha" sounds that function as disauthorization without insult.

· **Strategic pitch descents or rises:** closures that sound like verdicts, openings that sound like interrogation, or suspended endings that force the other person to complete the sense.

· **"Technical" register:** a flat, stable, depersonalized prosody that may be professional in some cases, but dehumanizing in others when applied to situations of pain or vulnerability.

Meta-signals

These are second-order patterns: regularities that become visible when episodes are accumulated and conditions are compared. Examples include:

· **Systematic asymmetry of pauses:** long silences after one specific person's interventions, but not after others'.

· **Selective stability:** cordial tone with some people, cold tone with others, consistently.

· **Non-random interruptions and overlaps:** recurrent cut-offs directed at certain voices or certain topics, especially when dissent appears.

· **Rhythm changes tied to sensitive content:** acceleration when responsibilities are mentioned, slowing down when precision is demanded, or extreme "calm" when harm is being narrated.

· **Repeated closure structures:** ending an interlocutor's turns with short phrases + falling contour + pause, as a mode of shutting things down.

The important methodological shift is this: a micro-signal is rarely diagnostic on its own; a sustained meta-signal can be. The reading of hidden coherences does not consist in "guessing" intention from an isolated tone, but in observing regularities that persist despite situational variation.

In operative terms, this distinction prepares the ground for measurable analysis: what can be quantified more robustly is not "bad intention," but the structure of signal distribution (timing, turns, variability, asymmetries) and its relationship to observable effects (closure of voice, increased caution, avoidance of dissent).

5.4. Subtext as a field: meaning "between" words

Subtext is often imagined as a "hidden message" that the speaker carries inside. That metaphor is useful, but insufficient. In interaction, subtext works better as a field[57]: a zone of meaning that emerges between what is said, what is left unsaid, what is assumed, and what is marked by form.

Several sources converge within that field:

1. Context and roles. A sentence does not mean the same thing when uttered by an equal as when uttered by an authority figure. Power does not merely add external consequences; it reorganizes the pragmatic meaning of the utterance.

2. Relational history. The same prosody may be playful in a trusting relationship and threatening in one marked by prior retaliation. Interactional memory turns form into an anticipatory signal.

3. Environmental norms. In some systems, dissent is rewarded; in others, it is punished. Where punishment is likely, subtext tends to become charged with risk: "what I say may be turned against me." There, prosody, even unintentionally, becomes an indicator of real margin.

4. Prosody as an inferential operator. Prosody does not merely "accompany" subtext; it constructs it. It marks irony, marks distance, marks mild threat, marks condescension, marks care. This is especially clear in highly semantically ambiguous expressions ("sure," "okay," "perfect," "I understand," "don't worry") where pragmatic function depends almost entirely on the way they are said.

5. Sound framing (music, effects, and absence). In recorded materials and, above all, in edited pieces (podcasts, videos, "produced" voice notes), the voice almost never arrives "raw": it arrives inside a sonic container. That container includes not only what is added (music, ambiance, effects), but also what is cut or rearranged (noise cleaning, compression, "clean" silences, removed breaths). These layers function as contextualization cues: they predispose the listener's affective state and, with it, the pragmatic reading of the utterance. In audiovisual sound theory, this is referred to as "added value": sound can make the same verbal material seem more tense, warmer, or more ironic without changing a single word[58]. And because music is especially effective at communicating and evoking emotion, its presence (or absence) can act as a prime that makes certain hidden coherences more or less perceptible[59]. In this sense, the soundtrack can operate as a subterranean narration: it unifies, punctuates, and orients evaluation without declaring itself.

In music psychology, it has been shown that a soundtrack can guide the comprehension and memory of a scene, operating as an interpretive schema[60]. And in studies of audiovisual congruence, the interpretation of ambiguous actions changes according to the (semantic and temporal) congruence between music and action, and even according to whether there is music or

(57) Stalnaker (2002) formulates the concept of common ground in order to model the background of assumptions that makes what lies "between the lines" inferable, connecting subtext with roles, history, and norm. Stalnaker, R. (2002). "Common ground." *Linguistics and Philosophy, 25*(5-6), 701-721.

(58) Chion develops the idea of "audio-vision" and "added value": how sound design (including music) reconfigures perceived meaning without changing the text. Chion, M. (1994). *Audio-Vision: Sound on Screen. Edited and translated by Claudia Gorbman.* New York: Columbia University Press.

(59) A key compendium on how music communicates and evokes emotion (mechanisms, evidence, and nuances), useful for supporting the affective priming effect of background music. Juslin, P. N., & Sloboda, J. A. (Eds.). (2010). *Handbook of Music and Emotion: Theory, Research, Applications.* Oxford: Oxford University Press.

(60) Experimental evidence that a soundtrack guides schemas of comprehension and memory for narrative material (it does not merely "decorate"): it alters what is understood and what is remembered. Boltz, M. G. (2001). "Musical soundtracks as a schematic influence on the cognitive processing of filmed events." *Music Perception, 18*(4), 427-454.

not[61]. Within the framework of this essay, this matters because:

· It can enhance coherences: when the soundtrack or effects are congruent with the direction of the act, they reinforce the "audibility" of the tension between literalness and treatment.

· It can conceal them: through rhetorical covering (a warm aesthetic that makes a cold prosody less visible), and also for perceptual reasons; the ear organizes mixtures, and certain components can mask vocal micro-signals, reducing their detectability for some listeners[62].

From this follows a caution: in mediated material, coherence "in the voice" and coherence "as perceived" do not always coincide. Sometimes the absence of music makes breathing, pause, and energy more salient—and therefore reveals more; at other times, the absence leaves more room for interpretive projection. In both cases, the coherence may still be there: what changes is its visibility/audibility within the sonic container.

From this perspective, subtext is not reduced to a private intention, but becomes an emergent property of the exchange: the way in which an utterance becomes interpretable under certain conditions of power, history, and norm. For that reason, the analysis of subtext requires two cautions:

· **Interpretive caution:** do not immediately attribute psychological meaning to a signal.

· **Structural caution:** do not minimize the real effect of a pattern simply because "there is no explicit insult."

Subtext is precisely the place where a large part of social life is decided: belonging or symbolic expulsion, permission or prohibition, recognition or ridicule, repair or closure.

5.5. The role of silences: containment / punishment / complicity / reputation

If prosody is the temporal dimension of language, silence is its most charged edge. In many interactions, silence is not the absence of communication: it is an act with a function. Analytically, it can be treated as part of the prosodic-pragmatic repertoire.

Within the framework of this essay, it is useful to distinguish at least four frequent functions of silence, which may coexist or alternate:

1) Silence as containment. This is the silence that holds: it creates space for the other person to think, regulate themselves, find words, or not be pushed to answer immediately. In contexts of care, this silence may be a sign of respect for the interlocutor's rhythm. In terms of effect, it tends to increase safety and agency.

2) Silence as punishment. This is the silence that sanctions: it appears as emptiness after an intervention, as intentional delay, as non-response, or as suspension that leaves the other person

(61) A classic study of semantic and formal congruence: music can "shift" the interpretation of ambiguous visual actions; it includes conditions in which the presence or absence of music changes the reading. Bolivar, V. J., Cohen, A. J., & Fentress, J. C. (1994). "Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action." *Psychomusicology, 13*, 28-59.

(62) A foundational framework for the perceptual organization of sound (auditory scene analysis): it explains why, in mixtures, some components mask others and change what is "detectable" in the voice. Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

"hanging" before the group. Its power comes from its ambiguity: it does not need to declare itself as punishment; it is enough that it functions as a warning. In terms of effect, it tends to reduce future voice, increase self-censorship, and activate hypervigilance.

3) Silence as complicity. This is the silence that confirms a harmful norm: when a dehumanizing comment or a disauthorization occurs and no one marks it, the group's silence may operate as tacit ratification. It does not require a conscious intention to harm; it may arise from fear, conformity, or calculation. Its effect is cultural: it produces social learning about what is tolerable.

4) Silence as reputational management (institutional). This is administered silence: unexplained delays, lack of updates, "this will be looked into," diversion into opaque channels, promises of "internal" treatment without traceability. In organizational settings, this kind of silence is not only interpersonal: it can be a strategy for avoiding records, diffusing responsibility, or cushioning image crises. Its effect is double: on the one hand, it generates uncertainty; on the other, it opens space for rumor and conversational closure.

These functions are not essences: they depend on context and pattern. The same silence may be containment in a safe relationship and punishment in an asymmetrical one. The analytical key lies in the observable fruit: does the silence enable expression or inhibit it? Does it organize care or administer fear? Does it increase clarity or increase fog?

With these distinctions—levels (semantics/pragmatics/prosody), temporality as architecture, micro-signals versus meta-signals, subtext as field, and silence[63] as act—the ground is prepared for the next step: to examine how certain prosodic configurations make it possible to detect hidden coherences (sarcasm, irony, exaggerated neutrality, double bind) without turning analysis into an exercise in divination.

(63) Ephratt (2008) proposes a typology of the pragmatic functions of silence and shows that it can operate as communicative action (affiliation, rejection, control, etc.), aligning with containment, punishment, complicity, and reputational management. Ephratt, M. (2008). "The functions of silence." *Journal of Pragmatics, 40*(11), 1909-1938.

6. Prosody as a Pathway for Detecting Hidden Coherences

The previous chapter distinguished levels (semantics, pragmatics, and prosody) in order to avoid a common mistake: treating meaning as though it resided solely in literal content. With that distinction in hand, this chapter turns to a specific phenomenon: hidden coherences. This is not about "reading minds" through tone, but about identifying alignments between the temporal form and the social function of speech that, by design or by habit, remain deniable if one attends only to the text.

Here, "hidden coherence" will refer to the stable convergence between a pragmatic intention (for example, ridiculing, controlling, expelling, deactivating, disciplining) and a prosodic pattern (rhythm, pauses, stress, energy, melodic contours, silences) that carries out that intention while the verbal content maintains a neutral, cordial, or technically correct surface. In other words, what is hidden is not necessarily "the emotional" as such, but the direction of the communicative act when literalness offers plausible deniability(64).

The aim of this chapter is twofold. First, to describe prototypical forms in which prosody tends to function as the preferred channel of subtext (sarcasm, irony, exaggerated neutrality, double bind). Second, to delimit the role of the listener: prosody does not "prove" anything on its own; it activates resonances and biases that must be handled methodologically if one wishes to move from intuition to responsible observation.

6.1. General principle: content can lie; temporal patterning filters intention

In human interaction, propositional content (what is literally asserted) is only one layer. Institutions—and people as well—can "clean up" the text: choose correct words, polite formulas, technical language, euphemisms. Speech, however, is not mere diction; it is action unfolding in time. The communicative act occurs as a sequence with rhythm, cuts, intensifications, delays, and silences that directly affect the interlocutor's experience.

Hence the operative principle of this chapter: when content and function diverge, temporal patterning tends to filter pragmatic intention, because it is less easy to control without cost (attentional, emotional, or relational) and because it operates through relatively automatic bodily registers. This does not mean that prosody is infallible, nor that it always betrays a conscious purpose. It means that, in contexts where the text is hyper-regulated (protocols, impeccable emails, institutional formality), prosody becomes a place where the system—human or institutional—tends to display its real "direction" with less possibility of editing.

There are at least four reasons why temporal patterning is relevant:

1.- Prosody modulates relational permission. The same sentence can open space or close

(64) Pinker, Nowak, and Lee (2008) analyze the logic of indirect speech and how it enables forms of plausible deniability and relational negotiation without explicitly committing the literal content. Pinker, S., Nowak, M. A., & Lee, J. J. (2008). "The logic of indirect speech." *Proceedings of the National Academy of Sciences of the United States of America, 105*(3), 833-838. https://doi.org/10.1073/pnas.0707192105

it, enable a reply or block it, recognize dignity or reduce the other person to a mere procedural object.

2.- Prosody marks hierarchy and control. Whoever interrupts(65), whoever speeds up to dominate, whoever uses punitive pauses or selective silences is not merely "speaking": they are distributing power.

3.- Prosody operates as a signal of internal alignment. When intention and form are aligned, a unity of the act is felt (coherence). When they are misaligned—kind wording, hostile tone—a dissonance appears that the listener registers as threat or falsity.

4.- Prosody creates effects even without insult. Relational harm can be produced without explicit words: a repeated pattern that humiliates, infantilizes, invalidates, or isolates is enough.

This chapter does not propose a metaphysics of tone, but a practical criterion: prosody is a reading hypothesis that is validated (or corrected) through context, repetition of patterns, comparison with baselines, and analysis of social effect. Within that framework, sarcasm(66), irony(67), exaggerated neutrality(68), and the double bind can be understood as forms of hidden coherence: mechanisms in which the real act "travels" more through form than through content.

6.2. Sarcasm: saying A with the prosodic structure of not-A

Sarcasm is a particularly instructive case because it shows, with great clarity, the fracture between literalness and act. The general structure is simple: A is said, but prosody performs "not-A" (contempt, mockery, disbelief, disauthorization). The text preserves an acceptable appearance; the form introduces the pragmatic charge.

Functionally, sarcasm allows one to attack without fully owning the attack. Literalness provides an alibi ("I didn't say anything"), while prosody communicates the real judgment. For that reason, sarcasm is a frequent vehicle of aggression with plausible deniability: the violence remains in a channel that is less traceable if analysis is reduced to transcription.

Typical prosodic signals of sarcasm (not exhaustive) may include:

· **Shifted emphasis:** stressing an "innocent" word as though it were the ironic nucleus (for example, emphasizing perfect or sure in order to invert its meaning).

· **Lengthenings:** prolonging vowels or syllables in a way that introduces evaluative distance ("suuuure…"), often associated with disauthorization.

· **Strategic pauses:** micro-breaks before or after the key term, as though one were letting the

(65) Zimmerman and West (1975) describe how interruptions and silences in conversation function as mechanisms of dominance and turn control, providing a classic framework for reading the distribution of power in interaction. Zimmerman, D. H., & West, C. (1975). "Sex roles, interruptions and silences in conversation." In B. Thorne & N. Henley (Eds.), *Language and Sex: Difference and Dominance* (pp. 105-129). Rowley, MA: Newbury House.

(66) Cheang and Pell (2008) identify acoustic cues associated with the production of sarcasm and show that sarcastic "tone" relies on measurable prosodic patterns (e.g., changes in F0, intensity, and timing). Cheang, H. S., & Pell, M. D. (2008). "The sound of sarcasm." *Speech Communication, 50*(5), 366-381. https://doi.org/10.1016/j.specom.2007.11.003

(67) Clark and Gerrig (1984) propose the pretense theory of irony to explain irony as an act in which the speaker performs a voice/position so that the listener can recover the real attitude, highlighting the role of shared context. Clark, H. H., & Gerrig, R. J. (1984). "On the pretense theory of irony." *Journal of Experimental Psychology: General, 113*(1), 121-126. https://doi.org/10.1037/0096-3445.113.1.121

(68) Hochschild (1983) develops the concept of emotional labor and the ways institutions regulate affective expression, providing a framework for understanding "professional neutrality" as a form of emotional management with social effects. Hochschild, A. R. (1983). *The Managed Heart: Commercialization of Human Feeling*. Berkeley: University of California Press.

blow fall without naming it.

· **Marked melodic contours:** exaggerated rises or falls (or, in some styles, a deliberate flattening) that communicate disbelief or contempt.

· **Minimal laughter or breath:** not always a full laugh; sometimes a paralinguistic micro-indicator is enough to degrade the interlocutor without breaking formality.

It is worth stressing that these signals are not isolated "proofs": they become relevant when they appear as a pattern, when they are consistent with context, and when they produce repeated effects (closure of participation, confusion, shame, inhibition of reply).

The social function of sarcasm. Sarcasm can operate as discipline: it sanctions an intervention without arguing against it. It can also build alliance: if third parties laugh or validate the tone, an in-group that "gets" the mockery is consolidated, while an out-group is exposed. In hierarchical contexts, sarcasm is especially effective because it requires no explicit statement: the subordinate perceives the attack, but bearing the burden of complaint places them at a disadvantage ("that's your interpretation," "don't be dramatic").

From the perspective of hidden coherences, sarcasm is coherence because it aligns intention (to degrade) with form (mockery), while masking that alignment in the text.

6.3. Irony: distance between literalness and cultural code (belonging/exclusion)

Irony shares with sarcasm the distance between the literal and the intended, but it usually operates according to a different social logic. Whereas sarcasm tends to be more directly aggressive, irony can function as a marker of perspective: saying something in order to mean something else, while relying on a shared code. Irony does not always humiliate; it can be gentle criticism, complicity, or aesthetic elaboration. Yet in contexts of power, irony can become a mechanism of exclusion.

Two dimensions are key here:

1.- Irony as belonging ("those who get it, get it"). Prosody can signal little winks: a slight inflection, an audible smile, a micro-shift in rhythm. What matters is not the gesture itself, but that it activates mutual recognition within the group.

2.- Irony as exclusion. When the code is not shared by everyone (or is deliberately used to leave someone out), irony becomes a status filter. The person who does not "get it" is marked as an outsider, naive, or incompetent.

In irony, prosody often functions as a marker of distance: it indicates that the speaker does not fully identify with the literalness, that there is an active metacommunicative layer. That distance may be warm (shared humor) or cold (intellectualized contempt). In institutional settings, cold irony makes it possible to exercise criticism or disauthorization without entering into open conflict: it

installs the gesture of "this is obvious" or "this is not worth discussing" without having to state it.

From an analytical point of view, irony requires additional caution: it is heavily dependent on culture, community, and context. For that reason, when coding irony as a possible hidden coherence, the criterion should not be "it sounds ironic," but rather: what relational effect does it produce, and who is kept inside or outside the code?

6.4. Exaggerated neutrality: cold violence

Professional neutrality exists and may be desirable in many domains (healthcare, justice, mediation, teaching, in certain moments). Not all affective restraint is hostility. The problem arises when neutrality ceases to be a resource of fairness and becomes an aesthetic of power: a way of dehumanizing, deactivating, or disciplining under impeccable appearances.

Exaggerated neutrality is not merely "speaking without emotion," but withdrawing minimal signals of recognition while maintaining (or even increasing) control. It is characterized by an asymmetry: the speaker preserves their position of authority, but reduces the interlocutor to an object of procedure, preventing them from obtaining basic validation (being heard, being understood, being taken into account).

Frequent indicators of exaggerated neutrality may include:

· **Sustained prosodic flattening:** minimal variation in pitch and stress, with the sensation of a communicative "wall."

· **Excessively chopped or mechanical rhythm:** phrases that sound like templates; pauses that are not for listening but for closure.

· **Absence of feedback ("backchannels"[(69)]):** lack of "mm-hm," "I understand," or minimal follow-up signals; the other person is left speaking into a vacuum.

· **Formality as a weapon:** technicalization or legalistic phrasing in moments where the human dimension would require at least some gesture of care (not sentimentalism, but recognition).

· **Closure without repair:** ending the exchange with correct formulas that, in practice, leave the interlocutor with no way forward ("it has been recorded," "the procedure will continue," "it will be looked into," without real avenues).

The subjective effect is often paradoxical: no one has insulted, no one has shouted, and yet the interlocutor leaves with a feeling of humiliation, invalidation, or threat. This happens because the harm does not travel through content, but through the mode of presence: the way the institution (or its representative) affectively withdraws while still demanding compliance.

In terms of hidden coherence, exaggerated neutrality can enact an intention of control or expulsion through a prosody that does not "sign" the violence. The text protects reputation; the form carries out the act.

(69) Yngve (1970) introduces the concept of the backchannel as a "secondary" channel of listener feedback (minimal responses, acknowledgments, listening signals), which is key for analyzing the presence or absence of interactive confirmation. Yngve, V. H. (1970). "On getting a word in edgewise." *In Papers from the Sixth Regional Meeting of the Chicago Linguistic Society* (pp. 567-578). Chicago: Chicago Linguistic Society.

6.5. Double bind: "I care for you" with a prosody of control

The double bind[70] describes situations in which contradictory messages are issued at different levels, so that any response from the receiver is penalized. Its force does not lie in an isolated sentence, but in a repeated structure: care is offered, but threat is communicated; openness is promised, but disagreement is sanctioned; one is invited to speak, but speech is punished.

A typical form, at the level of discourse, is: "I care for you" or "I'm doing this for your own good." At the prosodic level, however, a pattern of control may appear: coldness, condescension, disciplinary emphasis, pauses that mark superiority, or a rhythm that imposes closure. The result is a two-layered message:

- **Explicit content:** protection, concern, rationality, "good manners."
- **Implicit act:** surveillance, domination, disauthorization, reduction of agency.

In hierarchical contexts, the double bind is especially effective when it is combined with affective asymmetry: warmth toward some, ice toward others, or alternation between "affection" and sanction. That variability is not accidental: it serves to maintain uncertainty, increase dependency, and make complaint more difficult ("sometimes they are kind; maybe the problem is me"). The receiving person is trapped between interpreting the content and responding to the act.

From a prosodic perspective, the double bind often manifests through the coexistence of incompatible signals: words of care delivered with a prosody of closure; inclusive formulas accompanied by punitive silences; "open" questions with an intonation that does not actually admit an answer; invitations to dialogue combined with systematic interruptions. The key analytic issue is not the occasional "strange tone," but the structural impossibility of a safe response.

As a hidden coherence, the double bind produces a specific effect: it confuses attribution. The recipient cannot easily name what is happening because the text denies the act. That very unsayability is part of the mechanism.

6.6. Coherence as perceived by the listener: resonance and biases (how not to fool ourselves)

Speaking of hidden coherences requires taking the role of the listener seriously. Prosody has an impact because it is not processed only as linguistic information: it activates bodily memory, social learning, and affective associations. This has analytical value (prosody "resonates" with prior experiences), but also a risk: the listener may project patterns onto ambiguous signals.

Three processes should be distinguished here:

1.- Resonance[71]*:* the form of speech evokes prior experiences and facilitates an immediate reading of the act ("this sounds like mockery," "this sounds like threat").

(70) Bateson, Jackson, Haley, and Weakland (1956) formulate the double bind framework as a communicative structure of incompatible messages in which any response is sanctioned, which is useful for describing coherences in which the content declares care while the form enacts control. Bateson, G., Jackson, D. D., Haley, J., & Weakland, J. H. (1956). "Toward a theory of schizophrenia." *Behavioral Science, 1*(4), 251-264. https://doi.org/10.1002/bs.3830010402

(71) Juslin and Laukka (2003) review evidence on how emotion is communicated vocally through acoustic patterns and how listeners recover states/attitudes from voice signals, providing a basis for affective "resonance." Juslin, P. N., & Laukka, P. (2003). "Communication of emotions in vocal expression and music performance: Different channels, same code?" *Psychological Bulletin, 129*(5), 770-814. https://doi.org/10.1037/0033-2909.129.5.770

2.- Inferential filling-in: faced with incomplete information, the mind fills in meaning using its history, culture, and expectations.

3.- Attribution bias[72]*:* there is a tendency to interpret the other person's behavior as a trait ("they are hostile") rather than as a state or context (fatigue, pain, professional training, situational pressure), or to confirm prior hypotheses ("I knew they were going to do that").

In an essay that seeks to build a bridge toward measurement, these dynamics cannot be treated as irrelevant noise: they are part of the phenomenon. The question is how to manage them so that prosodic reading does not become an oracle.

A sober approach can rest on five methodological rules[73]:

• **Rule 1:** Separate observation from interpretation. First describe observable traits (pause, interruption, acceleration, drop in energy, absence of response); only then propose a pragmatic hypothesis. Do not reverse the order.

• **Rule 2:** Privilege patterns over events. An isolated signal is ambiguous; a repeated pattern under comparable conditions (same context, different moments) has greater force.

• **Rule 3:** Compare against a baseline. Every speaker and every culture has styles. The question is rarely "does it sound cold?" but rather "does it sound significantly colder here than in that speaker's own norm, and how is that distributed toward whom?"

• **Rule 4:** Triangulate with effect and context. If a supposedly neutral prosody systematically produces closure of participation, fear of speaking, self-censorship, and asymmetry of turns, the hypothesis of covert hostility gains plausibility. If there is no such effect, and context offers an alternative explanation, the hypothesis should weaken.

• **Rule 5:** Maintain epistemic humility. Prosody is not a verdict. It is a signal that orients questions: What structure is being produced here? What margin for reply exists? What is rewarded or punished? Who is allowed to speak, and who is not?

There is one decisive point in this final rule: the objective is not to "label" people, but to describe dynamics. This shift in focus—from "what kind of person is this?" to "what kind of interaction is being produced here?"—reduces the risk of interpretive injustice and opens a more suitable terrain for analysis, discussion, and eventually measurement.

With these cautions in place, hidden coherences cease to be a field of suspicions and become a field of observation: a zone where prosody helps read what the text does not sign, but the time of speech carries out.

(72) Ross (1977) describes systematic attribution biases (e.g., over-attributing internal traits relative to context), offering a framework for understanding how the listener may turn a situated prosodic signal into a "trait" of the speaker. Ross, L. (1977). "The intuitive psychologist and his shortcomings: Distortions in the attribution process." In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 173-220). New York: Academic Press. https://doi.org/10.1016/S0065-2601(08)60357-3

(73) Bakeman and Gottman (1997) present principles and techniques for observing interaction and analyzing sequences, prioritizing repeated patterns, comparison, and explicit rules, which is useful as a methodological basis for a non-"oracular" prosodic reading. Bakeman, R., & Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis* (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511527685

7. Prosody, Power, and Organization

Up to this point, prosody has been presented as a way of reading "hidden coherences" in interaction: tensions between literalness and act, between proper form and hostile effect. This chapter shifts the focus from the isolated episode to the organization as a system. The question is no longer only what a person does with their tone, but what an institution does with the tones it tolerates, rewards, or punishes.

In hierarchical contexts[74], prosody does not operate solely as individual expression, but as a component of social regulation: it manages turns, distributes credibility, defines what is sayable and what ways of sounding are acceptable, and helps produce stable climates (of trust or fear). For that reason, when one speaks of covert hostility or symbolic violence in institutions, prosody functions both as symptom and as mechanism: it reveals asymmetries and, at the same time, can reproduce them.

7.1. Harassment as a systemic phenomenon (not merely dyadic)

To understand why prosody matters in contexts of power, it is helpful to view persistent hostility not as an isolated episode between two people, but as a pattern that becomes stabilized and learned within a system. What is decisive, for this essay, is that this pattern does not travel primarily in "what is said" (as text), but in the temporal enactment of treatment: turns, pauses, interruptions, selective coldness, closing cadences, and implicit permissions to exist in the conversation.

From this perspective, harassment can be understood as a systemic dynamic[75] with at least three layers:

1.- Micro layer (interaction): concrete episodes in which micro-signals appear (tones of disauthorization, punitive silences, interruptions, selective coldness, "condescending correctness," etc.). Here prosody acts as a channel of the act: it turns into a social gesture what, in text, might seem neutral.

2.- Meso layer (recurrent scene): meetings, hallways, classrooms, hearings, video calls. Repetition is key: it is not an isolated gesture, but a pattern that consolidates because it recurs in the same bodies and the same roles (who interrupts, who gets cut off, who is granted pause, who receives curt replies).

3.- Macro layer (climate and culture): an organization "learns" what is permitted and what is not. It learns who can speak forcefully without cost and who is labeled "problematic" for showing the slightest tension. It learns which tone is considered professional and which is considered "excess emotionality," even when the content is identical. That "informal pedagogy" ends up creating climate.

Within this framework, prosody and the temporal management of turn-taking are a mechanism of

(74) Pfeffer (1992) develops how power and organizational influence structure which behaviors become legitimate and which dynamics become normalized in the everyday life of institutions. Pfeffer, J. (1992). *Managing with Power: Politics and Influence in Organizations*. Boston, MA: Harvard Business School Press.

(75) Salin (2003) reviews how workplace bullying can be sustained by structures and processes in the environment (conditions that enable, motivate, or precipitate it), supporting a reading that goes beyond the episode between two people. Salin, D. (2003). "Ways of explaining workplace bullying: A review of enabling, motivating and precipitating structures and processes in the work environment." *Human Relations, 56*(10), 1213-1232.

high efficacy and high deniability[76]: they can reduce agency without signing it in the content. This is why they offer a privileged way of observing how permission to speak is distributed, how the chill settles in, and which voices are left without air.

7.2. Voice as a tool of hierarchy: turns, permissions, interruption

Hierarchy is not expressed only in organizational charts. It is expressed in the management of turn-taking and in the management of tone. Everyday conversation contains implicit rules of alternation (who speaks, when one yields, when one resumes). In institutions, those rules are rewritten in the key of power[77]: certain people acquire tacit permission[78] to cut in, accelerate, close topics, or demand answers; others, by contrast, operate under prosodic constraints[79] (to speak "carefully," not to sound "irritated," not to "make the atmosphere tense").

There are at least four especially frequent mechanisms in which prosody and hierarchy are coupled:

a) Control of the opening and closing of the turn. Some expressions are semantically neutral but prosodically decisive: "okay," "right," "perfect," "as I was saying...," "let's leave it here." With a closing prosodic contour (falling tone, exit speed, clipping of syllables), these formulas function as a seal of authority: they do not describe, they close. In strong hierarchies, closure becomes a tool: it defines which topics "do not deserve" air.

b) Interruption and overlap as capture of the floor. To interrupt is not merely to talk over someone; it is to reconfigure the right to exist in the conversation. Interruption may be explicit, but it may also be prosodic: entering before the closure point, raising energy and speed, or using a tone that forces the other person to yield (because it imposes urgency or superiority). When this is distributed asymmetrically (always toward the same people, always from the same people), it ceases to be style and becomes structure.

c) Pauses and latency as sanction. The time between turns communicates. A silence after someone's intervention may function as a space for reflection... or as punishment. In hierarchical contexts, "latency" (taking time to respond, replying with dry monosyllables, cutting continuity) operates as a micro-sanction: it discourages that person from exposing themselves again. Silence here is not absence; it is action.

d) Institutional register as a filter of humanity. In organizations, the "technical" register is not in itself problematic: it may be necessary for clarity. The problem emerges when it is used as a weapon: an excessively aseptic tone in the face of suffering, a rigid courtesy that invalidates, a "correct" voice that produces humiliation. Dehumanization does not require insult: it can be produced through the contrast between what is being lived and the coldness of the response.

(76) Andersson and Pearson (1999) conceptualize incivility as low-intensity deviant behavior with ambiguous intent, explaining why certain forms of hostility can operate with room for denial and at the same time escalate spirally within the work environment. Andersson, L. M., & Pearson, C. M. (1999). "Tit for tat? The spiraling effect of incivility in the workplace." *Academy of Management Review, 24*(3), 452-471.

(77) Keltner, Gruenfeld, and Anderson (2003) propose the approach-inhibition theory of power, describing how power tends to disinhibit and orient behavior, whereas lack of power increases inhibition and social vigilance, with implications for interactional patterns. Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). "Power, approach, and inhibition." *Psychological Review, 110*(2), 265-284.

(78) Detert and Burris (2007) analyze employee voice and how certain leadership styles facilitate or block people from raising problems or disagreements, providing empirical grounding for the issue of real permission to speak. Detert, J. R., & Burris, E. R. (2007). "Leadership behavior and employee voice: Is the door really open?" *Academy of Management Journal, 50*(4), 869-884.

(79) Van Dyne, Ang, and Botero (2003) conceptualize silence and voice as multidimensional constructs, distinguishing different motives and forms of remaining silent or speaking up in organizations. Van Dyne, L., Ang, S., & Botero, I. C. (2003). "Conceptualizing employee silence and employee voice as multidimensional constructs." *Journal of Management Studies, 40*(6), 1359-1392.

Taken together, these mechanisms show a practical thesis: power is not only spoken; it is timed. Hierarchy administers conversational time (who has space, who loses it) and does so, to a great extent, through prosodic resources that are socially effective and difficult to challenge in formal terms.

7.3. Prosody as social discipline: who is allowed to sound irritated and who is not

Institutions do not only regulate the content of what is expressed; they regulate the manner(80) of expressing it. This regulation operates as social discipline: an implicit pedagogy that indicates which tones are acceptable for each role.

Within this logic, prosody becomes a marker of legitimacy. To "sound" a certain way enables or disables credibility. In practice, this produces at least three structural effects:

a) Tone policing. Attention shifts from content to form: what matters is not so much what is being denounced, but how it is being denounced. "Tone" becomes an excuse for disregarding substance. The consequence is not aesthetic, but political: when the focus is systematically placed on the prosody of the person pointing to a problem, the problem goes without genuine engagement.

b) Unequal distribution of affective permission. In many organizations, certain positions may display irritation(81), irony, or forcefulness without cost; by contrast, other positions must always sound "pleasant," "light," or "rational." This asymmetry is structural: it defines who may introduce tension into the field without being sanctioned and who pays the price for it, even when they are right. Here prosody functions as a boundary: it delimits who may exercise agency.

c) Normalization of "professional tone" as a mask. The ideal of professionalism may operate as a legitimate criterion (clarity, respect, order), but also as a device of concealment: an aesthetic of correctness that allows symbolic violence to appear under the guise of neutrality. In that case, the organization rewards not truth or care, but the ability to sound impeccable while harm is being done.

Prosodic discipline is not always explicit. It is learned through consequences: who receives a response, who is ridiculed "gently," who is frozen by silence, who is labeled "conflictive." Over time, that learning(82) produces self-censorship: not only are words chosen; energy is reduced, nuance is clipped, and anything that might sound like dissent is avoided.

7.4. When the organization learns: from episode to cultural contagion

One of the most decisive properties of toxic climates is not the existence of isolated episodes, but

(80) Rafaeli and Sutton (1987) describe emotional expression as part of the work role and how organizations establish normative expectations regarding which emotion (and which form of expressing it) is considered appropriate in performance. Rafaeli, A., & Sutton, R. I. (1987). "Expression of emotion as part of the work role." *Academy of Management Review, 12*(1), 23-37.

(81) Brescoll and Uhlmann (2008) show that the expression of anger has different consequences depending on gender and status, providing evidence of differential costs when someone expresses tension or irritation in workplace contexts. Brescoll, V. L., & Uhlmann, E. L. (2008). "Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace." *Psychological Science, 19*(3), 268-275.

(82) Argyris and Schön (1978) develop the framework of organizational learning (including the logic of learning from errors and routines), which is useful for describing how a system incorporates patterns and turns them into practical norms. Argyris, C., & Schön, D. A. (1978). *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison-Wesley.

their capacity to become a "school": the institution incorporates the pattern and stabilizes it.

In that process, prosody plays a central role because it transmits norms without making them explicit. There is no need to declare, "here, one does not question": it is enough for whoever questions to receive, systematically, a discouraging prosodic treatment (ice, cut-offs, technicalization, minimal laughter, punitive latency).

That "organizational learning" usually manifests in five linked dynamics:

1.- Communicative hypervigilance: people begin to measure every word and, above all, every tone. They do not speak from clarity; they speak from avoidance. The goal ceases to be understanding; it becomes not exposing oneself.

2.- Alignment with power: the patterns of the dominant role are imitated, not always out of adherence, but out of survival. The same coldness is adopted, the same way of closing turns, the same style of mild mockery or "technical" correction. In this way, prosody becomes contagious.

3.- Rewriting of the narrative[83]: the institution learns to narrate the conflict in an acceptable key. What had been harassment becomes a "misunderstanding," a "communication problem," "sensitivity," or "perception." Prosody acts here as a moral background: the official account usually arrives in a neutral, stable tone, with calm authority, making it difficult to contradict without seeming "exaggerated."

4.- Organizational silence[84]: people stop speaking, or speak only in safe and informal channels. Public conversation becomes impoverished: there is less visible disagreement, but not because there is more coherence; rather, because there is more fear. It is a kind of "peace" that does not reduce conflict: it hides it.

5.- Secondary revictimization: when someone tries to seek repair, they encounter a second layer of harm: a cold, legalistic, or bureaucratic[85] response, in a register that reduces lived experience to mere procedure. The effect is double: the harm is denied, and the experience is disauthorized through "correct" form.

Taken together, these dynamics show why prosody is a particularly revealing thermometer of climate: not because it reveals the "inner truth" of each individual, but because it registers how the system distributes warmth and cold. When affective asymmetry becomes stable (warmth toward some, ice toward others) and is reproduced in repeated scenes, prosody ceases to be a personal trait: it becomes culture.

By the end of this chapter, prosody appears no longer only as an interpretive signal, but as an organizational vector: it participates in the production of moral and relational order (or disorder). This prepares the ground for the next step of the essay: to formulate, from a structural framework, how coherence (or incoherence) between discourse, procedure, and treatment is to be understood,

(83) Weick (1995) formulates the sensemaking approach to explain how organizations construct shared narratives and meanings about "what happened," providing a basis for understanding the institutional re-narration of conflicts. Weick, K. E. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA: Sage.

(84) Morrison and Milliken (2000) propose the concept of organizational silence as a collective phenomenon sustained by beliefs and practices that discourage speaking, thereby limiting learning and change. Morrison, E. W., & Milliken, F. J. (2000). "Organizational silence: A barrier to change and development in a pluralistic world." *Academy of Management Review, 25*(4), 706-725.

(85) Bies and Moag (1986) introduce interactional justice as a criterion of fairness tied to interpersonal and communicative treatment, relevant for describing additional harm when the institutional response becomes cold and formalistic in situations requiring repair. Bies, R. J., & Moag, J. S. (1986). "Interactional justice: Communication criteria of fairness." In R. J. Lewicki, B. H. Sheppard, & M. H. Bazerman (Eds.), *Research on Negotiation in Organizations* (Vol. 1, pp. 43-55). Greenwich, CT: JAI Press.

and why that coherence can be evaluated by its effects.

8. Bridge to Universal Coherence

So far, we have seen that prosody is temporal architecture: it turns content into act and organizes experience (the opening of voice or its closure). In hierarchical settings, that architecture allows literalness to remain "correct" while treatment becomes corrosive. This chapter introduces a bridge to Universal Coherence (UC) in order to name coherences and incoherences across layers(86) (intention-means-effect; discourse-treatment-fruit) and to sustain a practical criterion: what matters is not only what is declared, but what is produced.

8.1. Prosody as "form" that reveals coherence across layers

A cross-cutting thesis of this essay is that an interaction is not defined solely by its semantic content, but by the set of layers that converge in the communicative act. For analytical purposes, it is useful to distinguish at least four planes:

1.- Literal content (what is said).

2.- Pragmatic act (what it is said for here, to this person, within this frame of power).

3.- Prosodic form (how what is said is temporalized and embodied: rhythm, pauses, emphasis, energy, timbre).

4.- Effect (what it produces in the field: clarity or confusion, agency or submission, openness or silence).

The bridge to UC appears when the question is formulated not only as "is the content true?" but as "are the layers aligned?"

Within this framework, coherence is not understood as a rhetorical virtue or as isolated internal logical consistency, but as relational alignment between what is declared, what is done, and what is produced. Put simply: communication is more coherent when content, form, and effect sustain the same ethical and functional direction, and less coherent when text and effect systematically contradict one another.

At this point, prosody operates as a privileged indicator because it belongs to the layer of form and, at the same time, participates in the act: prosody turns a sentence into a social gesture(87). Two utterances may be identical in wording and radically different in relational structure.

· "I understand your point" may be an act of recognition or an act of closure.

· "Thank you for letting me know" may open a channel of safety or signal that speaking up will be costly.

· "It's not personal" may be protection or a threat with plausible deniability.

(86) Watzlawick, Beavin, and Jackson (1967) distinguish the content level and the relational level of communication, showing how the relational component can redefine the meaning of a message beyond its literal wording. Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1967). *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies, and Paradoxes.* New York, NY: W. W. Norton & Company.

(87) Wennerstrom (2001) develops a framework of prosody and discourse analysis in which temporal patterns (prominence, rhythm, intonation) guide listener inferences and connect sound form with interactive function. Wennerstrom, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis.* Oxford, UK: Oxford University Press.

In UC terms, prosody often reveals where the system's real coherence lies when content has been "cleaned up" by institutional norms. It may reveal:

A) Positive coherence, when intention, care, form, and effect are coupled: the listener perceives legibility, grounded calm, and capacity for action.

B) Dark coherence, when the system is aligned with control or disauthorization: the content remains neutral, but the form organizes inequality (turns, punitive silences, selective coldness), and the effect is a reduction of voice.

This point matters because it avoids a common simplification: the alternative is not "true prosody" versus "lying text." What often happens in complex environments is something else: the system may be coherent, but with a purpose different from the one it declares. Prosody helps locate that purpose.

8.2. The container error applied to institutional language

The UC framework introduces a useful distinction for institutional analysis: source and container.

· By *source* is meant that which justifies and guides a practice: care, truth, repair, justice, learning, protection of the vulnerable, integrity of process.

· By *container* is meant the visible form that wraps and conveys that source: protocols, procedures, formal language, communication registers, an aesthetic of neutrality, templates, "professional tone."

The container error appears when the container[(88)] substitutes for the source: when the organization confuses formal compliance with ethical compliance, or the aesthetic of correctness with actual care. At the level of language, this becomes especially delicate because the institution can perfect the text until it is impeccable while the exchange remains hostile in its dynamics.

Applied to prosody, the container error takes recurrent forms:

1. Performative neutrality as shielding: the "technical" or "aseptic" tone is used to keep the human dimension from entering the field.

2. Formality as weapon: the correct register becomes a surface that allows violence to be denied ("I haven't said anything offensive") while disauthorization is carried out through prosodic means and turn management.

3. Procedure without repair: the institutional channel focuses on "processing" rather than clarifying, protecting, and correcting; communication adopts a prosody of administered wear and tear (prolonged silences, minimal responses, closure through fatigue).

The thesis is not anti-protocol. The container is necessary: without form, the source disperses and the organization becomes arbitrary. The problem appears when the container is absolutized and begins to function as a moral substitute: "we have been correct" is used as equivalent to "we have exercised care."

(88) Weber (1922/1978) describes modern bureaucracy as an order of regulated procedures and impersonality, in which administrative form can acquire primacy over the concrete relation to persons. Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology* (G. Roth & C. Wittich, Eds.). University of California Press. (Original work published 1922)

From this perspective, prosody matters because the institution speaks through its containers. The same procedure may sound like an opening or like a threat[89] depending on how its temporality is articulated:

· *Is there real space to respond, or only a formal turn?*

· *Is the rhythm maintained so the account can be completed, or is it accelerated to cut it off?*

· *Are silences used as containment or as punishment?*

· *Is time given to disagreement, or is it cooled down through technification?*

When the container becomes a substitute, a paradox appears: the organization may display external coherence (documents, statements, "values") while its internal layers become misaligned (treatment, turns, decisions, protection). In that scenario, prosodic analysis does not "guess" intentions; it points to the way the container is being used: as a channel of care or as a device of control.

8.3. Fruit criterion: what it produces in the climate (clarity vs. fear, stability vs. cynicism)

The bridge to UC contributes an evaluative criterion that fits the essay's methodological caution: communication is not evaluated by what it declares, but by what it produces in a sustained way[90].

This criterion—called here the fruit criterion—is especially useful for preventing prosody from becoming an "emotional polygraph." Instead of asserting "this person had bad intentions," the analysis shifts toward a more replicable question:

What fruit does this communicative pattern leave over time?

In interaction and organizational climate, the relevant fruits are usually observable:

· **Clarity:** legibility around what is being addressed increases; fog decreases; people can understand what is expected and why.

· **Psychological safety:** people can disagree without paying a disproportionate cost; self-censorship decreases.

· **Stability without rigidity:** the system sustains limits and norms, but can repair without humiliating.

· **Reduced internal friction:** the relational noise that forces people to justify themselves, protect themselves, or anticipate attacks goes down; focus on task and learning goes up.

· **Shareable sensemaking[91]:** the account of what happened can circulate without being

(89) Heritage and Clayman (2010) show how institutional interaction is organized through asymmetries, formats, and turn design, and how those forms distribute access, reply, and authority in regulated settings. Heritage, J., & Clayman, S. (2010). *Talk in Action: Interactions, Identities, and Institutions*. Wiley-Blackwell.

(90) Edmondson (1999) defines psychological safety and links it to the real willingness to speak, disagree, and report problems in teams. Edmondson, A. (1999). "Psychological safety and learning behavior in work teams." *Administrative Science Quarterly, 44*(2), 350-383.

(91) Weick (1995) formulates organizational sensemaking as the shared construction of meaning around "what happened" and explains how organizations turn experience into common narratives. Weick, K. E. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA: Sage.

punished; reputational closure does not become the dominant priority.

The fruit criterion also makes it possible to distinguish between styles and hostility. A dry prosody may be a professional trait, local culture, fatigue, or neurodiversity; it becomes a risk indicator only when, combined with turn distribution, silences, and decisions, it systematically produces fear, asymmetry, and closure.

In other words: *fruit helps distinguish occasional dissonances from stable incoherences. And it prepares the ground for the next chapter, where patterns are translated into observable and comparable indicators without collapsing human complexity.*

8.4. Eloquence as convergence: when layers align and it "becomes audible"

A central contribution of the UC framework is to recover the notion of eloquence not as "the ability to speak beautifully," but as a structural quality[92]: what happens when several layers converge in such a way that coherence becomes perceptible and shareable.

In the context of this essay, eloquence can be understood as the condition in which:

· Literal content is clear and does not hide behind euphemisms.

· The pragmatic act is congruent with what is declared (if listening is promised, listening occurs; if investigation is promised, investigation occurs).

· Prosody sustains the direction of the act (it neither theatricalizes care nor enacts contempt under the cover of formality).

· The effect on the listener and the group is an opening of agency: people can speak, correct, and repair.

In more operational terms: when eloquence is present, the exchange leaves less ambiguous residue. There is no need to read between the lines to know whether one is being invited to participate or merely tolerated. Coherence is "heard" because the temporal form of speech does not contradict the declared purpose.

The idea matters by contrast: in contexts of covert hostility, an inverted eloquence often appears. The text is formal, but the system becomes eloquent in another direction: it makes audible, to those who suffer it, that dissent will be costly. That negative eloquence does not rely on insults; it relies on the convergence of micro-signals:

· Asymmetrical turns.

· Punitive pauses.

· Ridiculing emphasis.

(92) Quintilian (2001) emphasizes the importance of actio—delivery, voice, pauses, and cadence—in rhetorical effectiveness, locating part of eloquence in the execution of speech and not only in verbal content. Quintilian. (2001). *The Orator's Education* (Institutio Oratoria) (D. A. Russell, Trans.). *Harvard University Press.* (Original work published c. 95 CE)

· Differential coldness.

· Systematic closure of the account.

The bridge to UC makes it possible to name this difference with a useful distinction: *coherence as alignment* does not automatically equal *coherence as care*. That is why the fruit criterion is indispensable: it prevents an efficient coordination of control from being mistaken for ethical coherence.

8.5. UC hypothesis: institutional hostility as stable incoherence between discourse and treatment

In light of the foregoing, a working hypothesis(93) can be formulated that connects directly with the purpose of the essay:

Applied UC hypothesis: *in environments of institutional hostility, a stable incoherence tends to appear between official discourse (values, neutrality, respect, "care") and actual treatment (prosodic and pragmatic micro-signals that produce asymmetry and silence). This incoherence is sustained over time through formal containers that permit plausible deniability and reputational management.*

The hypothesis has several analytical implications:

1. Conflict does not play out only in content: there may be "neutral" messages that, by virtue of their form and sequence, operate as punishment.

2. The institution may protect its narrative at the expense of its internal coherence: the system prefers to close narratives rather than repair relationships.

3. Prosody functions as an early sensor: before the climate becomes openly hostile, temporal patterns usually appear that indicate a restriction of agency (interruptions, accelerating closures, selective silences, differential coldness).

This formulation avoids a frequent error in social diagnosis: reducing hostility to "bad people" or to "misunderstandings." In UC reading, institutional hostility is a structural phenomenon: a configuration in which several layers (rules, hierarchy, incentives, communicative style) converge so that the cost of speaking is high(94) and the cost of abusing is low or deniable.

Put differently: *the focus is not to moralize the voice, but to describe when a system maintains stability through dissonance: it declares one thing and does another, and sustains that gap through means that avoid explicit signature. Prosody, precisely because it is temporal form, is often one of the places where that stability most continuously gives itself away.*

(93) Argyris and Schön (1978) develop a theory of action and organizational learning centered on how institutions incorporate routines, correct errors, and stabilize ways of acting that may drift away from their explicit discourse. Argyris, C., & Schön, D. A. (1978). *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison-Wesley.

(94) Morrison and Milliken (2000) conceptualize organizational silence as a collective phenomenon sustained by beliefs and practices that discourage speaking, with effects of self-censorship, closure, and blocked learning. Morrison, E. W., & Milliken, F. J. (2000). "Organizational silence: A barrier to change and development in a pluralistic world." *Academy of Management Review, 25*(4), 706-725.

8.6. Prosody and ethics: congruence between intention, means, and care

Introducing the bridge to UC makes it necessary to spell out a normative dimension that has been implicit since the prelude: this essay does not study prosody only out of descriptive curiosity, but because of its relation to dignity, agency, and interpersonal justice[95] in contexts where power can operate without declaring itself.

From this perspective, the ethics of speech is not exhausted by "not insulting." It includes at least three levels:

A) Intention: the inner direction of the act (to care, to clarify, to dominate, to protect image).

B) Means: the concrete way of carrying out that intention (words, prosody, turns, silences, timing).

C) Care: attention to the effects on the other person and on the field (whether voice is reduced, whether fear is produced, whether repair is enabled).

Ethical congruence appears when these levels align: when the means does not betray the declared intention and when care is verified by fruit (clarity, stability without violence, possibility of reply). Ethical betrayal appears when the means is used to deny the real act: when content declares respect but form organizes humiliation; when protocol declares neutrality but temporality administers wear and tear; when a "correct" tone functions as a container that absorbs the truth.

This point has a practical consequence: measuring prosody in institutional contexts requires an explicit ethics. Without it, any instrument may become a tool of control. That is why, even in an indicator-oriented approach, the criterion cannot be "detect lies" or "classify people," but rather:

· Detect interactional patterns that tend to produce silence or harm.

· Open space for correction and institutional learning.

· Protect those left in a vulnerable position.

· Uphold due process[96] (prosody as a signal, not a verdict).

With this, the bridge is established: UC provides a language for speaking about coherence across layers and a criterion of verification by fruit; prosody provides an observable channel through which that coherence becomes audible. The next chapter takes this convergence and translates it into a methodological program: how to move from "it is felt" to "it is observed," without falling into determinisms or technologies of suspicion.

(95) Goffman (1967) introduces the notion of face and the work of social image in interaction, showing how speech protects or erodes dignity, status, and recognition beyond literal content. Goffman, E. (1967). *Interaction Ritual: Essays on Face-to-Face Behavior*. Garden City, NY: Anchor Books.

(96) Bies and Moag (1986) introduce interactional justice as a criterion of fairness tied to the mode of treatment and show that the experience of justice is also at stake in the way one responds, listens, and explains. Bies, R. J., & Moag, J. S. (1986). "Interactional justice: Communication criteria of fairness." In R. J. Lewicki, B. H. Sheppard, & M. H. Bazerman (Eds.), *Research on Negotiation in Organizations* (Vol. 1, pp. 43-55). Greenwich, CT: JAI Press.

9. From “it is felt” to “it is observed”

The methodological problem arises wherever corrosiveness requires plausible deniability: when conflict is discussed solely at the level of the text, the experience of harm tends to be dismissed as “interpretation,” “oversensitivity,” or “misunderstanding.”

This chapter proposes a cautious transition: moving from “it is felt” to “it is observed” without turning the voice into an oracle or measurement into a surveillance device. The idea is not to replace contextual, ethical, or legal judgment, but to introduce an intermediate layer of observables that makes it possible to discuss patterns with less ambiguity. Instead of asserting “this person had bad intentions,” the approach is organized around more testable questions: which temporal patterns repeat themselves, how turns and silences are distributed, which asymmetries become stabilized, and what “fruit” that pattern leaves over time.

9.1. Basic problem: hostility is deniable; what is measurable must be replicable

Covert hostility is sustained by a paradox: it often produces intense effects (reduced agency, shame, self-censorship, fear) without leaving an explicit trace in literal content. Conflict thus becomes epistemologically unequal. The affected person contributes experience; the institution contributes form: minutes, emails, protocols, discursive “neutrality.” In that confrontation, harm becomes trapped in an ambivalence: it is perceived with bodily and relational clarity, yet it is discussed using tools that tend to erase precisely the bodily and the relational.

This tension is aggravated by three factors:

1.- The centrality of the written record. Institutions trust what can be archived and cited. Language that can be formally signed off on is rewarded; what happens in form (intonation, pauses, rhythm, turns) falls outside the record.

2.- The probabilistic nature of prosodic interpretation. Prosody does not offer inner certainties[98]. It signals patterns, not direct readings of intention. This makes the harm both real and disputable at the same time.

3.- The possibility of denial without contradicting the text. Plausible deniability does not require lying; it only requires remaining within what is defensible. That is why the most harmful phenomena in hierarchical environments rarely appear as direct insults: they appear as temporal sequences that restrict, cool, or humiliate without breaking formal correctness.

In this scenario, “measuring” does not mean turning the human into a number, but constructing replicable descriptions[97]: if several people observe the same episode with a shared guide, they should be able—with some margin of variation—to identify shareable elements (durations, interruptions, silences, accelerations, changes in energy, asymmetries). Replicability does not

(97) Bakeman and Gottman (1997) present principles and techniques for observing interaction and analyzing sequences, with attention to repeated patterns, comparison, and explicit rules. Bakeman, R., & Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis* (2nd ed.). Cambridge: Cambridge University Press.

(98) Ladd (2008) develops the phonology of intonation and underscores the contextual and probabilistic character of its interpretation. Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press.

eliminate interpretation; it frames it. It makes it possible to move from "I felt it" to "we can point to observable elements of the exchange," while preserving caution: an indicator is not a verdict.

9.2. Design principle: measure patterns, not "souls"

Prosodic measurement becomes dangerous when it tries to infer mental states as though they were observable objects ("they were lying," "they were being hostile," "they had bad intentions"). This essay proposes another principle: measure temporal and relational patterns, not "souls." The goal is to describe how the interaction is organized, not to assign internal psychology.

This principle translates into four design decisions:

(a) Prioritize physical and temporal variables. Prosody can be operationalized through acoustic and time-based features(99): pitch (estimated F0), energy (RMS/dB), duration, rhythm, pauses, voice/non-voice ratio, and turn distribution. These variables do not exhaust the phenomenon, but they provide a minimum basis for discussing it without relying solely on impressions.

(b) Work across multiple scales: from the micro-event to the climate. An isolated micro-signal may be style, fatigue, or context. What matters for institutional hostility is not the single detail but its stability: repeated sequences, persistent asymmetries, regularities that condition who may speak and at what cost.

(c) Measure asymmetries and distributions, not "ideal values." A common trap is to impose a universal standard of "good prosody." Here the focus is on functional differences:

· Who receives longer silences after speaking?

· Who is interrupted more often?

· Who is responded to with systematic coldness?

· Which voices are accelerated to cut them off, and which are allowed to finish?

The reference point is not "how a person should sound," but how treatment is distributed within the same environment.

(d) Triangulate by "fruit." In order to avoid turning measurement into an emotional polygraph, interpretive closure must be grounded in sustained effects: clarity, psychological safety, real possibility of reply, reduced fear, openness to repair. If a prosodic and pragmatic pattern coincides with a "fruit" of closure (silence, self-censorship, isolation), its relevance as a risk indicator increases. If it coincides with a fruit of care (clarity, agency, repair), the reading changes.

In summary: *the design does not seek to "read" the individual; it seeks to make visible what the system tolerates, rewards, or punishes through form. Prosody thus becomes a dimension of institutional ecology: not because of mysticism, but because of structure.*

(99) Kent and Read (2002) synthesize methods for acoustic analysis of speech and discuss energy/intensity measures (RMS/dB) together with practical limits related to channel, distance, and noise. Kent, R. D., & Read, C. (2002). *The Acoustic Analysis of Speech* (2nd ed.). Singular/Thomson Learning.

9.3. Minimum pipeline (PWA prototype)

Moving from the conceptual framework to a path of observation requires a minimum pipeline: a sequence of steps that transforms audio into analyzable units, and those units into comparable variables. The goal of a prototype is not to "be right," but to make the phenomenon inspectable[100]: to allow others to review, replicate, critique, and improve the procedure.

A minimum pipeline, implementable in a PWA prototype, can be described in eight stages:

1) Audio capture or upload (input). Two scenarios are considered:

· Direct recording (microphone) for exploratory analysis or training.

· File upload for deferred study (meetings, hearings, classes, interviews).

In both cases, the goal is to obtain a signal that is sufficiently clean for temporal analysis (not necessarily "perfect," but usable).

2) Acoustic preprocessing. In order to stabilize the analysis, standard operations are applied:

· Conversion to mono;

· Normalization of sampling frequency (e.g., 16 kHz);

· Temporal segmentation into windows (to facilitate transcription and avoid blocking with long files).

This step does not "interpret"; it only prepares the material so that the following modules are reproducible.

3) Automatic transcription with time stamps. The core of the pipeline is a transcription that includes word-level timestamps. Without time markers, prosody floats free; with them, the signal can be aligned with linguistic units.

It is important to emphasize that transcription is not used only for "text," but to anchor time: the start/end of each word, the gaps between words, and phrase boundaries.

4) Segmentation into pragmatic units. From timestamped words, segments (phrases/turns) are built using simple rules:

· Punctuation.

· Silences above a threshold (gaps).

· Window limits if the audio was processed in chunks.

The purpose is not to reconstruct "perfect grammar," but to organize the temporal flow in an inspectable way: what was said, when, with what pauses, and in what blocks.

(100) Couper-Kuhlen and Selting (1996) bring together studies showing prosody as an interactional resource for delimiting units, projecting closures, and managing turn-taking and alignment. Couper-Kuhlen, E., & Selting, M. (Eds.). (1996). *Prosody in Conversation: Interactional Studies*. Cambridge, UK: Cambridge University Press.

5) Extraction of prosodic features at the word level. For each word (or temporal token), basic acoustic and temporal variables are extracted:

· Duration (s): end - start.

· Energy (RMS/dB): proxy for intensity/emphasis (with caution: it depends on microphone and distance).

· Estimated pitch (Hz): proxy for tonal height (with limits: noise, non-modal intonation, extreme registers).

· Voiced ratio (0-1): proportion of time with voice versus noise/silence; useful for detecting whispering, fatigue, clipping, or interference.

· Local speed: approximately, syllables per second or temporal variation within word/segment.

· Internal pauses (maximum pause within the interval): indicators of clipping, hesitation, containment, or micro-breaks.

· Gaps between words (preceding and following silence): a critical variable for punitive silences or turn-cutting.

These variables are deliberately low-level: they are not moral interpretations; they are measures of the time and energy of speech.

6) Extraction of syllable-level features (approximation). To refine rhythm, an approximate syllabic segmentation can be based on energy (intensity peaks) and temporal transitions. At the syllable level, one can obtain:

· Mean duration and variability.

· Mean energy and range.

· Mean pitch and range.

· Mean voiced ratio.

· Syllabic rate (syllables/second).

· Estimated number of syllables per word/segment.

This module should be declared for what it is: a useful approximation for rhythm and pattern, not a perfect phonological analyzer. Its value lies in capturing temporal structure, not in "guessing the syllable" as a linguistic object.

7) Encoding into a compact vector + normalization. To compare episodes without depending on absolute values, the prototype encodes each segment into a fixed set of temporal/acoustic features

(duration, energy, estimated pitch, voice/non-voice ratio, preceding/following silences, etc.) with explicit normalization by speaker and by context (baseline).

The detailed specification of the vector (the "16Si" proposal), together with ranges and conversion/normalization rules, is included in Appendix 13.7.

8) Pattern-analysis layer: similarity, asymmetries, tendencies. Once segments are represented as sequences of vectors, second-order analyses become possible:

· Similarity between contiguous fragments (temporal coherence).

· Abrupt breaks (changes in register or energy).

· Asymmetries by person/role.

· Persistence of silences or interruptions.

· Repeated motifs (patterns that "return").

At this point, intuitions can be translated into observables without collapsing them into a verdict: patterns are detected, and then interpreted with context.

The value of the prototype does not lie in "automating judgment," but in creating a bridge between human inspection and quantitative record: the audio can be heard, the segment can be seen, the variables can be reviewed, the ranges can be discussed, and episodes can be compared. That methodological transparency is part of the ethics of the approach.

9.4. Units of analysis (Phase 1 / Phase 2)

So that measurement does not fall into reductionism, it is crucial to state the units of analysis and the jump between levels. This essay proposes two phases, not as a rigid chronology, but as a methodological distinction:

Phase 1: micro-meso units (interaction)

This phase focuses on describing concrete episodes with the greatest possible temporal anchoring.

Typical units:

· Token/word: acoustic features and immediate gaps.

· Segment/phrase: aggregates by temporal block (means, ranges, variability).

· Speaking turn: when it is possible to distinguish "who has the floor" and how they are allowed to sustain it.

· Pragmatic event: interruption, correction, closure[101], minimal laughter, silence after an intervention, abrupt register shift.

(101) Goodwin (1981) analyzes conversational organization and how local markers acquire meaning within broader sequential patterns. Goodwin, C. (1981). *Conversational Organization: Interaction between Speakers and Hearers.* New York, NY: Academic Press.

Objective: to build a "map" of micro-signals with enough detail to be debatable, and enough structure to be comparable.

Phase 2: macro units (climate and trajectory)

This phase focuses on how patterns stabilize and what effects they produce.

Typical units:

· Meeting/classroom/hearing: distribution of turns and silences, asymmetries, density of events.

· Team/institution: regularities across multiple scenes (not a single event).

· Temporal trajectory: evolution of the pattern (does it intensify? normalize? get repaired? shift to informal channels?).

Objective: to distinguish "style" from "climate." A style may be constant and neutral; a hostile climate is characterized by repetition, asymmetry, and a fruit of closure (silence, withdrawal, fear).

This distinction avoids a common mistake: taking an isolated scene as a structural diagnosis. Institutional hostility, if it exists, tends to be stable and multi-scene. Therefore, useful measurement is not the one that detects "a tone," but the one that detects patterns sustained over time.

9.5. Taxonomy of prosodic-pragmatic indicators (draft)

A taxonomy is a working tool: it does not aim to fix reality, but to organize observation and discussion. The following proposal is explicitly presented as a draft, because its usefulness depends on iteration, testing, and review.

A) Turn indicators (conversational power)[102]

· Interruptions (frequency, distribution by role).

· Overlaps (who talks over whom).

· Abrupt closures (cutting off an idea before it is completed).

· Topic reassignment (agenda shift after a participant's intervention).

· "Formal turn" without real response (the person is allowed to speak, but is ignored or frozen out).

These indicators connect directly with hierarchy: who is allowed to sustain a turn, and who pays a cost for trying.

B) Silence indicators (temporality as control)

(102) Zimmerman and West (1975) describe how interruptions and silences in conversation function as mechanisms of dominance and turn control. Zimmerman, D. H., & West, C. (1975). "Sex roles, interruptions and silences in conversation." In B. Thorne & N. Henley (Eds.), *Language and Sex: Difference and Dominance* (pp. 105-129). Rowley, MA: Newbury House.

· Long gaps after someone's intervention (post-intervention silence).

· Selective silences (not general: applied to a specific person or kind of intervention).

· Punitive pauses (silence as disapproval).

· Silence as erasure (the content is not taken up again; it is allowed to drop).

· Temporal compression (speech is accelerated so as not to leave room for reply).

Silence is central because it is a tool of control that does not need words[103].

C) Intonation and contour indicators (act markers)

Although automatic measurement of full contour is complex, some proxies can be observed:

· Falls that "close" (sentencing intonation).

· Rises that question (a doubtful intonation that disauthorizes).

· Compressed ranges (technical coldness).

· Exaggerated ranges in mockery or irony.

· Displaced emphasis (sarcasm: prosody "contradicts" the literal).

The key is not "which exact pitch," but the use of contour as a signature of the act[104] (care, closure, threat, mockery).

D) Rhythm indicators (temporal dominance)

· Acceleration to dominate (speaking quickly to prevent reply).

· Slowing down to humiliate (stretching speech to infantilize or exasperate).

· Irregular variability under pressure (tension, containment).

· Syllabic rate and its shifts at critical moments (corrections, plausible deniability, sanction).

Rhythm matters because it determines whether the other person has real room to exist in the interaction.

E) Energy/intensity indicators (warmth/coldness)

· Systematically reduced energy toward someone (ice).

· Elevated energy in corrections (dominance).

(103) Ephratt (2008) proposes a typology of the pragmatic functions of silence and shows that it can operate as communicative action of containment, rejection, or control. Ephratt, M. (2008). "The functions of silence." *Journal of Pragmatics, 40*(11), 1909-1938.

(104) Gussenhoven (2004) describes how tonal contours and F0 range are organized in intonation systems and how they relate to pragmatic interpretations such as continuity, closure, and attitude. Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge, UK: Cambridge University Press.

· Abrupt intensity changes when naming, correcting, or disauthorizing.

· "Institutional whisper" (low energy with high authority: threat without volume).

Control of confounders is essential here: microphone, distance, environmental noise. For that reason, comparative analyses within the same scene are recommended.

F) Voice-quality indicators (voice/non-voice)[105]

· Increased noise or breathiness (fatigue, containment, tension).

· Reduced voiced ratio at moments of control.

· Frequent breaks (micro-breaks as a sign of internal vigilance or pressure).

These indicators usually work better as signals of the "condition of the interaction" than as individual signals.

G) Pragmatic indicators of plausible deniability (text + prosody)[106]

Literal content may include typical markers: "I'm just…," "it's not that serious," "that's your interpretation," "don't be dramatic," "we are professionals here." What matters is how they fit with form: tone, pause, emphasis, sequence.

A robust approach combines:

· Detection of linguistic markers.

· Prosodic observation of the utterance.

· Analysis of the effect: closure, humiliation, blockage.

H) Indicators of affective asymmetry (differential treatment)

The most powerful structural indicator is usually comparison: warmth toward some and ice toward others, within the same space and role. This asymmetry may appear in:

· Mean energy.

· Tonal range.

· Subsequent silences.

· Response speed.

· Turn availability.

(105) Kreiman and Sidtis (2011) review voice quality from the perspectives of production and perception and show its multicausal nature, including physical state, technique, and context. Kreiman, J., & Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception.* Chichester, UK: Wiley-Blackwell.

(106) Pinker, Nowak, and Lee (2008) analyze the logic of indirect speech and how it enables forms of plausible deniability without explicitly committing the literal content. Pinker, S., Nowak, M. A., & Lee, J. J. (2008). "The logic of indirect speech." *Proceedings of the National Academy of Sciences of the United States of America, 105*(3), 833-838.

· Correction style.

When affective asymmetry repeats itself, it turns prosody into culture: it stops being a personal trait and becomes an organizational pattern.

9.6. Indices and thresholds: when it stops being "style" and becomes "climate"

An isolated indicator is rarely enough. That is why it is more useful to speak of indices: combinations of variables that point to a structural phenomenon. The idea is not to invent a magic number, but to define explicit rules for discussing "how much pattern" appears and with what persistence.

A useful criterion for distinguishing style from climate can be stated as follows:

· **Style:** a relatively stable and general trait (not selective), with variation explainable by context and without a systematic fruit of fear/closure.

· **Hostile climate:** a repeated, selective, or asymmetrical pattern[107] that produces a stable fruit of silencing, reduced agency, or revictimization.

From this distinction, provisional indices can be defined:

(1) Turn Asymmetry Index (TAI). Measures the distribution of interruptions, overlaps, and speaking time by role/person. Signal of performative hierarchy.

(2) Temporal Sanction Index (TSI). Measures silences following interventions (gaps) and their selectivity. Signal of punishment or freezing.

(3) Differential Coldness Index (DCI). Compares energy, tonal range, and rhythm directed at different interlocutors. Signal of affective asymmetry.

(4) Reputational Closure Index (RCI). Combines pragmatic markers (plausible deniability, technification) with closure patterns (acceleration, lack of reply). Signal of institutional priority to close the narrative.

(5) Composite Prosodic-Pragmatic Risk Index[108] (provisional). Integrates the preceding indices with explicit weightings. Its usefulness depends on calibration with samples and peer review; it should be treated as a heuristic, not as proof.

With respect to thresholds, it is best to avoid the "universal threshold." Instead, three strategies are recommended:

· **Intra-scene relative threshold:** compare within the same meeting or context (controlling for microphone and noise).

(107) Morrison and Milliken (2000) conceptualize organizational silence as a collective phenomenon sustained by beliefs and practices that discourage speaking and block learning and change. Morrison, E. W., & Milliken, F. J. (2000). "Organizational silence: A barrier to change and development in a pluralistic world." *Academy of Management Review, 25*(4), 706-725.

(108) DeVellis (2017) describes the development of scales and indicators, including item generation, refinement, and the evaluation of reliability and validity. DeVellis, R. F. (2017). *Scale Development: Theory and Applications* (4th ed.). Thousand Oaks, CA: SAGE.

· **Longitudinal threshold:** observe persistence across scenes (trajectory).

· **Fruit-based threshold:** increase interpretive weight when the pattern coincides with observable effects (organizational silence, turnover, withdrawal, complaints, deterioration of psychological safety).

In this way, the indicator ceases to be an isolated number and becomes a structure of reading: it quantifies pattern, compares it over time, and places it in relation to effects.

9.7. Mixed measurement: quantitative + qualitative + triangulation by "fruit"

Purely quantitative measurement carries two risks: over-reduction and false authority. Purely qualitative measurement carries two others: remaining at the level of impressions and becoming non-replicable. For that reason, the appropriate methodological proposal is mixed.

An operational scheme:

Quantitative layer (patterns)

· Counts of interruptions, overlaps, and silences.

· Turn distributions.

· Means and ranges of prosodic variables by segment.

· Asymmetry and persistence indices.

· Detection of breaks and repeated motifs.

Qualitative layer (meaning and sequence)

· Coding of episodes (what happens, in what order, what function it serves).

· Narrative analysis: how the conflict is rewritten, how the account is frozen.

· Pragmatic analysis: what the act does in context (close, humiliate, care, repair).

Triangulation by "fruit"[109] (verification)

Interpretation gains strength when it is connected to observable effects:

· Psychological safety or self-censorship.

· Real availability of reply.

· Quality of repair.

(109) Denzin (1978) formalizes the triangulation of data, methods, researchers, and theories as a strategy for strengthening inferences and reducing the biases of single-interpretation readings. Denzin, N. K. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods* (2nd ed.). New York, NY: McGraw-Hill.

· Circulation of the account without punishment.

· Team stability without cynicism.

Within this scheme, the number does not "prove" intentions; it points to locations in the pattern that deserve attention. And qualitative analysis prevents the number from becoming dogma.

9.8. Ethics: how to measure without turning it into a tool of persecution

Any communicative measurement tool can be used either to care or to control. In organizations already inclined toward surveillance, a prosodic system can become a weapon: penalizing styles, forcing institutional "smiling," persecuting dissent, or refining plausible deniability. That is why ethics is not an appendix; it is part of the design.

The following minimum ethical conditions are proposed:

1) Explicit and limited purpose. Measurement should exist in order to prevent harm and improve conditions, not to punish or purge without due process. If its real purpose is disciplinary, the system loses legitimacy.

2) Consent and transparency. In non-public settings, capture and analysis require clear rules: what is analyzed, for what purpose, who has access, how long it is stored, how it is protected, and how objections can be raised. Opacity turns the tool into a threat.

3) Data minimization and protection.

reduce retention of raw audio when unnecessary;

privilege local extraction (on-device) whenever possible;

export only aggregated/anonymous variables for climate analysis;

avoid direct identification unless there is a legitimate procedure and guarantees.

4) Do not pathologize styles: difference ≠ harassment[111]. Prosodic norms vary by culture, class, age, gender, region[110], and neurodiversity. An ethical approach requires:

calibration by context;

intra-scene comparison;

caution regarding bias against accents or "non-normative" styles;

evaluation by sustained pattern and by fruit, not by isolated tone.

(110) Jun (2005) compiles prosodic typologies across multiple languages and shows systematic variation in patterns of intonation and phrasing. Jun, S.-A. (Ed.). (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

(111) McCann and Peppé (2003) review the evidence on prosody in autism spectrum disorders and discuss atypical prosodic profiles relevant to considering neurodiversity in the analysis. McCann, J., & Peppé, S. (2003). "Prosody in autism spectrum disorders: a critical review." *International Journal of Language & Communication Disorders, 38*(4), 325-350.

5) Right to reply and review[112]. If an analysis is used to guide intervention, it should include mechanisms for human review, explanation of criteria, and the right to disagreement. Measurement should be auditable, not a "black box" of authority.

6) Principle of repair. The ultimate goal is not to "point out culprits," but to improve the communicative environment: open real space for response, reduce punitive silences, adjust turn-taking norms, train listening, and facilitate repair. If the system is oriented toward punishment without repair, it reproduces the very problem it claims to solve.

In sum: measuring prosody is defensible only if it is designed to protect dignity and agency. Otherwise, it becomes a more sophisticated version of the same container that, in UC reading, substitutes itself for the source.

(112) Detert and Burris (2007) analyze employee voice and how certain leadership styles facilitate or block people from raising problems or disagreements. Detert, J. R., & Burris, E. R. (2007). "Leadership behavior and employee voice: Is the door really open?" *Academy of Management Journal, 50*(4), 869-884.

10. Applications by Domain

The preceding methodological proposal is not intended to be universal in its implementation, but it is transferable in its logic: prosody and the temporality of turn-taking matter wherever text is regulated and symbolic violence requires plausible deniability. Each domain, however, comes with its own biases and specific risks: performative styles in politics, procedural norms in justice, pedagogical asymmetries in education, contractual hierarchies in organizations, and operational pressure in customer service and healthcare.

This chapter describes possible applications without turning the tool into a total solution. Prosody can help us observe patterns; intervention always requires context, safeguards, and ethics.

10.1. Politics

Politics is a laboratory of public prosody[113]. The political act does not take place only in what is said, but in how the scene is organized. Prosody serves functions of belonging, exclusion, and dominance:

· Ridicule under plausible deniability: attacks presented as "jokes," "irony," or "just a comment," sustained by prosodic contours that leave little room for reply without making the respondent seem exaggerated.

· Turn control as agenda control[114]: strategic interruptions, accelerations that prevent development, theatrical pauses that mark hierarchy.

· Technocratic neutrality as moral deactivation: a cold register in the face of social harm, turning human problems into procedures and deactivating empathy without denying the content.

Possible applications:

1.- Analysis of debates and parliamentary sessions in order to measure the distribution of turns and interruptions, and compare how different actors are treated (affective asymmetry).

2.- Moderation training: time rules not only as formal fairness, but as protection of real space for reply.

3.- Research on polarization: detecting patterns of performative contempt and their relation to the public perception of legitimacy.

Specific risk: *politics includes deliberate performance. For that reason, the indicator must be read as a pattern of interaction, not as a psychological diagnosis.*

(113) Jamieson (1988) analyzes the transformation of political discourse in the electronic age. Jamieson, K. H. (1988). *Eloquence in an Electronic Age: The Transformation of Political Speechmaking*. New York: Oxford University Press.

(114) Bull (2003) analyzes debates, interviews, and political speeches through microanalysis, with attention to turn-taking and the fine-grained organization of public interaction. Bull, P. (2003). *The Microanalysis of Political Communication: Claptrap and Ambiguity.* London: Routledge.

10.2. Legal and judicial frameworks

In legal contexts, prosody can radically alter the experience of justice. Two scenes matter especially:

· **Questioning and hearings:** the way of asking may be care or coercion[115]. The same question, with a different rhythm, can become a trap.

· **Complaint and repair processes:** procedural coldness can revictimize[116] when tone turns lived experience into "paperwork" and reply into an annoyance.

Possible applications:

1.- Auditing interview practices (police, expert examination, judicial inquiry): identifying patterns of interruption, temporal pressure, punitive silences, or intonations of disauthorization.

2.- Training for legal operators: distinguishing necessary neutrality (impartiality) from aggressive neutrality (dehumanization).

3.- Designing living protocols: ensuring that procedure does not replace repair (container vs. source), and that procedural language does not function as reputational closure.

Specific risk: *privacy and due process. Measurement cannot become "automatic evidence"; it can only guide human review with safeguards.*

10.3. Education

School and university are contexts in which prosody organizes learning and dignity[117]. The teacher-student hierarchy can produce either care or humiliation[118]. Typical indicators include:

· **Subtle ridicule:** corrections delivered with a condescending tone, minimal laughter, pauses that turn the student into an object.

· **Punitive silences:** ignoring questions, not responding, freezing interventions until the person stops speaking.

· **Dehumanized evaluation:** a technical tone that reduces effort and vulnerability to "performance," without relational consideration.

Possible applications:

1.- Classroom observation: combining a qualitative guide with temporal variables (who receives more silence, who gets interrupted, who has space).

(115) Eades (2010) examines how language operates in the legal process, and how formulation, register, and interpretation shape questioning, testimony, and decisions. Eades, D. (2010). *Sociolinguistics and the Legal Process.* Bristol: Multilingual Matters.

(116) Tyler (1990) examines the relationship between legitimacy, procedural treatment, and obedience to the law. Tyler, T. R. (1990). *Why People Obey the Law*. New Haven and London: Yale University Press.

(117) Cazden (2001) describes the classroom as a discursive space of teaching and learning, with attention to how participation is organized through interaction. Cazden, C. B. (2001). *Classroom Discourse: The Language of Teaching and Learning* (2nd ed.). Portsmouth, NH: Heinemann.

(118) Noddings (2005) places care at the center of the school experience. Noddings, N. (2005). *The Challenge to Care in Schools* (2nd ed.). New York, NY: Teachers College Press.

2.- Teacher training: cultivating a hospitality of communicative forms - clarity, limits without humiliation, correction without disauthorization.

3.- Prevention of fear-based climates: detecting early patterns that produce withdrawal and abandonment.

Specific risk: *confusing a demanding pedagogical style with hostility. For that reason, longitudinal comparison and attention to fruit are needed: does it produce learning and agency, or fear and silence?*

10.4. Business and organizations

In organizations, prosody is a device of everyday hierarchy. Many dynamics of pressure are never written down; they are enacted in meetings, hallways, calls, and evaluations. Here the essay's central themes converge strongly:

· **Asymmetrical turns** (who may speak and who is cut off).

· **Selective ice** (differential coldness).

· **Technical register as a weapon** (formality used to disauthorize).

· **Plausible deniability** ("that's your interpretation," "I'm just being professional").

Possible applications:

1.- Auditing communicative climate: not in order to monitor individuals, but to identify meeting patterns that generate organizational silence(119).

2.- Meeting interventions: turn-taking rules, facilitation, decision logs, and explicit spaces for disagreement without punishment(120).

3.- HR and compliance processes: reviewing whether the "container" (procedure) is serving repair or reputational closure. Prosodic analysis can detect when a formal process is being carried out as technical humiliation.

Specific risk: *managerial instrumentalization. In companies already biased toward control, measuring prosody can become a device of emotional discipline. That is why design ethics - purpose, transparency, minimization - is a condition of possibility.*

10.5. Customer service / call centers / healthcare

These environments combine three factors: high emotional load, frequent interaction, and pressure from metrics. That makes them prone both to dehumanization and to excessive surveillance.

In call centers and customer service:

(119) Detert and Edmondson (2011) describe tacit beliefs about when speaking up at work is risky or inappropriate. Detert, J. R., & Edmondson, A. C. (2011). "Implicit voice theories: Taken-for-granted rules of self-censorship at work." *Academy of Management Journal, 54*(3), 461-488.

(120) Van Dyne, Ang, and Botero (2003) conceptualize employee voice and organizational silence as multidimensional constructs. Van Dyne, L., Ang, S., & Botero, I. C. (2003). "Conceptualizing employee silence and employee voice as multidimensional constructs." *Journal of Management Studies, 40*(6), 1359-1392.

· Prosody becomes a "product" (sounding friendly)[121], and may turn into an obligation that hides abuse toward workers.

· Performance metrics may push toward a coercive rhythm[122] (speeding up in order to close), increasing friction and degrading the quality of treatment.

· Hostility may come from the customer or from the system (rigid scripts, sanctions for stepping outside the script).

In healthcare:

· There is a real tension between professionalism (containment) and human care[123]**.**

· Technical neutrality may be necessary in emergencies, but when it becomes a pattern in the face of vulnerability it can produce dehumanization. · Team prosody (among professionals) also matters: clinical hierarchies may generate fear and errors through silence[124].

Possible applications:

1.- Relational quality without punitive surveillance: using aggregated indicators to detect operational pressure that produces cold treatment, both to protect users and to protect workers.

2.- Detection of communicative burnout: when clipped delivery rises, voiced ratio drops, and temporal irritability increases; not as a clinical diagnosis, but as a signal of operating conditions.

3.- Design of breaks and repair protocols: if pressure imposes an impossible rhythm, the system produces hostility as a byproduct. The useful intervention is not to demand "friendliness," but to correct the conditions.

Specific risk: *these sectors already suffer from high levels of surveillance. Any additional measurement must be justified by protection and improved conditions, not by control. If it is used to punish the operator for "tone," it institutionalizes the very violence it claims to combat.*

10.6. Media (advertising, film/TV series, news, podcasts, and social media)

Media is a key domain because here "voice" is rarely a pure phenomenon of the speaker: it is a production construct[125]. Prosody matters, yes, but it is coupled to another equally decisive layer:

(121) Hochschild (1983) describes the commercialization of feeling in service work. Hochschild, A. R. (1983). *The Managed Heart: Commercialization of Human Feeling*. Berkeley, CA: University of California Press.

(122) Korczynski (2003) analyzes communities of coping and collective emotional labor in the service sector. Korczynski, M. (2003). "Communities of coping: Collective emotional labour in service work." *Organization, 10*(1), 55-79.

(123) Street, Makoul, Arora, and Epstein (2009) explain pathways through which clinician-patient communication is linked to health outcomes. Street, R. L., Jr., Makoul, G., Arora, N. K., & Epstein, R. M. (2009). "How does communication heal? Pathways linking clinician-patient communication to health outcomes." *Patient Education and Counseling, 74*(3), 295-301.

(124) Edmondson (2003) studies how team leadership shapes willingness to speak up in the operating room. Edmondson, A. C. (2003). "Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams." *Journal of Management Studies, 40*(6), 1419-1452.

(125) Chion (1994) formulates the notion of audio-vision and examines the joint perception of sound and image. Chion, M. (1994). *Audio-Vision: Sound on Screen* (C. Gorbman, Ed. & Trans.). New York, NY: Columbia University Press.

the sound frame (music, effects, mixing, equalization, compression, ambience… and also its absence). That frame can heighten hidden coherences (make them more audible), but it can also conceal them (make them less perceptible) without causing them to disappear; they merely become less detectable for part of the audience.

In advertising and commercials:

· **Music as narrative glue:** a soundtrack can turn an ambiguous statement into something "epic," "trustworthy," or "urgent," generating emotional coherence even when the content is weak.

· **Voice + sonic branding as authority[126]:** the same text, with different prosody (vocal smile, restraint, paternal tone, technical tone) and a different background, produces opposite readings (care vs. pressure).

· **Disclaimers and perceptual asymmetry:** the main message is usually musicalized and supported; what limits, corrects, or reduces it (conditions, fine print) is often accelerated, lowered, or stripped of framing, thereby concealing the real coherence of the exchange.

In news and informational formats:

· **Ambience as framing:** a "raw" piece (without music) can produce an effect of authenticity and gravity; a piece with music or stings can produce alarm, suspicion, or epicness without stating it literally.

· **Editing as external prosody:** cuts, added silences, volume changes, compression, and the selection of breaths alter the perception of safety, doubt, aggressiveness, or control, even if the original speaker did not "sign" that intention.

· **Differential coldness by treatment:** the difference is not always in what is said, but in how different subjects or topics are treated sonically (warmth, airtime, associated music, silence, editing rhythm), which can create "selective ice" at the media level.

In film, TV series, documentaries, and entertainment:

· **Music makes subtext legible:** it can reveal the "hidden" coherence of a scene (threat, irony, tenderness) or impose a reading (moralize) that covers over the dialogue's intentional ambiguities.

· **Silence as an active resource:** the absence of music is not neutral; it can intensify intimacy, discomfort, or truth, and for that very reason it can function either as an amplifier of coherence or as a technique of emotional pressure.

Possible applications:

1.- Media literacy and perceptual training: using A/B comparisons (the same voiceover with different music / with and without effects / with a different mix) to show how inferences of intention, emotion, and credibility change without any change in the text.

(126) Van Leeuwen (1999) explores resources shared across voice, music, and sound, such as rhythm, perspective, timbre, and modality. Van Leeuwen, T. (1999). *Speech, Music, Sound*. Basingstoke: Macmillan.

2.- Audit of sound treatment (editorial ethics): reviewing pieces to identify where the sound frame is replacing the source (generating emotional certainty without evidence), or where it produces reputational closure (installing an "already settled" reading without allowing real reply).

3.- Comparative analysis of framing biases: studying systematic patterns of prosody and sound frame associated with topics, actors, or groups - not to psychologize the journalist, but to describe product practices: the distribution of warmth/coldness, rhythm, silences, music, and so forth.

4.- Responsible production guidelines: establishing minimum standards of transparency (what is editing/archive/voice-over, what has been remixed), and explicitly separating the informative from the emotional when the format claims to be "neutral."

Specific risk: *in media, the material is performance plus post-production. This introduces two dangers:*

(a) Attributing to the speaker what belongs to the edit (blaming "intention" where what exists is "mix").

(b) Turning the analysis into aesthetic policing (punishing styles) instead of using it to describe framing structures and their effects.

For that reason, here more than anywhere else, the unit of analysis must be the piece itself (the complete sonic product), and the reading must be triangulated: text + prosody + sound frame + effect on the audience.

11. Discussion: risks, limits, and objections

The central argument of this essay is deliberately ambivalent: on the one hand, it holds that prosody—and, more broadly, the temporal organization of speech—makes it possible to read coherences that the literal text does not sign; on the other, it insists that such reading cannot be turned into an oracle, a verdict, or a psychological "proof." That ambivalence is not a defect: it is the minimum condition for working with prosody in real environments (institutions, teams, education, justice) without falling into two symmetrical extremes: mysticism ("the voice never lies") or cynicism ("everything is interpretation").

This chapter sets out, explicitly, the most plausible objections to the approach and proposes practical criteria for sustaining it with sobriety: how to avoid false positives, how to protect communicative diversity, how to prevent punitive uses, and how to safeguard the methodology so that it serves repair rather than control.

11.1. False positives

A false positive occurs when an indicator suggests covert hostility, but the real explanation is different or insufficient. The risk is obvious: if the aim is to move from "it is felt" to "it is observed," a poor design may end up moving from "it is observed" to "it is accused."

There are four typical sources of false positives in prosody:

1.- Inherent ambiguity of signals[127]. A long pause may be punishment, but it may also be a search for words; a flat voice may be coldness, but it may also be fatigue or style; an increase in speed may be dominance, but it may also be nerves or genuine urgency. Prosody is rich precisely because it is multicausal[128].

2.- Situational noise (physical and technical context). Bad microphones, audio compression, ambient noise, latency in video calls, or turn-taking poorly managed by technology can imitate patterns that would otherwise be interpreted as interruption, indifference, or closure.

3.- Bodily and biographical noise (the speaker's state). Fatigue, pain, medication, anxiety, shyness, grief, or professional training (for example, styles of "technical neutrality") may produce signals similar to those that would be attributed to hostility.

4.- Observer and listener bias. The listener "fills in" from their own history; the analyst may project moral hypotheses too early. In contexts of conflict, interpretation tends to polarize.

To reduce false positives without denying the phenomenon, the approach needs operational rules. The strongest are:

· **Pattern, not event rule[129].**(An isolated signal is ambiguous; a repeated pattern under

(127) Ladd (2008) presents intonation as a system whose interpretation depends on structure and context, and underscores the margin of ambiguity inherent in prosodic contours. Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press.

(128) Kreiman and Sidtis (2011) review voice quality as a multicausal phenomenon shaped by physiology, perception, and context. Kreiman, J., & Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Malden, MA: Wiley-Blackwell.

(129) Bakeman and Gottman (1997) develop methods of sequential observation oriented toward repeated patterns and explicit comparisons, rather than isolated episodes. Bakeman, R., & Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis* (2nd ed.). New York, NY: Cambridge University Press.

comparable conditions (same task, similar roles, different episodes) carries greater interpretive force.

· **Intra-speaker baseline rule.** The question is rarely "does it sound cold?", but rather "does it sound consistently colder here than in the speaker's own norm, and how is that distributed toward whom?" Comparison against the speaker's own baseline reduces moralism and cultural bias.

· **Relational asymmetry rule.** Covert hostility does not usually appear as a general "strange tone," but as differential warmth, differential time, differential permission: the issue is not that someone always "sounds bad," but that their manner changes systematically according to addressee or position.

· **Minimum convergence rule.** A single indicator is not enough. The analysis gains legitimacy when several layers converge: prosody (rhythm/intonation/pause), pragmatics (acts of invalidation, plausible deniability, closure), and outcome (effect on participation, turn-taking, and safety to speak).

· **Reversible-interpretation rule.** Every prosodic reading should be able to formulate plausible alternative hypotheses and make explicit what evidence would weaken them. This rule protects against confirmation bias.

One additional point: besides false positives, there is the symmetrical risk of false negatives. Real harm may occur with "controlled" or formally impeccable prosody (through training, calculation, or institutional habituation). For that reason, prosody should be treated as a pathway of reading, not as an exhaustive detector: it helps us see, but it does not exhaust the visible.

11.2. Culture, gender, class, neurodiversity (legitimate variation)

Prosody is not universal in its form or in its interpretation[131]. Legitimate styles vary by language, region, class, age, gender, professional trajectory, and neurocognitive profile. Ignoring this would turn the proposal into a technology of normalization: a machine for rewarding those who "sound as expected" and penalizing those who fall outside the mold.

There are three main risks:

1.- Confusing difference with hostility. A direct style may be read as aggressive in an environment where indirectness is the norm; an intense prosody may be read as lack of control in environments that punish emotion; a monotone voice may be interpreted as contempt when it is simply a stable trait[132].

2.- Reproducing hierarchies under a technical appearance. If the "professional voice" is defined through narrow patterns, those who do not fit—because of accent, class, gender[130], or neurodiversity—may be labeled problematic. Prosodic analysis would then become an apparently neutral container that conceals symbolic violence.

(130) Foulkes and Docherty (2006) describe sociophonetic variation as a constitutive part of the social life of speech and relate it to age, gender, class, and trajectory. Foulkes, P., & Docherty, G. J. (2006). "The social life of phonetics and phonology." *Journal of Phonetics, 34*(4), 409-438.

(131) Jun (2005) brings together descriptions of multiple languages and shows systematic variation in intonation and phrasing within a common analytical framework. Jun, S.-A. (Ed.). (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

(132) McCann and Peppé (2003) review the evidence on prosody in autism spectrum disorders and show heterogeneous profiles, not reducible to a single pattern. McCann, J., & Peppé, S. (2003). "Prosody in autism spectrum disorders: A critical review." *International Journal of Language & Communication Disorders, 38*(4), 325-350.

3.- Turning care into surveillance. In the name of "healthy communication," one may intensify control over expressive styles, impoverishing diversity and raising the cost of speaking.

The way to avoid this is to return to a guiding principle already outlined in the essay: do not measure "what is correct"; measure asymmetries and effects. In practical terms:

· **Calibration by context and role.** A courtroom is not a classroom; an ICU is not an assembly; a call center is not an intimate conversation. Each context imposes legitimate constraints on tone and time. The question, then, is not "does it meet a universal ideal?" but "in this context, is form being used to care or to reduce?"

· **Focus on distributions, not traits.** What matters is not only that coldness exists, but to whom it is applied, when, with what consistency, and with what effect. Covert hostility is often selective.

· **Right to style and accommodation.** Any instrumentation must include the possibility of registering diverse communicative styles as a normal part of the ecosystem. Prosodic divergence cannot be treated as pathology by default.

· **Caution with psychological inferences.** Prosody is a bodily and relational correlate; attributing inner states is tempting, but methodologically risky. Where a psychological judgment is needed, another kind of evidence and another kind of professional practice are required.

The ethical implication is strong: *a prosodic-reading tool that fails to incorporate diversity runs the risk of becoming, itself, a form of institutional harassment with "technical" vocabulary.*

11.3. Prosody is not stand-alone proof: signal, not verdict

In real conflicts—and even more so in legal or disciplinary settings—a reasonable demand appears: "Can it be proven?" This essay offers a cautious answer: prosody is not a verdict, but it can be a signal[133].

To sustain that distinction seriously, three theses should be fixed:

1.- Prosody does not demonstrate intention. It may suggest probability, style, or relational stance; but intention is an opaque variable. Prosodic analysis can orient questions, not close diagnoses.

2.- Prosody does participate in harm (and in care). Even if intention is not "proven," the effect can be real: punitive silences, closed turns, humiliation through technification, mockery with plausible deniability. In hierarchical environments, the effect is amplified by the cost of reply.

3.- The minimum ethical unit is the pattern with effect. The operative question is not "what did they mean?", but "what structure did this produce, and with what regularity?" If a consistent pattern reduces participation, induces self-censorship, and creates asymmetry[135], there is an institutional problem even if subjective intention is not accessible.

(133) Eades (2010) examines how linguistic evidence in legal contexts depends on wording, interpretation, and context, and how a decontextualized reading can distort what occurred. Eades, D. (2010). *Sociolinguistics and the Legal Process*. Bristol: Multilingual Matters.

(135) Edmondson (1999) links psychological safety to the real willingness to speak up, disagree, and report problems in teams. Edmondson, A. (1999). "Psychological Safety and Learning Behavior in Work Teams." *Administrative Science Quarterly, 44*(2), 350-383.

For that reason, the appropriate place of prosodic analysis is that of auxiliary evidence within a broader triangulation[134]:

· **Prosodic observation** (time, intonation, pauses, energy, turns).

· **Pragmatic observation** (acts of invalidation, closures, disauthorization, plausible deniability).

· **Contextual and documentary evidence** (roles, rules, relational history, decisions, consequences).

· **Result in the climate** (safety to speak, clarity, repair, or closure).

When prosody is isolated as "proof," it becomes unjust. When it is integrated as a "signal" within a framework of care, it becomes useful.

11.4. Instrumentalization: measuring to control vs. measuring to protect

Every proposal for social measurement faces a dilemma: what can serve to protect can also serve to monitor. With prosody, the risk is especially acute because it affects the body: measuring voice can become a device of control[136] over presence, emotion, and obedience.

There are at least four typical modes of instrumentalization:

1.- Individual surveillance and punishment. Scoring people for "tone" and using it to sanction, label, or expel them. This use turns the approach into a disciplinary technology and distorts its original aim (to make visible what is denied so that it can be repaired).

2.- Aesthetic standardization ("the correct voice") as power. Imposing a narrow prosodic ideal under the banner of "quality" or "professionalism." It becomes a form of covert social selection.

3.- Institutional reputation over repair. Using metrics to demonstrate "compliance" and close the issue without addressing the harm. Here the container error appears clearly: the procedure substitutes itself for the source (repair).

4.- Blaming the victim ("it was their perception"). An organization may use incomplete measurement to invalidate accounts: "according to the system, there was nothing." Once again, the container displaces the source: the instrument becomes an excuse not to look at the human pattern.

The only serious way to address this risk is to make an explicit normative principle: if measurement reduces the freedom to speak or increases fear, it has failed, even if it is "accurate." The criterion of verification is not only statistical; it is ethical-structural: the fruit in the climate.

For that reason, any applied tool must rest on minimum conditions:

(134) Denzin (1978) formalizes the triangulation of data, methods, researchers, and theories as a safeguard against overly closed inferences. Denzin, N. K. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods* (2nd ed.). New York, NY: McGraw-Hill.

(136) Foucault (1995) examines how modern institutions shift visible punishment toward more diffuse forms of discipline and control. Foucault, M. (1995). *Discipline and Punish: The Birth of the Prison*. New York, NY: Vintage Books.

· **Declared and limited purpose:** prevention and improvement, not automatic sanction.

· **Data minimization:** measure what is necessary, keep the minimum, avoid unnecessary accumulation.

· **Clear governance[137]:** who has access, for what, and under what controls.

· **Preference for aggregate analysis:** patterns of teams, meetings, or processes, not personal "profiles."

· **Protection against retaliation:** if speaking carries a cost, measurement becomes a weapon.

11.5. Safeguards: transparency, peer review, right to reply

To keep the approach from degrading into a tool of suspicion or control, it needs methodological and institutional safeguards. "Safeguarding" here does not mean shielding it from criticism, but exactly the opposite: designing it so that it can be criticized, reviewed, and corrected[138].

A minimum set of safeguards can be organized into five layers:

1.- Operational transparency. Definitions, variables, thresholds, assumptions, and limits must be explicit. If an indicator is used, one should be able to answer: "what exactly does it measure?", "how is it calculated?", "what cannot be concluded from it?".

2.- Peer review (in the broad sense)[139]. In research: coding with multiple evaluators, comparison across teams, publication of protocols. In institutional application: periodic audits, independent oversight, bias review.

3.- Right to reply and contextualization[140]. No one should be evaluated through prosody without the possibility of contextualizing it (state of health, load, technical conditions, role, episode). Reply is not a courtesy gesture; it is an epistemic guarantee.

4.- Mandatory triangulation. Explicit prohibition on using prosody as a sole criterion. Any signal must be contrasted with pragmatics, context, and effects. This reduces injustice and improves validity.

5.- Principle of repair. The output of the analysis should not "close" a problem, but open it carefully: describe a pattern so it can be discussed and corrected. If measurement produces fear, opacity, or silence, it has inverted its purpose.

With these layers in place, prosodic reading can be sustained as a tool of clarity: not to moralize individuals, but to identify where communicative form is producing a climate of reduced voice. Without them, the approach becomes indistinguishable from that which it meant to criticize.

(137) Nissenbaum (2009) develops the idea of contextual privacy and situates the problem in information flows that are improper relative to context, actors, and purpose. Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford Law Books.

(138) Munafò, Nosek, Bishop, and colleagues (2017) call for transparency, documentation, and open review as basic conditions of reproducible science. Munafò, M. R., Nosek, B. A., Bishop, D. V. M., et al. (2017). "A manifesto for reproducible science." *Nature Human Behaviour, 1, 0021.*

(139) Krippendorff (2018) systematizes coding, reliability, and validity in content analysis, with attention to explicit protocols and comparison across evaluators. Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology* (4th ed.). Los Angeles, CA: SAGE Publications.

(140) Tyler (2007) summarizes procedural justice around voice, neutrality, respect, and trust, and places the possibility of being heard at the center of process legitimacy. Tyler, T. R. (2007). "Procedural Justice and the Courts." *Court Review, 44*(1/2), 26-31.

12. Conclusion

This essay began from a contemporary scene: impeccable texts, correct protocols, and sanitized discourse that nevertheless leave behind a residue of closure, fear, or corrosion. The hypothesis was that, when a system learns to "clean up" content, the real dynamics of treatment shift toward less archivable channels: prosody, silences, turns, and micro-signals.

12.1. Synthesis

Throughout the essay, prosody was proposed as a temporal architecture: not an ornament of language, but the plane on which a sentence becomes an act. That is why it can sustain hidden coherences: stable alignments between intention, emotion (as bodily dynamic), and social structure (permissions, costs, and conversational hierarchies), even when the text remains defensible.

The bridge to Universal Coherence (UC) made it possible to name the phenomenon in integrative terms: when the container (form, protocol, correctness) replaces the source (care, truth, repair), incoherence is not always signed in words, but is enacted in the time of speech. The fruit criterion helped preserve a practical mode of verification: what matters is not whether something sounds "nice" or "professional," but what it produces in the climate (clarity or fog; safety or self-censorship; possibility of reply or closure).

Finally, the methodological chapter proposed a cautious transition: moving from "it is felt" to "it is observed" without turning the voice into an oracle or measurement into surveillance. Prosody here is not a verdict; it is a signal for orienting testable questions about patterns of interaction.

12.2. Contribution: from "I felt it" to "observable patterns"

The central contribution can be formulated as a shift in focus: from moralizing individuals ("what kind of person is this?") to describing dynamics ("what kind of interaction is being produced?"). That shift reduces the risk of interpretive injustice and opens a more suitable terrain for analysis, discussion, and, eventually, measurement.

The proposed program insists on five practical ideas:

1.- Separate observation from interpretation. First describe the features: pauses, latencies, interruptions, contours, energy, distribution of turns. Then propose a pragmatic hypothesis.

2.- Privilege patterns over events. An isolated signal is ambiguous; a repeated pattern distributed asymmetrically is more informative.

3.- Compare against a baseline. The question is rarely "does it sound cold?" but rather "does it sound significantly colder here than in its own norm, and toward whom?"

4.- Triangulate with effect and context. If the pattern produces consistent closure

(silence, withdrawal, fear of speaking), it gains plausibility. If there is no fruit, the hypothesis weakens.

5.- Maintain due process. To measure is not to accuse. An indicator should serve learning and repair, not punishment.

***In sum**, the essay offers a language and a framework for translating diffuse experiences—often disqualified for lack of "textual proof"—into more verifiable discussions about temporal organization, treatment, and conversational power.*

12.3. Hospitality of communicative forms

One ethical consequence of the approach is the need for a hospitality of forms: not turning differences of style (culture, class, profession, shyness, neurodiversity, fatigue, language) into moral suspicion. Prosody is multicausal. For that reason, any reading that ignores baseline, context, and fruit is condemned to generate false positives.

Hospitality does not mean relativism. It means precision: protecting communicative diversity while still detecting patterns that produce harm. The operative criterion is twofold:

A) Distributed pattern: it appears consistently and selectively (toward certain people, on certain topics, at critical moments).

B) Observable fruit: the pattern reduces agency (closure of participation), stabilizes asymmetry, or installs fear.

This combination makes it possible to sustain a sober position: neither "the voice never lies" nor "everything is interpretation." Prosody becomes useful when it is used to increase clarity, not to manufacture culprits.

12.4. Research agenda

If this essay is to become an empirical program, there are at least seven tasks:

1.- Build ethical corpora: with consent, anonymized, with data minimization and the right to withdraw.

2.- Establish coding reliability: clear guides, training, and inter-rater evaluation.

3.- Define baselines by speaker and by context: so as not to confuse style with patterns of treatment.

4.- Validate the link with fruit: correlate patterns with climate measures (psychological safety, organizational silence, participation) and with observable outcomes (turnover, absenteeism, complaints).

5.- Compare across domains: justice, education, business, healthcare, politics. Each has its

own biases.

6.- Test interventions: if you change turn-taking rules or meeting practices, does the pattern decrease and the fruit improve?

7.- Shield the design against punitive uses: transparency, auditability, reparative purpose, and governance.

The final question is not "*can we detect intentions?*" but rather: "*can we detect in time patterns of treatment that reduce voice and erode humanity in systems where the text is no longer enough to make them visible?*"

13. Appendices

13.1. Glossary

· **Prosody:** the temporal organization of speech (intonation, rhythm, pauses, stress, energy, phrasing).

· **Turn:** an interactive unit of speech; it includes taking, maintaining, yielding, and contesting the "floor."

· **Response latency:** the time between the end of one turn and the beginning of the next (gap).

· **Overlap:** simultaneous speech; it may be cooperative or interruptive.

· **Punitive silence:** silence used as sanction (it withdraws recognition or blocks reply).

· **Selective chill:** differential coldness according to person/topic (it is not simply "being serious"; it is a matter of distribution).

· **Plausible deniability:** the possibility of denying the act without contradicting the text ("I didn't say that").

· **Fruit:** observable effect on climate/participation (clarity vs. fog; agency vs. closure).

· **Baseline:** the speaker's/group's habitual pattern against which deviations are compared.

13.2. Coding guide

Rule 0: first describe, then interpret.

A) Turn management (mark with approximate time)

· INT: interruption (a cut-off that prevents completion).

· OVL: overlap (simultaneous speech).

· GAP: latency/gap (short/medium/long).

· YLD: explicit turn-yielding (invites the other person to speak).

B) Silences

· SIL-SEARCH: search/thinking pause.

· SIL-TECH: technical (lag, mic, connection).

· SIL-PUN: punitive (withholds response, freezes the exchange).

C) Prosody/style

· FLAT: sustained flattening.

· FALL-CUT: abrupt fall with clipping ("verdict"-like ending).

· RATE-UP / RATE-DOWN: acceleration / slowing down.

· EN-DROP: drop in energy toward a person/topic.

D) Interpretation (only if there is pattern + fruit)

· CLOSURE: reduces participation.

· CONTROL: restricts room for reply.

· CARE: opens space, repairs, legitimizes.

13.3. Observation template

- Context: (domain, roles, channel: in-person/video call, duration).

- Participants: (A/B/C...).

- Key episodes (table):

1) Time (min:sec)

2) Who is speaking

3) Brief literal text

4) Observable features (pauses, latency, interruption, energy, contour)

5) Distribution (to whom is it directed?)

6) Immediate fruit (does participation open up or close down?)

13.4. Perception questionnaire

1) "In this space I can disagree without paying a disproportionate cost."

2) "If I point out a problem, it is likely to be heard and responded to clearly."

3) "Some people are interrupted more than others."

4) "There are silences that function as punishment."

5) "Formal correctness here is usually accompanied by humane treatment."

6) “When someone expresses discomfort, the response tends to be cold or technical.”

7) “I feel my contributions have room to breathe: I can finish what I begin.”

8) “When a mistake is made, there is repair rather than humiliation.”

13.5. Fictional cases / vignettes

Vignette A (meeting):

- Correct wording + closing cadence + subsequent silences that freeze questions.

- Objective of analysis: to see whether the pattern is distributed (always toward the same people) and what fruit it produces.

Vignette B (classroom/interview):

- Condescending correction: minimal laughter + pause + shifted emphasis.

- Objective of analysis: to distinguish pedagogical demand from hostility (baseline + fruit).

13.6. Ethics and consent

- Purpose: repair/learning, not surveillance.

- Transparency: what is measured, how, for what purpose, and who has access.

- Minimization: do not store audio if aggregated features are sufficient.

- Right to withdraw: a real possibility of opting out without penalty.

- Auditability: explainable metrics, not a black box.

- Prohibition of punitive individual use (except under a clear due-process framework and with triangulation).

13.7. Minimum technical specification: variables, ranges, export formats

Proposal: a compact feature vector by segment/turn (normalized by speaker and context).

Example of 16 dimensions (“16Si”, indicative):

- *Temporal dimensions* (syllable/rhythm): mean duration, duration variability, speech rate, internal pauses, latency.

- *Energy dimensions:* mean energy, variability, abrupt drops, final energy.

- *Pitch dimensions:* estimated mean level, range, final slope, micro-jumps.

- *Voiced/unvoiced:* voiced/unvoiced ratio, silence rate, silence duration.

- *Turn:* interruptions received/emitted, overlaps, explicit yields.

Normalization:

- *By speaker:* compare against their baseline (z-score or percentiles).

- *By context:* adjust by channel (in-person/online), noise, microphone, fatigue.

Export:

- *CSV/JSON with:* session ID, speaker ID, timestamps, features, context notes.

Specific example in the OptimeFlow(s) Re-Pasa prototype (PWA)

Encoding into a compact vector (16Si) and conversion to UC. The prototype can organize the variables into a 16-dimensional vector (16Si) that combines syllabic features and word-level features. A typical operational scheme:

Syllabic dimensions (8):

· Mean energy.

· Mean voiced ratio.

· Mean pitch.

· Pitch range.

· Mean duration.

· Syllabic rate.

· Syllable count.

· Energy range.

Word/interval dimensions (8):

· Word energy.

· Word pitch.

· Local speed.

· Word voiced ratio.

· Word duration.

· Preceding silence.

· Following silence.

· Maximum internal pause.

These magnitudes can also be mapped onto a UC scale with explicit ranges: the point of interest here is not “the number,” but the public rule of conversion. By declaring ranges and method, measurement becomes auditable: another person can discuss whether the range is appropriate, whether normalization introduces biases, or whether it would be preferable to separate by context.

14. Final note and invitation to collaborate

If this essay is useful to you, the most valuable contribution you can make is not to "agree" with it, but to help make it more verifiable and more human: examples, critiques, counterexamples, domains where it fails, and ethical safeguards that would prevent it from becoming a tool of control.

Prosody must not become tone policing. If a form of measurement ever deserves to exist, it should be to protect voice, open space for reply, and facilitate repair: so that what today remains at the level of "it feels like" can be discussed without humiliation and corrected without fear.

I am open to:

· Comments, critiques, and suggestions for possible future versions.

· Collaborations in the design of tools that may support the proposed methodology and analysis.

· Translations or cultural adaptations.

· Any honest dialogue that helps make visible and prevent prosody being used as a widespread tool of control.

You can write to me directly at ace@optimeflow.com or leave a public comment on the Zenodo page (DOI: 10.5281/zenodo.18913483).

This work belongs to the shared domain of knowledge and care. The more we use it together, the more coherent and more human our organizations will become.

Thank you for reading and for being part of this.

Andrés Calvo Espinosa

March 2026

Universal Coherence Project

Bibliography

(1) Lehiste (1970) systematizes suprasegmental features (duration, pitch, intensity, pauses) and their linguistic function in the organization of speech. Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press. — p. 1

(2) Gumperz (1982) introduces the notion of contextualization cues and shows how signals such as prosody and rhythm guide inferences about framing and intention beyond lexical content. Gumperz, J. J. (1982). *Discourse Strategies*. Cambridge: Cambridge University Press. — p. 1

(3) Austin (1962) formulates speech act theory and distinguishes between what is said and what is done in saying it, providing the basis for treating the "performative" as a pragmatic dimension. Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press. — p. 1

(4) Bourdieu (1991) analyzes the symbolic power of language and how social authority is inscribed in forms of speech capable of exercising domination without declaring it explicitly. Bourdieu, P. (1991). *Language and Symbolic Power*. Cambridge, MA: Harvard University Press. — p. 1

(5) Pinker, Nowak, and Lee (2008) explain the logic of indirect speech and the role of plausible deniability when the literal content remains defensible while another intention is being communicated. Pinker, S., Nowak, M. A., & Lee, J. J. (2008). *"The logic of indirect speech." Proceedings of the National Academy of Sciences of the United States of America, 105*(3), 833-838. — p. 1

(6) Couper-Kuhlen and Selting (1996) bring together studies of prosody in conversation and show how intonation, rhythm, and pauses operate as resources for social action and interactional organization. Couper-Kuhlen, E., & Selting, M. (Eds.). (1996). *Prosody in Conversation: Interactional Studies*. Cambridge: Cambridge University Press. — p. 2

(7) Scherer (2003) reviews research paradigms on vocal communication of emotion and relates acoustic parameters (F0, intensity, tempo, voice quality) to affective states and evaluations. Scherer, K. R. (2003). "Vocal communication of emotion: A review of research paradigms." *Speech Communication, 40*(1-2), 227-256. — p. 2

(8) Sacks, Schegloff, and Jefferson (1974) describe the system of turn organization in conversation and its practical rules, which are key to conceptualizing "turn order" as social structure. Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). "A simplest systematics for the organization of turn-taking for conversation." *Language, 50*(4), 696-735. — p. 2

(9) Edmondson (1999) defines psychological safety in teams and links it to learning behaviors and voice, including the willingness to speak up, disagree, and report problems. Edmondson, A. (1999). "Psychological safety and learning behavior in work teams." *Administrative Science Quarterly, 44*(2), 350-383. — p. 2

(10) Ladd (2008) develops the phonology of intonation and emphasizes the contextual and probabilistic character of its interpretation, avoiding direct readings of intention from isolated signals. Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press. — p. 3

(11) Jun (2005) compiles prosodic typologies across multiple languages and shows systematic variation in patterns of intonation and phrasing, supporting the idea that there is no single universal reading. Jun, S.-A. (Ed.). (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. — p. 3

(12) McCann and Peppé (2003) review the evidence on prosody in autism spectrum disorders, discuss atypical prosodic profiles, and highlight interpretive challenges relevant to considering neurodiversity in prosodic analysis. McCann, J., & Peppé, S. (2003). "Prosody in autism spectrum disorders: a critical review." *International Journal of Language & Communication Disorders, 38*(4), 325-350. — p. 3

(13) Miles, Huberman, and Saldaña (2014) systematize qualitative analysis and coding techniques, including pattern coding, in order to convert observations into comparable and traceable categories. Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). Thousand Oaks, CA: SAGE. — p. 4

(14) Denzin (1978) formalizes triangulation (of data, methods, researchers, and theories) as a strategy for strengthening inferences in social research and reducing the bias of single-interpretation readings. Denzin, N. K. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods* (2nd ed.). New York, NY: McGraw-Hill. — p. 4

(15) DeVellis (2017) describes the development of scales and indicators (generation and refinement of items; evaluation of reliability and validity), which is useful for operationalizing constructs without turning them into verdicts. DeVellis, R. F. (2017). *Scale Development: Theory and Applications* (4th ed.). Thousand Oaks, CA: SAGE. — p. 4

(16) Crystal (1969) systematizes "prosodic systems" and describes prosody as the suprasegmental organization of speech (beyond the segment), providing a classic basis for an operational definition such as the one used in this essay. Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge, UK: Cambridge University Press. — p. 5

(17) Gussenhoven (2004) describes how tonal contours and F0 range are organized in intonation systems and how they relate to

pragmatic interpretations (continuity, closure, attitudes), which is useful for supporting the description of contours and ranges. Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge, UK: Cambridge University Press. — p. 5

(18) Couper-Kuhlen and Selting (1996) gather studies showing prosody as an interactional resource (delimiting units, projecting closures, managing turn-taking and alignment), supporting the idea that it is not merely a "property of the speaker." Couper-Kuhlen, E., & Selting, M. (Eds.). (1996). *Prosody in Conversation: Interactional Studies*. Cambridge, UK: Cambridge University Press. — p. 5

(19) Ephratt (2008) analyzes silence as a pragmatic behavior with conversational functions (e.g., suspension, containment, sanction/withdrawal), which is useful for treating pauses and silences as part of the prosodic-pragmatic repertoire. Ephratt, M. (2008). "The functions of silence." *Journal of Pragmatics, 40*(11), 1909-1938. — p. 6

(20) Kreiman and Sidtis (2011) review voice quality (timbre/voice quality) from the perspectives of production and perception, showing its multicausal nature (physical state, technique, context), which is key to interpreting "roughness" and other qualities without turning them into a diagnosis. Kreiman, J., & Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Chichester, UK: Wiley-Blackwell. — p. 6

(21) Poyatos (1993) describes paralanguage as a systematic component of oral communication (non-lexical sounds, vocalizations, micro-signals) that is integrated into the turn and alters the communicative act beyond literal content. Poyatos, F. (1993). *Paralanguage: A Linguistic and Interdisciplinary Approach to Interactive Speech and Sound*. Amsterdam/Philadelphia: John Benjamins. — p. 5

(22) Wennerstrom (2001) develops a framework of prosody and discourse analysis in which temporal patterns (prominence, rhythm, intonation) guide listener inferences and connect sound form with interactional function, providing a basis for the idea of observable alignments beyond content. Wennerstrom, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford, UK: Oxford University Press. — p. 7

(23) Bryant and Fox Tree (2005) investigate whether there is an "ironic tone of voice" and which prosodic signals help listeners recognize irony/sarcasm in speech, supporting the idea that the prosodic contour can contradict literal content. Bryant, G. A., & Fox Tree, J. E. (2005). "Is there an ironic tone of voice?" *Language and Speech, 48*(3), 257-277. — p. 7

(24) Bateson, Jackson, Haley, and Weakland (1956) introduce the concept of the double bind as a pattern of incompatible messages that traps the receiver's response, which is useful for describing verbal care delivered with a prosody of control/punishment and producing submission without an explicit order. Bateson, G., Jackson, D. D., Haley, J., & Weakland, J. (1956). "Toward a theory of schizophrenia." *Behavioral Science, 1*(4), 251-264. — p. 7

(25) Drew and Heritage (1992) compile studies of interaction in institutional settings, showing how roles, procedures, and goals organize speech (turn rights, formats, constraints), which provides a basis for analyzing covert hostility where the text is regulated. Drew, P., & Heritage, J. (Eds.). (1992). *Talk at Work: Interaction in Institutional Settings*. Cambridge, UK: Cambridge University Press. — p. 8

(26) Eisenberg (1984) describes strategic ambiguity in organizational communication as a resource for maintaining flexibility and avoiding explicit commitments, which connects with the notion of plausible deniability when the content remains "correct" and the asymmetry shifts into form. Eisenberg, E. M. (1984). "Ambiguity as strategy in organizational communication." *Communication Monographs, 51*(3), 227-242. — p. 8

(27) Titze (1994) develops a physical-physiological explanation of voice production (respiration, phonation, and laryngeal control) and its relation to parameters such as F0, intensity, and voice quality. Titze, I. R. (1994). *Principles of Voice Production*. Prentice Hall. — p. 9

(28) Stevens (2000) presents an acoustic theory of speech production (including the source-filter model) and describes the physical basis of measures such as fundamental frequency (F0) and its perceptual correlate. Stevens, K. N. (2000). *Acoustic Phonetics*. The MIT Press. — p. 9

(29) de Cheveigné and Kawahara (2002) propose the YIN F0 estimator and describe typical errors in automatic estimation, including "octave errors." de Cheveigné, A., & Kawahara, H. (2002). "YIN, a fundamental frequency estimator for speech and music." *The Journal of the Acoustical Society of America, 111*(4), 1917-1930. — p. 10

(30) Gussenhoven (2004) explains how intonation is interpreted as a configuration of contours, range, and prominence, and how its temporal alignment contributes to pragmatic meaning. Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press. — p. 10

(31) Kent and Read (2002) synthesize methods for acoustic analysis of speech and discuss energy/intensity measures (e.g., RMS/dB), together with practical limitations related to channel, distance, and noise. Kent, R. D., & Read, C. (2002). *The Acoustic Analysis of Speech* (2nd ed.). Singular/Thomson Learning. — p. 10

(32) Sacks, Schegloff, and Jefferson (1974) formulate a model of turn-taking in conversation, showing how pauses, overlaps, and

latencies structure rights and obligations of participation. Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). “A simplest systematics for the organization of turn-taking for conversation.” *Language, 50*(4), 696-735. — p. 11

(33) Jaworski (1993) analyzes silence as a pragmatic and social resource (containment, power, sanction), emphasizing that its function depends on the interactive frame. Jaworski, A. (1993). *The Power of Silence: Social and Pragmatic Perspectives*. Sage. — p. 11

(34) Kreiman and Van Lancker Sidtis (2011) offer an interdisciplinary framework on the production and perception of voice quality, including the distinction between periodic/aperiodic components and their perceptual correlates. Kreiman, J., & Van Lancker Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Wiley-Blackwell. — p. 11

(35) Ong (1982) analyzes the difference between orality and literacy and describes how, prior to written record, memory and vocal repetition function as a social “archive” in oral cultures. Ong, W. J. (1982). *Orality and Literacy: The Technologizing of the Word*. Methuen. — p. 13

(36) McNeill (1995) studies how rhythmic synchrony (dance, marching, “drill”) produces group cohesion and coordination, linking shared time with social organization. McNeill, W. H. (1995). *Keeping Together in Time: Dance and Drill in Human History*. Harvard University Press. — p. 13

(37) Durkheim (1912/1995) explains ritual as a mechanism of social integration and the generation of collective effervescence, which is useful for grounding the cohesive function of shared and repeated patterns. Durkheim, E. (1995). *The Elementary Forms of Religious Life* (K. E. Fields, Trans.). *Free Press*. (Original work published 1912) — p. 13

(38) Lord (1960) describes the transmission of oral poetry through formulas and repetition, showing how rhythmic and formulaic patterns make content transportable without external support. Lord, A. B. (1960). *The Singer of Tales*. Harvard University Press. — p. 13

(39) Aristotle (2007) develops the triad ethos-pathos-logos and the role of credibility and emotion in public persuasion, providing the classical basis for linking manner of speaking and audience effect. Aristotle. (2007). *On Rhetoric: A Theory of Civic Discourse* (G. A. Kennedy, Trans., 2nd ed.). Oxford University Press. — p. 14

(40) Quintilian (2001) details the role of actio (delivery) and the training of voice, pauses, and cadence in oratory, as a historical codification of prosodic control in rhetorical education. Quintilian. (2001). *The Orator’s Education* (Institutio Oratoria) (D. A. Russell, Trans.). Harvard University Press. (Original work published c. 95 CE) — p. 14

(41) Bourdieu (1991) conceptualizes the symbolic power of language and shows how “legitimate” styles and registers can exercise domination without explicit insult, supporting the idea of symbolic violence through form. Bourdieu, P. (1991). *Language and Symbolic Power* (J. B. Thompson, Ed.; G. Raymond & M. Adamson, Trans.). Harvard University Press. — p. 15

(42) Austin (1962) formulates speech act theory and performatives, clarifying how certain institutional or ritual formulas “do” things (swear, absolve, consecrate) in addition to describing them. Austin, J. L. (1962). *How to Do Things with Words*. Oxford University Press. — p. 15

(43) Bernardi et al. (2001) show that repetitive recitation (rosary/mantras) can couple with breathing rhythm and modulate autonomic cardiovascular rhythms, supporting the link between rhythm, breathing, and state. Bernardi, L., Sleight, P., Bandinelli, G., Cencetti, S., Fattorini, L., Wdowczyc-Szulc, J., & Lagi, A. (2001). “Effect of rosary prayer and yoga mantras on autonomic cardiovascular rhythms: Comparative study.” *BMJ, 323*(7327), 1446-1449. doi:10.1136/bmj.323.7327.1446 — p. 15

(44) Goody (1986) analyzes how writing, record-keeping, and textualization reconfigure social organization and institutions, linking archive/documentation to new forms of control and administration. Goody, J. (1986). *The Logic of Writing and the Organization of Society*. Cambridge University Press. — p. 16

(45) Weber (1922/1978) describes bureaucracy and rational-legal authority, emphasizing regulated procedures and impersonality as the basis of the ideal/appearance of institutional impartiality. Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology* (G. Roth & C. Wittich, Eds.). University of California Press. (Original work published 1922) — p. 16

(46) Heritage and Clayman (2010) synthesize research—especially from conversation analysis—on institutional interaction, showing how asymmetries, turn design, and question formats structure contexts such as interviews and other regulated situations. Heritage, J., & Clayman, S. (2010). *Talk in Action: Interactions, Identities, and Institutions*. Wiley-Blackwell. — p. 16

(47) Austin (1962) develops speech act theory and shows that to utter is to perform actions within a situational framework, providing the basis for treating conversation as a sequence of situated acts. Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press. — p. 18

(48) Lyons (1977) presents semantics as the study of linguistic meaning and propositional content, which is useful for delimiting

the level of "what is said" as distinct from contextual use. Lyons, J. (1977). *Semantics* (Vols. 1-2). Cambridge: Cambridge University Press. — p. 18

(49) Huang (2014) synthesizes pragmatics as the study of meaning in use: how context, roles, and inferences turn a sentence into a social act. Huang, Y. (2014). *Pragmatics* (2nd ed.). Oxford: Oxford University Press. — p. 18

(50) Grice (1975) formulates the cooperative principle and conversational maxims in order to explain implicatures, that is, what is communicated without being literally stated. Grice, H. P. (1975). "*Logic and conversation*." In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics* (Vol. 3: Speech Acts, pp. 41-58). New York, NY: Academic Press. — p. 18

(51) Goffman (1967) introduces the notion of face and face-work in interaction, describing how dignity/status is maintained or eroded beyond literal content. Goffman, E. (1967). *Interaction Ritual: Essays on Face-to-Face Behavior*. Garden City, NY: Anchor Books. — p. 18

(52) Cutler, Dahan, and van Donselaar (1997) review evidence showing that prosody (stress, rhythm, intonation, pauses) guides speech comprehension and alters interactive effect in real time. Cutler, A., Dahan, D., & van Donselaar, W. (1997). "Prosody in the comprehension of spoken language: A literature review." *Language and Speech, 40*(2), 141-201. — p. 19

(53) Lambrecht (1994) develops the framework of information structure (topic/focus) and its encoding, supporting the idea that focus/background and hierarchy are also marked through prosody. Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press. — p. 20

(54) Duncan (1972) describes signals and rules for taking and yielding turns in conversation, including prosodic cues that project continuation, closure, or yielding, and that organize access to conversational space. Duncan, S. (1972). "Some signals and rules for taking speaking turns in conversations." *Journal of Personality and Social Psychology, 23*(2), 283-292. — p. 20

(55) Juslin and Laukka (2003) review how the voice communicates emotion through patterns of tempo, pitch, intensity, and timbre, supporting prosody as a calibration of threat/safety and affective climate. Juslin, P. N., & Laukka, P. (2003). "Communication of emotions in vocal expression and music performance: Different channels, same code?" *Psychological Bulletin, 129*(5), 770-814. — p. 20

(56) Goodwin (1981) analyzes conversational organization and how local marks acquire meaning within sequential patterns, which is useful for distinguishing punctual micro-signals from meta-signals as second-order regularities. Goodwin, C. (1981). *Conversational Organization: Interaction between Speakers and Hearers*. New York, NY: Academic Press. — p. 20

(57) Stalnaker (2002) formulates the concept of common ground in order to model the background of assumptions that makes what lies "between the lines" inferable, connecting subtext with roles, history, and norm. Stalnaker, R. (2002). "Common ground." *Linguistics and Philosophy, 25*(5-6), 701-721. — p. 22

(58) Chion develops the idea of "audio-vision" and "added value": how sound design (including music) reconfigures perceived meaning without changing the text. Chion, M. (1994). *Audio-Vision: Sound on Screen. Edited and translated by Claudia Gorbman*. New York: Columbia University Press. — p. 22

(59) A key compendium on how music communicates and evokes emotion (mechanisms, evidence, and nuances), useful for supporting the affective priming effect of background music. Juslin, P. N., & Sloboda, J. A. (Eds.). (2010). *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford: Oxford University Press. — p. 22

(60) Experimental evidence that a soundtrack guides schemas of comprehension and memory for narrative material (it does not merely "decorate"): it alters what is understood and what is remembered. Boltz, M. G. (2001). "Musical soundtracks as a schematic influence on the cognitive processing of filmed events." *Music Perception, 18*(4), 427-454. — p. 22

(61) A classic study of semantic and formal congruence: music can "shift" the interpretation of ambiguous visual actions; it includes conditions in which the presence or absence of music changes the reading. Bolivar, V. J., Cohen, A. J., & Fentress, J. C. (1994). "Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action." *Psychomusicology, 13*, 28-59. — p. 23

(62) A foundational framework for the perceptual organization of sound (auditory scene analysis): it explains why, in mixtures, some components mask others and change what is "detectable" in the voice. Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press. — p. 23

(63) Ephratt (2008) proposes a typology of the pragmatic functions of silence and shows that it can operate as communicative action (affiliation, rejection, control, etc.), aligning with containment, punishment, complicity, and reputational management. Ephratt, M. (2008). "The functions of silence." *Journal of Pragmatics, 40*(11), 1909-1938. — p. 24

(64) Pinker, Nowak, and Lee (2008) analyze the logic of indirect speech and how it enables forms of plausible deniability and relational negotiation without explicitly committing the literal content. Pinker, S., Nowak, M. A., & Lee, J. J. (2008). "The logic of indirect speech." *Proceedings of the National Academy of Sciences of the United States of America, 105*(3), 833-838. https://doi.org/10.1073/pnas.0707192105 — p. 25

(65) Zimmerman and West (1975) describe how interruptions and silences in conversation function as mechanisms of dominance and turn control, providing a classic framework for reading the distribution of power in interaction. Zimmerman, D. H., & West, C. (1975). “Sex roles, interruptions and silences in conversation.” In B. Thorne & N. Henley (Eds.), *Language and Sex: Difference and Dominance* (pp. 105-129). Rowley, MA: Newbury House. — p. 26

(66) Cheang and Pell (2008) identify acoustic cues associated with the production of sarcasm and show that sarcastic “tone” relies on measurable prosodic patterns (e.g., changes in F0, intensity, and timing). Cheang, H. S., & Pell, M. D. (2008). “The sound of sarcasm.” *Speech Communication, 50*(5), 366-381. https://doi.org/10.1016/j.specom.2007.11.003 — p. 26

(67) Clark and Gerrig (1984) propose the pretense theory of irony to explain irony as an act in which the speaker performs a voice/position so that the listener can recover the real attitude, highlighting the role of shared context. Clark, H. H., & Gerrig, R. J. (1984). “On the pretense theory of irony.” *Journal of Experimental Psychology: General, 113*(1), 121-126. https://doi.org/10.1037/0096-3445.113.1.121 — p. 26

(68) Hochschild (1983) develops the concept of emotional labor and the ways institutions regulate affective expression, providing a framework for understanding “professional neutrality” as a form of emotional management with social effects. Hochschild, A. R. (1983). *The Managed Heart: Commercialization of Human Feeling*. Berkeley: University of California Press. — p. 26

(69) Yngve (1970) introduces the concept of the backchannel as a “secondary” channel of listener feedback (minimal responses, acknowledgments, listening signals), which is key for analyzing the presence or absence of interactive confirmation. Yngve, V. H. (1970). “On getting a word in edgewise.” *In Papers from the Sixth Regional Meeting of the Chicago Linguistic Society* (pp. 567-578). Chicago: Chicago Linguistic Society. — p. 28

(70) Bateson, Jackson, Haley, and Weakland (1956) formulate the double bind framework as a communicative structure of incompatible messages in which any response is sanctioned, which is useful for describing coherences in which the content declares care while the form enacts control. Bateson, G., Jackson, D. D., Haley, J., & Weakland, J. H. (1956). “Toward a theory of schizophrenia.” *Behavioral Science, 1*(4), 251-264. https://doi.org/10.1002/bs.3830010402 — p. 29

(71) Juslin and Laukka (2003) review evidence on how emotion is communicated vocally through acoustic patterns and how listeners recover states/attitudes from voice signals, providing a basis for affective “resonance.” Juslin, P. N., & Laukka, P. (2003). “Communication of emotions in vocal expression and music performance: Different channels, same code?” *Psychological Bulletin, 129*(5), 770-814. https://doi.org/10.1037/0033-2909.129.5.770 — p. 29

(72) Ross (1977) describes systematic attribution biases (e.g., over-attributing internal traits relative to context), offering a framework for understanding how the listener may turn a situated prosodic signal into a “trait” of the speaker. Ross, L. (1977). “The intuitive psychologist and his shortcomings: Distortions in the attribution process.” In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 173-220). New York: Academic Press. https://doi.org/10.1016/S0065-2601(08)60357-3 — p. 30

(73) Bakeman and Gottman (1997) present principles and techniques for observing interaction and analyzing sequences, prioritizing repeated patterns, comparison, and explicit rules, which is useful as a methodological basis for a non-“oracular” prosodic reading. Bakeman, R., & Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis* (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511527685 — p. 30

(74) Pfeffer (1992) develops how power and organizational influence structure which behaviors become legitimate and which dynamics become normalized in the everyday life of institutions. Pfeffer, J. (1992). *Managing with Power: Politics and Influence in Organizations*. Boston, MA: Harvard Business School Press. — p. 31

(75) Salin (2003) reviews how workplace bullying can be sustained by structures and processes in the environment (conditions that enable, motivate, or precipitate it), supporting a reading that goes beyond the episode between two people. Salin, D. (2003). “Ways of explaining workplace bullying: A review of enabling, motivating and precipitating structures and processes in the work environment.” *Human Relations, 56*(10), 1213-1232. — p. 31

(76) Andersson and Pearson (1999) conceptualize incivility as low-intensity deviant behavior with ambiguous intent, explaining why certain forms of hostility can operate with room for denial and at the same time escalate spirally within the work environment. Andersson, L. M., & Pearson, C. M. (1999). “Tit for tat? The spiraling effect of incivility in the workplace.” *Academy of Management Review, 24*(3), 452-471. — p. 32

(77) Keltner, Gruenfeld, and Anderson (2003) propose the approach-inhibition theory of power, describing how power tends to disinhibit and orient behavior, whereas lack of power increases inhibition and social vigilance, with implications for interactional patterns. Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). “Power, approach, and inhibition.” *Psychological Review, 110*(2), 265-284. — p. 32

(78) Detert and Burris (2007) analyze employee voice and how certain leadership styles facilitate or block people from raising problems or disagreements, providing empirical grounding for the issue of real permission to speak. Detert, J. R., & Burris, E. R. (2007). “Leadership behavior and employee voice: Is the door really open?” *Academy of Management Journal, 50*(4),

869-884. — p. 32

(79) Van Dyne, Ang, and Botero (2003) conceptualize silence and voice as multidimensional constructs, distinguishing different motives and forms of remaining silent or speaking up in organizations. Van Dyne, L., Ang, S., & Botero, I. C. (2003). "Conceptualizing employee silence and employee voice as multidimensional constructs." *Journal of Management Studies, 40*(6), 1359-1392. — p. 32

(80) Rafaeli and Sutton (1987) describe emotional expression as part of the work role and how organizations establish normative expectations regarding which emotion (and which form of expressing it) is considered appropriate in performance. Rafaeli, A., & Sutton, R. I. (1987). "Expression of emotion as part of the work role." *Academy of Management Review, 12*(1), 23-37. — p. 33

(81) Brescoll and Uhlmann (2008) show that the expression of anger has different consequences depending on gender and status, providing evidence of differential costs when someone expresses tension or irritation in workplace contexts. Brescoll, V. L., & Uhlmann, E. L. (2008). "Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace." *Psychological Science, 19*(3), 268-275. — p. 33

(82) Argyris and Schön (1978) develop the framework of organizational learning (including the logic of learning from errors and routines), which is useful for describing how a system incorporates patterns and turns them into practical norms. Argyris, C., & Schön, D. A. (1978). *Organizational Learning: A Theory of Action Perspective.* Reading, MA: Addison-Wesley. — p. 33

(83) Weick (1995) formulates the sensemaking approach to explain how organizations construct shared narratives and meanings about "what happened," providing a basis for understanding the institutional re-narration of conflicts. Weick, K. E. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA: Sage. — p. 34

(84) Morrison and Milliken (2000) propose the concept of organizational silence as a collective phenomenon sustained by beliefs and practices that discourage speaking, thereby limiting learning and change. Morrison, E. W., & Milliken, F. J. (2000). "Organizational silence: A barrier to change and development in a pluralistic world." *Academy of Management Review, 25*(4), 706-725. — p. 34

(85) Bies and Moag (1986) introduce interactional justice as a criterion of fairness tied to interpersonal and communicative treatment, relevant for describing additional harm when the institutional response becomes cold and formalistic in situations requiring repair. Bies, R. J., & Moag, J. S. (1986). "Interactional justice: Communication criteria of fairness." In R. J. Lewicki, B. H. Sheppard, & M. H. Bazerman (Eds.), *Research on Negotiation in Organizations* (Vol. 1, pp. 43-55). Greenwich, CT: JAI Press. — p. 34

(86) Watzlawick, Beavin, and Jackson (1967) distinguish the content level and the relational level of communication, showing how the relational component can redefine the meaning of a message beyond its literal wording. Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1967). *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies, and Paradoxes*. New York, NY: W. W. Norton & Company. — p. 36

(87) Wennerstrom (2001) develops a framework of prosody and discourse analysis in which temporal patterns (prominence, rhythm, intonation) guide listener inferences and connect sound form with interactive function. Wennerstrom, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford, UK: Oxford University Press. — p. 36

(88) Weber (1922/1978) describes modern bureaucracy as an order of regulated procedures and impersonality, in which administrative form can acquire primacy over the concrete relation to persons. Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology* (G. Roth & C. Wittich, Eds.). University of California Press. (Original work published 1922) — p. 37

(89) Heritage and Clayman (2010) show how institutional interaction is organized through asymmetries, formats, and turn design, and how those forms distribute access, reply, and authority in regulated settings. Heritage, J., & Clayman, S. (2010). *Talk in Action: Interactions, Identities, and Institutions*. Wiley-Blackwell. — p. 38

(90) Edmondson (1999) defines psychological safety and links it to the real willingness to speak, disagree, and report problems in teams. Edmondson, A. (1999). "Psychological safety and learning behavior in work teams." *Administrative Science Quarterly, 44*(2), 350-383. — p. 38

(91) Weick (1995) formulates organizational sensemaking as the shared construction of meaning around "what happened" and explains how organizations turn experience into common narratives. Weick, K. E. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA: Sage. — p. 38

(92) Quintilian (2001) emphasizes the importance of actio—delivery, voice, pauses, and cadence—in rhetorical effectiveness, locating part of eloquence in the execution of speech and not only in verbal content. Quintilian. (2001). *The Orator's Education* (Institutio Oratoria) (D. A. Russell, Trans.). *Harvard University Press.* (Original work published c. 95 CE) — p. 39

(93) Argyris and Schön (1978) develop a theory of action and organizational learning centered on how institutions incorporate routines, correct errors, and stabilize ways of acting that may drift away from their explicit discourse. Argyris, C., & Schön, D. A. (1978). *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison-Wesley. — p. 40

(94) Morrison and Milliken (2000) conceptualize organizational silence as a collective phenomenon sustained by beliefs and practices that discourage speaking, with effects of self-censorship, closure, and blocked learning. Morrison, E. W., & Milliken, F. J. (2000). "Organizational silence: A barrier to change and development in a pluralistic world." *Academy of Management Review, 25*(4), 706-725. — p. 40

(95) Goffman (1967) introduces the notion of face and the work of social image in interaction, showing how speech protects or erodes dignity, status, and recognition beyond literal content. Goffman, E. (1967). *Interaction Ritual: Essays on Face-to-Face Behavior*. Garden City, NY: Anchor Books. — p. 41

(96) Bies and Moag (1986) introduce interactional justice as a criterion of fairness tied to the mode of treatment and show that the experience of justice is also at stake in the way one responds, listens, and explains. Bies, R. J., & Moag, J. S. (1986). "Interactional justice: Communication criteria of fairness." In R. J. Lewicki, B. H. Sheppard, & M. H. Bazerman (Eds.), *Research on Negotiation in Organizations* (Vol. 1, pp. 43-55). Greenwich, CT: JAI Press. — p. 41

(97) Bakeman and Gottman (1997) present principles and techniques for observing interaction and analyzing sequences, with attention to repeated patterns, comparison, and explicit rules. Bakeman, R., & Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis* (2nd ed.). Cambridge: Cambridge University Press. — p. 42

(98) Ladd (2008) develops the phonology of intonation and underscores the contextual and probabilistic character of its interpretation. Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press. — p. 42

(99) Kent and Read (2002) synthesize methods for acoustic analysis of speech and discuss energy/intensity measures (RMS/dB) together with practical limits related to channel, distance, and noise. Kent, R. D., & Read, C. (2002). *The Acoustic Analysis of Speech* (2nd ed.). Singular/Thomson Learning. — p. 43

(100) Couper-Kuhlen and Selting (1996) bring together studies showing prosody as an interactional resource for delimiting units, projecting closures, and managing turn-taking and alignment. Couper-Kuhlen, E., & Selting, M. (Eds.). (1996). *Prosody in Conversation: Interactional Studies*. Cambridge, UK: Cambridge University Press. — p. 44

(101) Goodwin (1981) analyzes conversational organization and how local markers acquire meaning within broader sequential patterns. Goodwin, C. (1981). *Conversational Organization: Interaction between Speakers and Hearers.* New York, NY: Academic Press. — p. 46

(102) Zimmerman and West (1975) describe how interruptions and silences in conversation function as mechanisms of dominance and turn control. Zimmerman, D. H., & West, C. (1975). "Sex roles, interruptions and silences in conversation." In B. Thorne & N. Henley (Eds.), *Language and Sex: Difference and Dominance* (pp. 105-129). Rowley, MA: Newbury House. — p. 47

(103) Ephratt (2008) proposes a typology of the pragmatic functions of silence and shows that it can operate as communicative action of containment, rejection, or control. Ephratt, M. (2008). "The functions of silence." *Journal of Pragmatics, 40*(11), 1909-1938. — p. 48

(104) Gussenhoven (2004) describes how tonal contours and F0 range are organized in intonation systems and how they relate to pragmatic interpretations such as continuity, closure, and attitude. Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge, UK: Cambridge University Press. — p. 48

(105) Kreiman and Sidtis (2011) review voice quality from the perspectives of production and perception and show its multicausal nature, including physical state, technique, and context. Kreiman, J., & Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Chichester, UK: Wiley-Blackwell. — p. 49

(106) Pinker, Nowak, and Lee (2008) analyze the logic of indirect speech and how it enables forms of plausible deniability without explicitly committing the literal content. Pinker, S., Nowak, M. A., & Lee, J. J. (2008). "The logic of indirect speech." *Proceedings of the National Academy of Sciences of the United States of America, 105*(3), 833-838. — p. 49

(107) Morrison and Milliken (2000) conceptualize organizational silence as a collective phenomenon sustained by beliefs and practices that discourage speaking and block learning and change. Morrison, E. W., & Milliken, F. J. (2000). "Organizational silence: A barrier to change and development in a pluralistic world." *Academy of Management Review, 25*(4), 706-725. — p. 50

(108) DeVellis (2017) describes the development of scales and indicators, including item generation, refinement, and the evaluation of reliability and validity. DeVellis, R. F. (2017). *Scale Development: Theory and Applications* (4th ed.). Thousand Oaks, CA: SAGE. — p. 50

(109) Denzin (1978) formalizes the triangulation of data, methods, researchers, and theories as a strategy for strengthening inferences and reducing the biases of single-interpretation readings. Denzin, N. K. (1978). *The Research Act: A Theoretical*

Introduction to Sociological Methods (2nd ed.). New York, NY: McGraw-Hill. — p. 51

(110) Jun (2005) compiles prosodic typologies across multiple languages and shows systematic variation in patterns of intonation and phrasing. Jun, S.-A. (Ed.). (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. — p. 52

(111) McCann and Peppé (2003) review the evidence on prosody in autism spectrum disorders and discuss atypical prosodic profiles relevant to considering neurodiversity in the analysis. McCann, J., & Peppé, S. (2003). "Prosody in autism spectrum disorders: a critical review." *International Journal of Language & Communication Disorders, 38*(4), 325-350. — p. 52

(112) Detert and Burris (2007) analyze employee voice and how certain leadership styles facilitate or block people from raising problems or disagreements. Detert, J. R., & Burris, E. R. (2007). "Leadership behavior and employee voice: Is the door really open?" *Academy of Management Journal, 50*(4), 869-884. — p. 53

(113) Jamieson (1988) analyzes the transformation of political discourse in the electronic age. Jamieson, K. H. (1988). *Eloquence in an Electronic Age: The Transformation of Political Speechmaking*. New York: Oxford University Press. — p. 54

(114) Bull (2003) analyzes debates, interviews, and political speeches through microanalysis, with attention to turn-taking and the fine-grained organization of public interaction. Bull, P. (2003). *The Microanalysis of Political Communication: Claptrap and Ambiguity*. London: Routledge. — p. 54

(115) Eades (2010) examines how language operates in the legal process, and how formulation, register, and interpretation shape questioning, testimony, and decisions. Eades, D. (2010). *Sociolinguistics and the Legal Process*. Bristol: Multilingual Matters. — p. 55

(116) Tyler (1990) examines the relationship between legitimacy, procedural treatment, and obedience to the law. Tyler, T. R. (1990). *Why People Obey the Law*. New Haven and London: Yale University Press. — p. 55

(117) Cazden (2001) describes the classroom as a discursive space of teaching and learning, with attention to how participation is organized through interaction. Cazden, C. B. (2001). *Classroom Discourse: The Language of Teaching and Learning* (2nd ed.). Portsmouth, NH: Heinemann. — p. 55

(118) Noddings (2005) places care at the center of the school experience. Noddings, N. (2005). *The Challenge to Care in Schools* (2nd ed.). New York, NY: Teachers College Press. — p. 55

(119) Detert and Edmondson (2011) describe tacit beliefs about when speaking up at work is risky or inappropriate. Detert, J. R., & Edmondson, A. C. (2011). "Implicit voice theories: Taken-for-granted rules of self-censorship at work." *Academy of Management Journal, 54*(3), 461-488. — p. 56

(120) Van Dyne, Ang, and Botero (2003) conceptualize employee voice and organizational silence as multidimensional constructs. Van Dyne, L., Ang, S., & Botero, I. C. (2003). "Conceptualizing employee silence and employee voice as multidimensional constructs." *Journal of Management Studies, 40*(6), 1359-1392. — p. 56

(121) Hochschild (1983) describes the commercialization of feeling in service work. Hochschild, A. R. (1983). *The Managed Heart: Commercialization of Human Feeling*. Berkeley, CA: University of California Press. — p. 57

(122) Korczynski (2003) analyzes communities of coping and collective emotional labor in the service sector. Korczynski, M. (2003). "Communities of coping: Collective emotional labour in service work." *Organization, 10*(1), 55-79. — p. 57

(123) Street, Makoul, Arora, and Epstein (2009) explain pathways through which clinician-patient communication is linked to health outcomes. Street, R. L., Jr., Makoul, G., Arora, N. K., & Epstein, R. M. (2009). "How does communication heal? Pathways linking clinician-patient communication to health outcomes." *Patient Education and Counseling, 74*(3), 295-301. — p. 57

(124) Edmondson (2003) studies how team leadership shapes willingness to speak up in the operating room. Edmondson, A. C. (2003). "Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams." *Journal of Management Studies, 40*(6), 1419-1452. — p. 57

(125) Chion (1994) formulates the notion of audio-vision and examines the joint perception of sound and image. Chion, M. (1994). *Audio-Vision: Sound on Screen* (C. Gorbman, Ed. & Trans.). New York, NY: Columbia University Press. — p. 57

(126) Van Leeuwen (1999) explores resources shared across voice, music, and sound, such as rhythm, perspective, timbre, and modality. Van Leeuwen, T. (1999). *Speech, Music, Sound*. Basingstoke: Macmillan. — p. 58

(127) Ladd (2008) presents intonation as a system whose interpretation depends on structure and context, and underscores the margin of ambiguity inherent in prosodic contours. Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press. — p. 60

(128) Kreiman and Sidtis (2011) review voice quality as a multicausal phenomenon shaped by physiology, perception, and context. Kreiman, J., & Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice*

Production and Perception. Malden, MA: Wiley-Blackwell. — p. 60

(129) Bakeman and Gottman (1997) develop methods of sequential observation oriented toward repeated patterns and explicit comparisons, rather than isolated episodes. Bakeman, R., & Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis* (2nd ed.). New York, NY: Cambridge University Press. — p. 60

(130) Foulkes and Docherty (2006) describe sociophonetic variation as a constitutive part of the social life of speech and relate it to age, gender, class, and trajectory. Foulkes, P., & Docherty, G. J. (2006). "The social life of phonetics and phonology." *Journal of Phonetics, 34*(4), 409-438. — p. 61

(131) Jun (2005) brings together descriptions of multiple languages and shows systematic variation in intonation and phrasing within a common analytical framework. Jun, S.-A. (Ed.). (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. — p. 61

(132) McCann and Peppé (2003) review the evidence on prosody in autism spectrum disorders and show heterogeneous profiles, not reducible to a single pattern. McCann, J., & Peppé, S. (2003). "Prosody in autism spectrum disorders: A critical review." *International Journal of Language & Communication Disorders, 38*(4), 325-350. — p. 61

(133) Eades (2010) examines how linguistic evidence in legal contexts depends on wording, interpretation, and context, and how a decontextualized reading can distort what occurred. Eades, D. (2010). *Sociolinguistics and the Legal Process*. Bristol: Multilingual Matters. — p. 62

(134) Denzin (1978) formalizes the triangulation of data, methods, researchers, and theories as a safeguard against overly closed inferences. Denzin, N. K. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods* (2nd ed.). New York, NY: McGraw-Hill. — p. 63

(135) Edmondson (1999) links psychological safety to the real willingness to speak up, disagree, and report problems in teams. Edmondson, A. (1999). "Psychological Safety and Learning Behavior in Work Teams." *Administrative Science Quarterly, 44*(2), 350-383. — p. 62

(136) Foucault (1995) examines how modern institutions shift visible punishment toward more diffuse forms of discipline and control. Foucault, M. (1995). *Discipline and Punish: The Birth of the Prison*. New York, NY: Vintage Books. — p. 63

(137) Nissenbaum (2009) develops the idea of contextual privacy and situates the problem in information flows that are improper relative to context, actors, and purpose. Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford Law Books. — p. 64

(138) Munafò, Nosek, Bishop, and colleagues (2017) call for transparency, documentation, and open review as basic conditions of reproducible science. Munafò, M. R., Nosek, B. A., Bishop, D. V. M., et al. (2017). "A manifesto for reproducible science." *Nature Human Behaviour, 1, 0021*. — p. 64

(139) Krippendorff (2018) systematizes coding, reliability, and validity in content analysis, with attention to explicit protocols and comparison across evaluators. Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology* (4th ed.). Los Angeles, CA: SAGE Publications. — p. 64

(140) Tyler (2007) summarizes procedural justice around voice, neutrality, respect, and trust, and places the possibility of being heard at the center of process legitimacy. Tyler, T. R. (2007). "Procedural Justice and the Courts." *Court Review, 44*(1/2), 26-31. — p. 64

Editorial note and statements

Editorial note

This essay does not seek to settle any question or to present as definitively proven what still needs to be tested. Its purpose has been, rather, to lay a sufficient conceptual foundation for the experimental and methodological beginning of prosody analysis within the CU framework.

I do not claim here that this framework is already conclusively true, functional, or fully validated in its application. What I do maintain is that, throughout these pages, sufficient conceptual foundations and methodological criteria have been established to justify the attempt to bring this proposal into the practical realm.

That is precisely the next step I intend to take: to submit these foundations to the test of experience, analysis, and confrontation with the reality of the prosodic phenomenon.

Acknowledgments

To my father, may he rest in peace, and to my mother: not only for bringing me into the world, but for giving me the certainty that, wherever I may be and in whatever "when" I am called to inhabit, they will always be with me, sustaining and supporting me, even when the world seems to be elsewhere.

I also want to express my gratitude to the many great historical figures who, even with everything against them, left burned into history the lesson that the world does not need to agree with you for you to try to give your own meaning to things, as long as that impulse is born from a commitment to something higher than oneself: the good of consensus as the final horizon.

Because, in the end, respect and the absence of fear are two of the most important conditions for thinking, for creating, and for trying to open a path.

Funding

This work has received no specific funding from public agencies, commercial entities, or non-profit organizations.

Conflicts of interest

The author declares that there are no conflicts of interest related to the content of this work. This text does not represent the official position of any religious, academic, or business institution, but solely the responsibility of the author within the framework of the Universal Coherence project (UC) .

How to cite

Recommended citation (APA 7): Calvo Espinosa, A. (2026). The Voice That Leaves No Signature: Prosody and Hidden Coherences Between Intention, Emotion, and Power (1.0). Zenodo. https://doi.org/10.5281/zenodo.18913483

Short in-text citation: (Calvo Espinosa, 2026)