

PEDRO J S VIEIRA DE OLIVEIRA

“DAS HÄTTE NICHT PASSIEREN DÜRFEN” – RE-NARRATING BORDER VOCALITIES AND MACHINE LISTENING CALIBRATION

As a sound artist and researcher, I have been interested in the materiality of voice databases, their semantic and political value, the decision-making process for the collection of speakers, the sounds they produce, the cultural meaning of these sounds, and their continued existence as archives. The creation of *speech corpora* – that is, the assembling of an exhaustive archive of possibilities for human speech – is an arduous task. Yet this seeming unfeasibility does not prevent speech corpora from being deployed in systems that rely on and judge over incomplete assessments of the ambiguity and contingency of prosody, pronunciation, inflection, and timbre. These systems of voice biometry seek to normalize vocal traits, as well as to establish an alleged quantifiable ‘truth’ to how a voice might convey language – a physical but also cultural and social phenomenon. Biometric technologies, Heather Murray argues, are “constituted by the practices involved in [their] use”; its performativity “is bound to and produces cultural understandings of gender, authority, and criminality”¹. State-sponsored campaigns transfer biometric assessment from machines to citizens, encouraging peer-surveillance in train stations, airports, and public offices as well as police-enforced racial profiling and violence. Processes like these evince how sound is instrumentalized to act as a disciplinary mechanism, and how biometry is fundamentally a performative gesture: it seeks to pinpoint that which it has set itself to reveal.

The classification of human beings into categories (in terms of

¹ Heather Murray, “Monstrous Play in Negative Spaces: Illegible Bodies and the Cultural Construction of Biometric Technology”, *The Communication Review*, 10 (4), 2007, pp. 347–365, here: p. 349. Available at: <https://doi.org/10.1080/10714420701715415> [accessed February 24, 2019].

gender, age, language, etc.), constructed by human and machinic assessment, seeks to neutralize ambiguity and any ‘loose ends’. This endeavor, obscured in many biometric systems by design, is but another set of disciplinary devices aimed to contain and regulate black and brown bodies. Or, in Louise Amoore’s words, “a matter of biopolitics, as a mobile regulatory site through which people’s everyday lives can be made amenable to intervention and management.”² Biometric technologies are calibrated within a set of normative assumptions that, in effect, convey white supremacist modes of seeing and listening. Scholar Simone Browne argues that it is in the moment of measuring – observing and listening – as well as calibration and application that the racialized ethos of biometry is revealed.³

The main inspiration for this work is the disastrous case of German neo-Nazi soldier Franco A., who managed to impersonate a Syrian refugee and was granted asylum seeker status in Germany by the *Bundesamt für Migration und Flüchtlinge* (BAMF) – the authority in charge of Migration and Refugees. His intention was to carry out a series of violent attacks on public figures, left-wing politicians and activists, making use of his identity as a Syrian migrant.⁴ His case seemed to put the entire German asylum seeker system in jeopardy. Numerous reports, statements, and press inquiries sought to shed light on how a white German citizen with no knowledge whatsoever of Levantine Arabic, as well as a questionable backstory, could be granted a temporary asylum status. The solution found by BAMF to address this case was to begin, with utmost urgency, testing and deploying so-called ‘accent recognition’ software to allegedly avoid ‘human mistakes’ which might have led to Franco A.’s acceptance within the system.

This project thus focuses on the application of voice databases for ‘accent recognition technologies’, and their deployment within the migration and border industries of the European Union – particularly in Germany. My interest in the use of sound archives for these specific technologies seeks to challenge the idea that it might be feasible, for the purposes of biometric assessment, to divorce a sound’s materiality from its constitution as a cultural phenomenon. I ‘zoom in’ and slow down processes of software calibration, re-narrating them as decolonizing and pedagogical moments. This exploration focuses on two aspects of

² Louise Amoore, “Biometric Borders: Governing Mobilities in the War on Terror”, *Political Geography*, 25 (3), 2006, pp. 336–251, here: p. 337. Available at: <https://doi.org/10.1016/j.polgeo.2006.02.001> [accessed February 24, 2019].

³ Simone Browne, *Dark Matters: On the Surveillance of Blackness*, Durham, Duke University Press, 2015, pp. 108–109.

⁴ Deutsche Welle, “German Officer-Refugee Probe Shifts to ‘Hit-List’”, *Deutsche Welle*, April 30, 2017. Available at: <https://www.dw.com/en/german-officer-refugee-probe-shifts-to-hit-list/a-38647209> [accessed August 31, 2018].

datasets: the first being the semantic content of the so-called "elicitation cues", short sentences which speakers of a database are required to say. The apparent triviality of those cues – snippets of old books, shopping lists, meaningless sentences – are yet another component in the desire to divorce sound from meaning; they also convey interesting assumptions and expectations on by whom they ought to be spoken in the first place. The second explores the breaking down of words in phonemes using an arbitrary system of classification; as with many technologies of the past decades, most speech recognition technologies are based on research financed by the US Defense Advanced Research Projects Agency (DARPA) in collaboration with Texas Instruments and the MIT. DARPA-TIMIT – how this database is known – is a widely employed model for designing speech recognition systems in multiple languages other than English. Hence most of these systems will feature phonetic transcriptions of words based on their English pronunciation, regardless of the language being tested.⁵

The first iteration of this project took place in Berlin during February and March 2018 at ACUD Gallery. As a three-channel installation, it focused primarily on the official press releases by BAMF concerning Franco A.'s case, and the subsequent implementation of accent recognition software. This piece had machines reading bureaucratic language, but doing so in what they might understand as a 'neutral' German accent. The second installment was commissioned by the Goethe-Institute Brussels, in collaboration with the Brussels Experimental choir under the direction of Floris Lammens. For this piece I turned phonemes into scores meant to be sung by a 'trio of immigrant voices', highlighting the distinct processes of translation that happen when the task of machinic reading is performed by humans. Three singers – Moumy Chahou, Farida Lehyan, and Ghassem Mousavi –, performed the piece on April 14, 2018. Each of them was assigned a slightly different score with different notations for prosody, modulation, and amplitude; some sang multiple lines while others cycled through shorter sets of phonemes. They also read from bureaucratic texts, albeit employing a deliberate 'mocking' tone, as an act of insolence and insubordination.

The scores, compositions and recordings thereof are the beginning of a larger project critically inquiring the function of the database as an installment of a sound archive. It challenges and intervenes on processes of assembling mechanisms which rely on human imperfections, as well

⁵ Timothy Leonido, "How to Own a Pool and Like It", *Triple Canopy*, April 5, 2017. Available at: <https://www.canopycanopycanopy.com/contents/how-to-own-a-pool-and-like-it/> [accessed February 24, 2019].

as in a long thread of translations and (mis)interpretations which rule and decide over the lives of other-ed bodies. By adding melodic components and making a musical composition out of elicitation cues, I emphasized their contingency as ‘building blocks’ for voice assessment. My goal was to make these cues move and unfold into the performance space as a form of reclaiming back other spaces – delegated and obscured inside tightly and sleekly designed systems. In refusing to ‘take part’ in the sound archive’s own shortcomings, these performances did not solely denounce the coloniality embedded in accent and voice recognition systems; instead, they enunciated poetic rehearsals for different, multiple futures.