The coded ear: epistemic violence as aesthetic material in AI

Dimitris Barnias
(dbarnias@gmail.com)
tokeno.net | https://mta.hmu.gr/
Hellenic Mediterranean University (Department of Music Technology & Acoustics)
Greece

ABSTRACT

Exploration in soundscape composition using Large Language Models confronts the researcher with AI's political dimensions and how these systems constitute power relations through the architecture of computational perception. This paper argues that this mode of perception is inherently political, a type of epistemic violence that encodes power relations into audio classification. To address it, advances a theoretical framework that understands AI as a perceptual authority producing its own realities. Building on critical AI studies, decolonial theory and media archaeology, the focus shifts to the acoustic domain to examine how machine listening manifests the coded ear – a parallel to Joy Buolamwini's coded gaze, that naturalizes Western, urban, English-language sonic epistemologies as objective perception. The coded ear is more than a translation: sound's relationality, temporality and embodiment make its algorithmic abstraction a different ontological operation. Within this framework, aesthetic friction is suggested as a poetics of political praxis. This friction unfolds through practices that treat AI biases as compositional material rather than error, rendering the underlying premises of machine listening audible. An uncomfortable thesis also follows from this analysis: the incompatibility between human and algorithmic perception is not a temporary technical limitation but a permanent political condition. Meaningful action therefore requires not better algorithms alone, but critical frameworks that refuse to close the gap between embodied and machine listening, while rendering that gap as a site of political art. Drawing on Segments application, a custom-built platform for algorithmic détournement used here not as an object of analysis but as a conceptual point of departure, the paper shows how such approaches expose, negotiate and subvert the politics of machine listening.

1 THE POLITICS OF LISTENING MACHINES

1.1 Infrastructural Perception

In recent years, AI is not simply used as a tool to process information, but increasingly AI mediates what information exists to be processed. This shift from tool to infrastructure marks a fundamental reconfiguration of power. In sound studies, when machine listening systems analyze acoustic environments, they do not simply assist human perception – they produce specific sonic realities shaped by training data, architectural constraints and corporate obligations. The transition from "What do we hear?" to "What is rendered hearable and by what authority?" becomes an important political question. The emergence of computational perception – or as it is termed here, disembodied listening: the process by which computational systems perform acoustic analysis and interpretation detached from the biological, situated and meaningful frame of human audition – represents what philosopher Bernard Stiegler (1998) calls disorientation, a technical exteriorization that fundamentally alters the conditions of experience. In the acoustic domain, this manifests through the friction between two incompatible epistemologies: the embodied, culturally situated listening of

humans and the disembodied, statistically-driven perception of AI systems. This friction is structural, revealing the political architecture of computational perception.

1.2 Ethics Washing

The technocratic response to AI bias has been to improve datasets, adjust architectures and fine-tune parameters. This approach, illustrated by initiatives like FAT (Fairness, Accountability and Transparency), treats political problems as engineering challenges. It assumes that with sufficient data diversity and algorithmic refinement, AI systems can achieve neutral, objective perception. Critics of FAT argue that the framework is often too theoretical, focuses only on technical fixes and is co-opted by big tech companies for ethics washing (Bietti, 2020; Barabas et al., 2020). They advocate that the very act of classification encodes power relations, the taxonomies that organize datasets reflect specific cultural worldviews, and the technical affordances of neural network architectures privilege certain forms of pattern recognition over others. These are not design issues to be fine-tuned, but features of a technical system that produces knowledge through statistical abstraction suitable for algorithmic processing.

2 EPISTEMIC VIOLENCE IN MACHINE LISTENING

2.1 From Coded Gaze to Coded Ear

Joy Buolamwini's (2018) concept of the coded gaze describes how computer vision systems encode specific visual epistemologies – typically those of their predominantly white, male, Western creators – into seemingly objective processes of image recognition. This gaze constructs the visible world according to specific parameters, rendering some phenomena legible and others invisible (see also Buolamwini & Gebru, 2018). Disembodied listening demonstrates a parallel phenomenon, the coded ear. For example, systems like YAMNet trained on Google's AudioSet dataset do not simply process sound, but produce a specific sonic ontology shaped by their training data and architectural design. The 521 classes of AudioSet reflect a specific cultural and geographic distribution. Categories like 'airplane', 'car horn' and 'speech' predominate, while acoustic phenomena from non-Western environments are underrepresented or absent – such as the microtonal pitch modulations of Turkish maqam traditions, traditional water management systems of non-industrial agricultural practices or gender-diverse vocal expressions that exceed binary classification schemas. The taxonomy itself encodes a particular relationship to sound, organized by functional utility rather than relational meaning or cultural reference. Using the same model as an example, the feedforward convolutional neural network architecture processes each 0.96-second window independently, severing the narrative and ecological continuity that characterizes human listening (Sterne, 2003). Sound becomes a series of discrete, statistically independent events rather than a continuous flow of meaningful experience. Moreover, the system's output – probability distributions across categories rather than definitive classifications – transforms knowledge into statistical confidence, while uncertainty becomes a technical problem to be minimized rather than a hermeneutic resource to be engaged. However, the coded ear is not a failure of engineering but a successful implementation of a specific perceptual administration. It hears exactly what it was designed to hear, a world of discrete, classifiable objects suitable for surveillance, recommendation, commodification and control. When it encounters sounds that do not fit this ontology it does not malfunction; rather it performs exactly as intended by subsuming difference into its native predetermined categories or rendering it as statistical noise.

2.2 Epistemic Injustice and Relational Harm

Engaging with Miranda Fricker's (2007) concept of epistemic injustice and extending it through Abeba Birhane's (2021) relational ethics framework, this paper also argues that machine listening systems produce specific forms of infrastructural harm that exceed individual misclassification. They manifest testimonial injustice – prejudice causing deflation of credibility to specific knowers – and hermeneutical injustice – structural gaps in collective understanding rendering certain experiences unintelligible.

2.2.1 Silencing of Acoustic Knowledge.

The taxonomies and training data that structure machine listening encode specific interpretive frameworks rooted in Western acoustic epistemologies (Robinson, 2020). Indigenous sonic ontologies where sound might constitute active relationships with ancestors, mark territorial boundaries, maintain ecological balance or embody shared obligations between human and non-human entities lack hermeneutical resources within algorithmic classification. Following Fricker's (2007) analysis of how marginalized groups are excluded from the "practices by which social meanings are generated", machine listening systems show how interpretive injustice becomes automated. The system cannot hear what it has no concepts for, and in this absence, alternative ways of knowing are systematically marginalized. When a Cretan rizitika song is classified generically as 'singing' or 'Mediterranean music' the system commits testimonial injustice. It gives a reduced level of credibility to the singer due to the prejudice of computational taxonomy that privileges acoustic signal over cultural signification. The song's function as territorial mapping, historical narration and social cohesion is rendered epistemically invisible. The system responds only to the sound's acoustic properties (pitch, timbre, rhythm) while systematically silencing its cultural knowledge. Similarly, when an Amazonian shaman's chant is rendered as plain audio data for models, for the tribe it remains an active engagement with forest spirits and ecological processes. By reducing these to instances of 'music' or 'speech' the system enacts what Boaventura de Sousa Santos calls epistemicide – the systematic destruction of rival knowledge while maintaining the appearance of technical neutrality (cited in Robinson, 2020).

This silencing is amplified by the networked nature of algorithmic harm (Birhane, 2021). The deployment of machine listening in smart cities, surveillance systems and environmental monitoring creates new configurations of power where certain communities are disproportionately subjected to algorithmic interpretation, while having minimal influence over the systems that parse their sonic environments. When urban sound sensors classify street performances as 'noise pollution', they enact testimonial injustice on a structural level. The sonic practices of marginalized communities are afforded diminished credibility not by individual listeners, but by infrastructural systems that pathologize their acoustic presence. The prejudice is encoded in the very architecture of the classifier which assumes a politically normative sonic environment defined by Western standards.

2.2.2 Improbability of the Alternative.

The political violence of the coded ear is not located in individual errors but in the systematic production of ignorance. The system does not just mishear – it constructs an acoustic world where certain forms of knowledge are literally inaudible. This epistemic violence is particularly acute in the acoustic domain, as sound is fundamentally relational and contextual, qualities that resist the object-oriented logic of computational classification. The harm here is fundamentally relational. It damages not just individual epistemic agents but also the webs of knowers that constitute genuine knowledge production (Birhane, 2021). When diasporic communities see their sonic practices reduced to unclassified audio in smart city infrastructures, the injustice extends beyond hermeneutical marginalization. It constitutes

what Fricker (2007) terms a critical level of reflexive awareness failure – a systematic refusal to recognize the credibility of alternative epistemologies, now automated and scaled through technical systems that present as objective. Machine listening systems thus function as what Birhane (2021) calls human-mission systems: technical infrastructures that encode specific human values and power relations while masking cultural specificity behind their construction. The epistemic injustice they produce is doubly hidden, by the black-boxing of algorithmic decision-making and by the rationalist assumption that audition can be formalized without remainder.

2.3 The Politics of AI Resources

The politics of machine listening extends beyond algorithmic design to the material infrastructure that enables it. Training large audio models requires massive computational resources, extensive curated datasets and significant technical expertise. These resources are concentrated in specific geographic and institutional locations, predominantly in the United States, Europe and China (Crawford, 2021), suggesting a future of competing techno-imperialisms. This concentration creates infrastructural colonialism, a contemporary form of extractive relation where sonic data from the Global South is harvested and processed through Northern-controlled algorithms, then returned as insights or services that embed Western epistemological frameworks. These flows of data and capital mirror historical colonial patterns (Robinson, 2020). The infrastructural dimension of AI power is not therefore reducible to model performance. The concentration of computational resources in Northern data centres, the extraction of training data from the Global South and the circulation of technical expertise through a handful of corporations and universities persist regardless of how accurate a particular classifier becomes (Crawford, 2021).

3 ALGORITHMIC GOVERNMENTALITY

3.1 The Authoritative Act of Categorization

Machine listening systems exercise power through what Ian Hacking (1999) calls dynamic nominalism: the concept that the categories invented to describe people do not simply label them but they actually change them. In the acoustic domain, this is the capacity of classification categories to reshape the reality they are supposed to describe. When a model classifies a sound as ‘bird’ with, e.g., 82% confidence, it performs an authoritative act that establishes what counts as a recognizable sound event. This authority is particularly strong because it presents itself as technical rather than political. The confidence score of 0.82 suggests objective measurement rather than cultural interpretation. Yet this number masks a complex process of statistical inference shaped by training data selection, annotation practices and model architecture. It also involves human decisions that encode specific values and assumptions.

The political effect is what Foucault (1991) might recognize as a form of algorithmic governmentality: the conduct of conduct through computational systems that shape behavior by structuring the possible field of action – systems of power that decide what counts as universal truths. When urban planning departments use machine listening to monitor noise pollution, they operationalize specific definitions of sound and silence that privilege certain acoustic ecologies over others. When conservation biologists deploy automated species recognition, they enact particular relationships between human researchers and non-human sound-makers, often collecting human ‘bycatch’ data and vexatious surveillance suspicions.

3.2 The Violence of the Cut

Jonathan Sterne's (2003) media archaeological analysis reveals how sound reproduction technologies constitute new sonic ontologies through what he calls the cut: the technological separation of sound from its source and context. Disembodied listening transfers this violence through what this paper terms the analytical cut: the segmentation of continuous acoustic experience into discrete processing windows. The 0.96-second analysis window used in the YAMNet model is not technically necessary. Alternative architectures could process longer contexts. However, it represents a specific political ontology of time. Continuous, narrative, embodied experience is transformed into a series of identical, disembodied, interchangeable moments suitable for statistical comparison (Ernst, 2016). The temporal unity of a bird's territorial call – introduction, escalation, climax, resolution – is separated into independent classifications: [Bird: 82%], [Bird: 79%], [Bird: 85%]. This temporal violence is simultaneously epistemological and political as it renders certain phenomena legible while making others absent. The algorithm does not simply hear differently, but it constructs a simplified acoustic world where meaning is partially encoded and only certain forms of acoustic life matter.

3.3 Bias as Worldview

Standard approaches to AI ethics treat bias as a deviation from neutral objectivity and as a technical flaw to be corrected through better data and cleverly written algorithms (Bietti, 2020; Barabas et al., 2020). This paper argues that this framework fundamentally misunderstands the nature of computational listening perception. Every system performing machine listening embodies what philosopher of science Thomas Kuhn called a Kuhnian paradigm – a coherent set of assumptions, methods and standards that constitute a way of seeing the world (Kuhn, 1962). The model's bias toward Western urban sounds is not an engineering error, but the necessary condition of its perceptual coherence. A system trained on globally diverse sounds with locally-appropriate taxonomies would not be less biased. It would simply manifest a different worldview, equally particular and equally political. Recognizing bias as a worldview rather than error has serious implications for critical practice. It means that technical fixes are insufficient. Improving dataset diversity cannot eliminate the political dimensions of algorithmic perception but only redistribute them. Ideally, multiple AI systems are needed that make their specific biases explicit and available for critical comparison. Additionally, although transparency is necessary it is not sufficient. Understanding how a system works technically does not reveal its political assumptions. Critical analysis must examine training data curation and taxonomic decisions, work already in progress by scientists investigating AI ethics (Birhane, 2021; Bietti, 2020). Error is epistemologically productive. The moments when AI systems fail by misclassifying, reveals the boundaries of their perceptual paradigms, thus making audible what is normalized as objective knowledge.

4 AESTHETIC FRICTION

4.1 From Critique to Practice

If the politics of machine listening cannot be solved through better engineering, this paper proposes a concept based on aesthetic friction: a framework of creative practices that work within algorithmic systems to expose, negotiate, and subvert their political assumptions. This approach draws on traditions of participatory and relational art [Bishop, 2012; Bourriaud, 2002] while engaging directly with computational systems. A custom-built platform developed through practice-based research enables this approach. Rather than treating AI

biases as obstacles to overcome, it treats them as compositional material. This strategy involves:

1. **Exposition:** Making the coded ear audible by foregrounding the specific ways AI systems parse, classify, and reorganize sound.
2. **Negotiation:** Creating compositional situations where human and algorithmic perception are placed in deliberate friction, making the gap between embodied experience and disembodied listening productive.
3. **Subversion:** Using algorithmic outputs against their intended purposes and reorganizing classified sounds to create new meanings that exceed the system's categorical logic.

4.2 Strategies of Algorithmic Détournement

The Situationist International advanced détournement – the diversion or hijacking of existing cultural materials – as a tactic against the commodification of experience (Debord, 1967). By recontextualizing elements from their original function, one negates their authority while repurposing their force. Against machine listening's claims to neutral perception, algorithmic détournement seeks an aesthetically-political action through strategic misdirection – taking systems designed for classification and control and rerouting them toward indeterminacy, failure and critical revelation.

For the composer, the primary compositional act shifts from arranging sounds in time to navigating the maps generated by AI analysis, thus performing a semantic reorganization. By grouping sounds based on algorithmic classifications rather than acoustic properties or cultural significance, composers reveal the internal logic of the coded ear. When 'Water' sounds are arranged by confidence score rather than by source (stream, rain, sea), the composition exposes how the system generates a false homogeneous reality across phenomena that humans experience as distinct. Composers can also use temporal exploitation: the atemporal, windowed perception of machine listening as a formal constraint. Compositions built from the 0.96-second segments make audible the violence of the analytical cut. Violent transitions between segments become rhythmic elements and the destruction of narrative continuity becomes a structural principle. This strategy refuses the restoration of organic unity and instead makes fragmentation itself aesthetically meaningful. Moreover, errors can be refined. Rather than filtering out low-confidence classifications, failures can be amplified and organized. A composition consisting entirely of sounds where the AI shows less than 50% confidence across all categories makes audible the system's uncertainty. Finally, by deliberately juxtaposing sounds from categories that the system treats as unrelated – 'Airplane' and 'Bird', 'Laughter' and 'Crying' – composers expose the inconsistent nature of the classification scheme. The composition becomes a critique of the taxonomy itself.

4.3 Toward a Posthuman Politics

These aesthetic strategies are not just formal experiments. They constitute a political practice that reconfigures the relationship between human and machine agency. By treating the algorithm as a collaborator with its own biased mode of perception, composers refuse the technophilic fantasy of seamless human-machine integration (Hayles, 1999; Braidotti, 2013) without rejecting the models altogether. This position draws on Donna Haraway's (2016) concept of staying with the trouble – engaging with asymmetric, partial and risky relationships rather than seeking transcendent solutions. In the context of disembodied listening, this means acknowledging complicity. Composers working with AI systems participate in the political economies of data extraction and computational resource consumption. Aesthetic friction cannot claim purity, but must work within complicity

(Crawford, 2021). It also means valuing incompatibility by accepting that the productive friction between human and machine perception is not a temporary limitation to be overcome, but a permanent condition to be refined (Braidotti, 2013). Political art in the age of AI does not seek harmony but stages productive conflict. Composition becomes the configuration of hybrid assemblages where neither human nor machine has final authority. Composers function simultaneously as practitioners, critics and configurators of these assemblages, constantly challenging the power relations embedded in their tools and their own position within those relations.

5 Beyond technical progress

A common objection to critical AI studies holds that current limitations are temporary features of an immature technology. It is believed that as datasets become more diverse, architectures more sophisticated, and processing more efficient, the political dimensions of machine listening will simply dissolve through better engineering. This objection misunderstands the nature of computational perception. The political issues illustrated in this paper are not technical flaws but structural features of algorithmic classification as such. Even a perfect AI system would still instantiate a specific perceptual regime with its own political commitments (Birhane, 2021; Crawford, 2021). The violence traced through the coded ear, are not bugs awaiting a patch, but foundational elements of the system's design. The fundamental reason is that human listening is embodied, affective, temporally continuous and meaning-seeking, while algorithmic listening is statistical, operational, atemporal and correlation-executing (Sterne, 2003; Stiegler, 1998). This difference cannot be engineered away because it is fundamental to computation itself. The cut is not a flaw, it is the method, making the ontological gap between human and algorithmic perception permanent. Importantly, engineering cannot eliminate the politics of classification. Any system that organizes the world into categories must decide what the categories are, where their boundaries lie, and how to treat uncertainty (Hacking, 1999). At the very bottom, these are political decisions. A perfectly representative global dataset would still require someone to choose the taxonomy – and every taxonomy privileges some ways of knowing while silencing others (Fricker, 2007; Robinson, 2020). Finally, the infrastructural dimension of AI power is not reducible to model performance. The concentration of computational resources in Northern data centers, the extraction of training data from the Global South and the circulation of technical expertise through a handful of corporations and universities persist regardless of how accurate a particular classifier becomes (Crawford, 2021). The politics of AI will not be easily solved, and they will persist as long as perception is authorized to machines. However, persistence is not paralysis. It is an invitation to listen more carefully, create differently, and to refuse by actions the false promise of a final neutral ear.

6 The uncloseable gap

As argued, machine listening systems are neither neutral tools nor failed human replacements. Against the technocratic fantasy of a fairer, more inclusive algorithm, this paper has proposed aesthetic friction through practices that do not try to close the gap between human and machine listening, but instead bring that gap creatively to the surface and expose the coded ear's violence. However, friction alone does not dismantle infrastructure, nor change policies in an instant. No composition, however inspired, will shut down a data center, redistribute GPU clusters or stop a smart city from using systems to police street musicians. As illustrated previously, the gap between human and algorithmic perception is a permanent political condition. It will not be resolved by better engineering, but neither will it be resolved by better art. What, then, does aesthetic friction achieve? Audibility. The violence of the coded

ear is normally silent. It operates in server farms, black-boxed APIs and confidence scores that hide as objectivity. When a composer takes every sound classified with less than 30% confidence and arranges those fragments into a forty-minute piece, the system's hidden ontology becomes acoustic. That creative friction becomes a powerful method against the coded ear's dominion. When the audience hears the creative fails, they do not learn how to fix the algorithm, but they learn that the algorithm has a worldview and that worldview is not theirs. As composers, using these tools the task is to make them betray their own assumptions in a joyful, relentless sabotage from within.

REFERENCES

[1] E. Bietti. 2020. From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT '20). ACM, New York, 210–219.

[2] C. Barabas, C. Doyle, J. B. Rubinovitz, and K. Dinakar. 2020. Studying up: Reorienting the study of algorithmic fairness around issues of power. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT '20). ACM, New York, 167–176.

[3] A. Birhane. 2021. Algorithmic injustice: A relational ethics approach. Patterns 2, 2 (2021).

[4] C. Bishop. 2012. Artificial Hells: Participatory Art and the Politics of Spectatorship. Verso, London.

[5] N. Bourriaud. 2002. Relational Aesthetics. Les Presses du réel, Dijon.

[6] R. Braidotti. 2013. The Posthuman. Polity Press, Cambridge.

[7] J. Buolamwini and T. Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research 81 (2018).

[8] K. Crawford. 2021. Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press, New Haven.

[9] G. Debord. 1967. The Society of the Spectacle. (English translation 1995).

[10] W. Ernst. 2016. Chronopoetics: The Temporal Being and Operativity of Technological Media. Rowman Littlefield, London.

[11] M. Foucault. 1991. Governmentality. In The Foucault Effect: Studies in Governmentality, G. Burchell, C. Gordon, and P. Miller (Eds.). University of Chicago Press, Chicago, 87–104.

[12] M. Fricker. 2007. Epistemic Injustice: Power and the Ethics of Knowing. Oxford University Press, Oxford.

[13] I. Hacking. 1999. The Social Construction of What? Harvard University Press, Cambridge, MA.

[14] D. J. Haraway. 2016. Staying with the Trouble: Making Kin in the Chthulucene. Duke University Press, Durham.

[15] N. K. Hayles. 1999. How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics. University of Chicago Press, Chicago.

[16] T. S. Kuhn. 1962. The Structure of Scientific Revolutions. University of Chicago Press, Chicago.

[17] D. Robinson. 2020. Hungry Listening: Resonant Theory for Indigenous Sound Studies. University of Minnesota Press, Minneapolis.

[18] J. Sterne. 2003. The Audible Past: Cultural Origins of Sound Reproduction. Duke University Press, Durham.

[19] B. Stiegler. 1998. Technics and Time, 1: The Fault of Epimetheus. Stanford University Press, Stanford.

[20] The Segments Application. 2026. https://ihearcolors.online/segments

Copyright: © 2026 Dimitris Barnias. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.