English Studies in Canada

Distant Listening and Resonance

Tanya E. Clement

Volume 46, Number 2-3-4, June–September–December 2020

New Sonic Approaches in Literary Studies

URI: https://id.erudit.org/iderudit/1111329ar
DOI: https://doi.org/10.1353/esc.2020.a903548

See table of contents

Publisher(s)

Association of Canadian College and University Teachers of English (ACCUTE)

ISSN

0317-0802 (print)
1913-4835 (digital)

Explore this journal

Cite this document

Clement, T. (2020). Distant Listening and Resonance. *English Studies in Canada*, *46*(2-3-4), 279–284. https://doi.org/10.1353/esc.2020.a903548

Distant Listening and Resonance

Tanya E. Clement
University of Texas Austin

For speech recordings, sound is text—the words people speak, but also other sounds that indicate a speaking and listening context: tone and laughter, coughing and crying, bird song, car engines and horns, a baby crying, thunder clapping, gun shots, the needle dropping, the needle scratching, to name a few. Using computation to analyze many texts at once in big data sets has been called "distant reading" in Digital Humanities (Underwood). I have described "distant *listening*" to sound texts as using computing to "distill the many-layered four-dimensional space of the text in performance (i.e., embodied within the performance network of interpretations with the listener in time and space) into a two-dimensional script called 'code'" (Clement, "Distant Listening"). Distant listening, like distant reading, implies a lack of granular observation based on proximity in terms of space as well as a removal in terms of emotion, experience, and individual or subjective knowledge. Sound travels differently than light; what is lacking is made up for in other ways. What is too close can be too loud. What is far can be communicated loud and clear. Resonance is both an embodied, physical experience as well as a cultural hermeneutic.

Specifying sound computationally is a process of discretization. Without going too far down the mathematical rabbit hole, discretization, it is safe to say, is a means of mathematically representing a continuous signal through samples that indicate the whole without actually capturing it fully. Sound is air pressure variation over time. Ears turn the pressure differences into neural activations while microphones create digital sound by translating pressure differences into voltage differences. An audio signal is a sequence of mathematical abstractions that map voltage (or pressure) over time in a wave, and frequency is the number of times per second that a sound pressure wave repeats itself (McFee, "Signals"). Audio signal processing tools "cannot work directly with continuous signals," so, before being processed by a computer, the sound pressure wave must be discretized. Signal discretization includes sampling and quantization (McFee, "Digital Sampling"). The sampling process is more or less precise when more or fewer discrete samples are used to represent a signal across a period of time, but all the information is never represented. Sampling implies absence.

Tanya E. Clement is an Associate Professor in English and the Director of the Initiative for Digital Humanities at the University of Texas Austin. Her research includes textual studies, sound studies, and infrastructure studies as these concerns impact academic research, research libraries, and the creation of DH tools and resources.

An ontology for modeling textuality through computers requires a balance between what's computable and what is meaningful: the model should be "internally consistent, and as much as possible avoid clashes with commonsense beliefs" (Floyd and Renear). In *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*, Ben Gold, Nelson Morgan, and Dan Ellis similarly describe this balance between meaning and matter within the history of speech transmission:

> If we think, for the moment, of speech as being a mode of transmitting word messages, and telegraphy as simply another mode of performing the same action, this immediately allows us to conclude that the intrinsic information rate of speech is exactly the same as that of a telegraph signal generating words at the same average rate. Speech, however, conveys emphasis, emotion, personality, etc., and we still don't know how much bandwidth is needed to transmit these kinds of information. (21)

A computationally tractable model of a text, is much like a bandwidth—"a range of frequencies or wave-lengths that falls between two given limits" ("band, n.2.")—it must be explicit, consistent, and manipulable (McCarty), yet it remains always partial and inexact.

Distant listening is at root a technically complex matter of fitting a mathematical abstraction of sound to a lived experience about what that sound means. When signal processing scientists talk about sound, they consider damping ratios, gain, frequencies, spectra, energy, and pitch energy and talk about how these features influence sound fidelity. When humanists talk about sound, they talk about language dynamics (tempo, pitch, tone/timbre, volume, pace, laughter, silence, applause, moans, screams, dialects, changing speakers, gender, age, changing genres), environment (fan hums, car horns, chickens, train whistles, bird calls, frogs mating), and materiality (recording noises such as changing tracks, distortion, the electronic grid, and needle drops). When humanists talk about sound, they then abstract from these features of sound to consider how these features influence meaning. For example, "matters of special significance for poetry" include "the cluster of rhythm and tempo (including word duration), the cluster of pitch and intonation (including amplitude), timbre, and accent" (Bernstein, 126) that indicate different styles of prosody (MacArthur, Marit J., et al.; Mustazza, "Machine-Aided Close Listening"). Other sounds can point to the recording venue such as crowd sounds (Clement and McLaughlin) or the distribution and preservation process including the machine noises, drops in the recordings, or distortions (Mustazza, "Provenance report," "Vachel Lindsay and the W. Cabell Greet Recordings"). Creating a computational model of a "meaningful" sound text with the explicitness and consistency that computation requires is generative: it "forc[es] us to confront the radical difference between what we know and what we can specify computationally, leading to the epistemological question *of how we know what we know*" (McCarty).

Researchers remain in the early days of epistemological questions surrounding issues of accuracy and efficacy of large-scale sound analysis, especially with historical speech recordings. Possibilities for using computational, cultural-analytic tools to develop knowledge about large caches of recorded audio is a primary objective of the High Performance Sound Technologies for Access and Scholarship project. In collaboration with machine learning scientists at the Illinois Informatics Institute at the University of Illinois at Urbana-Champaign, the Texas Advanced Computing Center at the University of Texas, and the NYU Center for Data Science, HiPSTAS researchers have investigated machine learning processes for detecting resonant patterns across large data sets of poetry collections (PennSound and SpokenWeb), oral histories (StoryCorps), radio recordings (WGBH), and university archives (Indiana University) to consider prosody as well as the basic and essential (but more mundane) tasks of event detection, keyword extraction, speaker disambiguation (diarization), speaker recognition, and quality. Such work reveals basic features of sound texts that point to more abstract and sophisticated modes of access and analysis with large datasets, but the prevailing literature indicates that these methods remain inaccurate. Recording quality, accents, or presence of background noise continue to influence accuracy. Recent work in the broader field of acoustics, speech, and signal processing uses features generated or learned as a byproduct of training large-scale deep networks for some general tasks like acoustic scene classification or source identification (Baevski, Alexei, et al. and Cramer, et al.), but inductive bias—assumptions about the data that are encoded in the model to learn the target function and to generalize beyond training data—are obfuscated in these black-box methods that resist critical intervention.

As both a physical property and a cultural hermeneutic, resonance serves as a useful theory for articulating how distant listening can make meaning differently. Resonance is both an embodied, physical experience as well as a cultural hermeneutic. In the physical realm, resonance occurs when an event creates sound waves with frequencies that match a receiving object's resonant frequency or the frequency at which that object naturally vibrates. Deaf studies scholars make clear that sound is "multimodal" (Mills 2015, 52; Friedner and Helmreich 2015). Without resonance, there is no sound, but resonant frequencies can also be experienced physically through vibrations or be aurally imperceptible such as those required to generate a gravitational pull between orbiting bodies in space and complex clusters of electrons or to make a child's swing push higher or a bridge collapse. While there is no sound without resonance, there is resonance without sound. As a cultural hermeneutic, resonance is multitudinous. As defined by the *Oxford English Dictionary resonance* evokes "images, memories, and emotions" ("resonance, n.") and *to resonate* ("resonate, v.") is "to respond in a sympathetic or corresponding manner; to react emotionally or positively." When I say some person, place, thing, or event *resonates* with me, I use the term as a placeholder for a sense or a feeling of significance for which the variables of causation are too numerous or too complex to articulate exactly. As a physical and cultural phenomenon, resonance seems to occur in a liminal, processual space between still and vibrating, between knowing and known. Resonance—where inexactitude, multitudity, the subjective, and the personal coexist with the tangible and physical—as a theory for framing meaning making with computational sound analysis offers an opportunity to reimagine the possible proximities of distant listening.

Works Cited

"band, n.2." OED Online, Oxford UP. June 2022. www.oed.com/view/Entry/15113.

Baevski, Alexei, Henry Zhou, Abdelrahman Mohamed, and Michael Auli. "wav2vec 2.0: A framework for self-supervised learning of speech representations." *Advances in Neural Information Processing Systems* 33 (2020): 12449–60.

Clement, Tanya. "Distant Listening: On Data Visualisations and Noise in the Digital Humanities." *Digital Studies / Le Champ Numérique* 3.2 (July 2013).

———. *Dissonant Records: Close Listening to Cultural Resistance in Audio Archives.* MIT Press (forthcoming 2024).

———, and Stephen McLaughlin. "Measured Applause: Toward a Cultural Analysis of Audio Collections." *Journal of Cultural Analytics* 1.1 (May 2016).

Cramer, Jason, Ho-Hsiang Wu, Justin Salamon, and Juan Pablo Bello. "Look, Listen, and Learn More: Design Choices for Deep Audio Embeddings." *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 2019.

Floyd, I., and Renear, A.H. "What Exactly is an Item in the Digital World?" Poster presented at the Annual Meeting of the Association for Information Science and Technology, Milwaukee, Wisconsin, 19–24 October 2007.

Gold, Ben, Nelson Morgan, and Dan Ellis. *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*, second edition. Wiley, 2011.

High Performance Sound Technologies for Access and Scholarship. www.hipstas.org.

MacArthur, Marit J., Georgia Zellou, and Lee M. Miller. "Beyond Poet Voice: Sampling the (Non-) Performance Styles of 100 American Poets." *Journal of Cultural Analytics* 3.1 (April 2018).

McCarty, Willard. "Modeling: A Study in Words and Meanings." *Companion to Digital Humanities* (Blackwell Companions to Literature and Culture). Eds. Susan Schreibman, Ray Siemens, and John Unsworth. Blackwell, 2004.

McFee, Brian. "Digital Sampling." *Digital Signals Theory*. https://brianmcfee.net/dstbook-site/content/ch02-sampling/intro.html.

———. “Signals.” *Digital Signals Theory*. www.brianmcfee.net/dstbook-site/content/ch01-signals/Intro.html.

Mustazza, Chris. “Machine-Aided Close Listening: Prosthetic Synaesthesia and the 3D Phonotext.” *Digital Humanities Quarterly* 12.3 (December 2018).

———. “Provenance report: William Carlos Williams’s 1942 reading for the NCTE” *Jacket2.* 21 May 2014. https://jacket2.org/article/provenance-report.

———. “Vachel Lindsay and the W. Cabell Greet Recordings.” *Chicago Review* 59 (2016): 98.

“resonance, n.” OED Online, Oxford UP. September 2022. www.oed.com/view/Entry/163743.

“resonate, v.” OED Online, Oxford UP. September 2022. www.oed.com/view/Entry/163745.

Underwood, Ted. “A Genealogy of Distant Reading.” *Digital Humanities Quarterly* 11.2 (June 2017).